

TECNOLOGIAS, MÉTODOS E TEORIAS NA ENGENHARIA DE COMPUTAÇÃO



ERNANE ROSA MARTINS
(ORGANIZADOR)

 **Atena**
Editora

Ano 2020

TECNOLOGIAS, MÉTODOS E TEORIAS NA ENGENHARIA DE COMPUTAÇÃO



ERNANE ROSA MARTINS
(ORGANIZADOR)

Atena
Editora

Ano 2020

Editora Chefe

Profª Drª Antonella Carvalho de Oliveira

Assistentes Editoriais

Natalia Oliveira

Bruno Oliveira

Flávia Roberta Barão

Bibliotecário

Maurício Amormino Júnior

Projeto Gráfico e Diagramação

Natália Sandrini de Azevedo

Camila Alves de Cremona

Karine de Lima Wisniewski

Luiza Alves Batista

Maria Alice Pinheiro

Imagens da Capa

Shutterstock

Edição de Arte

Luiza Alves Batista

Revisão

Os Autores

2020 by Atena Editora

Copyright © Atena Editora

Copyright do Texto © 2020 Os autores

Copyright da Edição © 2020 Atena Editora

Direitos para esta edição cedidos à Atena Editora pelos autores.



Todo o conteúdo deste livro está licenciado sob uma Licença de Atribuição *Creative Commons*. Atribuição 4.0 Internacional (CC BY 4.0).

O conteúdo dos artigos e seus dados em sua forma, correção e confiabilidade são de responsabilidade exclusiva dos autores, inclusive não representam necessariamente a posição oficial da Atena Editora. Permitido o *download* da obra e o compartilhamento desde que sejam atribuídos créditos aos autores, mas sem a possibilidade de alterá-la de nenhuma forma ou utilizá-la para fins comerciais.

A Atena Editora não se responsabiliza por eventuais mudanças ocorridas nos endereços convencionais ou eletrônicos citados nesta obra.

Todos os manuscritos foram previamente submetidos à avaliação cega pelos pares, membros do Conselho Editorial desta Editora, tendo sido aprovados para a publicação.

Conselho Editorial

Ciências Humanas e Sociais Aplicadas

Prof. Dr. Álvaro Augusto de Borba Barreto – Universidade Federal de Pelotas

Prof. Dr. Alexandre Jose Schumacher – Instituto Federal de Educação, Ciência e Tecnologia do Paraná

Prof. Dr. Américo Junior Nunes da Silva – Universidade do Estado da Bahia

Prof. Dr. Antonio Carlos Frasson – Universidade Tecnológica Federal do Paraná

Prof. Dr. Antonio Gasparetto Júnior – Instituto Federal do Sudeste de Minas Gerais

Prof. Dr. Antonio Isidro-Filho – Universidade de Brasília

Prof. Dr. Carlos Antonio de Souza Moraes – Universidade Federal Fluminense
Profª Drª Cristina Gaio – Universidade de Lisboa
Prof. Dr. Daniel Richard Sant’Ana – Universidade de Brasília
Prof. Dr. Deyvison de Lima Oliveira – Universidade Federal de Rondônia
Profª Drª Dilma Antunes Silva – Universidade Federal de São Paulo
Prof. Dr. Edvaldo Antunes de Farias – Universidade Estácio de Sá
Prof. Dr. Elson Ferreira Costa – Universidade do Estado do Pará
Prof. Dr. Eloi Martins Senhora – Universidade Federal de Roraima
Prof. Dr. Gustavo Henrique Cepolini Ferreira – Universidade Estadual de Montes Claros
Profª Drª Ivone Goulart Lopes – Istituto Internazionele delle Figlie de Maria Ausiliatrice
Prof. Dr. Jadson Correia de Oliveira – Universidade Católica do Salvador
Prof. Dr. Julio Candido de Meirelles Junior – Universidade Federal Fluminense
Profª Drª Lina Maria Gonçalves – Universidade Federal do Tocantins
Prof. Dr. Luis Ricardo Fernandes da Costa – Universidade Estadual de Montes Claros
Profª Drª Natiéli Piovesan – Instituto Federal do Rio Grande do Norte
Prof. Dr. Marcelo Pereira da Silva – Pontifícia Universidade Católica de Campinas
Profª Drª Maria Luzia da Silva Santana – Universidade Federal de Mato Grosso do Sul
Profª Drª Paola Andressa Scortegagna – Universidade Estadual de Ponta Grossa
Profª Drª Rita de Cássia da Silva Oliveira – Universidade Estadual de Ponta Grossa
Prof. Dr. Rui Maia Diamantino – Universidade Salvador
Prof. Dr. Urandi João Rodrigues Junior – Universidade Federal do Oeste do Pará
Profª Drª Vanessa Bordin Viera – Universidade Federal de Campina Grande
Prof. Dr. William Cleber Domingues Silva – Universidade Federal Rural do Rio de Janeiro
Prof. Dr. Willian Douglas Guilherme – Universidade Federal do Tocantins

Ciências Agrárias e Multidisciplinar

Prof. Dr. Alexandre Igor Azevedo Pereira – Instituto Federal Goiano
Profª Drª Carla Cristina Bauermann Brasil – Universidade Federal de Santa Maria
Prof. Dr. Antonio Pasqualetto – Pontifícia Universidade Católica de Goiás
Prof. Dr. Cleberton Correia Santos – Universidade Federal da Grande Dourados
Profª Drª Daiane Garabeli Trojan – Universidade Norte do Paraná
Profª Drª Diocléa Almeida Seabra Silva – Universidade Federal Rural da Amazônia
Prof. Dr. Écio Souza Diniz – Universidade Federal de Viçosa
Prof. Dr. Fábio Steiner – Universidade Estadual de Mato Grosso do Sul
Prof. Dr. Fágner Cavalcante Patrocínio dos Santos – Universidade Federal do Ceará
Profª Drª Girlene Santos de Souza – Universidade Federal do Recôncavo da Bahia
Prof. Dr. Jael Soares Batista – Universidade Federal Rural do Semi-Árido
Prof. Dr. Júlio César Ribeiro – Universidade Federal Rural do Rio de Janeiro
Profª Drª Lina Raquel Santos Araújo – Universidade Estadual do Ceará
Prof. Dr. Pedro Manuel Villa – Universidade Federal de Viçosa
Profª Drª Raissa Rachel Salustriano da Silva Matos – Universidade Federal do Maranhão
Prof. Dr. Ronilson Freitas de Souza – Universidade do Estado do Pará
Profª Drª Talita de Santos Matos – Universidade Federal Rural do Rio de Janeiro
Prof. Dr. Tiago da Silva Teófilo – Universidade Federal Rural do Semi-Árido
Prof. Dr. Valdemar Antonio Paffaro Junior – Universidade Federal de Alfenas

Ciências Biológicas e da Saúde

Prof. Dr. André Ribeiro da Silva – Universidade de Brasília
Prof^ª Dr^ª Anelise Levay Murari – Universidade Federal de Pelotas
Prof. Dr. Benedito Rodrigues da Silva Neto – Universidade Federal de Goiás
Prof^ª Dr^ª Débora Luana Ribeiro Pessoa – Universidade Federal do Maranhão
Prof. Dr. Douglas Siqueira de Almeida Chaves -Universidade Federal Rural do Rio de Janeiro
Prof. Dr. Edson da Silva – Universidade Federal dos Vales do Jequitinhonha e Mucuri
Prof^ª Dr^ª Eleuza Rodrigues Machado – Faculdade Anhanguera de Brasília
Prof^ª Dr^ª Elane Schwinden Prudêncio – Universidade Federal de Santa Catarina
Prof^ª Dr^ª Eysler Gonçalves Maia Brasil – Universidade da Integração Internacional da Lusofonia Afro-Brasileira
Prof. Dr. Ferlando Lima Santos – Universidade Federal do Recôncavo da Bahia
Prof^ª Dr^ª Gabriela Vieira do Amaral – Universidade de Vassouras
Prof. Dr. Gianfábio Pimentel Franco – Universidade Federal de Santa Maria
Prof. Dr. Helio Franklin Rodrigues de Almeida – Universidade Federal de Rondônia
Prof^ª Dr^ª Iara Lúcia Tescarollo – Universidade São Francisco
Prof. Dr. Igor Luiz Vieira de Lima Santos – Universidade Federal de Campina Grande
Prof. Dr. Jefferson Thiago Souza – Universidade Estadual do Ceará
Prof. Dr. Jesus Rodrigues Lemos – Universidade Federal do Piauí
Prof. Dr. Jônatas de França Barros – Universidade Federal do Rio Grande do Norte
Prof. Dr. José Max Barbosa de Oliveira Junior – Universidade Federal do Oeste do Pará
Prof. Dr. Luís Paulo Souza e Souza – Universidade Federal do Amazonas
Prof^ª Dr^ª Magnólia de Araújo Campos – Universidade Federal de Campina Grande
Prof. Dr. Marcus Fernando da Silva Praxedes – Universidade Federal do Recôncavo da Bahia
Prof^ª Dr^ª Mylena Andréa Oliveira Torres – Universidade Ceuma
Prof^ª Dr^ª Natiéli Piovesan – Instituto Federaci do Rio Grande do Norte
Prof. Dr. Paulo Inada – Universidade Estadual de Maringá
Prof. Dr. Rafael Henrique Silva – Hospital Universitário da Universidade Federal da Grande Dourados
Prof^ª Dr^ª Regiane Luz Carvalho – Centro Universitário das Faculdades Associadas de Ensino
Prof^ª Dr^ª Renata Mendes de Freitas – Universidade Federal de Juiz de Fora
Prof^ª Dr^ª Vanessa Lima Gonçalves – Universidade Estadual de Ponta Grossa
Prof^ª Dr^ª Vanessa Bordin Viera – Universidade Federal de Campina Grande

Ciências Exatas e da Terra e Engenharias

Prof. Dr. Adélio Alcino Sampaio Castro Machado – Universidade do Porto
Prof. Dr. Alexandre Leite dos Santos Silva – Universidade Federal do Piauí
Prof. Dr. Carlos Eduardo Sanches de Andrade – Universidade Federal de Goiás
Prof^ª Dr^ª Carmen Lúcia Voigt – Universidade Norte do Paraná
Prof. Dr. Douglas Gonçalves da Silva – Universidade Estadual do Sudoeste da Bahia
Prof. Dr. Eloi Rufato Junior – Universidade Tecnológica Federal do Paraná
Prof. Dr. Fabrício Menezes Ramos – Instituto Federal do Pará
Prof^ª Dra. Jéssica Verger Nardeli – Universidade Estadual Paulista Júlio de Mesquita Filho
Prof. Dr. Juliano Carlo Rufino de Freitas – Universidade Federal de Campina Grande
Prof^ª Dr^ª Luciana do Nascimento Mendes – Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Norte
Prof. Dr. Marcelo Marques – Universidade Estadual de Maringá

Profª Drª Neiva Maria de Almeida – Universidade Federal da Paraíba
Profª Drª Natiéli Piovesan – Instituto Federal do Rio Grande do Norte
Prof. Dr. Takeshy Tachizawa – Faculdade de Campo Limpo Paulista

Linguística, Letras e Artes

Profª Drª Adriana Demite Stephani – Universidade Federal do Tocantins
Profª Drª Angeli Rose do Nascimento – Universidade Federal do Estado do Rio de Janeiro
Profª Drª Carolina Fernandes da Silva Mandaji – Universidade Tecnológica Federal do Paraná
Profª Drª Denise Rocha – Universidade Federal do Ceará
Prof. Dr. Fabiano Tadeu Grazioli – Universidade Regional Integrada do Alto Uruguai e das Missões
Prof. Dr. Gilmei Fleck – Universidade Estadual do Oeste do Paraná
Profª Drª Keyla Christina Almeida Portela – Instituto Federal de Educação, Ciência e Tecnologia do Paraná
Profª Drª Miranilde Oliveira Neves – Instituto de Educação, Ciência e Tecnologia do Pará
Profª Drª Sandra Regina Gardacho Pietrobon – Universidade Estadual do Centro-Oeste
Profª Drª Sheila Marta Carregosa Rocha – Universidade do Estado da Bahia

Conselho Técnico Científico

Prof. Me. Abrãao Carvalho Nogueira – Universidade Federal do Espírito Santo
Prof. Me. Adalberto Zorzo – Centro Estadual de Educação Tecnológica Paula Souza
Prof. Me. Adalto Moreira Braz – Universidade Federal de Goiás
Prof. Dr. Adaylson Wagner Sousa de Vasconcelos – Ordem dos Advogados do Brasil/Seccional Paraíba
Prof. Dr. Adilson Tadeu Basquerote Silva – Universidade para o Desenvolvimento do Alto Vale do Itajaí
Prof. Me. Alexsandro Teixeira Ribeiro – Centro Universitário Internacional
Prof. Me. André Flávio Gonçalves Silva – Universidade Federal do Maranhão
Profª Ma. Anne Karynne da Silva Barbosa – Universidade Federal do Maranhão
Profª Drª Andrezza Lopes – Instituto de Pesquisa e Desenvolvimento Acadêmico
Profª Drª Andrezza Miguel da Silva – Faculdade da Amazônia
Prof. Dr. Antonio Hot Pereira de Faria – Polícia Militar de Minas Gerais
Prof. Me. Armando Dias Duarte – Universidade Federal de Pernambuco
Profª Ma. Bianca Camargo Martins – UniCesumar
Profª Ma. Carolina Shimomura Nanya – Universidade Federal de São Carlos
Prof. Me. Carlos Antônio dos Santos – Universidade Federal Rural do Rio de Janeiro
Prof. Ma. Cláudia de Araújo Marques – Faculdade de Música do Espírito Santo
Profª Drª Cláudia Taís Siqueira Cagliari – Centro Universitário Dinâmica das Cataratas
Prof. Me. Clécio Danilo Dias da Silva – Universidade Federal do Rio Grande do Norte
Prof. Me. Daniel da Silva Miranda – Universidade Federal do Pará
Profª Ma. Daniela da Silva Rodrigues – Universidade de Brasília
Profª Ma. Daniela Remião de Macedo – Universidade de Lisboa
Profª Ma. Dayane de Melo Barros – Universidade Federal de Pernambuco
Prof. Me. Douglas Santos Mezacas – Universidade Estadual de Goiás
Prof. Me. Edevaldo de Castro Monteiro – Embrapa Agrobiologia
Prof. Me. Eduardo Gomes de Oliveira – Faculdades Unificadas Doctum de Cataguases
Prof. Me. Eduardo Henrique Ferreira – Faculdade Pitágoras de Londrina

Prof. Dr. Edwaldo Costa – Marinha do Brasil
Prof. Me. Eliel Constantino da Silva – Universidade Estadual Paulista Júlio de Mesquita
Prof. Me. Ernane Rosa Martins – Instituto Federal de Educação, Ciência e Tecnologia de Goiás
Prof. Me. Euvaldo de Sousa Costa Junior – Prefeitura Municipal de São João do Piauí
Profª Ma. Fabiana Coelho Couto Rocha Corrêa – Centro Universitário Estácio Juiz de Fora
Prof. Dr. Fabiano Lemos Pereira – Prefeitura Municipal de Macaé
Prof. Me. Felipe da Costa Negrão – Universidade Federal do Amazonas
Profª Drª Germana Ponce de Leon Ramírez – Centro Universitário Adventista de São Paulo
Prof. Me. Gevair Campos – Instituto Mineiro de Agropecuária
Prof. Dr. Guilherme Renato Gomes – Universidade Norte do Paraná
Prof. Me. Gustavo Krahl – Universidade do Oeste de Santa Catarina
Prof. Me. Helton Rangel Coutinho Junior – Tribunal de Justiça do Estado do Rio de Janeiro
Profª Ma. Isabelle Cerqueira Sousa – Universidade de Fortaleza
Profª Ma. Jaqueline Oliveira Rezende – Universidade Federal de Uberlândia
Prof. Me. Javier Antonio Albornoz – University of Miami and Miami Dade College
Prof. Me. Jhonatan da Silva Lima – Universidade Federal do Pará
Prof. Dr. José Carlos da Silva Mendes – Instituto de Psicologia Cognitiva, Desenvolvimento Humano e Social
Prof. Me. Jose Elyton Batista dos Santos – Universidade Federal de Sergipe
Prof. Me. José Luiz Leonardo de Araujo Pimenta – Instituto Nacional de Investigación Agropecuaria Uruguay
Prof. Me. José Messias Ribeiro Júnior – Instituto Federal de Educação Tecnológica de Pernambuco
Profª Drª Juliana Santana de Curcio – Universidade Federal de Goiás
Profª Ma. Juliana Thaisa Rodrigues Pacheco – Universidade Estadual de Ponta Grossa
Profª Drª Kamilly Souza do Vale – Núcleo de Pesquisas Fenomenológicas/UFPA
Prof. Dr. Kárpio Márcio de Siqueira – Universidade do Estado da Bahia
Profª Drª Karina de Araújo Dias – Prefeitura Municipal de Florianópolis
Prof. Dr. Lázaro Castro Silva Nascimento – Laboratório de Fenomenologia & Subjetividade/UFPR
Prof. Me. Leonardo Tullio – Universidade Estadual de Ponta Grossa
Profª Ma. Lillian Coelho de Freitas – Instituto Federal do Pará
Profª Ma. Liliani Aparecida Sereno Fontes de Medeiros – Consórcio CEDERJ
Profª Drª Lívia do Carmo Silva – Universidade Federal de Goiás
Prof. Dr. Lucio Marques Vieira Souza – Secretaria de Estado da Educação, do Esporte e da Cultura de Sergipe
Prof. Me. Luis Henrique Almeida Castro – Universidade Federal da Grande Dourados
Prof. Dr. Luan Vinicius Bernardelli – Universidade Estadual do Paraná
Prof. Dr. Michel da Costa – Universidade Metropolitana de Santos
Prof. Dr. Marcelo Máximo Purificação – Fundação Integrada Municipal de Ensino Superior
Prof. Me. Marcos Aurelio Alves e Silva – Instituto Federal de Educação, Ciência e Tecnologia de São Paulo
Profª Ma. Maria Elanny Damasceno Silva – Universidade Federal do Ceará
Profª Ma. Marileila Marques Toledo – Universidade Federal dos Vales do Jequitinhonha e Mucuri
Prof. Me. Ricardo Sérgio da Silva – Universidade Federal de Pernambuco
Profª Ma. Renata Luciane Polsaque Young Blood – UniSecal

Prof. Me. Robson Lucas Soares da Silva – Universidade Federal da Paraíba
Prof. Me. Sebastião André Barbosa Junior – Universidade Federal Rural de Pernambuco
Profª Ma. Silene Ribeiro Miranda Barbosa – Consultoria Brasileira de Ensino, Pesquisa e Extensão
Profª Ma. Solange Aparecida de Souza Monteiro – Instituto Federal de São Paulo
Prof. Me. Tallys Newton Fernandes de Matos – Faculdade Regional Jaguaribana
Profª Ma. Thatianny Jasmine Castro Martins de Carvalho – Universidade Federal do Piauí
Prof. Me. Tiago Silvio Dedoné – Colégio ECEL Positivo
Prof. Dr. Welleson Feitosa Gazel – Universidade Paulista

Tecnologias, métodos e teorias na engenharia de computação

Editora Chefe: Profª Drª Antonella Carvalho de Oliveira
Bibliotecário Maurício Amormino Júnior
Diagramação: Karine de Lima Wisniewski
Edição de Arte: Luiza Alves Batista
Revisão: Os Autores
Organizador: Ernane Rosa Martins

Dados Internacionais de Catalogação na Publicação (CIP) (eDOC BRASIL, Belo Horizonte/MG)

T255 Tecnologias, métodos e teorias na engenharia de computação [recurso eletrônico] / Organizador Ernane Rosa Martins. – Ponta Grossa, PR: Atena, 2020.

Formato: PDF

Requisitos de sistema: Adobe Acrobat Reader

Modo de acesso: World Wide Web

Inclui bibliografia

ISBN 978-65-5706-361-3

DOI 10.22533/at.ed.613200409

1. Computação – Pesquisa – Brasil. 2. Tecnologia.
I. Martins, Ernane Rosa.

CDD 004

Elaborado por Maurício Amormino Júnior – CRB6/2422

Atena Editora

Ponta Grossa – Paraná – Brasil

Telefone: +55 (42) 3323-5493

www.atenaeditora.com.br

contato@atenaeditora.com.br

APRESENTAÇÃO

A Engenharia de Computação é a área que estuda as técnicas, métodos e ferramentas matemáticas, físicas e computacionais para o desenvolvimento de circuitos, dispositivos e sistemas. Esta área tem a matemática e a computação como seus principais pilares. O foco está no desenvolvimento de soluções que envolvam tanto aspectos relacionados ao software quanto à elétrica/eletrônica. O objetivo é a aplicação das tecnologias de computação na solução de problemas de Engenharia. Os profissionais desta área são capazes de atuar principalmente na integração entre software e hardware, tais como: automação industrial e residencial, sistemas embarcados, sistemas paralelos e distribuídos, arquitetura de computadores, robótica, comunicação de dados e processamento digital de sinais.

Dentro deste contexto, esta obra aborda os mais diversos aspectos tecnológicos computacionais, tais como: desenvolvimento de um método de verificação biométrica de indivíduos; uma abordagem para encontrar evidências de fraude aplicando técnicas de mineração de dados a bancos de dados públicos das licitações do governo federal brasileiro; o desenvolvimento de um método computacional para a classificação automática de melanomas; a aplicação de algoritmos recentes de aprendizagem de máquina, denominados XGBoost e Isolation Forest, para predição de irregularidades no consumo de energia elétrica; um modelo de receptor 5-HT_{2C} humano que foi criado através de modelagem por homologia e estudos de acoplamento molecular com os ligantes ácido fúlvico, paroxetina, citalopram e serotonina; a análise do uso do Controlador Lógico Programável (CLP), apresentando sua composição (estrutura, programação e linguagem Ladder), montagem, vantagens e desvantagens, exemplo de tipos e fabricantes; uma sugestão de melhoria das etapas de análise de negócios e engenharia de requisitos, por meio do uso de conceitos viáveis de metodologias ágeis; a construção de um aplicativo, denominado QEnade, para a disponibilização de questões do ENADE para os estudantes; uma síntese conceitual do PC voltada para âmbito educacional referente à educação básica brasileira; um sistema de localização híbrido capaz de usar diferentes tecnologias para fornecer a localização interna e externa de robôs ou de outros dispositivos móveis; um sistema de sumarização multidocumento de artigos de notícias escritos em português do Brasil; o emprego de duas técnicas de aprendizado de máquinas para prever se parte do público infantojuvenil da cidade de Monte Carmelo está suscetível a algum risco ou situação constrangedora nas redes sociais; a identificação das principais tecnologias que estão sendo utilizadas no contexto de Transformação Digital no cenário mundial; os elementos utilizados na construção de um sistema computacional, sem custo financeiro para a instituição e de fácil compreensão para o usuário, que utiliza os conhecimentos estatísticos para realizar a descrição, a apresentação e análise dos dados coletados; uma discussão acerca da confiabilidade das informações disseminadas na internet, para

entender os riscos e a importância da avaliação dos conteúdos encontrados no ambiente virtual; uma proposta de estratégia para a navegação de robôs semiautônomos baseada apenas em informações locais, obtidas pelos sensores instalados no robô e um planejador probabilístico que gera caminhos a serem seguidos localmente por ele, garantindo assim o desvio de obstáculos.

Sendo assim, esta obra é significativa por ser composta por uma gama de trabalhos pertinentes, que permitem aos seus leitores, analisar e discutir diversos assuntos importantes desta área. Por fim, desejamos aos autores, nossos mais sinceros agradecimentos pelas significativas contribuições, e aos nossos leitores, desejamos uma proveitosa leitura, repleta de boas reflexões.

Ernane Rosa Martins

SUMÁRIO

CAPÍTULO 1..... 1

BIOMETRIA PERIOCLAR USANDO TECNOLOGIA SMART APLICADA EM VISÃO DE ROBÔS

Victor Fagundes Stein Rosa
Alceu de Souza Britto Júnior
Dierone César Foltran Júnior
Ariangelo Hauer Dias

DOI 10.22533/at.ed.6132004091

CAPÍTULO 2..... 8

BRAZILIAN GOVERNMENT PROCUREMENTS: AN APPROACH TO FIND FRAUD TRACES IN COMPANIES RELATIONSHIPS

Rebeca Andrade Baldomir
Gustavo Cordeiro Galvão Van Erven
Célia Ghedini Ralha

DOI 10.22533/at.ed.6132004092

CAPÍTULO 3..... 20

CLASSIFICAÇÃO AUTOMÁTICA DE MELANOMAS USANDO DICIONÁRIOS VISUAIS PARA APOIO AO DIAGNÓSTICO CLÍNICO

Renata Francelino de Souza
Glauco Vitor Pedrosa

DOI 10.22533/at.ed.6132004093

CAPÍTULO 4..... 30

EMPLOYING GRADIENT BOOSTING AND ANOMALY DETECTION FOR PREDICTION OF FRAUDS IN ENERGY CONSUMPTION

Ricardo Nascimento dos Santos
Sami Yamouni
Beatriz Albiero
Estevão Uyrá
Ramon Vilarino
Juliano Andrade Silva
Tales Fonte Boa Souza
Renato Vicente

DOI 10.22533/at.ed.6132004094

CAPÍTULO 5..... 42

IN SILICO STUDY OF THE INTERACTION BETWEEN HUMAN 5-HT_{2C} RECEPTOR AND ANTIDEPRESSANT DRUG CANDIDATES

Rômulo Oliveira Barros
Jhonatan Matheus Sousa Costa
Wildrimak de Souza Pereira
Diego da Silva Mendes
Fábio Luis Cardoso Costa Júnior
Ricardo Martins Ramos

DOI 10.22533/at.ed.6132004095

CAPÍTULO 6.....	50
MODELO PARA DETERMINAR PERFIS DE DESEMPENHO ACADÊMICO NA UNNE COM MINERAÇÃO DE DADOS EDUCACIONAIS	
Julio César Acosta David Luis La Red Martínez	
DOI 10.22533/at.ed.6132004096	
CAPÍTULO 7.....	59
O USO DO CONTROLADOR LÓGICO PROGRAMÁVEL (CLP)	
Viviane Alencar Marques Araújo do Nascimento	
DOI 10.22533/at.ed.6132004097	
CAPÍTULO 8.....	72
PRÁTICAS ÁGEIS NA ELICITAÇÃO DE REQUISITOS PARA DESENVOLVIMENTO DE SOFTWARE EM UMA COOPERATIVA DE SAÚDE	
Mariangela Catelani Souza Bruno Cardoso Maciel José Alexandre Ducatti Paulo Sérgio Gaudêncio Mauro Leonardo Mendes de Souza Lygia Aparecida das Graças Gonçalves Corrêa Elizângela Cristina Begido Caldeira Bruna Grassetti Fonseca Patrícia Cristina de Oliveira Brito Cecconi Ana Paula Garrido de Queiroga Humberto Cecconi Carlos Alípio Caldeira	
DOI 10.22533/at.ed.6132004098	
CAPÍTULO 9.....	86
QENADE: APLICATIVO MÓVEL PARA PREPARAÇÃO DE ESTUDANTES PARA O ENADE	
Helder Guimarães Aragão	
DOI 10.22533/at.ed.6132004099	
CAPÍTULO 10.....	93
SÍNTESE DOS CONCEITOS DO PENSAMENTO COMPUTACIONAL VOLTADA PARA EDUCAÇÃO BÁSICA BRASILEIRA	
Nayara Poliana Massa	
DOI 10.22533/at.ed.61320040910	
CAPÍTULO 11.....	109
SISTEMA DE LOCALIZAÇÃO HÍBRIDO BASEADO EM NUVEM PARA AMBIENTES INTERNOS E EXTERNOS	
Raul de Queiroz Mendes Roberto Santos Inoue Tatiana de Figueiredo Pereira Alves Taveira Pazelli Rafael Vidal Aroca	
DOI 10.22533/at.ed.61320040911	

CAPÍTULO 12.....	131
SUMARIZAÇÃO AUTOMÁTICA DE ARTIGOS DE NOTÍCIAS EM PORTUGUÊS USANDO PROGRAMAÇÃO LINEAR INTEIRA E REGRESSÃO	
Hilário Tomaz Alves de Oliveira Laerth Bruno de Brito Gomes	
DOI 10.22533/at.ed.61320040912	
CAPÍTULO 13.....	144
TÉCNICAS DE APRENDIZADO DE MÁQUINA APLICADAS NA PREVISÃO DE VULNERABILIDADES QUANTO AO USO DA INTERNET PELO PÚBLICO INFANTOJUVENIL	
Franciele Cristina Espanhol Ferreira Alves Fernanda Maria da Cunha Santos	
DOI 10.22533/at.ed.61320040913	
CAPÍTULO 14.....	156
TECNOLOGIAS DISRUPTIVAS NO CONTEXTO DA TRANSFORMAÇÃO DIGITAL	
Rejane Maria da Costa Figueiredo Leonardo Sagmeister de Melo John Lenon Cardoso Gardenghi Ricardo Ajax Dias Kosloski	
DOI 10.22533/at.ed.61320040914	
CAPÍTULO 15.....	173
UM SISTEMA ESTATÍSTICO PARA APOIO AO ACOMPANHAMENTO DE DESEMPENHO ACADÊMICO	
Guilherme Álvaro Rodrigues Maia Esmeraldo Francisco Wilcley Lacerda de Lima Rennan Rodrigues Isídio Teles Francisca Alves de Souza Cícero Carlos Felix de Oliveira	
DOI 10.22533/at.ed.61320040915	
CAPÍTULO 16.....	186
UMA DISCUSSÃO ACERCA DA INTERNET: DESAFIOS PARA CONFIABILIDADE DA INFORMAÇÃO	
Breno Meirelles Costa Brito Passos Eli Shuab Carvalho Lima Bruno Soares Galdino Lívia Santos Lima Lemos	
DOI 10.22533/at.ed.61320040916	
CAPÍTULO 17.....	196
UMA ESTRATÉGIA PARA NAVEGAÇÃO DE ROBÔS DE SERVIÇO SEMIAUTÔNOMOS USANDO INFORMAÇÃO LOCAL E PLANEJADORES PROBABILÍSTICOS	
Elias José de Rezende Freitas Guilherme Augusto Silva Pereira	
DOI 10.22533/at.ed.61320040917	

SOBRE O ORGANIZADOR.....	210
ÍNDICE REMISSIVO.....	211

BIOMETRIA PERIOCLAR USANDO TECNOLOGIA SMART APLICADA EM VISÃO DE ROBÔS

Data de aceite: 27/08/2020

Victor Fagundes Stein Rosa

Universidade Estadual de Ponta Grossa -
UEPG

Alceu de Souza Britto Júnior

Universidade Estadual de Ponta Grossa -
UEPG
Departamento de Informática - DEINFO
ontificia Universidade Católica do Paraná -
PUC-PRP

Dierone César Foltran Júnior

Universidade Estadual de Ponta Grossa -
UEPG
Departamento de Informática - DEINFO

Ariangelo Hauer Dias

Universidade Estadual de Ponta Grossa -
UEPG
Departamento de Informática - DEINFO

RESUMO: Nesse projeto de pesquisa foi desenvolvido um método de verificação biométrica de indivíduos, baseado em características da região periocular, que consiste nas seguintes partes: olhos, as pálpebras, cílios, pele e parte das sobrancelhas. As características de texturas foram avaliadas através do método de variância local. Após os testes preliminares essa tecnologia foi embarcada em dispositivos móveis que são conectados a um servidor através do protocolo WAMP.

PALAVRAS-CHAVE: Extração de características, variância local, região periocular, reconhecimento de padrões.

ABSTRACT: In this research project, was developed a method of biometric verification of individuals based on characteristics of the periocular region, which consists of: eyes, eyelids, eyelashes, skin and part of the eyebrows. The texture characteristics were evaluated using the local variance method. After preliminary tests, this technology was shipped on mobile devices that are connected to a server using the WAMP protocol.

KEYWORDS: Feature extraction, local variance, periocular region, pattern recognition.

INTRODUÇÃO

O objetivo deste projeto de pesquisa foi de desenvolver de forma eficiente métodos de identificação biométrica de indivíduos baseados em características da região periocular. Após o desenvolvimento e testes o sistema foi embarcado em dispositivos móveis.

Nos dias atuais os *smartphones* tornaram-se parte fundamental da nossa rotina. Com o uso cada vez mais difundido dessa tecnologia é de máxima importância termos sistemas seguros e eficientes.

E a utilização de características biológicas para reconhecimento de pessoas torna-se uma ideia viável, pois cada pessoa possui características biométricas únicas. Características biológicas podem ser usadas como parâmetros para biometria desde que satisfaçam os seguintes princípios:

universalidade, todos devem possuir a característica, unicidade, a característica deve ser única para cada indivíduo, permanência, a característica deve ser constante com o passar do tempo, a coleta, a característica deve ser passível de mensuração e aceitabilidade, o recolhimento das características deve ser tolerado pelo indivíduo.

A região periocular é toda a região circundante os olhos. É composta por sobrancelha, íris, pupila, esclera, pálpebras e pele. Ela torna-se interessante para um sistema de verificação biométrico, pois, ao contrário de outras regiões do corpo, essa sofre menos desgaste com o envelhecimento, com marcas de expressão ou a presença de pelos faciais.

As características de textura foram avaliadas através do método de variância local. A variância local é método extrator de características baseado na textura das imagens. O método por ser expresso através da equação 1. As imagens foram analisadas em escala de cinza, tornado assim o processamento simplificado sem perdas de informação.

$$VAR(w) = \frac{1}{p} \sum_{p=0}^{p-1} (g_p - \mu) \quad (1)$$

Onde $\mu = \frac{1}{p} \sum_{p=0}^{p-1} g_p$. Tal método tem a vantagem de que o esforço computacional para seu cálculo é comparativamente menor que outros descritores de textura como *Local Binary Patterns*, *Local Phase Quantization* e *Gray Scale Co-occurrence Matrix*. O método pode também ser calculado em paralelo com o descritor *Local Binary Patterns*, tornando o esforço computacional ainda menor. Na figura 1 podemos visualizar uma imagem após o processamento da variância local.

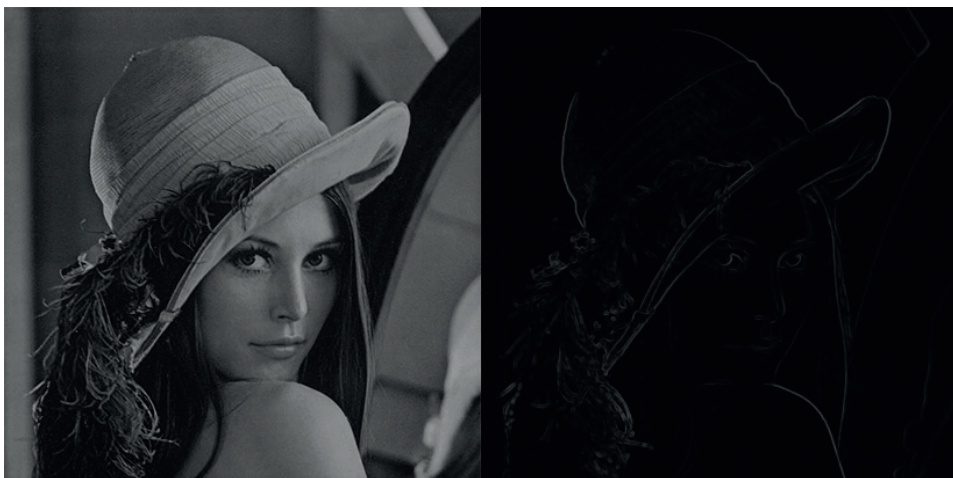


Figura 1 – Imagem original e imagem após processamento de variância local

Fonte: mostlymaths.net/2009/12/image-processing-history-lena.html

A textura de um segmento da imagem pode ser caracterizada a partir de um histograma. Esse histograma torna-se uma assinatura da imagem e com ele pode-se fazer verificações. Todo esse processamento gera um vetor característico que tem 255 posições, onde cada uma representa diferentes escalas de cinza. Na figura 2 podemos visualizar o histograma da imagem 1 após o processamento da variância local, nesse histograma foi plotado somente a escala de cinza de 0 – 50 para a melhor visualização.

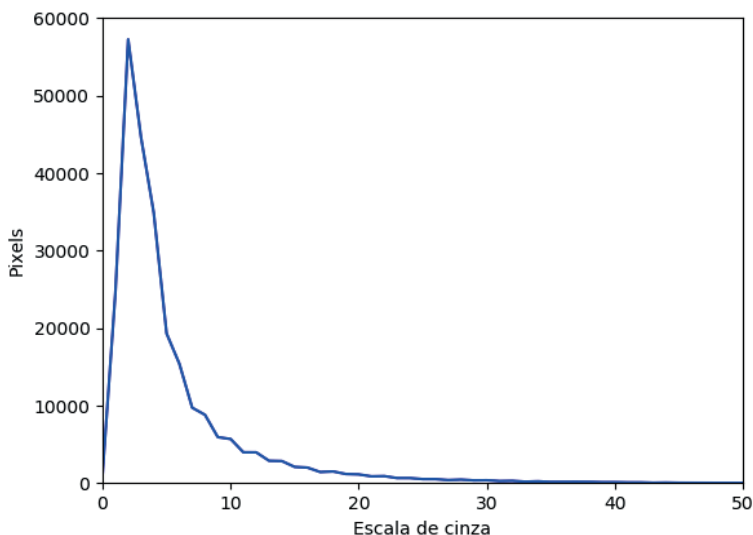


Figura 2 – Histograma da figura 1 após processamento

Fonte: O autor

Após gerado o vetor característico é necessário termos um método de comparação. Existem inúmeros métodos para se efetuar tal diferenciação, o utilizado nesse trabalho foi o da distância euclidiana. Tal método pode ser aplicado para vetores de n pontos, na equação 2 visualizamos essa solução para dois pontos.

$$distEuclid_{ab} = \sqrt{\sum_{j=1}^p (X_{aj} - X_{bj})^2} \quad (2)$$

Onde p varia de 1 até j , X_{aj} é o valor da variável j para o indivíduo a e X_{bj} é o valor da variável j para o indivíduo b .

MATERIAL E MÉTODOS

A solução apresentada neste projeto de pesquisa segue o modelo observado na figura 3.



Figura 3 – Esquema de funcionamento do sistema

Fonte: O autor

A imagem é obtida através de uma aplicação em um dispositivo móvel, figura 4 à direita, e então é enviada para um servidor, através de protocolo WAMP. Neste servidor foi feito todo o processamento da imagem, desde a obtenção da região periocular até o armazenamento dos vetores característicos.



Figura 4 – Algoritmo identificando face e região periocular e aplicativo desenvolvido para dispositivos móveis

Fonte: O autor

O reconhecimento da face é dado através de um *Haar Cascade*, um algoritmo de busca de objetos em imagens. Na figura 4 visualizamos o reconhecimento, no quadrado vermelho o algoritmo reconhece a face e no quadrado verde o mesmo reconhece a região periocular.

Após o envio da imagem para o servidor a mesma é convertida para tons de cinza

para o processamento da variância local. Na figura 5 observamos a imagem antes e após o processamento da variância local.

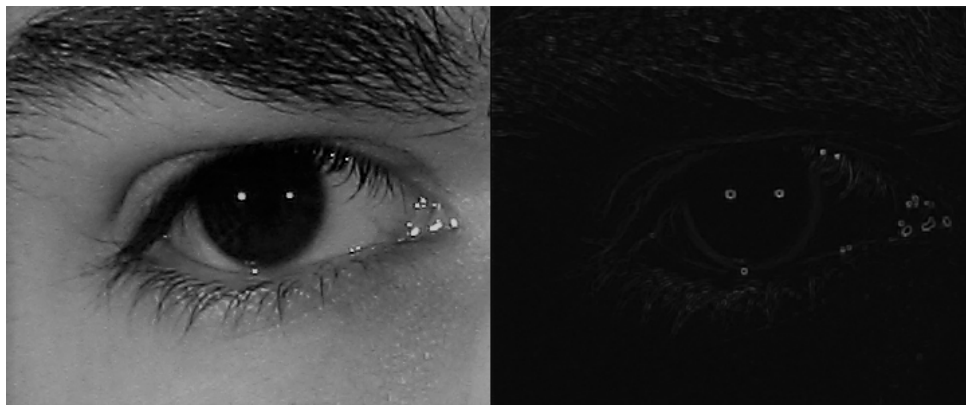


Figura 5 – Processamento da região periocular

Fonte: *Database UBIRIS V2*

Depois de gerado esse vetor, o mesmo foi armazenado no servidor em conjunto com um ID para cada indivíduo. O método utilizado para a comparação dos vetores característicos foi o da distância euclidiana, equação 2.

Após isso foi realizada a etapa de testes. Para isso foi utilizado uma parte do *dataset* UBIRIS v.2, que possui 261 indivíduos onde, 54,4% são do sexo masculino e 45,6% do Feminino, as imagens têm tamanho 400x300 pixels e são todas no formato *TIFF*. O formato *TIFF* possui a característica de armazenar imagens sem perda de qualidade.

A linguagem de programação escolhida para o projeto foi o Python, devido ao fato de ser uma linguagem muito utilizada atualmente no meio profissional e acadêmico além de possuir diversas bibliotecas úteis para o projeto.

Outra definição foi a escolha das bibliotecas que seriam usadas. Dentre as bibliotecas cita-se: OpenCV usada para o desenvolvimento de aplicativos na área de visão computacional. NumPy, a qual suporta *arrays* multidimensionais, possuindo diversas funções matemáticas para trabalhar com essas estruturas.

RESULTADOS E DISCUSSÃO

Na tabela 1 visualiza-se os resultados obtidos na etapa de teste. Foram realizados ao todo 396 (trezentos e noventa e seis) testes.

Testes	Resultado	%
Acertos	266	67,2
Falsos Positivos	88	22,2
Falsos Negativos	42	10,6

Tabela 1 – Resultados dos testes de verificação Biométrica

Fonte: O autor

O método apresentado possui taxa de acerto considerável. Porém os testes mostraram que o valor de *threshold*, comparado com outros métodos, é demasiado alto. Fazendo com que o sistema possua baixo índice de confiabilidade. Quanto menor a distância encontrada mais confiança o sistema tem para fazer a confirmação do indivíduo.

Na tabela 2, notamos os valores de distância euclidiana entre imagens de um mesmo indivíduo, dessa forma todas deveriam ser verificadas como corretos.

Imagem 1	Imagem 2	Distância Euclidiana	Verificação
11_1	11_2	1142,9	Verificado
11_1	11_3	4726,8	Verificado
11_1	11_4	3428,7	Verificado
11_1	11_5	1083,7	Verificado
11_1	11_6	10186,5	Falso Negativo
11_1	11_7	25198,7	Falso Negativo
11_1	11_8	24311,6	Falso Negativo
11_1	11_9	23856,7	Falso Negativo

Tabela 2 – Análise de Resultados do indivíduo 11

Fonte: O autor

Analisando a tabela percebemos que a diferença entre fotos de um mesmo indivíduo são muito altas, devido principalmente à distância da câmera e do rosto do indivíduo outro fator que influencia no resultado é a iluminação da imagem. Fazendo com que esse método se torne altamente dependente das condições locais em que o usuário capture sua foto.

CONCLUSÃO

Nesse projeto de pesquisa foi apresentada uma solução para a verificação biométrica em dispositivos móveis utilizando a variância local. Foi desenvolvida toda uma aplicação para dispositivos móveis para realizar a captura da imagem e o envio dela para um servidor. Nesse servidor é realizado todo processamento da imagem e armazenamento de vetores característicos de cada indivíduo.

Tal método mostra baixa robustez para ser embarcado em uma aplicação comercial. Para tal finalidade o método deve ser aperfeiçoado ainda mais.

Nesse trabalho a região periocular foi tratada separadamente. O tratamento de ambas as regiões pode surtir efeitos positivos no desempenho do sistema. Outra possibilidade de projeto futuro é a combinação de extratores de características para vários fatores de análise.

AGRADECIMENTOS

Agradeço aos professores Dierone César Foltran Júnior, Alceu de Souza Britto Júnior e Ariangelo Hauer Dias pela orientação do projeto, a Universidade Estadual de Ponta Grossa pela concessão de recursos e à Fundação Araucária pela bolsa concedida para realização deste projeto.

REFERÊNCIAS

CLARKE, R. “*Human identification in information systems: management challenges and public policy issues*”, Information Technology and People, v. 7, n. 4, pp. 6 – 37, 1994.

WOODARD, D., PUNDLIK, S., LYLE, J., et al. “*Periocular region appearance cues for biometric identification*”. In: Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on, pp. 162–169, 2010.

PARK, U., ROSS, A., JAIN, A. “*Periocular biometrics in the visible spectrum: A feasibility study*”. In: BTAS '09: Proceedings of the IEEE 3rd International Conference on Biometrics: Theory, Applications, and Systems, pp. 1–6, Sept 2009.

JUEFEI-XU, F., LUU, K., et al. “*Investigating age invariant face recognition based on periocular biometrics*”. In: Biometrics (IJCB), 2011 International Joint Conference on, pp. 1–7, 2011.

LYLE, J., MILLER, P., PUNDLIK, S., et al. “*Soft biometric classification using periocular region features*”. In: Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on, pp. 1–7, 2010.

BRAZILIAN GOVERNMENT PROCUREMENTS: AN APPROACH TO FIND FRAUD TRACES IN COMPANIES RELATIONSHIPS

Data de aceite: 27/08/2020

Data de submissão: 20/06/2020

Rebeca Andrade Baldomir

Universidade de Brasília, Departamento de
Ciência da Computação
Brasília – Distrito Federal
<http://lattes.cnpq.br/8920792435174654>

Gustavo Cordeiro Galvão Van Erven

Ministério da Transparência e Controladoria-
Geral da União
Brasília – Distrito Federal
<http://lattes.cnpq.br/9209495282522722>

Célia Ghedini Ralha

Universidade de Brasília, Departamento de
Ciência da Computação
Brasília – Distrito Federal
<http://lattes.cnpq.br/5632722847264046>

ABSTRACT: Data mining has been an area of high visibility in recent years and many researches have shown good efficiency in this area to find information in large databases. This paper presents an approach to find fraud traces applying data mining techniques to public databases of the Brazilian Federal Government bidings. The aim is to find evidence of fraud, such as stunts and cartels. The task of finding fraud evidences in large amount of data is complex for auditors since they have correlate data. The proposed approach was used to develop a prototype which has been used by auditors in the Ministry of Transparency and General Comptroller of the Union (CGU).

KEYWORDS: Stunts, fraud, public bidings, data mining.

LICITAÇÃO NO GOVERNO BRASILEIRO: UMA ABORDAGEM PARA ENCONTRAR EVIDÊNCIAS DE FRAUDE NO RELACIONAMENTO DE EMPRESAS

RESUMO: A mineração de dados tem sido uma área de alta visibilidade nos últimos anos e muitas pesquisas têm mostrado boa eficiência para encontrar informação em bancos de dados massivos. Este artigo apresenta uma abordagem para encontrar evidências de fraude aplicando técnicas de mineração de dados a bancos de dados públicos das licitações do governo federal brasileiro. O objetivo é encontrar evidências de fraude, como conluio e cartéis. A tarefa de encontrar evidências de fraude em grande quantidade de dados é complexa para os auditores considerando o volume de dados correlacionados. A abordagem proposta foi usada para desenvolver um protótipo que foi usado pelos auditores do Ministério da Transparência e Controladoria-Geral da União (CGU).

PALAVRAS-CHAVE: Conluio, fraude, licitação pública, mineração de dados.

1 | INTRODUCTION

In the context of procurement processes, Government Agencies use to order goods and services through some processes that might ensure a good (or the best) quality by fair price. In Brazil, there are several kind of the procurement process and

they are often controlled by an information system called SIASG – Integrated System for General Services Management in the Federal Government.

In the set of procurement processes, one of the most important is called Pregão [1], a reverse auction process. In this process, the suppliers present their initial proposals and then can start to give bids to lower the prices until the lowest offer remain and the other competitor give up.

Although this process encourages competition between companies, they can use some mechanisms to fraud it [2]. Companies can associate with each other and make collusion and cartels, that is, they will agree which one will win each bid item before the date of the auction and thus they can continue to alternate between them for the best contracts in each process.

Another unfair and common technique occurs when an individual, or a group of partners, own multiple companies from the same segment or industry. In this case, the companies are used to simulate competition between them, although the winner will always be a part of the same group of partners [3].

In case of collusion, fake companies or another type of fraud, the processes will be harmful, since the price can be manipulated to a very high value, or the products may come from a low quality brand.

A useful technique for finding these groups or cartels is called association rules. This technique makes it possible to search for relationships between items in a transaction [4]. In our context, items can be understood as a transaction and the companies are products of these transactions. Companies that usually show up together have more chance to be related as a collusion.

In this paper we propose an information system approach to search for possible irregularities in processes associated with a specific company, in order to improve the work performed by CGU auditors. The Apriori scope is pruned in database and the rules can execute online, improving the auditor's target analysis, and other systems can be integrated to make the rule more meaningful as well.

The rest of the paper is organized as follows: in Section 2 we present a short overview of background; in Section 3 we present related work; in Section 4 we present the developed prototype to search for collusion and the results gathered inside the Ministry of Transparency and General Comptroller of the Union¹ (CGU). Finally, the Section 5 presents conclusion and future work.

2 | BACKGROUND

This section presents the background required for a better contextualization of this work, including association rules, the Apriori algorithm and the application domain in the context of the procurement processes of the Brazilian Federal Government.

2.1 ASSOCIATION RULES

Association Rules is a data mining technique that aims to discover strong relationships between data, finding patterns in associated data. There are two important metrics that show whether the relationship found by the association rule is strong:

1. Confidence: represents the percentage of times that the rule appeared correctly. The higher the confidence, better is the quality of the rule.
2. Support: represents the chance of the rule appearing in the analyzed dataset.

Another very important variable in the association rules is lift, an affinity measure that tells whether the occurrence of itemset A is independent of the occurrence of itemset B if $P(A \cup B) = P(A) P(B)$. Otherwise, itemset A and B are dependent and correlated. This definition can easily be adjusted to more than two itemsets. The relation between the occurrence of A and B can be measured by Equation 1.

$$\text{lift}(A,B) = P(A \cup B) / P(A).P(B) \quad (1)$$

If the result of the equation is less than 1, the itemsets A and B are negatively correlated, in other words, an occurrence of one leads to an absence of the other. However, if the result is greater than 1, they are positively correlated, which means that the occurrence of one side leads to the occurrence of the other. If the value is exactly 1, the values are independent and there is no relationship between them [5].

2.2 APRIORI ALGORITHM

In order to find strong associations between data, the Apriori algorithm was proposed [6]. Apriori is an algorithm that extracts high trust rules finding relationships between the data. This algorithm uses an iterative approach consisting of k-itemsets to be used to explore (k+1)-itemsets. For each set it is necessary to check the presence of each item and collect the items that satisfy the minimum support. The resulting set is denoted by L_1 . Then L_1 is used to find L_2 which is used to find L_3 , and so on, until no more frequent k-itemsets are found. The discovery of each L_k requires a complete check of the database.

Table 1 illustrates 10 transactions that will be used to exemplify the Apriori algorithm. Initially, sets are created with each of the transaction items and their support calculated (Table 2).

The next set will be consist of creating all possible combinations of sets containing 2 transactions, and discarding the ones with a support value below the desired one represented by red lines in Table 2. Using a 20% support, all sets will be combined generating new sets. This process is repeated until there are no more sets to combine.

TRANSACTION	ITEM 1	ITEM 2	ITEM 3
1	1	1	0
2	0	1	1
3	0	1	0
4	1	1	1
5	1	0	0
6	0	1	0
7	1	0	0
8	1	1	0
9	1	1	0
10	0	0	0

Table 1. Transactions used to exemplify the Apriori algorithm.

ITERATION 1	
SET	SUPPORT
1	0.6
2	0.7
3	0.2
ITERATION 2	
1.2	0.4
1.3	0.1
2.3	0.2

Table 2. Support of each set using Apriori algorithm.

2.3 PROCUREMENT PROCESS

Procurement processes are the way the Government purchases products or services. It is the form that ensures full competition among participants seeking to ensure compliance with the constitutional principle of isonomy, where the selection of the most advantageous proposal for the administration is ensured [7]. There are several procurement processes modalities, with the Pregão being the most known one [1]. Procurement processes are a major target for corruption due to being linked to the financial system, besides containing flaws in its process. Therefore, auditing has been an important process in the search and prevention of irregularities in Brazilian procurement processes.

3 I RELATED WORK

Nowadays, in Brazil, corruption is a central subject due its economics and politics situation. According to the Organization for Economic Cooperation and Development (OECD) report² of 2016, public procurements are extremely vulnerable to corruption. Most of 50% of the foreign bribery cases aim public procurements contracts. Therefore, several works have been proposed addressing the corruption problem. Table 3 summarizes the papers found and described in this section.

REFERENCE	ASSOCIATION RULES	OTHER ML TECHNOLOGY	OTHER APPROACH	ONLINE SYSTEM	DOMAIN
Present work	X			X	procurement
Ralha & Silva (2012) [4]	X	X			procurement
Nguyen et al.(2017) [8]		X			supply chain
Hein et al. (2014) [9]		X			economic sector
De Padua et al. (2016) [10]	X	X			transaction
Minovic et al. (2014) [11]			X	X	procurement
Erven et al. (2018) [12]			X		procurement

Table 3. Related Work compared to the present work.

Ralha and Silva [4] propose a solution for cartel detection using data from ComprasNet, a Brazilian procurements database, using data mining alongside multiagent systems. The AGMI (AGent-MIning tool) is composed by autonomous agents arranged in three layers that deliberate about the data mining techniques and parameters to discover knowledge about suspicious groups of companies. The system generates several association rules with scores built from the business domain to identify which them have high quality.

Data mining is also used in Nguyen et al. [8], where the state-of-art of data mining in supply chain management is presented and in Nelson Hein et al. [9], where factorial analysis is used to select social performance indexes of the most representative companies in the cyclic consumption sector.

De Padua et al. [10] uses community detection between transactions' items to identify clusters to split the database among them, and reach better rules, since several of them do not usually implies in a good relation. This happens since rules with support and confidence higher would be obvious.

Minovic et al. [11] proposes introducing semantic technologies to procurement processes in Serbia to enable processing data by machines. In this model, the specialists would be able to define business rules related to suspicious situations being alerted when one of them occurs. Thus, a final application would previously recognize acquisitions that are potentially harmful and take actions to avoid them.

Erven et al. [12] presents a proposal of graph database model to identify relationships between companies that participate in Brazilian public procurement processes. Partners relationships and other data are inserted into a graph database and the shortest paths are found between companies that attend to the same acquisition process. A notation is proposed to model de database and it is compared against a relational database. At least, several queries is performed looking for insight that would be fraud evidences.

4 I ASSOCIATION RULES IN PROCUREMENT PROCESSES

A prototype was developed in order to assist the CGU auditors to find fraud in the relationship between companies participating in procurement processes. To achieve this goal, association rules between data were applied through the Apriori algorithm.

4.1 METHODOLOGY

The model presented in Figure 1 shows how the data mining phases were used in this work. The first step is the Business Understanding where you will have adequate understanding of the problem that needs to be solved. The Data Understanding is the phase where you will understand the data and leave them in the correct way to use. This work will be focused on the next steps of the model that are Data Preparation, Modeling and Evaluation. The Deployment step was done at the CGU where auditors were able to use and evaluate the application.



Figure 1. Data Mining Phases.

Data Preparation - the step where the interesting data was chosen from all available data in ComprasNet, a web site set up by Ministério do Planejamento, Desenvolvimento e Gestão (MPOG) in order to provide information on the procurement processes and contracts promoted by the Federal Government to society. Of all existing data available, only 6 variables were considered relevant for this project: the purchase identifier, the CNPJ of the company that participated in the procurement process, the reference date, the value of the purchase and the modality of the procurement process. These columns were chosen because they are the only ones required to identify relationships between companies.

Modeling - Data mining was done using the pre-processed data already available in the database. The algorithm used in this work was Apriori which will be detailed in 4.2.

Evaluation - Having discovered the patterns of relationship between companies through the association rules, they will be presented to CGU auditors through a web interface. Therefore, the interpretation and evaluation of this information will be done by CGU's own auditors.

4.2 PROTOTYPE

The prototype was developed using two programming languages, Python and R. This decision was made in order to integrate this tool with other tools already deployed in CGU which were also developed in Python. It would facilitate integration if this prototype was also developed in Python. The R language was used for running the Apriori algorithm, using an existing R package. In order to integrate both languages, the rpy2 library was used, enabling the execution of R code from the Python application. The prototype was subdivided into three main components: Web, Data and Apriori according to the Figure 2.

Web - the Web component is responsible for receiving a certain CNPJ through Hypertext Transfer Protocol (HTTP) requests and sending this CNPJ to be processed. Once the processing is finished, this component shows the results in the Hypertext Markup Language (HTML) format. This component was developed using the Python language. CherryPy was chosen to build it, as it is a framework that allows the creation of web applications in the same way as any other object-oriented Python program. This results in smaller source code developed in less time.

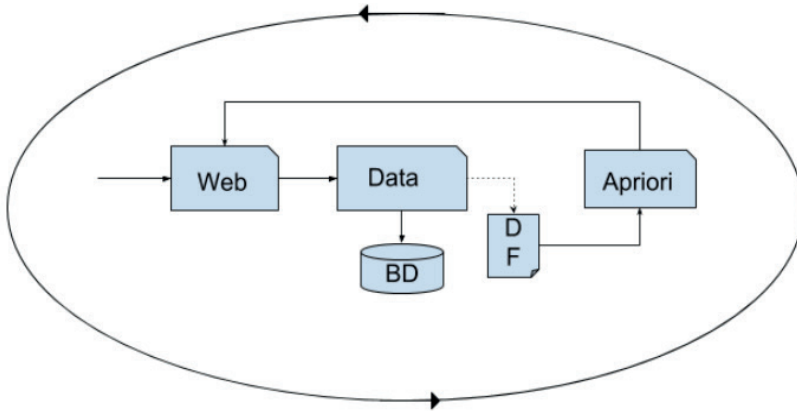


Figure 2. Prototype components.

Data - this component is responsible for accessing the database and returning a list with CNPJs that participated in a purchase alongside the informed CNPJ in the Web component. The component makes this list available

to the Apriori component through a data structure called dataframe. The Data component was also developed in Python. The database accessed by this component was modeled using the MySQL Workbench tool Community⁶ in version 6.3.9. The created database table contains a column for each column present in the data extracted on the pre-processing step.

Apriori - association rules were applied in the procurement data available at ComprasNet in order to identify the possible correlations between the companies. The algorithm used to apply these association rules was Apriori, which recognizes frequent items. For an item to be considered frequent, it must appear in the transactions in at least an amount equal to the minimum support. The set of items analyzed by the algorithm will be the CNPJs returned by the Data component. The Apriori algorithm was applied using an implementation in the R language. To exemplify the application of Apriori in the procurement process data, CNPJ 007XXXXXXXXXX was used. One of the rules resulting from the application was the following:

```

[1]
1      039XXXXXXXXXXXX
2      ->
3      676XXXXXXXXXXXX
4      0.3333333
5      1
  
```

Notice that in this rule, specifically in Lines 1 and 3, that the CNPJs related by the rule are presented. In Line 4, the support is presented and in Line 5, the confidence used in the application of the rule is presented. It can be concluded from this rule that in 33% of the purchases made between CNPJs 007XXXXXXXXXX and 039XXXXXXXXXX CNPJ 676XXXXXXXXXX was also contained in the set of buyers.

4.3 RESULTS

The prototype's interface is shown in Figure 3. Note that it contains the CNPJ of the companies participating in the rule, the number of victories of each company participating in the procurement process, the lift, confidence and support used in the Apriori algorithm.

The search field can be used to search for the CNPJ present on the left side of the rules generated by the algorithm. The column Company and the column Participated with are the companies that participated in the procurement process with the CNPJ searched and appear on the right side of the rules. There are no limits regarding the number of companies that can participate in the bids together with the CNPJ searched.

Empresa	Vitórias	Participou com	Vitórias	Suporte	Confiança	Lift
020	0	676	0	0.25	1.0	4.0
676	0	020	0	0.25	1.0	4.0
157	1	209	1	0.25	1.0	4.0

Figure 3. Prototype interface.

Initially, the prototype was executed locally with access to a base of 4,482,006 records (Table 4). Afterwards, the first tests in CGU were executed in order to collect feedbacks of the prototype's execution. Today, the developed prototype is executing in CGU with access to a base of 123,940,403 records that correspond to 20 years of procurement process data from 1997 to 2017.

EXECUTION ENVIRONMENT	NUMBER OF RECORDS
LOCAL	4.482.006
FIRST TESTS AT CGU	37.522.993
REAL ENVIRONMENT CGU	123.940.403

Table 4. Number of records in relation to the execution environment.

An interesting use case found for the prototype was to use it in conjunction with other existing tools on CGU. Particularly, the use of the prototype alongside Yggdrasil, a CGU tool that searches for links or relationships between CNPJs, made it possible to find a link between associates from two different companies who were participating in a procurement process together, which could be an indicative of fraud. The prototype built in this study was used to filter possible CNPJs in order to limit the search space, given that there are about 500 million CNPJs that could have a link. This way, the interesting CNPJs were narrowed down to those who appeared together in an association rule produced by the prototype, which made the search much easier.

Figure 4 illustrates the link found by Yggdrasil. The circles in red are companies' CNPJs and the link is represented by an edge between two vertices on the graph.

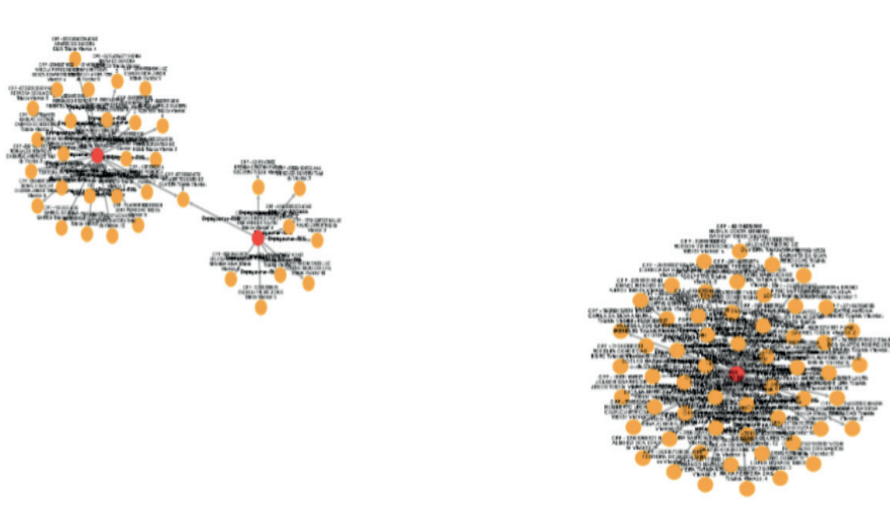


Figure 4. Link between two companies found in Yggdrasil system.

5 | CONCLUSION

The CGU has a responsibility to fight corruption and for that, it is necessary to consider frauds such as stunts and cartels. This work proposes the use of data mining in procurement process data in order to find relationships between the participating companies.

In order to discover the relationships between companies in the Federal Government bidding processes, CGU's auditors would have to manually analyze all procurement process data available on the ComprasNet, which makes this task unfeasible. Thus, this work was developed in order to assist the auditors work on the discovery of relationships between companies participating in procurement processes.

The auditor familia with such relationships can check if there are other indications of a fraudulent relationship using another system already built by CGU that verifies links between companies.

The prototype developed during this study allows for the scope of analysis to be broadened regarding a certain CNPJ, and can be complemented by the analysis of links between companies. Finding all the links, or smaller paths, coupled with a base of over 500 million relationships would be computationally costly and only a few relationships would really be relevant. However, in conjunction with the prototype, it may indicate which sets of companies are most interesting to analyse.

ACKNOWLEDGMENT

The authors would like to thank the CGU for support. C. G. Ralha would like to thank the Brazilian National Council for Scientific and Technological Development (CNPq) for the research grant (process 303863/2015-3).

REFERENCES

- [1] L. A. JOIA and F. ZAMOT, **Internet-based reverse auctions by the brazilian government**, The Electronic Journal of Information Systems in Developing Countries, vol. 9, no. 1, pp. 1–12, 2002.
- [2] S. ROSE-ACKERMAN and B. J. PALIFKA, **Corruption and government: Causes, consequences, and reform**. Cambridge University Press, 2016.
- [3] G. C. van ERVEN, M. HOLANDA, and R. N. CARVALHO, **Detecting evidence of fraud in the brazilian government using graph databases**, in World Conference on Information Systems and Technologies. Springer, 2017, pp. 464-473.
- [4] C. G. RALHA and C. V. S. SILVA, **A multi-agent data mining system for cartel detection in brazilian government procurement**, Expert Systems with Applications, vol. 39, no. 14, pp. 11 642–11 656, 2012.
- [5] J. HAN and M. KAMBER, **Data Mining: Concepts and Techniques**. Morgan Kaufmann Publishers, Inc, 2005.
- [6] R. AGRAWAL and R. SRIKANT, **Fast algorithms for mining association rules in large databases**, in Proceedings of the 20th International Conference on Very Large Data Bases, ser. VLDB '94. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1994, pp. 487–499.
- [7] BRASIL, **Lei no 8.666, de 21 de junho de 1993**, disponível em: [http://www.planalto.gov.br/ccivil03/leis=L8666cons.html](http://www.planalto.gov.br/ccivil03/leis/L8666cons.html). Acessado em: 07/11/2017.
- [8] T. NGUYEN, Z. LI, V. SPIEGLER, P. IEROMONACHOU, and Y. LIN, **Big data analytics in supply chain management: A state-of-the-art literature review**, Computers & Operations Research, 2017.

[9] N. HEIN and F. KREUZBER, **Aplicação da análise fatorial como ferramenta de data mining no desempenho social das empresas do setor de consumo cíclico listadas na bmf bovespa**, 2014.

[10] R. de PADUA, E. L. S. JUNIOR, L. P. do CARMO, V. O. de CARVALHO, and S. O. REZENDE, **Preprocessing data sets for association rules using community detection and clustering: a comparative study**.

[11] M. MIROSLAV, M. MILOS, Š. VELIMIR, D. BOZO, and L. ORD, **Semantic technologies on the mission: Preventing corruption in public procurement**, Computers in industry, vol. 65, no. 5, pp. 878–890, 2014.

[12] G. Van ERVEN, W. SILVA, R. CARVALHO, and M. HOLANDA, **Graphed: A graph description diagram for graph databases**, in World Conference on Information Systems and Technologies. Springer, 2018, pp. 1141–1151.

CLASSIFICAÇÃO AUTOMÁTICA DE MELANOMAS USANDO DICIONÁRIOS VISUAIS PARA APOIO AO DIAGNÓSTICO CLÍNICO

Data de aceite: 27/08/2020

Data de submissão: 16/06/2020

Renata Francelino de Souza

Universidade de Brasília (UnB), Faculdade do Gama (FGA)
Brasília, DF
<http://lattes.cnpq.br/3798771694444645>

Glauco Vitor Pedrosa

Universidade de Brasília (UnB), Faculdade do Gama (FGA)
Brasília, DF
<http://lattes.cnpq.br/9976045320641172>

RESUMO: O diagnóstico de doença de pele com o auxílio de imagens é uma maneira não invasiva de detectar a gravidade de uma lesão de pele. Este trabalho apresenta o desenvolvimento de um método computacional para a classificação automática de melanomas. A técnica conhecida como Bag-of-Features (BoF) foi investigada para extrair as características das imagens. Para isso, foi realizada uma extensa avaliação experimental usando diferentes parâmetros na construção dessa abordagem de modelagem. As combinações deste método foram feitas com vários classificadores e comparando tamanhos diferentes para o dicionário visual de palavras para a abordagem BoF. Foi obtido uma precisão de 77,6% de acurácia com um dicionário visual de 20 palavras visuais. De fato, os resultados experimentais mostram que a abordagem BoF é um descritor eficaz e eficiente no desenvolvimento de um sistema computacional para o auxílio na classificação

de lesões cutâneas.

PALAVRAS-CHAVE: Visão Computacional; Diagnóstico Assistido por Computador; Aprendizado de Máquinas; Câncer de Pele; Melanomas.

AUTOMATIC CLASSIFICATION OF MELANOMAS USING VISUAL DICTIONARIES TO SUPPORT CLINICAL DIAGNOSIS

ABSTRACT: The diagnosis of skin disease with the aid of images is a non-invasive way to detect the severity of a skin lesion. This work presents the development of a computational method for the automatic classification of melanomas. The technique known as Bag-of-Features (BoF) was investigated to extract the characteristics of the images. For this, an extensive experimental evaluation was carried out using different parameters in the construction of this modeling approach. The combinations of this method were made with several classifiers and comparing different sizes for the visual word dictionary for the BoF approach. A 77,6% accuracy was obtained with a visual dictionary of 20 visual words. In fact, the experimental results show that the BoF approach is an effective and efficient descriptor in the development of a computer system to aid in the classification of skin lesions.

KEYWORDS: Computer Vision; Computer Aided-Diagnosis; Machine Learning; Skin Cancer; Melanomas.

1 | INTRODUÇÃO

A incidência de câncer de pele tem aumentado nas últimas décadas. De acordo com o Instituto de Câncer INCA, o câncer de pele é, atualmente, o tipo mais comum no Brasil e no mundo, e ocorre com maior frequência em pessoas com pele branca. Dados da Organização Mundial da Saúde em 2019 apresentam que 1 em cada 3 casos de cânceres são câncer de pele.

O tempo em que a doença é diagnosticada faz diferença. De acordo com a Organização Mundial da Saúde, quando a doença é detectada precocemente, o paciente com melanoma tem 99% de chances de viver cinco anos a mais do que o tempo previsto. Além disso, um paciente que realiza a biópsia e inicia o tratamento após 30 a 59 dias tem 5% mais risco de morte em relação ao paciente que se trata em até 30 dias. Os pacientes que são tratados a partir de 119 dias após a realização da biópsia tem 41% mais risco de falecimento em relação àqueles que se trata dentro de um mês. Quanto maior o tempo para iniciar o tratamento para melanoma, maior o risco de óbito.

Várias pesquisas estão sendo desenvolvidas para melhorar os resultados obtidos com auxílio de sistemas de Diagnóstico Assistido por Computador (em inglês, *Computer Aided-Diagnosis*) na detecção de melanomas (ALFED, 2016). Isso só é possível porque os tipos de câncer de pele (benigno e maligno) possuem características visuais que possibilitam o uso de imagens como recurso para detectar e classificar a doença por meio de um sistema computacional.

O melanoma, em geral, apresenta um contorno assimétrico, com pigmento mais escuro, formato irregular, tom da cor não uniforme e diâmetro maior do que o câncer benigno. Esses aspectos visuais são denominados como Regras de Identificação ABCDE (Assimetria, Borda, Cor, Diâmetro, Evolução), de acordo com (NAMOZOV, 2018) e (SITU, 2018). A Figura 1 mostra um exemplo de uma lesão de pele diagnosticada como melanoma e outro como não-melanoma.

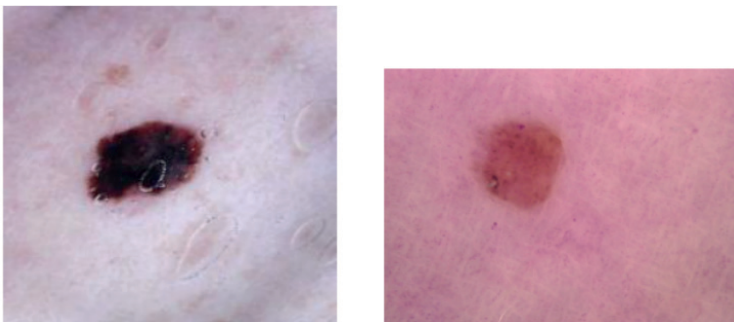


Figura 1. Exemplo de lesão de pele diagnosticada como melanoma (esquerda) e não melanoma (direita).

Os sistemas de Diagnóstico Auxiliado por Computador (CAD) são utilizados na área médica para apoiar a tomada de decisão de médicos diante de determinada doença com o uso de imagem. Esses sistemas não substituem o diagnóstico de um médico, porém contribuem para uma decisão mais assertiva do profissional (HALALLI, 2017).

Pesquisas buscam aliar o poder da computação à área de diagnósticos não invasivos de melanomas para melhorar a taxa de diagnósticos precoces. Os sistemas de análise de imagens por computador facilitam a detecção precoce dos melanomas e podem diminuir o número de biópsias realizadas em lesões benignas para descartar melanoma. De fato, técnicas de processamento de imagens vêm sendo utilizadas para a detecção e classificação de melanomas (ALFED, 2016). Por esse motivo, tem se tornado um método para o auxílio ao diagnóstico de câncer de pele, por se tratar de uma detecção não invasiva, ou seja, sem necessidade de extrair tecido da pele.

Este trabalho visa a apresentar o desenvolvimento de um sistema baseado em visão computacional que seja capaz de classificar uma doença de câncer de pele em melanoma ou benigno. Para tanto, foi investigado o uso da abordagem Bag-of-Features (BoF) para extração de características de imagens de lesão de pele. A proposta é investigar diferentes parâmetros da abordagem BoF e comparar seu desempenho para classificação de lesões de pele em melanomas ou não-melanomas. Os testes foram realizados usando uma base de imagens pública, composta por 1000 imagens de melanomas, sendo 500 imagens de melanomas e 500 de cânceres de pele benignos.

O texto deste trabalho está organizado da seguinte forma: a Seção 2 brevemente analisa os trabalhos relacionados e discute os detalhes da abordagem BoF para descrição de imagens digitais. Na Seção 3, mostra-se os resultados experimentais conduzidos e analisados e, finalmente, a Seção 4 apresenta as considerações finais deste trabalho.

2 | CONCEITOS E TRABALHOS CORRELATOS

Uma maneira não invasiva para detectar um câncer de pele é por meio da observação de aspectos visuais da lesão de pele como: Assimetria, Bordas, Cor e Diâmetro (ABCD). O melanoma caracteriza-se por ter um formato assimétrico, bordas irregulares, geralmente possui variação de coloração na mancha e o tamanho é maior que uma pinta comum. Além disso, outro aspecto visual que pode ser considerado é a Evolução da lesão na pele: se a mancha aumenta rapidamente de tamanho é importante ficar atento.

Ao longo da última década, o uso de sistemas de computação visual na detecção e diagnóstico de melanoma tem aumentado como um auxílio ao diagnóstico clínico de melanomas. Os trabalhos propostos para essa finalidade têm em comum a aplicação das seguintes fases em seus sistemas: aquisição de imagem, processamento e análise da imagem (pré-processamento e segmentação), extração e seleção de características e por fim, aplicação de métodos de classificação.

A etapa de extração de característica se refere ao procedimento de extrair e representar as características visuais de uma imagem em um vetor-de-características. Esse vetor-de-característica é o elemento que será usado nas atividades de classificação e/ou recuperação de imagens pelo sistema baseado em visão computacional. Por isso, é importante que o vetor represente de maneira sucinta e o mais fidedigno possível as características mais importantes da imagem.

A abordagem Bag-of-Features (BoF), também denominada de Bag-of-Visual-Words ou Bag-of-Keypoints, se tornou uma abordagem bastante popular para extração de características em imagens digitais (PEDROSA, 2015). A metodologia BoF é baseada na representação não-ordenada de descritores locais extraídos de uma imagem e é, portanto, conceitualmente e computacionalmente mais simples do que muitos métodos como Deep Learning. O processo para representar uma imagem em BoF está esquematizado pela Figura 2 e pode ser definido pelas seguintes fases: (i) detecção e representação das regiões-de-interesse; (ii) atribuição de cada região-de-interesse a uma palavra-visual de acordo com um dicionário (vocabulário) de palavras-visuais pré-definido; (iii) contagem da ocorrência (frequência) de cada palavra-visual contida na imagem.

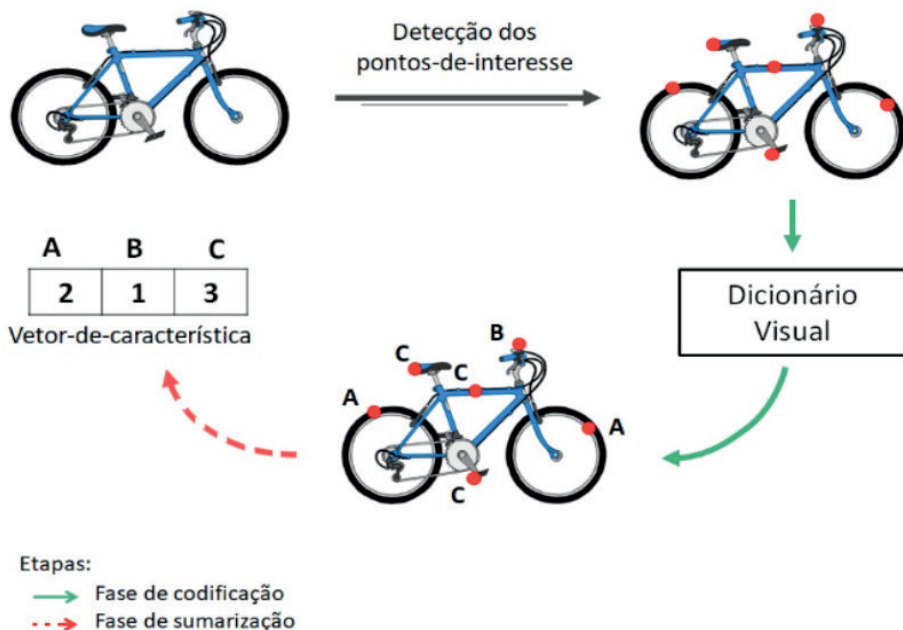


Figura 2. Passos para representar uma imagem em Bag-of-Features (BoF).

Vários trabalhos ao longo das últimas décadas utilizaram o descritor BoF para descrever as características visuais de uma lesão de pele. O trabalho de (BARATA, 2014)

apresenta uma comparação entre um sistema baseado em características visuais globais e outro sistema usando a abordagem Bag-of-Features. O desempenho de cada sistema foi analisado e ambos obtiveram resultados similares e satisfatórios, porém o sistema baseado no método Bag-of-Features obteve uma vantagem por ter menor custo computacional durante o processo de classificação da base de imagens. O trabalho de (SABBAGHI, 2016) combinou o Bag-of-Features para descrever características de cor e utilizou um algoritmo de Deep Learning para classificação. A acurácia atingida foi de 95%. O estudo de (BARATA, 2017) comparou a eficiência entre os descritores locais Bag-of-Features e Sparse Coding com a utilização de diversos tipos de histogramas. O resultado encontrado pelos autores evidenciou uma melhor precisão na classificação de melanoma por meio do uso do Sparse Coding, com 85,5% de sensibilidade. O trabalho de (ALFED, 2017) combinou o Histograma de Gradiente (HG), Histograma de Linhas (HL) e os momentos de Zernike na descrição de características. O sistema proposto atingiu 98,79% de acurácia, superior a outros índices encontrados na literatura.

A geração do dicionário de palavras-visuais é uma tarefa fundamental para a abordagem BoF. A estratégia para a geração desse dicionário está esquematizada na Figura 3 e é definida pelos seguintes passos: primeiramente um subconjunto de imagens do banco de dados é escolhido; para cada imagem, suas regiões-de-interesse são detectadas e descritas utilizando algum descritor gerando vetores-de-características; por fim, é realizado um agrupamento dos dados desse espaço de características utilizando algum algoritmo de agrupamento. O centroide de cada grupo é considerado uma palavra-visual do dicionário. O algoritmo de agrupamento e a quantidade de palavras-visuais são fatores que interferem diretamente na qualidade e no desempenho da geração do dicionário de palavras-visuais. O agrupamento pode ser realizado, por exemplo, através do algoritmo k-means, que é um método de agrupamento simples e bastante utilizado na literatura. Porém, os trabalhos na literatura tendem a escolher empiricamente a quantidade de palavras visuais do dicionário representadas pelo k-clusters obtidos através da execução do k-means. Esse é o fato que será investigado neste trabalho.

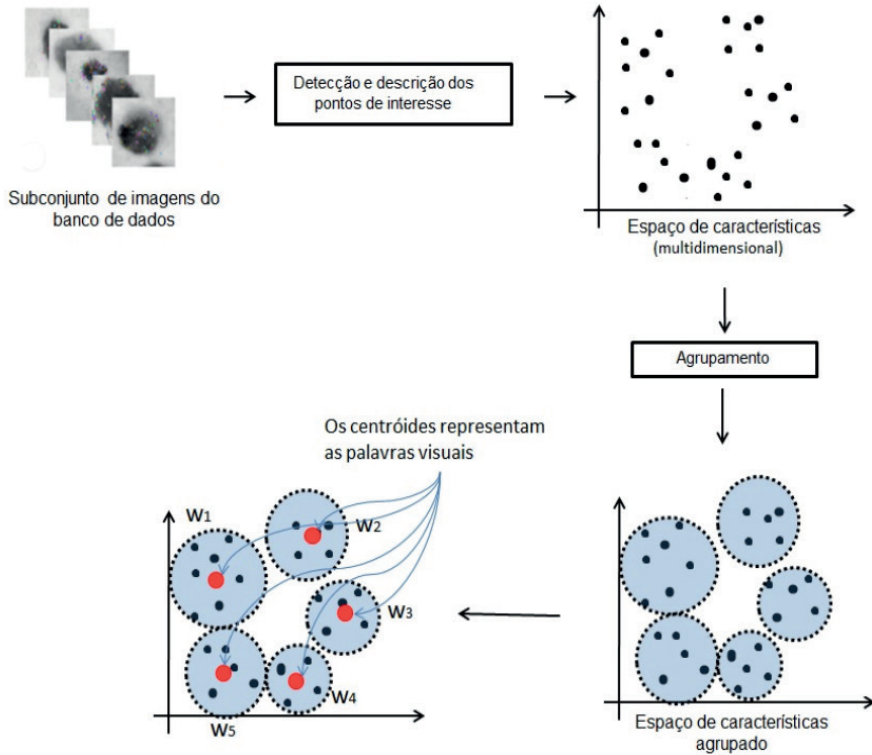


Figura 3. Metodologia para a geração de dicionários de palavras-visuais que será utilizado pela abordagem Bag-of-Features.

3 | RESULTADOS EXPERIMENTAIS

O objetivo deste trabalho é investigar a abordagem Bag-of-Features (BoF) como uma técnica para descrição de imagens de lesão de pele. A Figura 4 mostra o passo-a-passo utilizado para descrever uma imagem em BoF. Primeiramente a imagem colorida é convertida em escala de cinza, em seguida é aplicado o descritor SIFT para detecção e descrição de pontos-de-interesse na imagem. Cada ponto detectado pelo SIFT é tratado como uma palavra-visual de acordo com um dicionário visual previamente construído. Por fim, a representação final é a contagem da frequência com que cada palavra-visual ocorre na imagem de lesão de pele.

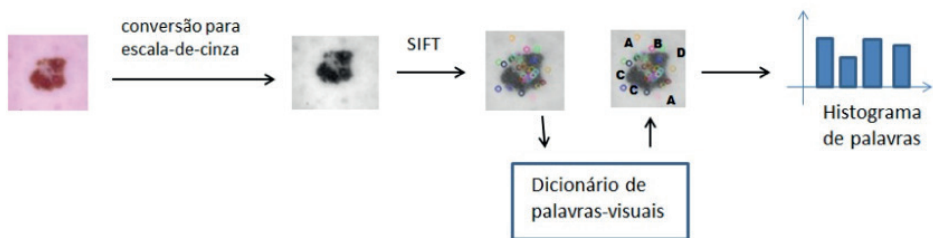


Figura 4. Metodologia utilizada para descrever uma imagem de lesão de pele em Bag-of-Features

Na tarefa de classificação de dados, o tamanho do vetor de características é um fator que afeta diretamente no custo computacional. Para a abordagem BoF o tamanho do vetor-de-características é o tamanho do dicionário visual de dados. Por isso, quanto maior o tamanho do dicionário, maior o tamanho do vetor-de-característica. Assim, busca-se o menor dicionário possível sem penalizar a acurácia na classificação das imagens.

As avaliações experimentais têm dois objetivos: (i) o primeiro é definir qual o tamanho ideal do dicionário visual para descrever lesões de peles; (ii) o segundo é investigar o melhor algoritmo de classificação, aquele que produz a melhor acurácia. A ideia é analisar o desempenho da abordagem BoF com base em custo e precisão computacional. Precisão refere-se à taxa de precisão obtida por um classificador e custo refere-se ao tempo de processamento exigido para classificar uma determinada imagem de lesão de pele.

3.1 Base de Dados

A base de dados de imagens de lesões de pele utilizada neste trabalho teve como fonte o sítio eletrônico do Kaggle. Para o desenvolvimento do trabalho foi considerado um total de 1000 imagens, sendo 500 imagens de melanoma e 500 imagens de não melanoma com o diagnóstico e perfil do portador da doença mapeados.

As pessoas mapeadas têm um perfil de idade entre 17 a 85 anos, sendo 54% homens e 45% mulheres. No geral, a pele desses indivíduos é branca e corresponde a partes do corpo como costas, abdômen, tronco, couro cabeludo, pés e outros. A amostra de imagens contida na plataforma é colorida, com recorte mais próximo da lesão, sendo possível identificar cores, variação de tamanho e outros a olho nu. Existem ruídos como a presença de pelos próximos à mancha, bolhas e o efeito da luz emitida pelo dispositivo de captura de imagens.

3.2 Classificação e Configuração Experimental

Foi avaliado o desempenho de sete diferentes algoritmos de classificação: Naive Bayes, Árvore de Decisão (C4.5), Bayes Network, MultiLayer Perceptron, LogitBoost, SVM e K-NN. Foi utilizada uma validação cruzada 10-folds usando cada algoritmo separadamente na avaliação experimental.

Em relação à representação da imagem, foi utilizada a técnica SIFT proposta por (LOWE, 2004) para extrair e representar pontos de interesse na imagem e o K-means como o algoritmo de agrupamento para gerar o dicionário visual. Foram utilizados diferentes valores de clusters para o algoritmo K-means, tal como será discutido a seguir.

3.3 Resultados

A gráfico da Figura 5 mostra a acurácia dos diferentes valores de K utilizados em relação aos classificadores analisados. Foram utilizados onze diferentes tamanhos de dicionários visual $K = \{ 3, 5, 10, 20, 40, 50, 60, 70, 80, 90, 100 \}$. O maior valor de acurácia (77,6%) foi obtido pelo classificador *MultiLayer Perceptron* com um tamanho de dicionário visual igual a $K=20$. Porém, com um dicionário visual com apenas 5 palavras visuais este mesmo classificador conseguiu uma acurácia de 76,9%, já para valores de dicionários maiores o valor da acurácia diminui bastante. O classificador LogitBoost que é baseado em Regressão Logística, se manteve com acurácia estável para todos os tamanhos de dicionários visuais.

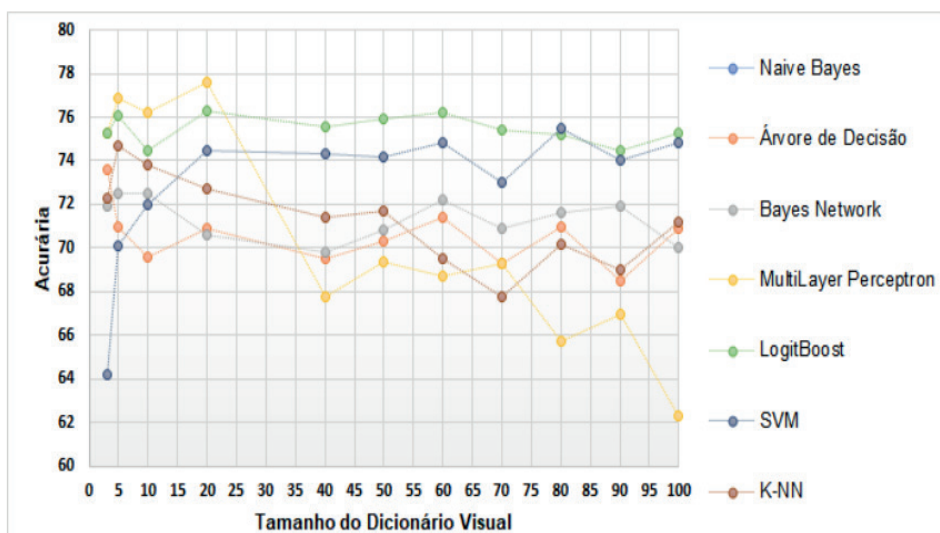


Figura 5. Acurácia obtida usando diferentes algoritmos de classificação e diferentes tamanhos de dicionário visual.

Percebeu-se que a acurácia poderia ser melhorada se alguma técnica de pré-processamento fosse aplicada. Em alguns casos, a classificação foi prejudicada devido a ruídos presentes nas imagens. A Figura 6 mostra alguns resultados classificados erroneamente: a classe predita foi melanoma e a classe esperada era não-melanoma. Para estes exemplos, os pêlos sobre as lesões prejudicaram a classificação correta das lesões.

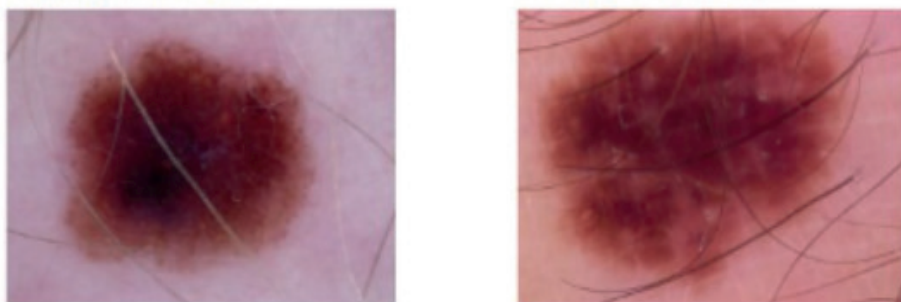


Figura 6. Algumas lesões de pele não-melanomas classificadas erroneamente como melanomas.

4 | CONCLUSÃO

Neste trabalho foi investigado o uso da abordagem Bag-of-Features para a classificação de lesões de pele em melanoma e não-melanoma. Apesar da popularidade da abordagem BoF, este descritor possui alguns parâmetros que podem afetar seu desempenho, por exemplo, o tamanho do dicionário visual. O desempenho depende muito desse tamanho, o que não é uma tarefa fácil configurar.

Neste trabalho foi mostrado que a escolha de diferentes tamanhos de dicionários visuais pode influenciar a precisão do desempenho da classificação da imagem. Resultados experimentais mostram que o algoritmo *MultiLayer Perceptron* com um tamanho de dicionário de 20 palavras visuais apresentou a melhor combinação para classificação de lesões de pele em melanomas e não-melanomas. Essa combinação alcançou 77,5% de acurácia. Isso indica que a abordagem BoF é, de fato, um descritor eficaz para representação de imagens e pode contribuir para o desenvolvimento de técnicas computacionais baseadas em visão computacional para a automatização do diagnóstico clínico.

REFERÊNCIAS

ALFED, N.; KHELIFI, F.; BOURIDANE, A. **Improving a bag of words approach for skin cancer detection in dermoscopic images**. In: 2016 International Conference on Control, Decision and Information Technologies (CoDIT). [S.l.: s.n.], 2016. p. 024–027

BARATA, C. et al. **Local Features Applied to Dermoscopy Images: Bag-of-Features versus Sparse Coding**. In: Iberian Conference on Pattern Recognition and Image Analysis. [S.l.: s.n.], 2017. p. 528–536.

BARATA, C. et al. **Two Systems for the Detection of Melanomas in Dermoscopy Images Using Texture and Color Features**. IEEE Systems Journal, v. 8, n. 3, p. 965–979, set.2014.

HALALLI, B.; MAKANDAR, A. **Computer Aided Diagnosis - Medical Image Analysis Techniques**. Breast Imaging, dez. 2017. Disponível em: <<https://www.intechopen.com/books/breast-imaging/computer-aided-diagnosis-medical-image-analysis-techniques>>.

LOWE, D. G. **Distinctive Image Features from Scale-Invariant Keypoints**. International Journal of Computer Vision, v. 60, n. 2, p. 91–110, nov. 2004. ISSN 0920-5691.

NAMOZOV, A.; CHO, Y. I. **Convolutional Neural Network Algorithm with Parameterized Activation Function for Melanoma Classification**. In:2018 International Conference on Information and Communication Technology Convergence (ICTC). [S.l.:s.n.], 2018. p. 417–419

PEDROSA, G.V.; TRAINA, A.J.M. **Making an image worth thousand visual words**. Proceedings of XI Workshop de Visão Computacional, 2015. p. 56-61

SABBAGHI, S.; ALDEEN, M.; GARNAVI, R. **A deep bag-of-features model for theclassification of melanomas in dermoscopy images**. In:2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). [S.l.:s.n.], 2016. p. 1369–1372

SITU, N. et al. **Malignant melanoma detection by Bag-of-Features classification**. In:2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. [S.l.: s.n.], 2008. p. 3110–3113

EMPLOYING GRADIENT BOOSTING AND ANOMALY DETECTION FOR PREDICTION OF FRAUDS IN ENERGY CONSUMPTION

Data de aceite: 27/08/2020

Data de submissão: 05/06/2020

Renato Vicente

Latam Datalab Serasa Experian
São Paulo – SP

<http://lattes.cnpq.br/1914965078179670>

Ricardo Nascimento dos Santos

Latam Datalab Serasa Experian
São Paulo – SP
<http://lattes.cnpq.br/8232259714089310>

Sami Yamouni

Latam Datalab Serasa Experian
São Paulo – SP
<http://lattes.cnpq.br/4444076421169709>

Beatriz Albiero

Latam Datalab Serasa Experian
São Paulo – SP
<http://lattes.cnpq.br/9288316852477235>

Estevão Uyrá

Latam Datalab Serasa Experian
São Paulo – SP
<http://lattes.cnpq.br/2326476575021426>

Ramon Vilarino

Latam Datalab Serasa Experian
São Paulo – SP
<http://lattes.cnpq.br/3200967512403660>

Juliano Andrade Silva

CPFL Energia
Campinas – SP
<http://lattes.cnpq.br/0659876755582449>

Tales Fonte Boa Souza

CPFL Energia
Campinas – SP
<http://lattes.cnpq.br/0846872049627609>

ABSTRACT: Energy fraud is a critical economical burden for electric power organizations in Brazil. In this paper we present the application of cutting-edge Machine Learning algorithms, namely XGBoost and Isolation Forest, for prediction of irregularities in electrical energy consumption. By using a Logistic Regression model as a benchmark, we show that the use of XGBoost results in a significant improvement in the F1-score for fraud predictions in two different scenarios: with and without inspection history features. Moreover, we also propose the use of the Isolation Forest algorithm for detection of anomalies in electrical energy consumption. We show that this approach may be useful in the case of lack of inspection history features, surpassing dummy classifiers.

KEYWORDS: machine learning, energy Fraud, XGBoost, anomaly detection.

UTILIZANDO GRADIENT BOOSTING
E DETECÇÃO DE ANOMALIAS
PARA PREDIÇÃO DE FRAUDES NO
CONSUMO DE ENERGIA

RESUMO: A Fraude em consumo de energia elétrica é um grande problema para distribuidoras de energia elétrica no Brasil. Neste trabalho é descrita a aplicação de algoritmos recentes de aprendizagem de

máquina, denominados XGBoost e Isolation Forest, para predição de irregularidades no consumo de energia elétrica. Tendo modelos de regressão logística como algoritmo base de comparação, mostra-se que o uso do método XGBoost permite obter um aumento de performance significativo no valor de F1-score para predição de fraudes em dois cenários distintos, com a presença ou ausência de dados históricos de inspeções de equipamentos *in loco*. Além disso, descreve-se o uso do algoritmo de Isolation Forest para detecção de anomalias no consumo de energia elétrica. Mostra-se que esta estratégia é útil na situação em que históricos de inspeções não estão disponíveis, resultando em uma capacidade preditiva superior à aleatoriedade.

PALAVRAS-CHAVE: Aprendizagem de máquina, fraude em distribuição de energia, XGBoost, detecção de anomalias.

1 | INTRODUCTION

The last few decades have witnessed a drastic increase in global energy consumption, driven by the accelerated growth of industry and technology. In special, world electricity consumption almost duplicated in the last quarter-century with new demands such as accessibility to home appliances and transportation [ENERDATA 2019]. In the future, despite the development of more efficient devices and production processes, it is expected a steady 1% annual growth in electricity demand [Antunes Lima 2019].

Competition for market share in electrical utility industry has raised the energy loss during distribution as a major concern for generation companies in order to increase efficiency [Management Solutions 2017]. The reasons for energy loss in distribution lines can be separated in two types: technical and non-technical [Antunes Lima 2019, Doukas et al. 2011, Management Solutions 2017]. The first case is inherent to the physical properties of electricity transport through grid, when a fraction of energy is converted and dissipated through heat or lost due to inductive and capacitive effects [Antunes Lima 2019, Doukas et al. 2011, Management Solutions 2017]. The second type of energy loss comprises in fraudulent practices by consumers that deliberately modify energy measuring devices in order to reduce household bills or perform new illegal electricity connections on main power cables in the neighborhood [Smith 2004, Ford et al. 2014, Cody et al. 2015a, Coma-Puig et al. 2016]. Despite being a crime in many countries, energy fraud is a widespread practice encouraged by the difficulty in verification, which relies mainly on *in situ* inspections.

In Brazil, the high incidence of energy fraud is a critical economical burden for electric power organizations. According to ANEEL (the Brazilian Electricity Regulatory Agency), it was estimated a loss of 14% of total electrical energy available for distribution on 2016, with fraud practice accounting for approximately half of this deficit and a total burden of \$1.2 billion at the same year [Antunes Lima 2019, Maia 2017]. This amounts to 3.6 times the last year's budget for the National Council for Scientific and Technological Development (CNPq).

1 <http://www.portaltransparencia.gov.br/orgaos/20501?ano=2018>

Recently, advancements in machine learning algorithms and computing power provided novel solutions to boost efficiency in detection of energy frauds. Based on patterns and anomalies identification in consumption, predictive models can highlight potential candidates for *in situ* inspection and reduce the cost of energy fraud detection [Ford et al. 2014, Cody et al. 2015a]. Herein, we describe the application of state-of-the art machine learning techniques for fraud detection in electrical energy consumption. These studies resulted in robust predictive models for fraud occurrence based on gradient boosting applied to registry profiles and energy consumption records. Moreover, we also propose a generalized and unsupervised model for fraud detection based on consumption anomalies inferred by isolation forests.

2 | RELATED WORK

Several applications of supervised and unsupervised machine learning algorithms for prediction of fraud and irregularity in electric utility can be found in literature. [Messinis and Hatzigargyriou 2018]. Examples of implementations of usual supervised methods include the application of support vector machines to identify customer's abnormal consumption behavior based on previous energy usage data [Nagi et al. 2010, Alfara et al. 2018]. Other case studies of well-established machine learning methods focused on fraud detection in electricity consumption are the use of decision trees [Monedero et al. 2012, Cody et al. 2015b], logistic regression, linear discriminant analysis ([Lawi et al. 2017]) and time series [Nogales et al. 2002]. Additionally, recent studies have provided new insights with the use of more complex machine learning models, such as Neural Networks [Nizar et al. 2008, Monedero et al. 2006, Costa et al. 2013] and rough set theory [Spiri et al. 2014]. Within the scope of Unsupervised Learning, Cabral et al. in [E. Cabral et al. 2008] present self-organizing maps that learns historical consumer energy consumption behavior. This study is focused in high voltage electricity consumers. Furthermore, [Angelos et al. 2011] proposes a two step methodology in order to find consumers with similar consumption profiles and hence potential fraudsters. This methodology consists of: (i) a C-means-based fuzzy clustering and (ii) a fuzzy classification system to rank users according to their irregular patterns.

3 | METHODOLOGY

This section describes the methodology used in the present research. First, we discuss the two datasets provided by CPFL Energia². The datasets basically consist of the same features and also share the same target distribution, but diverge with relation to historical information. The first dataset containing outdated information was used to fit the

² CPFL Energia, Rua Jorge de Figueiredo Correa, n 1632, Jardim Professora Tarclia CEP 13.087-397, Campinas/SP, Brazil.

model. Later, a second dataset containing more recent information was used in an out-of-time validation. Secondly, we introduce the model used for this fraud classification task, the XGBoost model. We also provide a list of hyperparameters in which we performed a grid search.

3.1 Datasets

In order to develop predictive models for fraud in energy consumption, the following primary datasets were considered: (i) reports of local inspections and (ii) history of energy consumption for each registry. The datasets were provided by CPFL Energia, a utility company distributing electricity. They contain data of a medium-size brazilian city with around 700 thousands customers between February of 2014 and September of 2018. Considering [Messinis and Hatziaargyriou 2018] definitions, the dataset contains low-resolution energy data, with a time resolution of one day, at consumer level. After feature engineering and categories aggregation a total of 64 features was employed in statistical modeling studies. For confidentiality reasons, the variables information are condensed in classes as described in Table 1. The resulting features are very similar to the ones described in [Messinis and Hatziaargyriou 2018]. The fraud event variable describes exclusively the fraudulent or non-fraudulent events observed by the company investigators. Any irregular behaviours with no proved malicious intent has been discarded.

Feature	Description
Fraud event	Target variable labelling fraud (1) and non-fraud events (0), as reported by local inspection
meter ID	Meter identifier related to the consumer
t_0	Inspection date used as reference for feature values
Inspection history	Variables describing the history of inspections results for a specific meter equipment.
Meter characteristics	Set of variables describing different meter characteristics (e.g.: age of equipment, model, manufacturer brand).
Geographic location	geographic variable calculated from the meter location.
Consumption at inspection date	Total electrical energy consumption in kwh, as measured at t_0
Consumption history	Historical energy consumptions before t_0
Consumption statistics	Set of statistic variables calculated from consumption history (e.g.: coefficient of variation).

Table 1. Fraud dataset description.

Figure 1 shows the fraud distribution in the two datasets used in the present article. The first one will be the basis to fit the classification model contains nearly 35 thousand

records, while the second one will be used to validate the model out-of-time and contains nearly 7 thousand records. It is straightforward to see the conservation of fraud proportion between both datasets. They both present an unbalanced dataset with around 72% of regular events and 28% of irregular or fraud events.

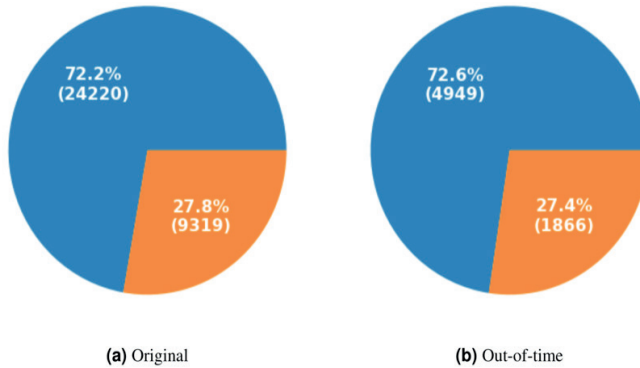


Figure 1: Fraud distribution in original (a) and out-of-time (b) datasets. Regular and irregular situations are displayed in blue and orange, respectively. The figures in bracket represent the true numbers of investigated customers for each category.

3.2 XGboost classifier for fraud consumption

We approach the problem of fraud detection as a supervised binary classification problem and set the fraud event feature (described in Section 3.1) as our target. The remaining features are used to fit a XGBoost model.

The XGBoost algorithm [Chen and Guestrin 2016] is a decision-tree-based ensemble model that has been first introduced in 2014. It uses gradient boosting, an iterative and additive approach where new models are trained to predict the residuals of prior models. Since we defined fraud detection as a binary classification problem, we use logistic regression (LR) for binary classification as our learning objective.

We perform a grid search over 5 hyperparameters: (i) Number of Estimators, (ii) Sub-sampling of Columns, (iii) Maximum Depth and regularization (iv) Gamma and (v) Min Child Weight (Minimum sum of instance weight (hessian) needed in a child leaf). We set our learning rate to 0.1. Remaining hyperparameters are set to default values. Table 2 shows the hyperparameters and the respective values on which we have performed the grid search.

Hyperparameter	Values
n_estimators	10-560

col_subsampling	0.3, 0.5
max_depth	1-12
gamma	1, 4, 10, 20
min_child_weight	1, 4, 10

Table 2. Hyperparameter grid search for XGBoost

The grid search is performed using 5-Fold [Raschka 2018] cross-validation. For evaluating the model performances, we use the F1-score. It is a metric defined as the harmonic mean between Precision and Recall, and is considered a parsimonious metric when dealing with unbalanced problems, which is our case.

3.3 Unsupervised analysis

We apply an anomaly detection analysis, making use of the less amount of data per observation. In special, we made no use of "fraud" labels when training the model, making it unsupervised. The anomaly detection model produces an "anomaly score" for each example, analogous to how a binary classifier would produce a score for the "True" label. We treated the unsupervised anomaly score as a fraud score, such that "common" examples (with low anomaly score) were considered as legitimate, and "odd" examples (with high anomaly score) were considered fraudulent. We fitted an Isolation Forest model ([Liu et al. 2008]), using all the data in which we had no label (no inspections were made), and evaluated the predictions on the inspected population. We made no parameter tuning, relying on the default parameters of the scikit-learn implementation [Liu et al. 2012]. After fitting, we ignored the default labels and have chosen the threshold that maximized the F1-score, based on the precision-recall curve.

4 | RESULTS AND DISCUSSION

In this section we compare the results obtained with the XGBoost model and a LR model. Both models were tested over two different datasets, one containing inspection history data, and a second without such features. We also show how we conducted feature selection to refine the XBoost model. Furthermore, there is a section dedicated to the topic of anomaly detection over energy consumption. In this, we show that anomaly scores may be used as indicative of fraud.

4.1. Models Comparison for Fraud Detection

Inspection history variables demonstrated to be among the most important features for fraud prediction. Hence, we created a new dataset by excluding inspection history data,

namely Newcomers. We chose to compare XGBoost with a LR model since it would be a good benchmark for a binary classification task. Therefore, we test XGBoost and LR performances over both datasets.

Three metrics were used to assess of these algorithms: F1-score, precision and recall. F1-score is the more adapted when addressing unbalanced classification problems, as it is the case here. In Table 3, results show that XGBoost outperforms LR in both settings, in particular for the newcomers customers. And, for all metrics, the "All customers" case displays higher performance than the "Newcomers"'s one. It is expected since customer historical data are taken into account.

	All customers		Newcomers	
	LR	XGBoost	LR	XGBoost
F1	0.67	0.8	0.13	0.61
Precision	0.79	0.9	0.37	0.81
Recall	0.58	0.71	0.08	0.48

Table 3. Fraud metrics for comparing performance of Logistic Regression (LR) and XGBoost models.

4.2 Out-of-time model validation

The out-of-time validation process checks the model robustness on a later dataset than the one on which the model has been fitted. It is useful when the application of a model to a population is changing over time such as the energy consumption. Results considering all customers as well as newcomers are displayed in table 4, both using XGBoost algorithm. Considering "All customers" column, we note a slight and consistent decrease in all the metrics. This trend is also observed for the newcomers except for the recall where the increase is not relevant regarding the standard deviation.

	All customers	Newcomers
F1	0.72	0.59
Precision	0.82	0.66
Recall	0.64	0.53

Table 4. Out-of-time validation.

4.3 Model refinement

The XGBoost model automatically provides a list of the features ranked by their importance on the predictive model problem. Following this list as an importance rank, we gradually increased the number of features in order to observe the smallest set of features

with higher importance that could provide a high predictive accuracy. We performed this study considering both the full dataset and the group of consumers with no previous inspection history (newcomers).

Figure 2 shows the change in F1-score for both cases with the gradual addition of features based in importance rank. As can be observed for the more general case considering all type of consumers (2(a)), F1-score drastically increases when the first 4 features are considered. These first 4 features consist with information of the coefficient of variation of energy consumption along last year, sum of previous fraud events for the location, and features that describes meter equipment age and geographical location.

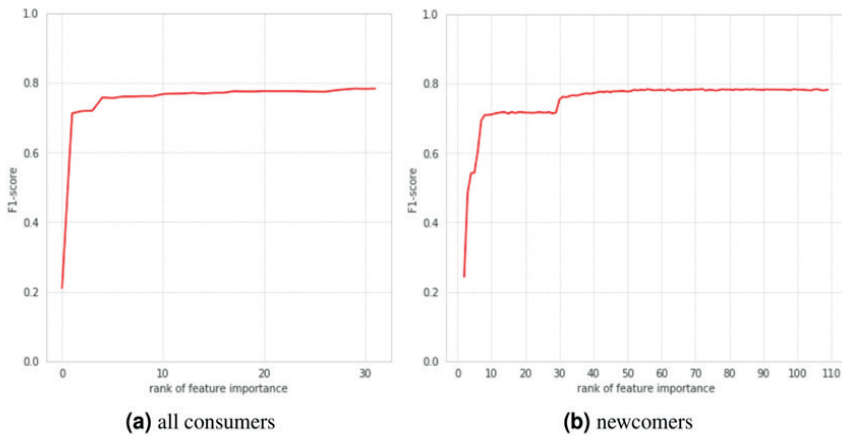


Figure 2: Refinement of predictive models in order to keep only the most relevant features for predictive efficiency for (a) all costumers and (b) costumers with no previous inspection history (newcomers).

When removing features related to historical inspection (Figure 2(b)), we observe a change in the pattern of F1-score increase, with a more gradual growth in performance and two main substantial increases: (i) after including the 8th and (ii) the 30th feature ranked by importance. In order to keep a predictive model with highest F1-score, we considered for the case of newcomers the top 42 features to be considered in the final model. Table 5 summarizes the results obtained for external cross-validation and external hold-out datasets.

	F1 external CV	F1 hold-out
All consumers	0.80	0.76
Newcomers	0.78	0.60

Table 5. Refinement of models.

Moreover, a pattern change in the final prediction models that is worthy to mention is the role of the most important features for fraud distinction. While historical fraud events are crucial features for the most general model considering all consumers, in the absence of historical features (newcomers), the obtained models still present considerable external predictive power with the geolocation and consumption coefficient of variation being the most important contributors for fraud discrimination.

4.4 Anomaly detection may help when there is no labeled data

Extending our analysis to contexts when even less data is available, we removed the correct labels from our training set, making the problem unsupervised. Assuming that data irregularities could be indicative of fraud, we use Isolation Forest ([Liu et al. 2008]), an anomaly detection that does not have label supervision, as described in 3.3. To test this, we collected a subset of the data in which the proportion of frauds was approximately 11%, comparing the results with the expected within our subset of data. Because we have no prior information about the target (it is unavailable in this setting), the precision does not raise above the true proportion, which is 11%. Hence, the best f1 score achievable in this setting is when all samples are categorized as frauds (at least reaching 100% recall).

The results, shown in Table 6, indicate that anomaly scores can be used as a proxy for fraud. Specially when the amount of unlabeled data is big, the anomaly results are a considerable improvement over a random baseline.

	Random	Random with same recall	Anomalies
F1	0.20	0.18	0.30
Precision	0.11	0.11	0.21
Recall	1.00	0.50	0.50

Table 6. Frauds detected as anomalies.

5 | CONCLUSIONS AND FUTURE WORK

In this paper, we presented both a supervised and an unsupervised approaches to detect fraud using data from CPFL Energia, a utility company distributing electricity.

Regarding the supervised study, models using XGBoost algorithms outperformed the benchmark logistic regression models displaying a F1-score of 0.8. We explained this difference by the fact that XGBoost models perform better on unbalanced datasets, as it is the case here. The resulting model has then been successfully validated on an out-of-time dataset and newcomers, which are populations without any investigation historic. A refinement study was also conducted, using the XGBoost feature importance list as a reference. For the model considering the general case of all type of consumers, the filtered

predictive models presented very low decrease in F1-score metric even when considering only the four most important features. Among these features, there are data that accounting consumption changes that suggests to be able to identify the changes in the customers behavior, historical fraud observations and a geographic variable related to the meter location.

On the other hand, for the unsupervised study, we ran an anomaly detection algorithm using Isolation Forest and it has shown promising results.

The results presented in this article only concern one mid-size city, and should be apply to other geographic regions. By doing so, we would potentially create a generic version of the fraud detection model.

ACKNOWLEDGMENTS

This work was supported by ANEELs research & development program (Project ID PD- 0063-3039/2018) in partnership with CPFL ENERGIA group companies.

REFERENCES

- Alfarra, H., Attia, A., and S. M. El Safty, C. (2018). **Nontechnical loss detection for metered customers in alexandria electricity distribution company using support vector machine.** *Renewable Energy and Power Quality Journal*, 1:468–474.
- Angelos, E., Saavedra, O., Carmona Cortes, O., and Souza, A. (2011). **Detection and identification of abnormalities in customer consumptions in power distribution systems.** *Power Delivery, IEEE Transactions on*, 26:2436–2442.
- Antunes Lima, D. (2019). **Perdas de energia - aneel (brazilian electricity regulatory agency).** <https://www2.camara.leg.br/atividade-legislativa/comissoes/comissoes-permanentes/cme/audiencias-publicas/2018/audiencia-publica-16-05-2018/ANEEL%20-%20%20Perdas%20Eletricas%20-%20Davi%20Lima.pdf>.
- Chen, T. and Guestrin, C. (2016). **XGBoost: A Scalable Tree Boosting System.** *arXiv e-prints*, page arXiv:1603.02754.
- Cody, C., Ford, V., and Siraj, A. (2015a). **Decision tree learning for fraud detection in consumer energy consumption.** In *2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA)*, pages 1175–1179. IEEE.
- Cody, C., Ford, V., and Siraj, A. (2015b). **Decision tree learning for fraud detection in consumer energy consumption.** *2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA)*, pages 1175–1179.
- Coma-Puig, B., Carmona, J., Gavalda, R., Alcoverro, S., and Martin, V. (2016). **Fraud detection in energy consumption: A supervised approach.** In *2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, pages 120–129. IEEE.

Costa, B., L. A. Alberto, B., M. Portela, A., W. M., and O. Eler, E. (2013). **Fraud detection in electric power distribution networks using an ann-based knowledge-discovery process.** *International Journal of Artificial Intelligence & Applications*, 4:17–23.

Doukas, H., Karakosta, C., Flamos, A., and Psarras, J. (2011). **Electric power transmission: An overview of associated burdens.** *International Journal of Energy Research*, 35(11):979–988.

E. Cabral, J., Pinto, J., M. Martins, E., and M. A. C. Pinto, A. (2008). **Fraud detection in high voltage electricity consumers using data mining.** pages 1 – 5.

ENERDATA (2019). **Global energy statistical yearbook 2019.** <https://yearbook.enerdata.net/electricity/electricity-domestic-consumption-data.html>.

Ford, V., Siraj, A., and Eberle, W. (2014). **Smart grid energy fraud detection using artificial neural networks.** In *2014 IEEE Symposium on Computational Intelligence Applications in Smart Grid (CIASG)*, pages 1–6. IEEE.

Lawi, A., Wungo, S. L., and Manjang, S. (2017). **Identifying irregularity electricity usage of customer behaviors using logistic regression and linear discriminant analysis.** *2017 3rd International Conference on Science in Information Technology (ICSITech)*, pages 552–557.

Liu, F., Ting, K., and Zhou, Z.-H. (2012). **Isolation-based anomaly detection.** *ACM Transactions on Knowledge Discovery from Data*, 6(1):1 – 39.

Liu, F. T., Ting, K. M., and Zhou, Z.-H. (2008). **Isolation forest.** In *2008 Eighth IEEE International Conference on Data Mining*, pages 413–422. IEEE.

Maia, C. (2017). **Perdas de energia custam mais de R\$8 bi aos consumidores em 2016.** <https://www.valor.com.br/empresas/5219107/perdas-de-energia-custam-mais-de-r-8-bi-aos-consumidores-em-2016>.

Management Solutions, M. (2017). **Fraud management in the energy industry.** <https://www.managementsolutions.com/sites/default/files/publicaciones/eng/fraud-management-in-the-energy-industry.pdf>.

Messinis, G. M. and Hatziaargyriou, N. D. (2018). **Review of non-technical loss detection methods.** *Electric Power Systems Research*, 158:250–266.

Monedero, I., Biscarri, F., Leon, C., Guerrero, J. I., Biscarri, J., and Millan, R. (2012). **Detection of frauds and other non-technical losses in a power utility using pearson coefficient, bayesian networks and decision trees.** *International Journal of Electrical Power & Energy Systems*, 34:90–98.

Monedero, I., Biscarri, F., Len, C., Biscarri, J., and Millan, R. (2006). **Midas: Detection of non-technical losses in electrical consumption using neural networks and statistical techniques.** pages 725–734.

Nagi, J., Yap, K. S., Tiong, S. K., Ahmed, S. K., and Mohamad, M. (2010). **Nontechnical loss detection for metered customers in power utility using support vector machines.** *IEEE Transactions on Power Delivery*, 25:11621171.

Nizar, A. H., Dong, Z. Y., and Wang, Y. (2008). **Power utility nontechnical loss analysis with extreme learning machine method.** *IEEE Transactions on Power Systems*, 23:946– 955.

Nogales, F., Contreras, J., J. Conejo, A., and Espinola, R. (2002). **Forecasting next-day electricity prices by time series models.** *Power Engineering Review, IEEE*, 22:58–58.

Raschka, S. (2018). Model Evaluation, **Model Selection, and Algorithm Selection in Machine Learning.** *arXiv e-prints*, page arXiv:1811.12808.

Smith, T. B. (2004). **Electricity theft: a comparative analysis.** *Energy policy*, 32(18):2067–2076.

Spiri, J. V., Stankovi, S. S., Doi, M. B., and Popovi, T. D. (2014). **Using the rough set theory to detect fraud committed by electricity customers.** *International Journal of Electrical Power & Energy Systems*, 62:727 – 734.

IN SILICO STUDY OF THE INTERACTION BETWEEN HUMAN 5-HT_{2C} RECEPTOR AND ANTIDEPRESSANT DRUG CANDIDATES

Data de aceite: 27/08/2020

Data de submissão: 25/05/2020

Rômulo Oliveira Barros

Laboratório de Pesquisa em Sistemas de Informação (LaPeSI) – Instituto Federal de Educação, Ciência e Tecnologia do Piauí (IFPI).
Teresina – Piauí

Jhonatan Matheus Sousa Costa

Laboratório de Pesquisa em Sistemas de Informação (LaPeSI) – Instituto Federal de Educação, Ciência e Tecnologia do Piauí (IFPI).
Teresina – Piauí

Wildrimak de Souza Pereira

Laboratório de Pesquisa em Sistemas de Informação (LaPeSI) – Instituto Federal de Educação, Ciência e Tecnologia do Piauí (IFPI).
Teresina – Piauí

Diego da Silva Mendes

Laboratório de Pesquisa em Sistemas de Informação (LaPeSI) – Instituto Federal de Educação, Ciência e Tecnologia do Piauí (IFPI).
Teresina – Piauí

Fábio Luis Cardoso Costa Júnior

Laboratório de Pesquisa em Sistemas de Informação (LaPeSI) – Instituto Federal de Educação, Ciência e Tecnologia do Piauí (IFPI).
Teresina – Piauí

Ricardo Martins Ramos

Laboratório de Pesquisa em Sistemas de Informação (LaPeSI) – Instituto Federal de Educação, Ciência e Tecnologia do Piauí (IFPI).
Teresina – Piauí

ABSTRACT: Depression is a major public health concern affecting 300 million people worldwide, according to WHO. The mechanism of depression is presumed to be related to the role of serotonin (5-hydroxytryptamine - 5-HT) and its receptors in the central nervous system. In this work a human 5-HT_{2C} receptor model was created by homology modeling, and molecular docking studies were performed with the ligands fulvic acid, paroxetine, citalopram, and serotonin itself. Fulvic acid had similar affinity parameters to paroxetine and s-citalopram, which are widely used drugs for the treatment of psychiatric disorders such as depression and anxiety.

KEYWORDS: Depression, Fulvic Acid, 5-HT_{2C}, Serotonin, Molecular Docking.

ESTUDO IN SILICO DA INTERAÇÃO ENTRE O RECEPTOR 5-HT_{2C} HUMANO E CANDIDATOS A DROGAS ANTIDEPRESSIVAS

RESUMO: A depressão é uma grande preocupação de saúde pública, que afeta 300 milhões de pessoas em todo o mundo, de acordo com a OMS. Presume-se que o mecanismo da depressão esteja relacionado ao papel da serotonina (5-hidroxitriptamina - 5-HT) e seus receptores no sistema nervoso

central. Neste trabalho, um modelo de receptor 5-HT_{2C} humano foi criado através de modelagem por homologia e foram realizados estudos de acoplamento molecular com os ligantes ácido fúlvico, paroxetina, citalopram e serotonina. O ácido fúlvico apresentou parâmetros de afinidade semelhantes aos da paroxetina e do s-citalopram, que são medicamentos amplamente utilizados no tratamento de distúrbios psiquiátricos, como depressão e ansiedade.

PALAVRAS-CHAVE: Depressão, Ácido Fúlvico, 5-HT_{2C}, Serotonina, Ancoramento Molecular.

1 | INTRODUCTION

Depression represents an important public health concern, affecting 300 million people worldwide, according to World Health Organization [WHO 2018]. The mechanism of depression is presumed to be strongly related to the role of serotonin (5-hydroxytryptamin – 5-HT) and its receptors in the central nervous system, especially in the dorsal raphe nucleus [Borroto-Escuela et al. 2018; Grandjean et al. 2019; Marek 2017; Yohn et al. 2017]. Antidepressant drugs like paroxetine, fluoxetine, and s-citalopram act by selectively inhibiting serotonin reuptake, and therefore making serotonin available inside synaptic ambient for a longer time [Grandjean et al. 2019]. It is known to exist 14 different types of serotonin receptors [Palacios 2016], and that 5-HT_{2C} is a prominent subject of studies on mental diseases like anxiety, major depression disorder, and bipolar disorder, and on neurodegenerative diseases like Alzheimer and Parkinson [Tohda 2014].

In traditional ayurvedic medicine a mineral called Shilajit, whose main compounds are fulvic acids (50–60% of the total) and humic acids, have been used for many centuries [Bhavsar et al. 2016]. Fulvic and humic acids are two of three fractions that can be obtained from the filtration of humic substances, which are components of humus, resulted from the degradation of organic matter in the soil and the water [Qin et al. 2019; Saleh et al. 1989; Thurman and Malcolm 1981]. Some beneficial activities are already related to fulvic acid, like antidiarrheal [Qin et al. 2019], anti-inflammatory [Chien et al. 2015], humoral immune stimulation [Vucskits et al. 2010], as potential drug against A β 17–42 mediated cytotoxicity and neurodegeneration [Verma et al. 2013], a promising topical remedy for drug-resistant wound infections [Zhao et al. 2015], and therapy for the management of oral biofilm infections [Sherry et al. 2012].

In silico studies play a major role in the process of drug discovery and development. Through computational calculations it is possible to evaluate stereochemical characteristics of the molecules involved in the biological functions and to predict the binding affinity of receptors and ligands. However, to the best of our knowledge, no in silico studies on the molecular interaction between 5-HT_{2C} and fulvic and humic acids were previously made.

Thus, the aim of this work is to in silico study the interaction of fulvic acid and humic acid ligands with a human 5-HT_{2C} receptor model created by homology modeling, and to

compare them with those of known antidepressants and serotonin, so that it may be possible to indicate antidepressant drug candidate for further *in silico*, *in vitro*, and *in vivo* studies.

2 | METHODS

2.1 Softwares

All softwares used in this study were chosen because of their high acceptance and referencing in the scientific community and because they are free. All of them were designed based on mathematical models of the biochemical properties of the atoms involved in the proteins constitution. The conformation of the proteins are mainly conceived by the means of genetic algorithm computation, especially that of molecular docking calculations.

2.2 Homology modeling

Structural models of 5-hydroxytryptamine 2C receptor isoform – a precursor (5-HT_{2C}) of *Homo sapiens* – were generated by homology modeling using the software MODELLER 9v21 (<https://salilab.org/modeller/>) [Eswar et al. 2006], through its graphical interface integrated to the software UCSF Chimera (<https://www.cgl.ucsf.edu/chimera/>) [Pettersen et al. 2004].

The search for homolog proteins to 5-HT_{2C} (amino acids sequences with accession code NP_000859.1 in NCBI's database) was done among experimental structures available in the Protein Data Bank (PDB) (<https://www.rcsb.org/>) [Berman 2000], from which three templates (Table 1) were chosen: 5-HT_{2C}, 5-HT_{2AR}, and 5-HT_{2B}-BRIL. One hundred independent models were generated and the model presenting the most negative value for the function of DOPE energy was chosen [Plácido et al. 2017].

Template proteins	PDB code	Resolution	Sequence Identity	Species
5-HT _{2C}	6BQG	3.0Å	76.5%	<i>Escherichia coli</i> , <i>Homo sapiens</i>
5-HT _{2AR}	6A93	3.0Å	56.4%	<i>Escherichia coli</i> , <i>Homo sapiens</i>
5-HT _{2B} -BRIL	4IB4	2.7Å	47.7%	<i>Escherichia coli</i> , <i>Homo sapiens</i>

Table 1. Details of the protein models used for homology modeling.

The 3D structures of fulvic acid (CID 5359407), humic acid (CID 90472028), paroxetine (CID 43815), and serotonin (CID 5202) were downloaded from PubChem Open Chemistry Database (<https://pubchem.ncbi.nlm.nih.gov/>) [Kim et al. 2019]. The structures of the ligands are shown in Figure 1.

2.3 Molecular Docking

The molecular docking calculations were done with the software AutoDock Vina (<https://vina.scripps.edu/>) [Trott and Olson 2010]. The ligand and the protein were prepared for the calculations with AutoDock Tools (ADT) 1.5.6 [Morris et al. 2009]. 5-HT2C model was set as rigid and all ligands was considered as flexible, with their torsion being added during the preparation process. Hydrogens were added to both receptor and ligands individually, Gasteiger charges were then calculated by ADT, and non-polar hydrogens were merged. The grid box was sized as 22.5 x 22.5 x 22.5 units of 1 Å for each axis. The grid box was centered in the coordinates of the oxygen of the residue Asp95, in the active site of 5-HT2C. The number of modes were set to 100, and exhaustiveness was set to 64. All other parameters were left in default. The conformations with the highest affinity were ranked by their root mean square deviation (RMSD). The conformation with the highest affinity was chosen to further visual inspection and analysis.

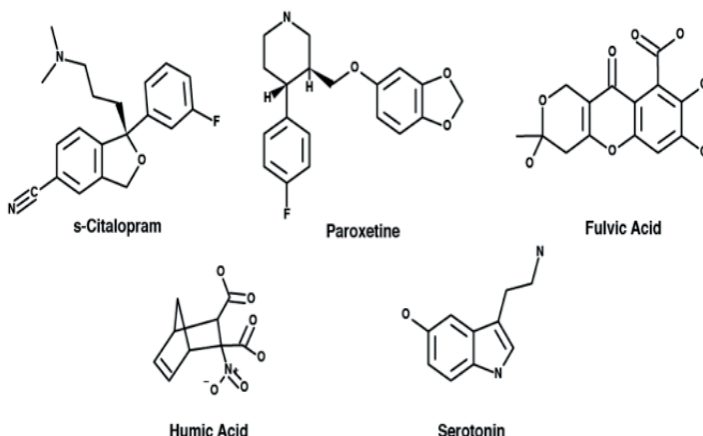


Figure 1. Structure of the ligands used in the study

3 | RESULTS AND DISCUSSION

The results of the molecular docking of the template-based model of 5-HT2C with fulvic acid, humic acid, paroxetine, s-citalopram, and serotonin are shown in Table 2.

Serotonin is the element used in this study to represent, alongside with paroxetine and s-citalopram, the standards to assess the results obtained with fulvic and humic acids. Trans-membranes proteins of the same family as 5-HT2C do the transportation of serotonin from the synaptic cleft back to the interior of neuron's cytoplasm. The binding energy showed by the complex of serotonin with the 5-HT2C model was $-6.5 \text{ kcal mol}^{-1}$. Therefore, these values were considered as reference for the typical physiological pattern for this enzyme

and its main ligand in the context of this template-based model of 5-HT2C.

Complex (Protein-Ligand)	ΔG_{bind} ^a (kcal/mol)	Amino acids interacting through hydrogen bonds ^b	Amino acids exerting hydrophobic interactions ^b
Paroxetine	-8.7	None	Asp95, Val96, Leu170, Trp91, Trp285, Phe288, Ile92, Ser99, Phe175, Val315
Fulvic Acid	-8.4	Ser99, Ser180, Thr100	Phe288, Leu170, Val96, Val176, Gly179, Phe289, Trp91, Ile92, Asp95, Ala183
S-citalopram	-8.0	Asn292, Ser71	Leu170, Val315, Trp91, Val96, Tyr79, Ile92, Asp95, Phe288, Ala74, Ile75, Leu311
Serotonin	-6.5	Asp95, Ser180, Val176, Ser99	Val96, Gly179, Phe289, Ile92, Leu170
Humic Acid	-6.3	Ser180, Ser99	Val96, Phe289, Gly179, Phe288, Leu170, Phe175, Val176, Ala183

Table 2. Parameters affinity of the molecular docking with 5-HT2C

^aEstimated free binding energy ^b Obtained with the Ligplot software.

The complex of paroxetine with the 5-HT2C model showed highest estimated free binding energy ($-8.7 \text{ kcal mol}^{-1}$) of all the complexes tested. Paroxetine formed no hydrogen bonds, but made hydrophobic interactions with 12 residues (Table 2). S-citalopram showed a binding affinity of $-8.0 \text{ kcal mol}^{-1}$ in complex with 5-HT2C. This ligand formed hydrogen bonds with the Asn292, and Ser71 residues of 5-HT2C, and made 13 hydrophobic interactions. These results are consistent with the fact that paroxetine and s-citalopram are well-known selective serotonin reuptake inhibitors (SSRI) [Matthäus et al. 2016], which are widely used for the treatment of several psychiatric disorders like depression and anxiety, with paroxetine being one of the most potent SSRI known [Davis et al. 2016].

Testing fulvic acid and humic acid was the main objective of this study. The complex of humic acid with 5-HT2C presented way less estimated binding affinity than any other compound in study ($-6.3 \text{ kcal mol}^{-1}$). It formed hydrogen bonds with Ser180 and Ser99, and made hydrophobic interactions with 10 residues. As for fulvic acid, it showed similar estimated free binding energy values of that of paroxetine ($-8.7 \text{ kcal mol}^{-1}$ versus $-8.4 \text{ kcal mol}^{-1}$), and higher affinity than serotonin ($-6.5 \text{ kcal mol}^{-1}$ versus $-8.4 \text{ kcal mol}^{-1}$). Fulvic acid made hydrogen bonds with Ser99, Ser180, Thr100 residues of 5-HT2C model, and made other 12 hydrophobic interactions within the active site. These results indicate that fulvic acid may be considered a potential drug candidate to the inhibition of 5-HT2C receptor, acting in similar way as paroxetine and s-citalopram, and having higher affinity than s-citalopram.

All hydrogen bonds and hydrophobic interactions are shown in Figure 2.

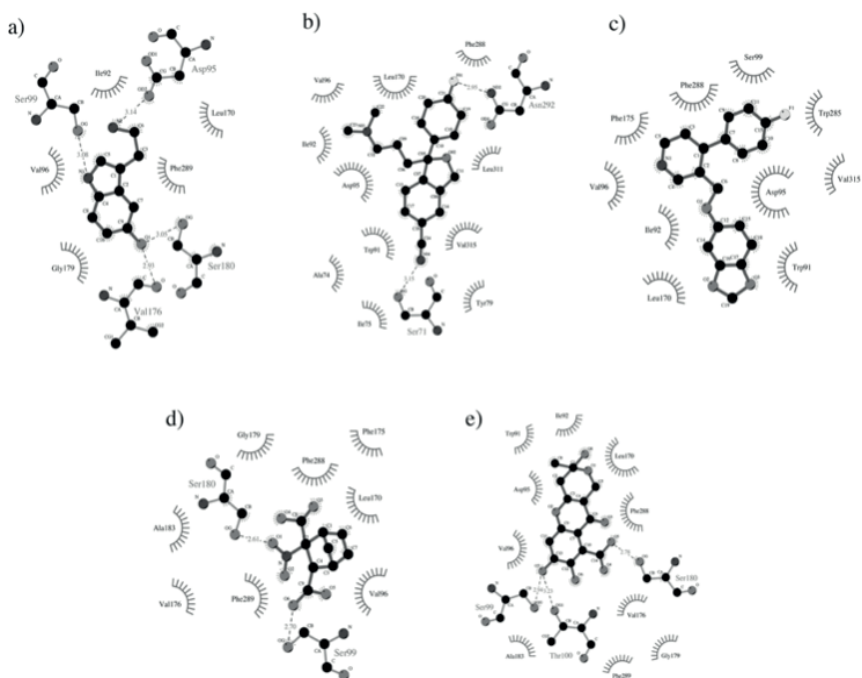


Figure 2. hydrogen bonds and hydrophobic interactions obtained with Ligplot+: a) Serotonin, b) s-citalopram, c) paroxetine, d) humic acid, and e) fulvic acid.

4 | CONCLUSION

Fulvic acid may be considered a potential drug candidate to the inhibition of 5-HT_{2C} receptor, once it showed similar binding affinity to paroxetine and s-citalopram, which are widely used drugs in the treatment of psychiatric disorders like depression and anxiety. On the other hand, humic acid had the lowest binding affinity of all ligands tested and may not have as great potential as fulvic acid.

Further studies, such as molecular dynamics simulation, in vitro, and in vivo studies are necessary to prove the efficacy of fulvic acid as an antidepressant drug.

REFERENCES

Berman, H. M. (1 jan 2000). The Protein Data Bank. *Nucleic Acids Research*, v. 28, n. 1, p. 235–242.

Bhavsar, S. K., Thaker, A. M. and Malik, J. K. (2016). *Shilajit. Nutraceuticals*. Elsevier. p. 707–716.

Borroto-Escuela, D., Narváez, M., Ambrogini, P., et al. (3 jun 2018). Receptor–Receptor Interactions in Multiple 5-HT_{1A} Heteroreceptor Complexes in Raphe-Hippocampal 5-HT Transmission and Their Relevance for Depression and Its Treatment. *Molecules*, v. 23, n. 6, p. 1341.

Chien, S.-J., Chen, T.-C., Kuo, H.-C., Chen, C.-N. and Chang, S.-F. (dec 2015). Fulvic acid attenuates homocysteine-induced cyclooxygenase-2 expression in human monocytes. *BMC Complementary and Alternative Medicine*, v. 15, n. 1.

Davis, B. A., Nagarajan, A., Forrest, L. R. and Singh, S. K. (1 apr 2016). Mechanism of Paroxetine (Paxil) Inhibition of the Serotonin Transporter. *Scientific Reports*, v. 6.

Eswar, N., Webb, B., Marti-Renom, M. A., et al. (sep 2006). Comparative Protein Structure Modeling Using Modeller. *Current Protocols in Bioinformatics*, v. 15, n. 1, p. 5.6.1-5.6.30.

Grandjean, J., Corcoba, A., Kahn, M. C., et al. (dec 2019). A brain-wide functional map of the serotonergic responses to acute stress and fluoxetine. *Nature Communications*, v. 10, n. 1.

Kim, S., Chen, J., Cheng, T., et al. (8 jan 2019). PubChem 2019 update: improved access to chemical data. *Nucleic Acids Research*, v. 47, n. D1, p. D1102–D1109.

Marek, G. J. (2017). Developing Serotonergic Antidepressants Acting on More Than the Serotonin Transporter. In: Handler, N.; Buschmann, H.[Eds.]. *Methods and Principles in Medicinal Chemistry*. Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA. p. 335–367.

Matthäus, F., Haddjeri, N., Sánchez, C., et al. (2016). The allosteric citalopram binding site differentially interferes with neuronal firing rate and SERT trafficking in serotonergic neurons. *European Neuropsychopharmacology: The Journal of the European College of Neuropsychopharmacology*, v. 26, n. 11, p. 1806–1817.

Morris, G. M., Huey, R., Lindstrom, W., et al. (dec 2009). AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *Journal of computational chemistry*, v. 30, n. 16, p. 2785–2791.

Palacios, J. M. (15 aug 2016). Serotonin receptors in brain revisited. *Brain Research*, v. 1645, p. 46–49.

Pettersen, E. F., Goddard, T. D., Huang, C. C., et al. (oct 2004). UCSF Chimera: A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, v. 25, n. 13, p. 1605–1612.

Plácido, A., Coelho, A., Abreu Nascimento, L., et al. (jul 2017). Cry1A(b)16 toxin from *Bacillus thuringiensis*: Theoretical refinement of three-dimensional structure and prediction of peptides as molecular markers for detection of genetically modified organisms: Cry1A(b)16 Structure and Prediction of Peptides for Detection of GMO. *Proteins: Structure, Function, and Bioinformatics*, v. 85, n. 7, p. 1248–1257.

Qin, Y., Zhang, M., Dai, W., et al. (sep 2019). Antidiarrhoeal mechanism study of fulvic acids based on molecular weight fractionation. *Fitoterapia*, v. 137, p. 104270.

Saleh, F. Y., Ong, W. A. and Chang, D. Y. (15 dec 1989). Structural features of aquatic fulvic acids. Analytical and preparative reversed-phase high-performance liquid chromatography separation with photodiode array detection. *Analytical Chemistry*, v. 61, n. 24, p. 2792–2800.

Sherry, L., Jose, A., Murray, C., et al. (2012). Carbohydrate Derived Fulvic Acid: An in vitro Investigation of a Novel Membrane Active Antiseptic Agent Against *Candida albicans* Biofilms. *Frontiers in Microbiology*, v. 3.

Thurman, E. M. and Malcolm, R. L. (apr 1981). Preparative isolation of aquatic humic substances. *Environmental Science & Technology*, v. 15, n. 4, p. 463–466.

Tohda, M. (2014). Serotonin 2C Receptor as a Superhero: Diversities and Talents in the RNA Universe for Editing, Variant, Small RNA and Other Expected Functional RNAs. *Journal of Pharmacological Sciences*, v. 126, n. 4, p. 321–328.

Trott, O. and Olson, A. J. (30 jan 2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry*, v. 31, n. 2, p. 455–461.

Verma, S., Singh, A. and Mishra, A. (jan 2013). The effect of fulvic acid on pre- and postaggregation state of A β 17–42: Molecular dynamics simulation studies. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, v. 1834, n. 1, p. 24–33.

Vucskits, A. V., Hullár, I., Bersényi, A., et al. (dec 2010). Effect of fulvic and humic acids on performance, immune response and thyroid function in rats: Effects of fulvic acid and humic acid in rats. *Journal of Animal Physiology and Animal Nutrition*, v. 94, n. 6, p. 721–728.

WHO (22 mar 2018). Depression. <https://www.who.int/news-room/fact-sheets/detail/depression>, [accessed on Jul 30].

Yohn, C. N., Gergues, M. M. and Samuels, B. A. (dec 2017). The role of 5-HT receptors in depression. *Molecular Brain*, v. 10, n. 1.

Zhao, Y., Paderu, P., Delmas, G., et al. (oct 2015). Carbohydrate-derived fulvic acid is a highly promising topical agent to enhance healing of wounds infected with drug-resistant pathogens. *The Journal of Trauma and Acute Care Surgery*, v. 79, n. 4 Suppl 2, p. S121-129.

MODELO PARA DETERMINAR PERFIS DE DESEMPENHO ACADÊMICO NA UNNE COM MINERAÇÃO DE DADOS EDUCACIONAIS

Data de aceite: 27/08/2020

Julio César Acosta

Facultad de Ciencias Exactas y Naturales
y Agrimensura - Facultad de Ciencias
Agrarias. Universidad Nacional del Nordeste.
Corrientes, Argentina.

David Luis La Red Martínez

Facultad de Ciencias Exactas y Naturales
y Agrimensura. Universidad Nacional del
Nordeste.
Corrientes, Argentina.

RESUMEN: En este trabajo se propone evaluar el rendimiento de los estudiantes mediante técnicas de Minería de Datos. La propuesta no se enfoca en analizar el perfil del estudiante solo a través de sus calificaciones, sino también, estudiar el desempeño académico en base a otras variables. Para definir los perfiles de los estudiantes y determinar patrones que conduzcan al éxito o fracaso académico, implementaremos un modelo que relaciona las calificaciones de los estudiantes con otras variables, tales como factores socioeconómicos, demográficos, actitudinales, entre otros; en base a lo cual clasificaremos los diferentes perfiles de alumnos. Describimos el modelo a implementar con el uso de Data Warehouse para determinar los perfiles de rendimiento académico en las asignaturas Algebra de la carrera Licenciatura en Sistemas de Información (LSI) de la Facultad de Ciencias Exactas y Naturales y Agrimensura (FaCENA) de la Universidad Nacional del Nordeste (UNNE) y Matemática I de la carrera Ingeniería

Agronómica (IA) de la Facultad de Ciencias Agrarias (FCA) de la UNNE (PI 16F002 acreditado por Res. N° 970/16 CS). Esperamos contribuir a encontrar una respuesta al bajo rendimiento académico de los alumnos observado históricamente, problema éste que es el disparador de nuestra investigación. Los modelos predictivos que buscamos, permitirán tomar acciones tendientes a evitar el fracaso académico, detectando los alumnos con perfil de riesgo de fracaso académico de manera temprana, a poco del inicio del cursado de las asignaturas; lo que permitirá concentrar en ellos los esfuerzos de tutorías y apoyos especiales.

PALABRAS CLAVE: rendimiento académico; almacenes de datos; minería de datos; modelos predictivos.

MODEL FOR DETERMINING ACADEMIC PERFORMANCE PROFILES AT UNNE WITH EDUCATIONAL DATA MINING

ABSTRACT: In this work it is proposed to evaluate the performance of the students using Data Mining techniques. The proposal does not focus on analyzing the student's profile only through their grades, but also studying academic performance based on other variables. To define the profiles of the students and determine patterns that lead to academic success or failure, we will implement a model that relates the students' qualifications with other variables, such as socioeconomic, demographic, and attitudinal factors, among

others; based on which we will classify the different student profiles. We describe the model to be implemented with the use of the Data Warehouse to determine the academic performance profiles in the Algebra subjects of the Bachelor's Degree in Information Systems (LSI) of the Faculty of Exact and Natural Sciences and Surveying (FaCENA) of the National University of the Northeast (UNNE) and Mathematics I of the Agricultural Engineering (IA) degree of the Faculty of Agricultural Sciences (FCA) of the UNNE (PI 16F002 accredited by Res. No. 970/16 CS). We hope to contribute to finding an answer to the historically observed low academic performance of students, a problem that is the trigger for our research. The predictive models we are looking for will allow us to take actions aimed at avoiding academic failure, detecting students with a risk profile of academic failure early, shortly after the start of the course; This will make it possible to focus on them the efforts of tutoring and special supports.

KEYWORDS: academic performance; data warehouses; data mining; predictive models.

INTRODUCCIÓN

En el mejoramiento de la calidad académica en la Universidad, no necesariamente debe enfocarse sólo en el sistema de enseñanza-aprendizaje, sino que debe atender otras variables, como por ejemplo, la sistematización de procesos de evaluación permanentes que permitan monitorear cuestiones ligadas a la calidad académica y retroalimente la propuesta de mejora para la Universidad (BRIAND, DALY y WÜST, 1999). Uno de los factores más críticos que debe evaluarse continuamente es el rendimiento académico. Se define al rendimiento académico como la productividad del sujeto, matizado por sus actividades, rasgos y la percepción más o menos correcta de los cometidos asignados (MALETIC, COLLARD y MARCUS, 2002). Al evaluar el rendimiento académico se analizarán elementos que influyen en el desempeño como: los factores socioeconómicos, la amplitud de programas de estudio, las metodologías de enseñanza, los conocimientos previos del alumno (MARCUS, 2003); por esto, no resulta adecuado evaluar el desempeño general de los alumnos a través de porcentajes de aprobación, notas obtenidas, etc., ya que este proceso de evaluación no brinda toda la información necesaria que pueda ser utilizada para detectar, y corregir problemas cognitivos, de aprehensión, de discernimiento, actitudinales. Implementamos un mecanismo que nos permite determinar las características propias del estudiante analizando la existencia de patrones de comportamiento y de condiciones de los estudiantes que posibiliten la definición de los perfiles de alumnos. Actualmente existen varios métodos para determinar y clasificar patrones que se utilizan en el área de la Inteligencia Artificial y del Aprendizaje de Máquinas (del inglés Machine Learning – ML) (MARCUS y MALETIC, 2003). La Minería de Datos (del inglés Data Mining - DM), son procesos de descubrimiento de nuevas y significativas relaciones, patrones y tendencias en grandes volúmenes de datos utilizando técnicas de AI y ML. Estas técnicas permiten extraer patrones y tendencias para describir y comprender mejor los datos y predecir comportamientos futuros. Un DW es una colección de datos orientada a un dominio, integrada, no volátil y variante en el tiempo para ayudar a tomar decisiones (SALTON,

1989). Los DW surgieron por dos razones: a) la necesidad de proporcionar una fuente única de datos limpia y consistente para propósitos de apoyo para la toma de decisiones; b) la necesidad de hacerlo sin afectar a los sistemas operacionales (MOLINA LOPEZ y GARCÍA HERRERO, 2006). En este trabajo se utilizaron de técnicas de DM, con volúmenes no muy grandes de datos que oscilan de cientos a miles, sobre la información del desempeño de los alumnos de las cátedras Álgebra (LSI) FaCENA- UNNE y Matemática I de la FCA- UNNE. Describimos el contexto de la experiencia y el modelo de DWH usado en la misma.

MATERIALES Y MÉTODOS

Trabajamos para detectar grupos de estudiantes en riesgo de fracaso en sus estudios, a fin de adoptar acciones proactivas frente al desgranamiento y el bajo rendimiento académico de los alumnos de primer año en la Universidad. La experiencia se realiza en las asignaturas Álgebra de la carrera LSI de la FaCENA de la UNNE y en Matemática I de la carrera IA de la FCA de la UNNE.

Si bien ambas asignaturas tienen régimen de acreditación similar, difieren en la carga horaria y los tiempos de dictado a saber: Álgebra (LSI) tiene 128 (ciento veintiocho) horas reloj de dictado de las cuales el 50% corresponde a teoría y el 50 % a trabajos prácticos en la modalidad cuatrimestral (corresponde al primer cuatrimestre de primer año de la carrera), mientras Matemática I (IA) tiene 96 (noventa y seis) horas reloj de dictado con idéntica distribución porcentual de tiempos de dictado de teoría y de trabajos prácticos, pero en la modalidad trimestral (corresponde al primer trimestre de primer año de la carrera).

En ambas asignaturas para alcanzar la condición de alumno regular, los alumnos deben asistir al menos al 75% de las clases de trabajos prácticos, que se dictan dos veces por semana en clases de 2 hs. cada una y deben aprobar 2 (dos) exámenes parciales cuyos contenidos son exclusivamente de trabajos prácticos; cada uno de ellos tiene su instancia de recuperación y para aquellos alumnos que hayan aprobado al menos 1 (uno) de los parciales en cualquiera de las 4 (cuatro) instancias disponibles, existe una instancia más para recuperar el examen que queda aún sin aprobar. Cualquiera de los exámenes parciales se aprueba con 60 (sesenta) puntos sobre 100 (cien) puntos posibles. La asistencia a clases de teoría es libre y se dictan dos veces por semana en clases de 2 hs. cada una.

Se acreditan las asignaturas con un examen final al que se accede en condición de alumno regular o de alumno libre; el alumno regular debe rendir en el examen final solamente los contenidos de teoría en un examen oral. El alumno que se presenta al examen final en condición de alumno libre, debe rendir un examen escrito de trabajos prácticos y tras aprobar esa instancia pasa al examen de teoría en condiciones similares a la antes mencionada.

Los porcentuales de los alumnos que regularizan las Algebra y Matemática I no son los deseados; en el caso de Algebra, de 320 alumnos inscriptos en los últimos 4 años, aproximadamente un 20% no alcanza a rendir el primer examen parcial en promedio y al final del cursado, regularizan la asignatura solo un 30% aproximadamente, en el caso de Matemática I el desgranamiento después del primer parcial no es tan evidente y el porcentual aproximado de alumnos regulares al final del cursado es del 40%.

La cantidad de alumnos que regularizan y/o que aprueban las asignaturas involucradas en este proyecto no es satisfactoria, consideramos que esa situación puede contribuir al desgranamiento y deserción de los alumnos en los primeros niveles de sus carreras. Es importante, por tanto, estudiar y determinar cuáles son las variables que inciden en el rendimiento académico a fin de poder establecer estrategias de acción pedagógicas que permitan mejorar dicho rendimiento.

Trabajamos principalmente en el desarrollo de métodos que contribuyan a encontrar técnicas para la detección temprana de los alumnos que tendrán dificultades en sus estudios, a fin ofrecerles una contención y acompañamiento especial en el inicio de sus estudios Universitarios. Indagaremos aspectos tales como: a) diferencia del nivel de aprendizajes de contenidos previos en los alumnos, b) situaciones particulares personales de los propios alumnos, c) la capacidad de las cátedras para el seguimiento del aprendizaje de los alumnos, d) escasa motivación para el estudio de ciencias básicas y otros que puedan revelarse como incidentes en la problemática que nos ocupa y otros que serán detallados adelante.

Para recuperar contenidos en los grupos de riesgo detectados trabajamos con materiales elaborados con nuevas tecnologías de la información (NTIC), que en la actual situación de pandemia se han revelado de gran utilidad. Mas allá del uso que actualmente se hace de dichos materiales, esto no debe desplazar ni sustituir las formas presenciales de enseñanza - aprendizaje, sino más bien situarnos en la posición de ofrecer alternativas diferentes para aquellos alumnos que requieren modelos diferentes para sus estudios y aprendizajes. Consideramos que las NTIC tienen el potencial para desempeñar un papel importante en la recuperación de contenidos al permitir un abordaje más eficaz, en el sentido de permitirnos procesos de aprendizaje más profundos y más persistentes (MOTSCHNIG-PITRIK y HOLZINGER, 2002), mientras el peso de un aprendizaje efectivo permanece con las personas, sus capacidades y valores interpersonales (DERNTL, HAMPEL, MOTSCHNIG-PITRIK y PITNER, 2011). En tal sentido, entendemos importante en nuestro trabajo el estudio que se efectuará en dos poblaciones aparentemente diferentes como son los alumnos de las carreras Licenciatura en Sistemas de Información y los alumnos de Ingeniería Agronómica, para determinar si los perfiles de los estudiantes varían según la elección de la carrera y medir las diferencias en la predisposición y adaptación para el trabajo y aprendizaje mediado con las NTICs, tarea en la que nos encontramos trabajando.

En los últimos años se han realizado numerosos trabajos relacionados con la

producción de contenidos; actualmente se tiene una concepción global e integral del e-learning (NICHOLS, 2003), en estos nuevos escenarios se incluyen la combinación del aprendizaje cara a cara y el soportado por medios tecnológicos (especialmente la Web), tal que las fortalezas de ambas configuraciones se puedan aprovechar y explotar. Este aprendizaje combinado (blended learning o b-learning) se considera de suma utilidad no sólo para las universidades sino también para la sociedad en general.

Desde nuestros trabajos previos, hemos podido corroborar lo que oportunamente hemos formulado, que los docentes del siglo XXI deben incorporar definitivamente las NTICs como recursos didácticos, sin abandonar los tradicionales de tiza y pizarrón, pero deben conocer el uso de las NTICs con al menos en parte del potencial que ellas ofrecen (ACOSTA y LA RED MARTINEZ, 2012); algunas teorías psicológicas y pedagógicas consideran necesaria la inclusión del e-moderador o e-moderador, docente con habilidades especiales en las actividades online (SALMON, 2000); la actividad del docente tutor se transforma a veces en un hecho fundamental, ya que la manera en que se usa la tecnología puede transformarse en un factor de gran influencia en la calidad de la EA- EV (enseñanza - aprendizaje en entornos virtuales). Se debe trabajar entonces para lograr una forma de EA-EV que tome en cuenta las necesidades individuales, los intereses y estilos (Wenger, White y Smith, 2009).

En este proyecto de investigación, las variables que inciden en el rendimiento académico de los alumnos se han detectados a fin de establecer, a través de los valores que ellas toman en cada caso, la población de alumnos en riesgo de fracaso, para establecer acciones tendientes a evitar el fracaso de cada uno de los alumnos, con las acciones que correspondan en cada caso particular y/o de cada grupo detectado y disminuir así el posterior desgranamiento.

DATA WAREHOUSE

Como soporte de los datos trabajamos con Data Warehouse (DW); en informática, un almacén de datos (DW), es un sistema especial de bases de datos utilizado para el almacenamiento de datos y el procesamiento de los mismos para la presentación de informes y análisis de información, es considerado como un componente central de la inteligencia de organizaciones.

Un DW es un repositorio de datos que proporciona una visión global, común e integrada de los datos (CURTO DIAZ, 2010) y presenta las siguientes características: a) Orientado a un tema: organiza una colección de información alrededor de un tema central. b) Integrado: incluye datos de múltiples orígenes y presenta consistencia de datos. c) Variable con el tiempo: se realizan fotos de los datos basadas en fechas o hechos. d) No volátil: sólo de lectura para los usuarios finales.

Detrás de la arquitectura de componentes del DW existe un conjunto de procesos básicos asociados: los ETL (del inglés Extract, Transform, Load – Extracción, Transformación y Carga). Los procesos ETL hacen referencia a la recuperación y transformación de los datos desde las fuentes orígenes cargándolos en el DW. En primer lugar los datos se analizan desde las fuentes y se extraen aquellos que serán de utilidad para el proceso en ejecución.

Luego de extraer los datos se los carga al DW pero, en muchas ocasiones, éstos requieren pasar por un proceso de transformación. La transformación de los datos significa un formateo y/o estandarización de los mismos convirtiendo ciertos números en fechas, eliminando campos nulos, etc.

Es necesario que antes de completar el DW con los datos se realicen controles para enviar información cualitativamente correcta. Luego se procede a aplicar alguna técnica para realizar el análisis de los datos almacenados en el DW. El método más utilizado es el proceso de DM que aplica la inteligencia artificial para encontrar patrones y relaciones dentro de los datos permitiendo la creación de modelos, es decir, representaciones abstractas de la realidad.

Existen varias alternativas del DM, por ejemplo la Minería de Datos en Educación (Educational Data Mining, EDM). El objetivo de la EDM es el desarrollo de métodos para la exploración de tipos de datos únicos provenientes de plataformas educativas y usándolos para entender mejor a los estudiantes en el aprendizaje (BAKER y YACEFF, 2009). Existen diversos estudios y publicaciones que abordan la evaluación de rendimiento académico utilizando técnicas de Minería de Datos (FORMIA y LANZARINI, 2013); (PEREIRA, ROMERO y TOLEDO, 2013); (LA RED MARTINEZ, ACOSTA, URIBE y RAMBO, 2012); (LA RED MARTINEZ, GIOVANNINI, BAEZ MOLINAS, TORRE y YACCUZZI, 2017).

Modelo propuesto: La estructura del DW consta de una tabla de hechos y varias tablas de dimensión. Una tabla de hechos o una entidad de hecho es una tabla o entidad que almacena medidas para medir el negocio como las ventas, el coste de las mercancías o las ganancias (IBM KNOWLEDGE CENTER, 2015).

Cada intersección de valores de las dimensiones y generalmente se trata de cantidades numéricas, continuamente evaluadas y aditivas. Se pueden distinguir dos tipos de columnas en una tabla de hechos, columnas de hechos y columnas llaves. Las columnas de hechos almacenan las medidas del negocio que se quieren controlar y las columnas llaves forman parte de la clave de la tabla. Una tabla de dimensiones o entidad de dimensiones es una tabla o entidad que almacena detalles acerca de hechos. Por ejemplo una tabla de dimensión de hora almacena los distintos aspectos del tiempo como el año, trimestre, mes y día.

Además incluye información descriptiva sobre los valores numéricos de una tabla de hechos. Las tablas de dimensiones para una aplicación de análisis de mercado, por ejemplo, pueden incluir el tipo de período de tiempo, región comercial y producto. Asimismo

las tablas de dimensiones describen los distintos aspectos de un proceso de negocio. Si se desea determinar los objetivos de ventas, se pueden almacenar los atributos de dichos objetivos en una tabla de dimensiones. Cada tabla de dimensiones contiene una clave simple y un conjunto de atributos que describen la dimensión.

En nuestro caso, las columnas de una tabla de dimensiones se utilizan para crear informes o para mostrar resultados de consultas. Por ejemplo las descripciones textuales de un informe se crean desde las etiquetas de las columnas de una tabla de dimensiones. El modelo que se presenta en este trabajo se compone de la tabla de hechos “ALUMNOS” y varias tablas de dimensiones asociadas a la misma que incluyen características que se desean estudiar.

Etapas de recolección de datos: Tal como se planteó, el estudio del desempeño académico de los estudiantes no sólo debe evaluarse teniendo en cuenta los resultados de las instancias de evaluaciones previstas por la asignatura sino que también deben analizarse otros factores culturales, sociales y/o económicos que afecten el rendimiento del alumno. Por ello para este trabajo es determinante la participación directa del estudiante, pues es necesario conocer datos sobre aspectos personales que no se pueden obtener de otra manera que no fuera a través de respuestas directas por parte de cada alumno. A tal fin se dispuso la elaboración de una aplicación web que permitió contar con una Encuesta On-Line compuesta por preguntas relacionadas a situación familiar e historial de estudios secundarios, entre otras cuestiones.

Etapas de depuración y preparación de datos: Para la realización de una correcta explotación del DW se debe asegurar que los datos obtenidos en la etapa anterior sean consistentes y mantengan la coherencia entre ellos. Así, en la etapa siguiente, se realiza un proceso de limpieza en los datos, que es la eliminación de aquellos registros con todos sus campos en blanco, corrección de errores tipográficos, llenado de algunos campos nulos, entre otros. La Encuesta no permite la carga, por parte de los estudiantes, de calificaciones de la asignatura en estudio. Esto se dispuso así para evitar errores en los datos ya sea por olvido, o confusión al momento de ingresar los valores. Por ello la carga de notas correspondientes al primer parcial, segundo parcial y sus recuperatorios, examen final y situación del alumno (regular, promovido o libre), es realizada por el equipo responsable de este trabajo de investigación. La información se obtiene a partir de la base de datos histórica de las cátedras continuará respecto a calificaciones de los alumnos. Con esta información depurada se procede a trabajar en las próximas etapas: - Carga de Datos al DW: Mediante la ejecución del flujo de datos, la información almacenada en la tabla *encuesta* se distribuirá a las tablas pertenecientes al modelo del DW.

RESULTADOS

Hasta el momento se completó el diseño del modelo del DW sobre el cual se implementaron técnicas de DM a fin de determinar perfiles de estudiantes vinculados a su desempeño académico en las asignaturas LSI-FaCENA e IA-FCA UNNE. En el avance que aquí se presenta respecto del Proyecto se pudo comprobar que la etapa de depuración y preparación de los datos ha demandado tiempo y esfuerzo debido principalmente a la poca integridad y coherencia en la información que se utilizará para realizar la evaluación final. Nos encontramos trabajando con el proceso de minería de datos para evaluar y comparar los patrones que se obtengan para definir los perfiles de estudiantes. La evaluación, análisis y utilidad de estos patrones con los que terminaremos construyendo un modelo predictivo de rendimiento académico permitirá soportar la toma de decisiones eficaces por parte del cuerpo docente de las asignaturas involucradas.

REFERENCIAS

ACOSTA, J. y LA RED MARTINEZ, D. **Un aula virtual no convencional de Algebra en la FaCENA-UNNE**. Saarbrücken: Editorial Académica Española. 2012

BAKER, R. y YACEF, K. The State of Educational Data Mining in 2009: A Review and Future Visions. **Journal of Educational Data Mining**, 1, 3-16. 2009.

BRIAND, L.; DALY, J. y WÜST, J. A unified framework for coupling measurement in objectoriented systems. **IEEE Transactions on Software Engineering**, 25 (1), 91-121. 1999.

CURTO DÍAZ, J. **Introducción al business intelligence**. UOC: Barcelona. 2010.

DERNTL, M.; HAMPEL, T.; MOTSCHNIG-PITRIK, R. y PITNER, T. Inclusive social tagging and its support in Web 2.0 services. **Computers in Human Behavior**, 27(4), 1460-1466. 2011.

FORMIA, S. y LANZARINI, L. Caracterización de la deserción universitaria en la UNRN utilizando minería de datos. **Revista Iberoamericana de Tecnología en Educación y Educación en Tecnología (TE&ET)**, (11):92-98. 2013.

LA RED MARTINEZ, D.; ACOSTA, J.; URIBE, V. y RAMBO A. Academic Performance: An Approach From Data Mining. **Journal of Systemics, Cybernetics and Informatics**, 10 (1), 66-72. 2012.

LA RED MARTINEZ, D.; GIOVANNINI, M.; BAEZ MOLINAS, M.; TORRE, J. y YACCUZZI, N. Academic performance problems: A predictive data mining-based model. **Academia Journal of Educational Research**; 5 (4), 61-75. 2017.

MALETIC, J.; COLLARD, M. y MARCUS, A. Source Code Files as Structured Documents. Proceedings 10th IEEE International Workshop on Program Comprehension (IWPC'02), 289- 292. París. 2002.

MARCUS, A. **Semantic Driven Program Analysis**, Kent, OH, USA, Kent State University Doctoral Thesis. 2003.

MARCUS, A. y MALETIC, J. Recovering Documentation-to-Source-Code Traceability Links using Latent Semantic Indexing. Proceedings 25th IEEE/ACM International Conference on Software Engineering (ICSE'03). 3(10), 125-137. Portland, Oregon. 2003.

MOLINA LOPEZ, J. Y GARCIA HERRERO, J. **Técnicas de Análisis de Datos**. Madrid: Universidad Carlos III. 2006.

MOTSCHING-PITRIK, R. y HOLZINGER, A. Student-centered teaching meets new media: concept and case study. **Journal of Educational Technology and Society**, 5(4), 160-172. 2002.

NICHOLS, M. A theory for e-Learning. **Journal of Educational Technology and Society**, 2, 1-10. 2003.

PEREIRA, R., ROMERO, A. y TOLEDO J. Descubrimiento de perfiles de deserción estudiantil con técnicas de minería de datos. **Vínculos**. (10) 1, 374-383 . 2013.

SALMON, G. **E-moderating: The key to teaching and learning online**. London: Kogan Page. 2000.

SALTON, G. **Automatic Text Processing: The Transformation, Analysis and Retrieval of Information by Computer**. Boston: Addison- Wesley Longman Publishing Co. 1989.

WENGER, E.; WHITE, D. y SMITH, J. **Digital habitats**. Stewarding technology for communities: Portland, OR, USA: Cpsquare. 2009.

CAPÍTULO 7

O USO DO CONTROLADOR LÓGICO PROGRAMÁVEL (CLP)

Data de aceite: 27/08/2020

Data de submissão: 04/08/2020

Viviane Alencar Marques Araújo do Nascimento

Especialista pela Universidade Candido Mendes – UCAM
Rio Branco – Acre
<http://lattes.cnpq.br/1206822539769154>

RESUMO: O Controlador Lógico Programável (CLP) é um dispositivo eletrônico onde são armazenados programas conforme a que se destina sua aplicação, também armazena comandos que são recebidos pela rede pelas entradas, e seu uso foi desenvolvido para as indústrias e automação residencial recentemente com o avanço da tecnologia. A pesquisa teve o objetivo de analisar o uso do CLP, apresentando sua composição (estrutura, programação e linguagem Ladder), montagem, vantagens e desvantagens, exemplo de tipos e fabricantes. A pesquisa caracteriza-se como descritiva, onde se utilizou dados documentais de meio eletrônico sobre o uso do CLP. A forma de abordagem adotada foi de estudo com aspecto qualitativo. O resultado da pesquisa apresenta uma abordagem do uso do CLP e descrevendo sua automatização nas indústrias. Alcançou-se com a pesquisa a análise do uso do CLP, em destaque para sua composição e sendo compreendidas as vantagens e desvantagens.

PALAVRAS-CHAVE: Controlador Lógico Programável. CLP. Ladder. Automação.

USE OF THE PROGRAMMABLE LOGIC CONTROLLER (PLC)

ABSTRACT: The Programmable Logic Controller (PLC) is an electronic device where programs are stored according to which their application is intended, it also stores commands that are received by the network through the inputs, and its use was developed for industries and home automation recently with the advancement of technology. The research aimed to analyze the use of PLC, presenting its composition (structure, programming and Ladder language), assembly, advantages and disadvantages, examples of types and manufacturers. The research is characterized as descriptive, using documentary data from electronic media on the use of the PLC. The approach adopted was a qualitative study. The result of the research presents an approach to the use of PLC and describing its automation in industries. The analysis of the use of the PLC was achieved with the research, with emphasis on its composition and the advantages and disadvantages being understood.

KEYWORDS: Programmable Logical Controller. CLP. Ladder. Automation.

INTRODUÇÃO

O presente artigo tem por objetivo analisar o uso do Controlador Lógico Programável (CLP). Escolhido na realização da pós-graduação em engenharia com ênfase em sistemas de automação, no ano de 2017. E trata-se de um artigo científico

onde apresenta relevâncias, científica, profissional e acadêmica.

Com essa concepção, apresentam-se questões que orientam o artigo como:

- Qual a história do CLP?
- Qual sua composição? Apresentando estrutura, programação e sua linguagem.
- Como e realizada sua montagem?

Quanto ao mercado de automação industrial é importante ter conhecimento sobre o controlador lógico programável, pois ele será utilizado durante o processo de máquinas e equipamentos dentro de uma empresa.

O olhar ao conceito de automação vem nas empresas uma preocupação para melhoria da produtividade, e redução de custos. Nota-se que o processo sendo automatizado, maior será o benefício da automação na regulação da qualidade do produto, na economia de energia, quebrando pela flexibilidade e segurança de funcionamento, para uma melhoria da produtividade (FERREIRA, 2007).

Matos, (2007, p. 2) define: “O CLP é um equipamento eletrônico composto, basicamente, por uma CPU, memória, módulos de entrada/saída (I/O) e é programável através de uma linguagem específica”.

Conforme o estudo de Angeloni (2013), ele diz que o elemento principal é o Controlador Lógico Programável (CLP), nela são armazenadas programações de endereços, horários e lógicas dos comandos. Onde gerencia e armazena no sistema suas variáveis, gerenciando também os comandos que são transformados em ação. Os comandos são recebidos pela rede ou pelos contatos ligados em suas entradas, analógicas ou digitais.

No que se refere ao Controlador Lógico Programável Azevedo (2013), em seu estudo, reflete que é um sistema eletrônico desenvolvido exclusivamente para uso na indústria, sendo que há outras aplicações atualmente na automação residencial. O hardware é formado de um CPU e memória RAM, semelhante ao computador, pois possui um software específico para seu hardware. Onde o sistema é bem mais rápido que o do computador. E como o CLP foi construído para o uso industrial ele possui grande robustez.

As aplicações do CLP são variadas, sendo representados através da linguagem *Ladder*, os exemplos diretos são as relacionadas ao acionamento de máquinas elétricas de circuitos de comando de força (TEXEIRA, 2007).

Fiala (2014) destaca em seu estudo sobre a linguagem de programação, que uma das primeiras linguagens para programar o CLP é a *Ladder*. Pois é uma linguagem gráfica, onde se baseia em símbolos semelhantes aos esquemas elétricos, como de contatos e bobinas, assim determinou que o CLP fosse aceito por técnicos e engenheiros habituado com sistema de controle por relé.

Nesse sentido, o objetivo do artigo é analisar o uso do Controlador Lógico Programável (CLP), através de fundamentações teóricas. Na pesquisa para alcançar o objetivo proposto, utilizou-se como recurso bibliográfico, uma revisão e a pesquisa bibliográfica, realizando estudos, análises e articulando as proposições teóricas das fontes de pesquisa divulgadas em meio eletrônico, como: livros, teses, dissertações, artigos e revistas.

METODOLOGIA

TIPO DE PESQUISA

O artigo tem como fundamentação a pesquisa bibliográfica, pois as fontes de pesquisa são de fundamentações teóricas de outras bibliografias de trabalhos já publicados, conforme Santos afirma:

O conjunto de materiais escritos/gravados, mecânica ou eletronicamente, que contém informações já elaboradas e publicadas por outros autores é uma bibliografia. São fontes bibliográficas os livros (de leitura corrente ou de referência, tais como dicionários, enciclopédias, anuários etc.), as publicações periódicas (jornais, revistas, panfletos etc.), fitas gravadas de áudio e vídeo, páginas de web sites, relatórios de simpósios/seminários, anais de congressos etc. A utilização total ou parcial de quaisquer destas fontes é o que caracteriza uma pesquisa bibliográfica (SANTOS, 2001, p.29).

TIPO DE ESTUDO

O artigo caracteriza-se com a pesquisa descritiva, pois foram utilizados dados documentais de meio eletrônico sobre o uso do CLP. A Abordagem do estudo envolve aspecto qualitativo, onde houve uma análise bibliográfica.

MATERIAL E NATUREZA

Os materiais e natureza utilizados como fonte de dados para pesquisa foram: livros técnicos, apostilas e produções científicas (monografia, dissertação, artigo científico e revistas), conforme disponível em ambientes eletrônicos (cf. na lista de referências).

PROCEDIMENTOS

As inclusões das produções científicas foram provenientes de buscas em ambientes eletrônicos e utilizando as seguintes palavras-chave: CLP, composição do CLP, controlador lógico programável, estrutura do CLP, linguagem ladder, montagem do CLP.

O procedimento da busca do material foi de forma manual e eletrônica (disponível em meio eletrônico, cf. lista de referências). Os materiais da pesquisa foram selecionados considerando os seguintes critérios de seleção:

- a) informações sobre o uso do CLP;
- b) descrição do CLP de forma direta e indireta nas produções científicas;
- c) analisar a composição do CLP, bem como estrutura, programação e linguagem de programação adequada;
- d) informações referente a montagem do CLP.

METODOLOGIA DE ANÁLISE DE DADOS

Realizou-se dois métodos na análise de dados:

I. Fase de análise de dados: após a coleta do material foi realizado, leitura previa do material, a seleção da literatura e a produção textual.

II. Fase do método de análise de dados: foi conforme a forma de abordagem da pesquisa qualitativa, que houve a produção textual considerando a leitura de estudo com produção de resumos, citações científicas e tópicos discursivos. Sendo que a produção envolveu paralelos entre autores.

O CLP

A área de automação cresce constantemente a cada dia, pois é bem divulgado nos meios de comunicação, e também sendo a realidade nas indústrias que veem a agilidade e automatização das tarefas. Ao realizar a automação, podem-se utilizar Controladores Lógicos Programáveis (CLP) eles são equipamentos designados para executar tarefas de Inter travamento, temporização, contagem, operações matemáticas, controle em malha aberta ou malha fechada e supervisão em máquinas com processos de médio e grande porte, controlando plantas industriais, redução do espaço nas instalações e facilidade em atividades de reprogramação de fábrica e na manutenção (BALLOCK, 2003).

A automação nos dias de hoje com o avanço da tecnologia vem se destacando das indústrias, e o uso do CLP em processos industriais vem sendo a solução para automatizar atividades com facilidade.

O CLP (figura 1), conforme Fiala (2014) é constituído por blocos, são eles: CPU, módulos I/Os, fonte de alimentação e base.



Figura 1 - CLP

Fonte: Fiala (2014) apud Siemens 2014.

HISTÓRIA DO CLP

Para a compreensão do que é um CLP, precisamos entender como foi seu surgimento? Qual a sua história? A seguir assuntos que resalvam esses questionamento.

Segundo Ballock (2003), o primeiro CLP surgiu na industria automobilítica, onde um usuário possível dos relés eletromagnéticos empregados para controlar operações de sequência e repetições em uma linha de montagem. Assim, a primeira geração de CLPs utilizou componentes como transistores e CIs com baixa escala de integração. Portanto o equipamento foi denominado Programmable Logic Controller (PLC) nos Estados Unidos, e Controlador Lógico Programável (CLP) tradução para português, onde o termo é registrado pelo fabricante de CLP's Allen Bradley.

A industria automobilistica procurou adqur suas operações na linha de montagem, isso mostra que a utilização do CLP na industria e importante. E seu surgimento não poderia ser melhor que sua propria aplicação.

Na década de 80 o CLP cresceu na indústria no Brasil, especialmente pela penetração da tecnologia usadas na matriz das multinacionais. Modernamente, com a redução alta no custo do CLP, nota-se o crescimento da utilização nas indústrias no geral, seja ela grande, pequena ou ramos diferentes de atividades (BALLOCK, 2003).

Ver-se que no brasil o CLP cresce assim como as tecnologias nas industrias se integram para melhorias nas produções e automação de processos.

A respeito da história do CLP Pertence Júnior (2010), classifica e resume o CLP em cinco gerações, sendo elas:

1ª - Na primeira geração os CLPs caracterizam-se por sua programação que é ligada ao hardware do equipamento. A Assembly era a linguagem utilizada no CLP de acordo com seu projeto porque ela vária, pois para poder programar tinha que conhecer a parte elétrica do projeto do CLP. Onde a programação era realizada por uma equipe técnica, gravando o programa na memória EPROM, era realizada no laboratório junto com o construído do CLP;

2ª - Onde aparecem as primeiras linguagens de programação sem dependência do hardware do equipamento, admissível pela inserção de um programa monitor no CLP, que compila as instruções, verifica as entradas, compara as instruções do usuário, assim aterando as saídas. Os programadores de memória EPROM eram chamdos de Maletas ou terminais de programação. Para o programa do usuário ser executado as memórias eram programadas e colocadas no CLP;

3ª - Aqui passam a ter entrada de programação os CLPs, são conectados um teclado ou um programador portátil, para alterar, apagar, gravar o programa do usuário, também, realiza testes tanto no equipamento quanto no programa. Na estrutura física houve alteração onde a tendencia foi os sistemas modulares com bastidores ou racks;

4ª - A diminuição dos preços com os clones do IBM PC e popularização dos microcomputadores, os CLPs passaram a incluir comunicação serial na entrada. A programação passou a ser realizada nos microcomputadores. Tendo como vantagem, a utilização de varias linguagens, possibilidade de testes e simulações, treinamento e ajuda do software, possibilidade de no micro armazenar vários programas, etc.;

5ª - Os CLPs na época atual tem-se uma preocupação na padronização dos protocolos de comunicação, de forma a oferecer que o equipamento de um fabricante comunica-se com outro equipamento de outro fabricante, sendo eles: controladores de processos, sistemas supervisórios, redes internas de comunicação, etc., assim propociona uma integração onde facilita a automação, para gerenciar plantas industriais mais flexíveis e normatizadas, sendo chamada de globalização.

Composição do CLP

A estrutura, programação, linguagem Ladder utilizada e montagem são o que compõe um CLP para uso em automação. A seguir os autores Gomes Júnior (2009), Silva; Soares; Rosa (2012), Azevedo (2013), Alievi (2008) e Fiala (2014) explanam esses conceitos.

Atualmente existem variedades de tipos de CLPs, determinados para aplicações diferentes, como mais poderosos, são verdadeiros computadores industriais de grande rendimento. Com aplicação em controle de processos para automatizar e monitorar com eficiência conforme os parâmetros do sistema e atuando de maneira a manter o controle do processo (GOMES JÚNIOR, 2009).

ARQUITETURA E ESTRUTURA

A arquitetura do CLP é constituída conforme Gomes Júnior (2009), de um processador principal e de uma memória de armazenamento, descritos abaixo e representados na figura 2 (exemplo de CLP GE FANUC VERSAMAX):

- Processador principal: onde se processa os sinais de entrada e emite os sinais de saída;
- Memória de armazenamento: em que é guardada toda a lógica de processamento dos sinais de entrada e cartões de entrada e saída, onde são ligados aos sensores e atuadores do sistema que está sendo controlado.



Figura 2 – Exemplo do CLP GE FANUC VERSAMAX

Fonte: Gomes Júnior (2009).

Um típico CLP é composto por: um processador ou CPU, onde executa o programa, e armazena as informações para a automação dos processos. Também as decisões serão tomadas com base nos sinais de entrada e de saída e no programa do seu usuário; pontos de entrada e pontos de saída, pontos de controle do sistema; e portas de comunicação, onde permitem a CPU conectar-se a dispositivos (exemplo o de comunicação) (SILVA, 2012).

Nesse sentido Silva; Soares; Rosa ressalta que:

A CPU é responsável por fazer a leitura dos *status* (condições, *status*) dos dispositivos de entrada por meio dos módulos de I/O. Os *status* são armazenados na memória RAM para serem processados pelo programa do usuário. Depois da execução do programa do usuário os *status* são atualizados pelo processador e a lógica de controle é realizada (SILVA; SOARES; ROSA, 2012, p.33).

O Projeto do CLP é de forma modular, vários tipos de módulos podem ser adicionados. Como módulos de comunicação, módulos de entradas e saídas analógicas, etc., O CLP por meio dos módulos pode programar várias funções específicas para controlar diversos processos. Na figura 3, apresenta-se o ciclo básico de funcionamento do CLP (AZEVEDO, 2013).

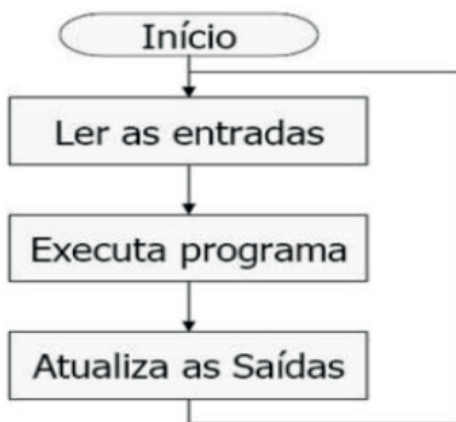


Figura 3 – Ciclo básico do CLP

Fonte: Azevedo (2013).

Quando inicia-se o CLP todos os sinais que existe nas entradas vão ser transferidos para a memória RAM. Assim iniciando o ciclo de scan do software (programa transferido para o CLP) contendo as informações de entrada e executando as lógicas que foram programadas. Quando a execução finaliza as saídas são atualizadas, portanto fechando o ciclo do scan (AZEVEDO, 2013).

PROGRAMAÇÃO E LINGUAGEM LADDER

O programa do CLP integra a lógica onde vai avaliar a condição dos pontos de entrada e os estados anteriores do CLP, fazendo as funções lógicas e acionando as saídas. Os sensores aplicados às entradas do CLP são lidos os sinais por ele mesmo, assim transferi para uma unidade memória (memória imagem). Onde são sinais associados entre si e aos sinais internos. No final do ciclo de varredura, há uma transferência de resultados para a memória imagem de saída, aí sim aplicados aos terminais de saída. Todo esse ciclo apresenta-se na figura 4 (ALIEVI, 2008).



Figura 4 – Processamento do CLP (ciclo)

Fonte: Alievi (2008) *apud* Silva Filho (2008).

Sobre a padronização de linguagem a norma define cinco delas, conforme Fiala (2014, p. 27) descreve: “A norma IEC 61131-3 define cinco linguagens de programação: Ladder; lista de instruções; texto estruturado; diagrama de blocos de função e diagrama funcional sequencial”.

A figura 5 expõe os principais comandos da linguagem Ladder, sendo eles: contato normalmente aberto e normalmente fechado, bobina vazia, bobina reset e bobina set.






- 
 Contato Normalmente Aberto: envia o sinal somente quando é acionado.
- 
 Contato Normalmente Fechado: envia o sinal durante o tempo todo até que seja acionado e, assim, deixar de enviar o sinal.
- 
 Bobina Vazia: a saída só se mantém ativada se um sinal for enviado constantemente para que continue ligada.
- 
 Bobina *Reset*: desliga uma saída quando é excitada com um sinal
- 
 Bobina *Set*: liga uma saída quando é excitada com um sinal.

Figura 5 – Comandos da linguagem Ladder

Fonte: Gomes Júnior (2009).

Para complementar a figura 6 apresenta um exemplo da representação de um programa que foi desenvolvido os comandos da linguagem Ladder, com composições básicas.

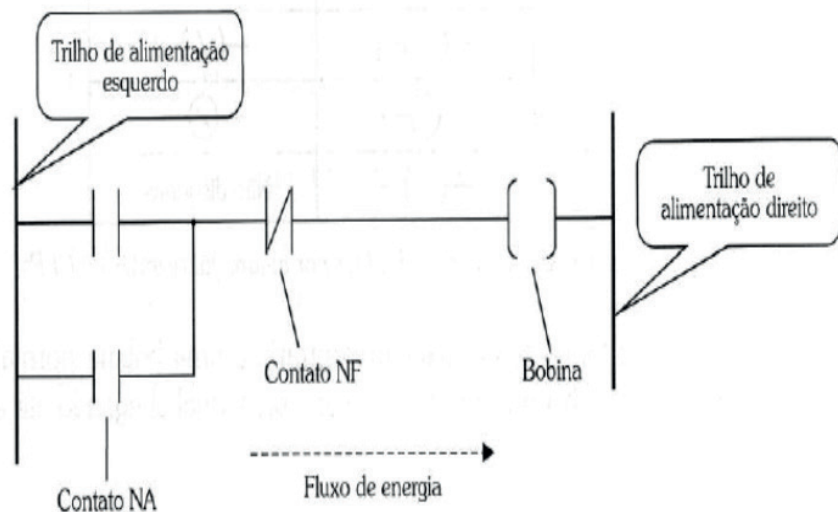


Figura 6 - Exemplo de programa em Ladder

Fonte: Alievi (2014) *apud* Franchi e Camargo (2008).

MONTAGEM

A montagem de um CLP foi destacada um exemplo já citado anteriormente como modelo de CLP, sendo o VersaMax.

O CLP GE Fanuc VersaMax é fabricado pela General Electric Company, sendo utilizado para aplicações industriais, sendo elas: controle e monitoramento de processos e acionamentos de equipamentos. Ele em específico é robusto com relação a ambientes que possui poeira, assim suporta temperaturas de -40°C a 85°C . Foi projetado para montagem em painéis de trilho DIN, onde permite a fixação de módulos e para fácil manutenção. A figura 7 observa-se a forma de montar o CLP (GOMES JÚNIOR, 2009).

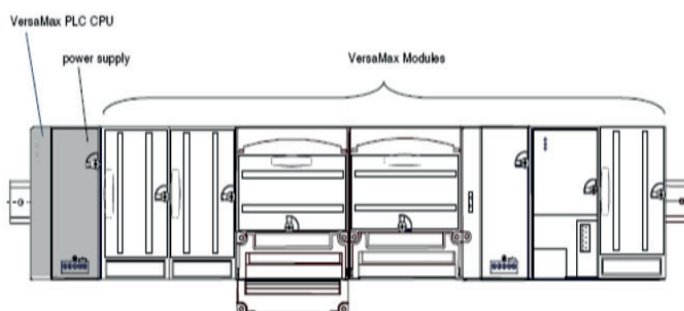


Figura 7 – Montagem do CLP

Fonte: Gomes Júnior (2009) *apud* GE FANUC (2006).

Vantagens e desvantagens do uso do CLP

Conforme Alievi (2008), onde compara os CLPs com os painéis eletromecânicos e destaca suas vantagens:

- Ocupam menor espaço;
- Requerem menor potência elétrica;
- Podem ser reutilizados;
- São programáveis;
- Apresentam maior confiabilidade;
- Manutenção mais fácil e rápida;
- Oferecem maior flexibilidade;
- Permite maior rapidez na elaboração do projeto do sistema.

As desvantagens são destacadas três e comparadas com sistemas domóticos centralizados de controle:

- Grande quantidade de cabeamento;
- Centralização das funções;
- Complicado interfaceamento homem-máquina.

Tipos e fabricantes de CLPs

Os tipos de fabricantes dos CLPs são de grande variedade é para Alievi (2008), atualmente no mercado existe uma quantidade maior de fabricantes e modelos de CLPs, onde cada um tem-se uma abrangência específica de aplicação em função de custo e integração. Assim destaca três categorias de emprego dos CLPs, são elas: de grande porte, médio porte e pequeno porte.

CONCLUSÃO

Sendo assim, considerando a bibliografia consultada foi possível dar-se a análise do uso do CLP no que se refere a sua história, composição, estrutura, programação, a linguagem utilizada Ladder, sua montagem como um exemplo, vantagens e desvantagens comparadas com outros sistemas e tipos e fabricantes.

No conteúdo identificou-se que o uso do CLP na indústria vem crescendo a cada dia com o uso da tecnologia para a automatização dos processos e produção industrial, e sua aplicação é bem utilizada, tanto com os CLPs de grande porte como os de médio e pequeno porte são de grande utilidade.

Com esse contexto sabe-se que para aplicar o uso do CLP tem-se que compreender sua estrutura, programação e linguagem a ser utilizadas de acordo com o as necessidades do projeto.

Além disso, fica evidente que são muitas vantagens para se utilizar o CLP nas indústrias e acrescentar a automação de seus processos, tendo um ambiente mais automatizado e estruturado.

REFERÊNCIAS

ALIEVI, César Adriano. **Automação residencial com utilização de controlador lógico programável**. 2008. 84p. Dissertação (monografia). Centro Universitário Feevale, Instituto de Ciência da Computação. Novo Hamburgo – RS, 2008.

ANGELONI, Guilherme Campos. **Automação residencial**. 2013. 91p. Dissertação (monografia). Departamento de Automação e Sistemas Centro Tecnológico, Universidade Federal de Santa Catarina. Florianópolis - SC, 2013.

AZEVEDO, Philipe Kretzer e Castro. **Desenvolvimento de um sistema supervisor e lógicas de CLP no ambiente de Geração de Energia**. 2013. 108p. Dissertação (monografia). Departamento de

Automação e Sistemas Centro Tecnológico, Universidade Federal de Santa Catarina. Florianópolis - SC, 2013.

BALLOK, Ivan Roberto. **Protótipo de um sistema para supervisão de equipamentos industriais**. 2003. 62p. Dissertação (monografia). Centro de Ciências Exatas e Naturais, Universidade Regional de Blumenau. Blumenau – SC, 2003.

FERREIRA, Flávio Meireles. **Controle de velocidade em motor CA trifásico utilizando controle PID do CLP**. 2007. 90p. Dissertação (monografia). Centro Universitário de Brasília, Faculdade de Ciências Exatas e Tecnologia. Brasília – DF, 2007.

FIALA, Leonardo K. **Automação de Caldeira Industrial por CLP**. 2014. 55p. Dissertação (monografia). Trabalho de Conclusão de Curso (Curso Superior de Tecnologia em Mecatrônica Industrial), Universidade Tecnológica Federal do Paraná. Curitiba – PR, 2014.

GOMES JÚNIOR, Edson. **Automação do painel de controle de um sistema de refrigeração industrial**. 2009. 35p. Dissertação (monografia). Escola de Minas e Colegiado do Curso de Engenharia de Controle de Automação, Universidade Federal de Ouro Preto. Ouro Preto – MG, 2009.

MATOS, Cláudio José Gonçalves. **Análise e Implementação da NORMA ANSI/ISA S88.01 no Controle Automatizado de um Sistema em Processo por Batelada**. 2007. 123p. Dissertação (pós-graduação). Programa de engenharia Elétrica, Universidade Federal do Rio de Janeiro. Rio de Janeiro – RJ, 2007.

PERTENÇA JÚNIOR, Antonio. **Desenvolvimento de um sistema de aquisição de dados para medições mecânicas através da utilização de controlador lógico programável (CLP)**. 2010. 123p. Dissertação (pós-graduação). Programa de pós-graduação em Engenharia Mecânica, Universidade Federal de Minas Gerais. Belo Horizonte – MG, 2010.

SANTOS, Antonio Raimundo dos. **A construção do conhecimento**. 4. Ed. Rio de Janeiro: DP&A editora, 2001.

SILVA, Bianca Romaniv; SOARES, Jefferson Wilhelm Meyer; ROSA, Victor Emanuel Correia de la. **Aplicação de um método de programação de CLP com base no diagrama SFC para automação de unidades geradoras hidrelétricas**. 2012. 96p. .Dissertação (monografia). Departamento Acadêmico de Eletrotécnica, Universidade Tecnológica Federal do Paraná. Curitiba – PR, 2012.

TEXEIRA, Maria de Lourdes Prado Ferreira. **Sistema de controle e supervisão de bombas de recalque de água para edifício residencial**. 2007. 70p. Dissertação (monografia). Centro Universitário de Brasília, Faculdade de Ciências Exatas e Tecnologia. Brasília – DF, 2007.

CAPÍTULO 8

PRÁTICAS ÁGEIS NA ELICITAÇÃO DE REQUISITOS PARA DESENVOLVIMENTO DE SOFTWARE EM UMA COOPERATIVA DE SAÚDE

Data de aceite: 27/08/2020

Data de submissão: 25/05/2020

Mariangela Catelani Souza

Faculdade de Tecnologia do Centro Paula
Souza – FATEC
São José do Rio Preto - SP

Bruno Cardoso Maciel

Faculdade de Tecnologia do Centro Paula
Souza – FATEC
São José do Rio Preto - SP

José Alexandre Ducatti

Faculdade de Tecnologia do Centro Paula
Souza – FATEC
São José do Rio Preto - SP

Paulo Sérgio Gaudêncio Mauro

Faculdade de Tecnologia do Centro Paula
Souza – FATEC
São José do Rio Preto - SP

Leonardo Mendes de Souza

Faculdade de Tecnologia do Centro Paula
Souza – FATEC
Catanduva - SP

Lygia Aparecida das Graças Gonçalves Corrêa

União das Faculdades dos Grandes Lagos -
UNILAGO
São José do Rio Preto - SP

Elizângela Cristina Begido Caldeira

União das Faculdades dos Grandes Lagos –
UNILAGO
São José do Rio Preto – SP

Bruna Grasseti Fonseca

União das Faculdades dos Grandes Lagos –
UNILAGO
São José do Rio Preto – SP

Patricia Cristina de Oliveira Brito Cecconi

União das Faculdades dos Grandes Lagos -
UNILAGO
São José do Rio Preto - SP

Ana Paula Garrido de Queiroga

União das Faculdades dos Grandes Lagos -
UNILAGO
São José do Rio Preto – SP

Humberto Cecconi

ETEC - Philadelpho Gouvêa Netto
São José do Rio Preto – SP

Carlos Alípio Caldeira

UEMG - Universidade do Estado de Minas
Gerais
Frutal – MG

RESUMO: A análise de negócios e a engenharia de requisitos estão hoje entre as tarefas mais árduas do processo de desenvolvimento de software. O objetivo deste artigo é sugerir uma melhoria dessas etapas por meio do uso de conceitos viáveis de metodologias ágeis, destacando os aspectos positivos e negativos através de um estudo de caso. Espera-se uma melhor comunicação entre todas as áreas de interesse (cliente, analista e equipe de desenvolvimento), implicando no levantamento de requisitos o mais próximo possível dos desejos do cliente, facilitando o processo de desenvolvimento de software.

PALAVRAS-CHAVE: Engenharia de requisitos, análise de negócios, software.

ABSTRACT: Business analysis and requirements engineering are today among the most arduous tasks of the software development process. The objective of this article is to suggest an improvement of these steps through the use of viable concepts of agile methodologies, highlighting the positive and negative aspects through a case study. Better communication between all the areas of interest (customer, analyst and development team) is expected, implying in the survey of requirements as close as possible to the client's wishes, facilitating the software development process.

KEYWORDS: Requirements engineering, business analysis, software.

1 | INTRODUÇÃO

A análise de negócios e engenharia de requisitos estão hoje entre as tarefas mais árduas no processo de desenvolvimento de software. O objetivo deste artigo é sugerir um aprimoramento destas etapas através da utilização de conceitos viáveis de metodologias ágeis, evidenciando os pontos positivos e negativos por meio de um estudo de caso. Espera-se melhor comunicação entre todas as áreas de interesse (cliente, analista e equipe de desenvolvimento), implicando num levantamento dos requisitos o mais próximo do desejo do cliente e de forma mais clara possível para o processo de desenvolvimento de software.

Com os avanços tecnológicos que vêm ocorrendo décadas, a informação passou a ser um recurso estratégico das empresas. O software se tornou, então, a força motora desta nova era. O primeiro passo na concepção de um projeto de software é a análise de negócio; de acordo com o IIBA (International Institute of Business Analysis) na 3ª versão do seu guia BABOK(2015) “a análise de negócios é a prática de permitir mudanças em uma empresa definindo necessidades e recomendando soluções que agreguem valor às partes interessadas. A análise de negócios permite que uma empresa articule as necessidades e as razões da mudança, e para projetar e descrever soluções que possam agregar valor”.

Baseado nesses fatos, podemos aplicar conceitos de metodologias ágeis, como o Scrum, Kanban, XP, dentre outros, na elicitação dos requisitos do projeto, tornando melhor a comunicação entre as áreas de interesse, contribuindo para um levantamento mais elaborado, com objetivos e requisitos bem definidos, tais pontos favorecem a equipe de desenvolvimento, aumentando as chances do projeto ser bem sucedido e entregue dentro dos prazos estabelecidos com qualidade destacada.

2 | JUSTIFICATIVA

Segundo o PMI (*Project Management Institute*), “Projeto é um esforço temporário empreendido para criar um produto, serviço ou resultado único, e são realizados para cumprir objetivos através da produção de entregas” (2017, p4). Assim, podemos entender que os projetos são únicos e apresentam começo e fim bem definidos. O maior problema

é hoje a comunicação desconexa entre cliente, analista e a equipe de desenvolvimento, definindo requisitos não bem elaborados, tal falta de sinergia entre as equipes corroboram para surgimento futuro de problemas no software, como por exemplo, não atender a real necessidade do cliente, conforme mostra figura 1.

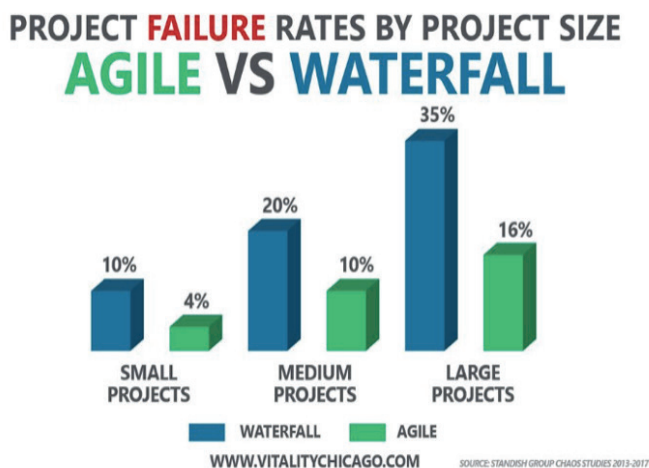


Figura 1: Taxa de falha por tamanho do Projeto
Fonte: Vitality Chicago (2018)

3 | OBJETIVO

- Aprimorar o processo de análise de negócios e engenharia de requisitos; Melhorar a produtividade da equipe de Inovação e Desenvolvimento; Aplicar boas práticas metodologias ágeis com foco em análise de negócios ágil e engenharia de requisitos; Melhorar a integração e comunicação entre as partes interessadas dos projetos; Entregar as soluções com qualidade no menor tempo.

4 | FUNDAMENTAÇÃO TEÓRICA

4.1 Metodologias Ágeis

De acordo com o Manifesto Ágil (2001), os métodos ágeis foram construídos para se buscar as melhores práticas para desenvolvimento de *software*, construídos sobre alguns princípios, como por exemplo:

- Pessoas relacionadas à negócios e desenvolvedores devem trabalhar em conjunto e diariamente, durante todo o curso do projeto; construir projetos ao redor de indivíduos motivados. Dando a eles o ambiente e suporte necessário, e confiar que farão seu trabalho; o método mais eficiente e eficaz de transmitir informações para, e por dentro de um time de desenvolvimento, é através de uma conversa cara a cara.

Podemos notar que os envolvidos em um projeto de software devem se sentir confiantes com uma boa comunicação e feedbacks constantes, sempre tendo em mente a prioridade é satisfazer o cliente com prazos cumpridos e software funcionando, e isso é possível como uma boa comunicação entre as partes envolvidas. Estes princípios sugerem maior sustentabilidade para todos os envolvidos no processo da construção do software.

As metodologias aplicadas a este estudo são:

1. Scrum
2. Extreme Programming
3. Kanban
4. Feature Driven Development

4.1.1 *Scrum*

O Guia do Scrum (2017) o define tal metodologia como um framework de suporte para as pessoas solucionarem problemas complexos e adaptativos, enquanto entregam os produtos de forma produtiva e criativa, contribuindo com o mais alto valor possível.

Scrum não é um processo, técnica ou um método definitivo. Em vez disso, é um framework dentro do qual você pode empregar vários processos ou técnicas. O Scrum deixa claro a eficácia relativa de suas práticas de gerenciamento de produto e técnicas de trabalho, de modo que você possa continuamente melhorar o produto, o time e o ambiente de trabalho. (Schwaber e Sutherland; 2017)

Conforme a Figura 2, observamos que a cada atividade ocorrem algumas tarefas a serem cumpridas e são eficazes até mesmo para quando se tem um prazo de entrega apertado e quando seus requisitos precisam ser mudados durante o processo.

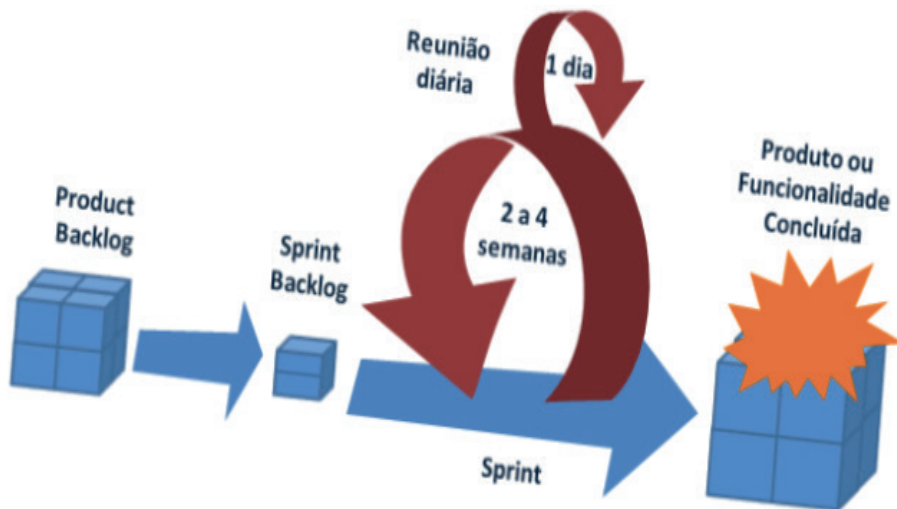


Figura 2: Fluxo do processo de Scrum.

Fonte: Mindmaster (2019)

Cerimônias:

- ***Sprint Planning***: é a reunião onde ocorre o planejamento da sprint, e levantamentos dos itens do *product backlog*.
- ***Daily Scrum***: Reunião rápida que acontece diariamente durante o sprint, tendo duração máxima de 15 minutos. Tem como objetivo o alinhamento das demandas do dia e revisão do progresso da equipe.
- ***Sprint Review***: Revisão realizada ao final da *sprint* onde o time demonstra tudo que foi desenvolvido.
- ***Retrospectiva***: É uma avaliação da *sprint* encerrada. Nesta reunião elenca-se os pontos positivos, negativos e o que deve ser melhorado.

Artefatos:

- ***Product Backlog***: Os funcionalidades a serem desenvolvidas ficam no *Product Backlog*. Cada item deve ser priorizados pelo valor de negócio.
- ***Sprint***: É um ciclo de desenvolvimento do Scrum. Tendo duração de 1 a 4 semanas.
- ***Sprint Backlog***: É a lista de tarefas que serão realizadas durante o sprint.
- ***Product Owner***: É o stakeholder chave do projeto. Tem como responsabilidade manter o *product backlog* priorizado.

- **Scrum Master:** É o líder da equipe, responsável por conduzir as cerimônias do scrum e dividir as tarefas.
- **Burndown Chart:** É um gráfico para acompanhamento do progresso da equipe.

4.1.2 Extreme Programming (XP)

Extreme Programming ou Programação Extrema é uma metodologia de desenvolvimento de software que possui foco em agilidade de equipes e qualidade de projetos. Pode ser aplicado em projetos de diversos portes. A XP assume que a volatilidade dos requisitos existe, em vez de tentar eliminá-la, trata o desenvolvimento do software a partir de uma abordagem flexível e colaborativa, na qual desenvolvedores e clientes fazem parte de uma única equipe que tem o propósito de produzir software de alto valor agregado (PRIKLADNICKI, WILLI, MILANI, 2014).

O processo XP é dividido em 4 etapas:

1. **Planejamento:** O primeiro passo no planejamento de um projeto XP é ouvir o cliente utilizando toda a abordagem da interação e comunicação entre as partes com a finalidade de entender qual a real necessidade do cliente. Utiliza-se um método chamado *User Stories*, ou Histórias do Usuário, onde é descrito a necessidade, atribuindo uma prioridade.
2. **Projeto:** Nesta etapa, o projeto é relacionado com o valor da Simplicidade, onde as histórias são avaliadas e organizadas. Caso exista um item de alta complexidade, utiliza-se a técnica de *Refactor* (Re-fabricar) para uma melhor avaliação da demanda com a finalidade de simplificar e aprimorar.
3. **Codificação:** Etapa de desenvolvimento das *User Stories* do projeto.
4. **Testes:** Integrado ao processo de codificação, os testes são realizados diariamente, cada unidade deve ser testada antes de ser liberada.

4.1.3 Kanban

Segundo o livro *Kanban em 10 Passos* (Boeg, 2011), o Kanban é um método de gestão de mudanças, com ênfase na visualização do andamento do trabalho, tal metodologia segue alguns princípios como:

- Medir e gerenciar o fluxo de processos
- Visualizar passo a passo a cadeia de atividades
- Limitar o trabalho em progresso
- Identificar oportunidades de melhorias

O foco do Kanban é conduzir mudanças evolucionárias no processo, sua aplicação é realizada por meio de cartões visuais destacados em um quadro. A utilização de cartões serve para controlar as tarefas que estão sendo executadas no momento. Uma das vantagens do Kanban é o controle mais do que detalhado da produção da empresa, já que ele fornece informações sobre quando, como e o que realizar em determinada tarefa.

4.1.4 Feature Driven Development (FDD)

Segundo Pressman (2011), o desenvolvimento dirigido a funcionalidades (*Feature Driven Development - FDD*), foi concebido por Peter Coad como um modelo apropriado para engenharia de *software* orientada a objetos. Posteriormente, o FDD foi aprimorado, apresentando um processo ágil adaptativo, o qual pode ser empregado em projetos de médio a grande porte.

Basicamente, o FDD possui duas fases: a Concepção e Planejamento, e a Construção, conforme evidenciado na figura 3 abaixo. Na fase de concepção e planejamento é que ocorre a triagem dos requisitos. Pode-se utilizar as técnicas tradicionais de elicitação de requisitos, mas sem perder o foco das funcionalidades, pois a ênfase do FDD é justamente nestas últimas (Pressman, 2011).

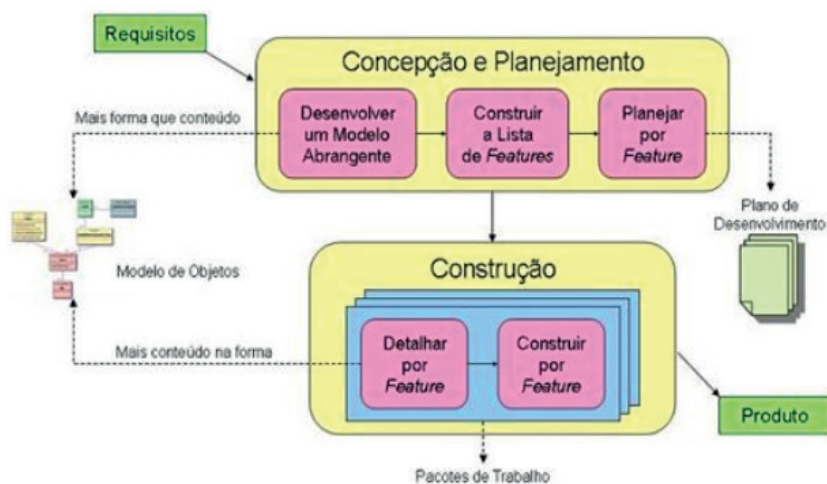


Figura 3: Exemplificação do processo de FDD

Fonte: Jorge Audy (2012)

Assim como outras metodologias ágeis, o FDD fomenta a colaboração entre os membros da equipe, gerencia problemas e complexidade de projetos através da segregação em funcionalidades e comunicação verbal, textual e gráfica para apresentação de detalhes

técnicos. Encorajando o desenvolvimento incremental na concepção de projetos, o desenvolvimento dirigido a funcionalidades enfatiza o controle de qualidade em um ciclo contínuo, através de inspeções do código e do projeto, aplicação de auditorias, coleta de métricas e utilização de padrões (Pressman, 2011).

5 | RESULTADOS E DISCUSSÕES

Por meio de uma busca e estudo sobre o tema, foram encontrados alguns trabalhos com algumas similaridades ao tema proposto por este artigo. Um dos artigos, destaca a importância dos métodos ágeis na engenharia de *software*, a autora salienta que as metodologias ágeis não surgiram para substituir por completo os métodos clássicos, mas sim para complementar, aprimorando a engenharia de *software* (Silva, 2016). Tal afirmação é condizente com a proposta deste artigo, que tem como objetivo aprimorar os processos de análise de negócios e engenharia de requisitos por meio da aplicação de conceitos chave de algumas metodologias ágeis.

Os métodos ágeis enfatizam a colaboração humana e a auto-organização das equipes. O método ágil é um ótimo método quando se trabalham com pequenas equipes e que a flexibilidade é necessária. Cada método ágil tem sua funcionalidade para cada tipo de projeto (Silva, 2016).

Conforme a citação, nota-se que os métodos ágeis, possuem particularidades para cada tipo de projeto, destacando-se alguns pontos positivos das metodologias abrangidas neste estudo, podemos por meio de um estudo de caso, aplicar tais métodos com a perspectiva de aprimoramento dos processos iniciais do desenvolvimento de *software*.

6 | A EMPRESA

A empresa selecionada e apresentada neste estudo foi fundada em 1971 por um grupo de médicos, é uma cooperativa médica com atuação em operadora de planos de saúde. Atualmente conta com mais de 1400 médicos cooperados, e possui mais de 260 mil beneficiários. É composta por 6 unidades situadas na região de São José do Rio Preto, sendo uma das unidades um pronto atendimento, e até o presente momento conta com um quadro de aproximadamente 1100 colaboradores.

Para gerir tanta informação envolvida nos processos-padrão do sistema de saúde, a empresa possui seu próprio departamento de TI, localizado na sede administrativa, responsável pelo ERP, website, infraestrutura, help desk e demais funções. Atualmente o setor conta com uma equipe de 25 profissionais, liderados por um coordenador e um gerente de TI.

O departamento de TI ainda é subdividido em 5 células:

1. Inovação e Desenvolvimento
2. Business Intelligence
3. Gestão do Beneficiário
4. Produção e Contas Médicas
5. Infraestrutura e Suporte

O célula-foco deste estudo é o grupo de Inovação e Desenvolvimento. Composto por 6 membros, sendo 2 analistas de negócios, 3 desenvolvedores full-stack, 1 desenvolvedor front-end.

6.1 Como era antes

As demandas e projetos da equipe de desenvolvimento eram geridas apenas pelo software ERP da empresa, no módulo de “Ordens de Serviço”. A equipe não possuía nenhuma metodologia de trabalho, as tarefas não tinham nenhuma priorização específica, apenas os casos urgentes tinham prioridade de execução instantânea, as demais atividades eram executadas conforme conhecimento da área dos membros do time, observado na figura.

ID Ordem	Di ordem	Projeto origem	Estágio atual	Descrição	Solicitante	Nome usuário prev	Completude
1308002	18/01/2019 09:41:28		Encerrado	Atualização de dados de competência de Sistema Med	JOSE CARLOS PEDROSSO	Felipe Rodrigues Bueno	Ab
1308001	18/04/2019 15:29:29		Encerrado	Atualização de dados de competência de Sistema Med	FELIPE RODRIGUES BUENO	Felipe Rodrigues Bueno	Ab
1308009	19/04/2019 09:43:42		Em análise (Qualificação)	Manutenção de Processos Endógenos de Saramento	DANIEL AUGUSTO DOS SANTOS	Felipe Rodrigues Bueno	Ab
1308106	16/04/2019 16:20:33		Em análise (Qualificação)	Integração de Sistema de Inscricao - RFE	AMARELO SILVA DE SIA OLIVEIRA	Felipe Rodrigues Bueno	Ab
1307139	05/10/2018 10:44:45		Pendente	Módulo de Planilhas para Projeções - Unimed Le	TERESINHA FERREIRA DE MELLO SOUSA	Felipe Rodrigues Bueno	Ab
1307138	05/10/2018 10:44:45		Pendente	Módulo de Planilhas para Projeções - Unimed Le	ISABELA DA SILVA RIBEIRO DE OLIVEIRA	Felipe Rodrigues Bueno	Ab
1307137	05/10/2018 10:44:45		Pendente	Módulo de Planilhas para Projeções - Unimed Le	MARCELA DA SILVA RIBEIRO	Felipe Rodrigues Bueno	Ab
1307166	05/11/2018 13:57:58		Pendente	PROPOSTAÇÃO SISTEMA MED - HOME CARE	ISABELA DA SILVA RIBEIRO DE OLIVEIRA	Felipe Rodrigues Bueno	Ab
1307160	05/11/2018 16:40:44		Pendente	Agendamento SOD para farmacia MED	DAIANE MARI SOUZA	Felipe Rodrigues Bueno	Ab
1307089	27/11/2018 10:28:52		Em Andamento	Criação de terminal Sistema Med	JULIANA DE FARIA CERON COURADO	Ederson Rodrigues de Silva	C
1307088	27/11/2018 10:28:52		Em Andamento	Criação de terminal Sistema Med	JULIANA DE FARIA CERON COURADO	Ederson Rodrigues de Silva	C
1307087	27/11/2018 10:28:52		Em Andamento	Criação de terminal Sistema Med	JULIANA DE FARIA CERON COURADO	Ederson Rodrigues de Silva	C
1307086	27/11/2018 10:28:52		Em Andamento	Criação de terminal Sistema Med	JULIANA DE FARIA CERON COURADO	Ederson Rodrigues de Silva	C
1307085	27/11/2018 10:28:52		Em Andamento	Criação de terminal Sistema Med	JULIANA DE FARIA CERON COURADO	Ederson Rodrigues de Silva	C
1307084	27/11/2018 10:28:52		Em Andamento	Criação de terminal Sistema Med	JULIANA DE FARIA CERON COURADO	Ederson Rodrigues de Silva	C
1307083	27/11/2018 10:28:52		Em Andamento	Criação de terminal Sistema Med	JULIANA DE FARIA CERON COURADO	Ederson Rodrigues de Silva	C
1307082	27/11/2018 10:28:52		Em Andamento	Criação de terminal Sistema Med	JULIANA DE FARIA CERON COURADO	Ederson Rodrigues de Silva	C
1308214	06/02/2019 12:14:50	13	Em análise (Foco)	PF - Solução Placa Postera - Esp Digital Beneficiario	VANE SIA NORRIS CAETANO	Marcelo A. Barbalho	Média
1308213	06/02/2019 12:14:50	13	Em análise (Foco)	PF - Fatur com seu Executivo de Contas - Esp Digital Beneficiario	VANE SIA NORRIS CAETANO	Bruno Cardoso Mascari	Ab
1308212	06/02/2019 12:14:50	13	Em análise (Foco)	PF - Fatur com seu Executivo de Contas - Esp Digital Beneficiario	VANE SIA NORRIS CAETANO	Felipe Rodrigues Bueno	Ab
1308211	06/02/2019 12:14:50	13	Em análise (Foco)	PF - Histórico de Agendamentos Endógena	VANE SIA NORRIS CAETANO	Felipe Rodrigues Bueno	Ab
1308667	13/02/2019 13:54:59	5	Em análise (Foco)	PIA e CD Dashboard / Relatório - Pós. Experiência cliente	DENNIS OLIVEIRA PEDROSSO	Bruno Cardoso Mascari	Ab
1308669	13/02/2019 14:05:02	3	Em análise (Foco)	PIA e CD Login - Pós. Experiência cliente	DENNIS OLIVEIRA PEDROSSO	Bruno Cardoso Mascari	Ab
1308693	11/02/2019 08:09:00	4	Em análise (Foco)	PIA e CD Solução de serviços adicionais na plataforma - Pós. Experiência cliente	DENNIS OLIVEIRA PEDROSSO	Bruno Cardoso Mascari	Ab
1308705	13/02/2019 16:58:46	6	Em análise (Foco)	PIA Automatizar envio de e-mail automático - Pós. Experiência cliente	DENNIS OLIVEIRA PEDROSSO	Bruno Cardoso Mascari	Média

Figura 4: Ordens de Serviço do ERP da empresa

Fonte: Elaborado pelo autor

A apresentação das demandas no sistema dificultava a organização e visualização das atividades pela equipe. Muitas vezes, os usuários abriam as OS's sem comunicar a equipe de Desenvolvimento, apenas depois de aberto o ticket que os analistas tomavam conhecimento dos projetos e tarefas. A partir daí era feito uma análise de requisitos, porém sem seguir nenhum tipo de metodologia ou padrão de projetos. Geralmente não havia nenhuma ou pouca reunião entre analistas e clientes para melhor entendimento das demandas solicitadas.

Notoriamente havia uma falha na organização e planejamento das atividades, e a comunicação era desconexa entre os envolvidos no projeto, isso tornava a visão dos projetos deturpada, o que gerava incertezas em relação ao resultado final do projeto.

6.2 Como ficou

Os principais negativos observados na metodologia de trabalho anterior foram:

- Priorização de projetos e demandas, Entendimento e comunicação
- Falta de visualização no processo, Organização e Planejamento das atividades

Baseado nestes fatos, o presente estudo teve como objetivo a melhora do processo de desenvolvimento de software com foco na elicitacão de requisitos ágeis.

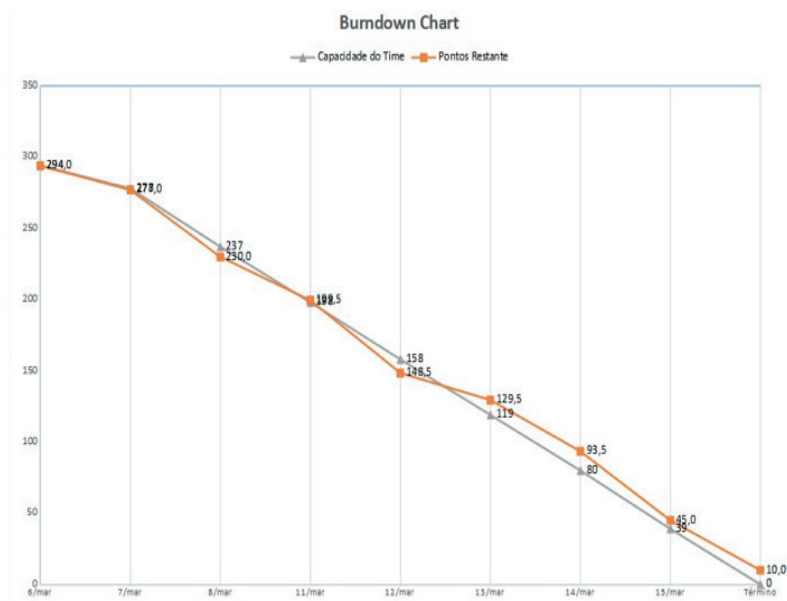


Figura 5: Gráfico Burndown da equipe de desenvolvimento

Fonte: Elaborado pelo autor

Outra técnica abordada no processo de desenvolvimento de software foram os valores do Extreme Programming, focando em interação e comunicação entre as partes envolvidas e o *feedback* do projeto. O *feedback* consiste na reavaliação dos requisitos e das demandas do projeto entre equipe e cliente, esta etapa foi primordial para melhorar o entendimento dos requisitos por parte dos analistas e desenvolvedores, pois diminui a probabilidade de erros na elicitação dos requisitos.

Por fim, a última necessidade era se ter uma visualização do andamento das atividades e efetividade da equipe, para sanar tal ponto, foi abordado o uso do *Kanban*. Aplicou-se o uso de um quadro virtual para realizar a gestão das tarefas, tal ferramenta se mostrou prática e simples de gerenciar, pois todos poderiam visualizar o quadro de cartões da equipe, e cada membro poderia visualizar o seu próprio, sendo mais organizado que um quadro físico com todos os cartões de atividades.

Definiu-se a adaptação do *Kanban* em 5 colunas, conforme observado na Figura 6:

1. *Product Backlog*: Todas as ordens de serviço que constavam no sistema ERP foram transpassadas para esta coluna.
2. *Sprint Backlog*: As tarefas definidas na reunião de planejamento da sprint, eram mapeadas nesta coluna com já com sua estimativa e informações relevantes.
3. Em Andamento: Tarefas que estavam em desenvolvimento pela equipe eram movidas da *sprint backlog* para Em Andamento.
4. Concluída: Ao finalizar uma tarefa e ela tiver sido validada pelo cliente, o cartão era concluído.
5. Pendência: Tarefas que possuíam alguma pendência que impossibilitava momentaneamente seu desenvolvimento e/ou conclusão eram alocadas nesta coluna.

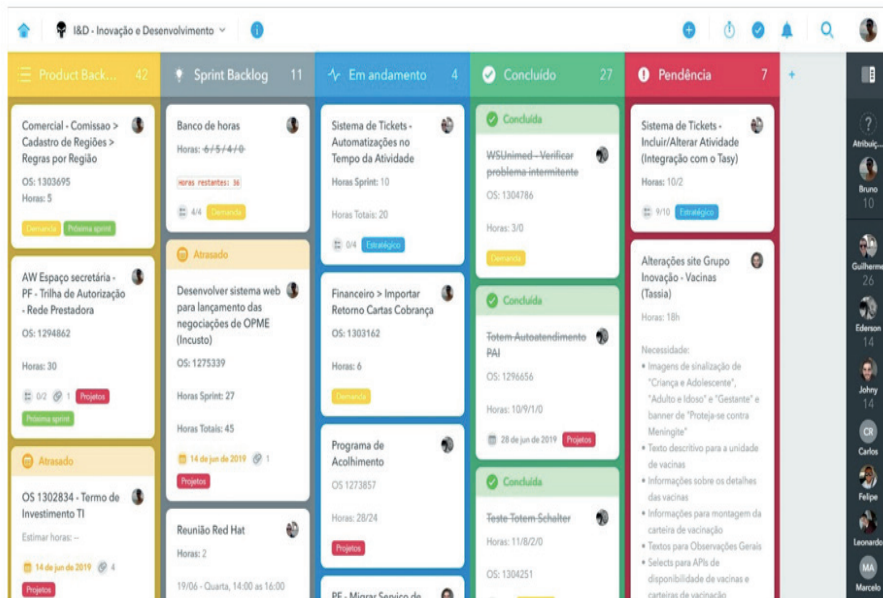


Figura 6: Kanban virtual da equipe

Fonte: Elaborado pelo autor

Deste modo, o time de desenvolvimento passou a ter visualização e controle das atividades planejadas na *sprint*, conseguindo notar pontos que prejudicam o progresso da equipe. O estudo foi aplicado por 06 meses até a data presente desta pesquisa.

71 CONCLUSÃO

A Engenharia de Requisitos e Análise de Negócio define, sem dúvida, são umas das mais importantes atividades a serem realizadas em projetos de desenvolvimento de software. Embora não garanta a qualidade dos produtos gerados, é um pré-requisito básico para que se obtenha sucesso no desenvolvimento.

É importante salientar que a elicitação de requisitos depende muito da interação entre os desenvolvedores de sistemas e as pessoas que efetivamente estão envolvidas com o processo de trabalho da organização, minimizando qualquer problema na definição de requisitos por parte do cliente.

Os maiores benefícios adquiridos com a aplicação de boas práticas ágeis foram a diminuição dos erros e mudanças drásticas nos requisitos das demandas, isso se deve a melhor visibilidade por parte dos envolvidos, e também pelo compartilhamento e colaboração ativa de todas as partes. Com isto, a equipe de desenvolvimento se tornou muito mais organizada, contribuindo com planejamento e cumprimento das metas. Por

conta de os membros estarem motivados e unidos, notou-se uma diminuição do retrabalho e a constante busca por melhoria do processo.

REFERÊNCIAS

AGILE MANIFESTO. Disponível em: <<http://agilemanifesto.org/>>. Acesso em: 09 de outubro de 2018.

BABOK O Guia para o Corpo de Conhecimento de Análise de Negócios. IIBA – International Institute of Business Analysis Versão 2.0, 2011.

BOEG, JESPER. Kanban em 10 passos. 2011.

DESENVOLVIMENTO ÁGIL. Extreme Programming XP.

IEEE - INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS. Standards glossary of software engineering terminology: STD 610.12, N.Y.,1990.

JORGE AUDY. Disponível em <<https://jorgeaudy.com/tag/metodologia-2/page/38/>>.

MACHADO, F. N. R.. Análise e Gestão de Requisitos de Software: Onde nascem os sistemas. 3 ed. São Paulo. Érica, 2016.

MINDMASTER. Disponível em <<http://www.mindmaster.com.br/scrum/>>.

O ANALISTA DE NEGÓCIOS. Disponível em <<http://oanalistadenegocios.com/>>.

PALUCHOWSKI, B. B.; BESKOW, L. Análise e estudo de aplicação das práticas do scrum no processo de desenvolvimento e manutenção de software. Três de Maio: SETREM, 2010.

PAULA FILHO, Wilson de Pádua. Engenharia de Software: fundamentos, métodos e padrões. 3 ed. São Paulo: LTC Editora, 2009.

PMBOK. Um Guia do Conhecimento em Gerenciamento de Projetos (Guia PMBoK). 6.ed. Pennsylvania, PMI, 2017.

PRESSMAN, R. S. Engenharia de Software: Uma abordagem profissional, 7 ed. São Paulo: Markon Books, 2011.

PRIKLADNICKI, R.; WILLI, R.; MILANI, F. Métodos ágeis para desenvolvimento de software. Porto Alegre: Bookman, 2014.

SCHWABER K., SUTHERLAND J. Guia do scrum. SCRUM.ORG. 2017.[S.L]

SILVA, A. G. A importância dos métodos ágeis na engenharia de software. 2016. UFF, Niterói, 2016.

SOARES, L. M. **Elicitação de requisitos baseado em métodos ágeis**: Um estudo de caso. 2017. Unisinos, São Leopoldo, 2017.

SOMMERVILLE, I. **Engenharia de Software**. 9 ed. São Paulo. Pearson Prentice Hall, 2011.

THE STANDISH GROUP. Chaos Report 2015.

YOUNG, R. **Recommended Requirements Gathering Practices**. 2002. Crosstalk The Journal of Defense Software Engineering.

QENADE: APLICATIVO MÓVEL PARA PREPARAÇÃO DE ESTUDANTES PARA O ENADE

Data de aceite: 27/08/2020

Data de submissão: 26/07/2020

Helder Guimarães Aragão

Centro Universitário Estácio da Bahia
Salvador – Bahia

<http://lattes.cnpq.br/9958289516496960>

RESUMO: O Exame Nacional de Desempenho dos Estudantes (ENADE) é, atualmente, o principal indicador que avalia o rendimento dos formandos dos cursos de graduação. Este Exame possui provas em consonância com os conteúdos programáticos previstos nas diretrizes curriculares dos cursos, focados no desenvolvimento de competências e habilidades necessárias para o estudante concluinte. Todo aluno de graduação é obrigado realizar esse Exame. As atuais formas de preparação do aluno, geralmente, envolvem a elaboração de provas com questões do ENADE ou simulados específicos disponibilizados no portal próprio da instituição. Apesar destas iniciativas, é raro o uso de tecnologias da informação nesse processo de preparação. É notório que as Instituições de Ensino precisam buscar tecnologias da informação mais próximas da realidade do aluno para melhor preparação ao ENADE. Dentre estas tecnologias, destacam-se os aplicativos. Neste contexto, o presente trabalho apresenta, portanto, a construção de um aplicativo, denominado QEnade, para a disponibilização de questões do ENADE para os estudantes do Centro Universitário Estácio da Bahia (Estácio FIB). O aplicativo

visa auxiliar a preparação dos alunos para a o Exame Nacional. Espera-se que o aplicativo promova uma maior adesão dos alunos na preparação para o ENADE, além de permitir autonomia aos estudantes em relação aos seus estudos.

PALAVRAS-CHAVE: ENADE, aplicativo, preparação;

QENADE: MOBILE APPLICATION FOR ENADE EXAM

ABSTRACT: The ENADE is currently the main performance indicator for undergraduate students. This exam has questions about curricular content that it is planned in the guidelines of the courses, focused on the development of competencies and skills necessary for the undergraduate students. All undergraduate students must be realizes this test. The current ways of student preparation generally involve the preparation of tests with questions from ENADE, which are made available on the institution's own portal. Despite these initiatives, it is rare to use information technologies in this preparation process. Then, the universities needs to find the information thecnologies near to student reality for better preparation enade test. In this context, this present projects aims the development applicacion, called QEnade, for provide questions of the ENADE for undegraduated students at Center University Estácio da Bahia. This mobile application aims help to undergraduated students for ENADE test. It is expected that the mobile application

will promote greater student adherence in studying for ENADE test, in addition to allowing students autonomy in relation to their studies.

KEYWORDS: ENADE, mobile application, test;

1 | INTRODUÇÃO

O Exame Nacional de Desempenho dos Estudantes (ENADE) é, atualmente, o principal indicador que avalia o rendimento dos formandos dos cursos de graduação. Este Exame possui provas em consonância com os conteúdos programáticos previstos nas diretrizes curriculares dos cursos, focados no desenvolvimento de competências e habilidades necessárias para o estudante concluinte. Todo aluno de graduação é obrigado a realizar esse Exame (ENADE, 2020).

Algumas instituições, incluindo a Estácio, entendem a importância do ENADE e investem na construção de instrumentos visando incentivar os alunos na participação do exame. Estas instituições buscam de forma contínua formas de prepará-los melhor para o referido exame.

As atuais formas de preparação do aluno, geralmente, envolvem a elaboração de provas com questões do ENADE ou simulados específicos disponibilizados no portal próprio da instituição. Apesar destas iniciativas, é raro o uso de tecnologias da informação nesse processo de preparação. É notório que as Instituições de Ensino precisam buscar tecnologias da informação mais próximas da realidade do aluno para melhor preparação ao ENADE. Dentre estas tecnologias, destacam-se os aplicativos. Apesar da profusão de aplicativos, uma simples busca na loja de aplicativos da Google, Google Play, pelo termo ENADE, traz como resultado poucos aplicativos voltados para o Exame. Aqueles que existem possuem limitações como, por exemplo, a escolha do curso. Muitas vezes os aplicativos existentes não podem ser adotados por uma instituição específica por serem generalistas.

Neste contexto, o presente trabalho apresenta a construção de um aplicativo, denominado QEnade para disponibilização de questões do ENADE focado nos estudantes do Centro Universitário Estácio da Bahia, Estácio FIB. O aplicativo visa auxiliar a preparação dos alunos para o Exame Nacional. Os objetivos específicos são: catalogar as provas do ENADE dos cursos, inicialmente, dos cursos de Tecnologia da Informação, documentar o processo de desenvolvimento do aplicativo com metodologia ágil, desenvolver o aplicativo e, por fim, disponibilizar para estudantes o aplicativo desenvolvido focando nos testes de uso e adesão.

Os resultados esperados são: maior adesão dos alunos ao ENADE, melhor preparação dos alunos para o Exame e inserção de tecnologias da informação no processo de preparação para o Exame. A viabilidade técnica e econômica é factível, pois as tecnologias a serem utilizadas são gratuitas.

2 | REFERENCIAL TEÓRICO

A educação móvel ou *mobile learning* consiste no conceito do uso de aplicativos e dispositivos móveis no processo de aprendizagem. A inserção destas tecnologias já vem sendo uma realidade, uma vez que cada vez mais a Internet está sendo adotada como plataforma de estudo por parte dos estudantes.

Totti, Gomes, Moreira e Souza (2011) afirmam que “A tecnologia antes vista como algo que tirava o sujeito do convívio social e do contato coletivo, torna-se cada vez mais customizadora, assim, os ambientes tornam-se individualizados, mas não individualistas”. Portanto, hoje pode-se notar que o uso de dispositivos móveis e aplicativos não se restringe a entretenimento, mas, também, para o desenvolvimento de aprendizagem.

Vários aplicativos e dispositivos móveis podem ser utilizados como instrumentos visando auxiliar a aprendizagem. *E-Readers, tablets e smartphones* são alguns exemplos de tecnologias móveis (SABOIA, VARGAS, VIVA, 2013).

Dentre as vantagens do uso de aplicativos e dispositivos móveis no processo ensino aprendizagem, destaca-se o incentivo do aprendizado independente de local e autonomia do aluno. Muitos trabalhos descrevem o desenvolvimento de aplicativos voltados para a aprendizagem.

O aplicativo Física In Mãos é um exemplo. Ele foi desenvolvido visando facilitar o ensino de física (FísicaInMãos, 2020). O aplicativo, por exemplo, Duolingo é voltado para o aprendizado de idiomas (Duolingo, 2020). Já o EduEdu é um aplicativo voltado para o ensino de alfabetização para crianças com dificuldades de ler e escrever. Este aplicativo, a partir de uma rápida avaliação, identifica quais os pontos a serem trabalhados na criança para então atividades personalizadas serem criadas (EduEdu, 2020).

Outro aplicativo que vale destacar é o RevisaApp que é focado no ENEM e no vestibular. Segundo a descrição do próprio aplicativo são mais de 260 conteúdos a respeito dos diversos assuntos (RevisaApp, 2020).

3 | METODOLOGIA

A pesquisa realizada para o projeto foi do tipo bibliográfica e aplicada. Ela foi dividida em fases. Na primeira fase, foi feita uma revisão da literatura da área de aplicativos móveis. Após esta revisão, foram estudados com maior profundidade os aplicativos existentes, que tiveram por objetivo a preparação de Estudantes. Este artigo mostra os resultados da primeira fase. Na Segunda fase, será feito um catálogo de provas dos últimos dois anos dos cursos de Tecnologia da Informação e a inserção em banco de dados próprio. Na terceira fase, serão estudadas a arquitetura de implementação de aplicativos móveis. Finalmente, na última fase, o aplicativo será disponibilizado para testes aos estudantes de Tecnologia da Informação do Centro Universitário Estácio da Bahia.

4 | RESULTADOS E CONCLUSÃO

O Aplicativo Qenade possui sua primeira versão com algumas questões de exames do ENADE anteriores. Conforme a Figura 1, o estudante precisa se identificar colocando seus dados básicos.


The image shows a mobile application interface for 'Enade Estácio'. At the top, there is a decorative banner with a desk scene including a notepad, pens, and a coffee cup. Below the banner, the title 'Enade Estácio' is displayed in a large, bold font. Underneath the title, the text 'Suas Informações' is followed by a red asterisk and the word 'Required'. The form consists of three input fields, each with a label and a red asterisk: 'Nome', 'Curso', and 'Matrícula'. Each field contains the placeholder text 'Your answer'. At the bottom of the form, there is a button with the text 'Tocar aqui para preencher a tela inteira'. The bottom of the screen shows the standard Android navigation bar with three icons: a square, a circle, and a triangle.

Figura 1: Cadastro básico do Estudante.

Após a realização do cadastro inicial, o estudante tem acesso ao banco de questões do ENADE, podendo responder pelo próprio aplicativo, conforme as Figuras 2 e 3.

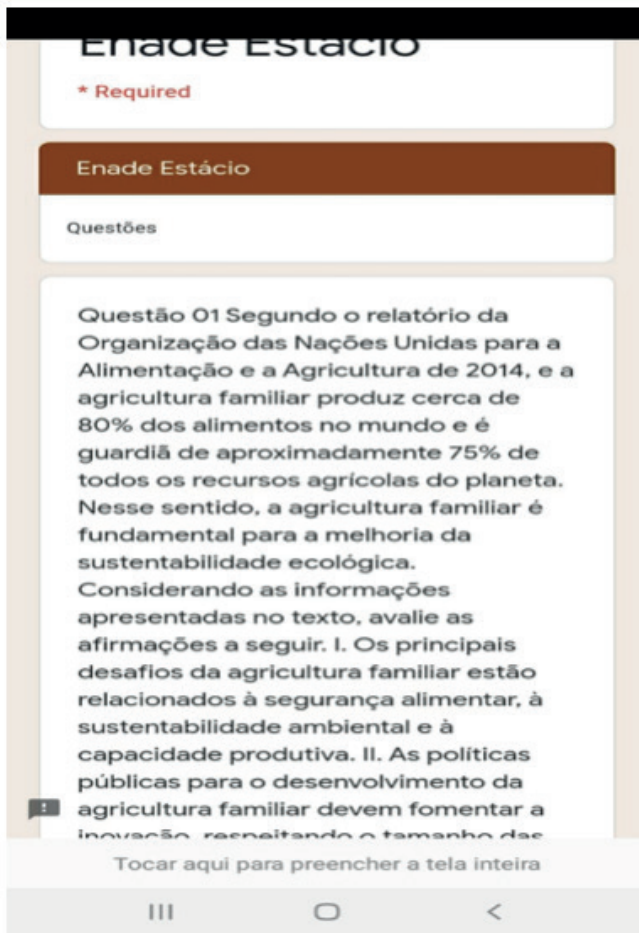


Figura 2: Acesso a uma questão do ENADE.

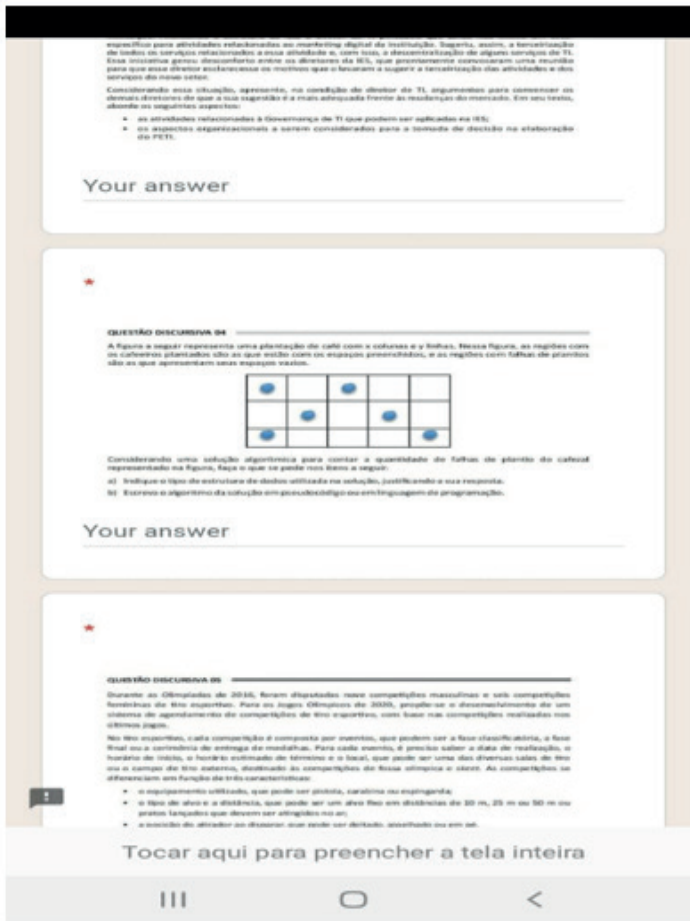


Figura 3: Visualização de Questão com imagem.

Para projetos futuros, pretende-se otimizar o banco de questões bem como melhorar a usabilidade do aplicativo. Espera-se que este aplicativo auxilie os estudantes na preparação para o ENADE, que é um importante exame realizado no final de cada curso. O aplicativo promoverá a autonomia e a liberdade ao estudante, permitindo que ele realize seus estudos em qualquer horário e lugar.

5 | RECONHECIMENTO

Este trabalho é resultado do Programa Pesquisa Produtividade do Centro Universitário Estácio da Bahia que concede bolsas para Professores Pesquisadores da Instituição. O autor agradece o incentivo a pesquisa.

REFERÊNCIAS

Duolingo. Disponível em: <https://pt.duolingo.com/>. Acesso em: 10/04/2020.

ENADE. Disponível em: <http://inep.gov.br/enade>. Acesso em: 24/07/2020.

EduEdu. Disponível em: <https://eduedu.com.br/>. Acesso em: 12/07/2020.

Física In mãos. Disponível em: <https://play.google.com/store/apps/details?id=raoni.ifrn.mestrado.fisicainmaos&hl=pt-BR>. Acesso em: 24/07/2020.

RevisaApp. Disponível em: <http://www.estuderevisapp.com/>. Acesso em: 12/07/2020.

SABOIA, J.; VARGAS, P.; VIVA, M. **O uso dos dispositivos móveis no processo de ensino e aprendizagem no meio virtual**. Revista Cesuca Virtual: conhecimento sem fronteiras. v.1, n.1, jul/2013. Disponível em: <http://www.abed.org.br/congresso2011/cd/181.pdf>. Acesso em: 20/04/2020.

TOTTI, A.R, GOMES, C.A.S, MOREIRA, S.P.T, SOUZA,W.G. **M-Learning: Possibilidades para a Educação a Distância**. Disponível em: <http://www.abed.org.br/congresso2011/cd/181.pdf>. Acessado em 16/08/2013.

SÍNTESE DOS CONCEITOS DO PENSAMENTO COMPUTACIONAL VOLTADA PARA EDUCAÇÃO BÁSICA BRASILEIRA

Data de aceite: 28/08/2020

Nayara Poliana Massa

Universidade Federal do Triângulo Mineiro
Uberaba – Minas Gerais
<http://lattes.cnpq.br/2154673448943624>

RESUMO: O ensino-aprendizagem do Pensamento Computacional (PC), ou seja, conceitos, habilidades e práticas da ciência da computação, é cada vez mais presente na vida escolar e fora dela, auxiliando as pessoas na resolução de problemas, no desenvolvimento de tecnologias e na compreensão de tendências do comportamento humano. Este artigo objetiva construir uma síntese conceitual do PC voltada para âmbito educacional referente à educação básica brasileira (da educação infantil ao ensino médio). Para isso, realizou-se pesquisa documental e bibliográfica, sistematizando os principais conceitos e aplicações. O artigo está organizado em três partes: 1) traz um levantamento bibliográfico baseado em documentos da Sociedade Brasileira de Computação, guias e publicações científicas a fim de descrever tipos de atividades, etapas, eixos das competências e habilidades do PC para cada nível de ensino; 2) ilustra e descreve algumas aplicações do PC por meio do Scratch, software lúdico que introduz a programação; 3) contextualiza a importância do curso de Licenciatura em Computação na multiplicação do PC. Esta sistematização é decorrente da necessidade de conhecer e compreender o PC de forma ampla, e de distinguir e aplicar os conceitos do

PC às atividades/práticas educativas.

PALAVRAS-CHAVE: Habilidades de Pensamento. Resolução de problemas. Scratch. Formação de Professores.

SYNTHESIS OF COMPUTATIONAL THINKING CONCEPTS FOR BRAZILIAN BASIC EDUCATION

ABSTRACT: The teaching-learning of Computational Thinking (PC), that is, concepts, skills and practices of computer science, is increasingly present in school life and beyond, assisting people in problem solving, technology development and understanding of human behavior trends. This article aims to construct a conceptual synthesis of the CP focused on the educational scope referring to Brazilian basic education (from kindergarten to high school). For this, documentary and bibliographic research was done, systematizing the main concepts and applications. The article is organized in three parts: 1) it brings a bibliographic survey based on documents of the Brazilian Society of Computation, guides and scientific publications in order to describe types of activities, stages, axes of the competences and abilities of the PC for each level of education; 2) illustrates and describes some PC applications through Scratch, playful software that introduces programming; 3) contextualize the importance of the course of Degree in Computing in the multiplication of the PC. This systematization is due to the need to know and understand the PC in a broad way, and to distinguish and apply the CP concepts

to the educational activities / practices.

KEYWORDS: Thinking skills. Problem solving. Scratch. Teacher training

1 | INTRODUÇÃO

O mundo tem enfrentado grandes avanços tecnológicos, principalmente com a facilidade de acesso à informação permitido pela internet. Tais avanços pressupõe a necessidade de pessoas com outra forma de pensamento, ou seja, aquelas que dão ênfase à cultura tecnológica e que vislumbrem sua importância, crescimento e utilidade, pessoas que pensem computacionalmente.

Wing (2006) percebeu a importância do Pensamento Computacional (PC) como forma de auxiliar todas as pessoas em várias áreas do conhecimento desde a infância, facilitando o entendimento e a resolução de problemas por meio dos conceitos da ciência da computação.

A programação de computadores pode preparar as pessoas para algo mais do que se tornarem somente desenvolvedores ou cientistas da computação. A programação auxilia no desenvolvimento do PC.

Como uma forma de mudança da cultura tecnológica, a programação começa a ser valorizada na educação desde a infância. Papert já compreendia a importância de se inserir o PC para crianças desde a década de 1980, como forma de aguçar os desafios que elas precisam para se prenderem a novos conhecimentos, ou seja, torná-las produtoras de tecnologia e não somente consumidoras. Assim, defendeu que “a melhor aprendizagem ocorre quando o aprendiz assume o comando” (PAPERT, 1994, p. 37). O interesse, naquela época, estava em formar adultos com uma visão diferente de futuro, alterando sua forma de pensar, tornando a cultura tecnológica como algo habitual e de grande importância.

Decorrente da necessidade de aliar a tecnologia com a educação, para acrescentar conhecimentos, muitas escolas, projetos sociais, faculdades, institutos e universidades aderiram ao desenvolvimento de habilidades do PC. Uma das ferramentas utilizadas para atrair a atenção das crianças e, ao mesmo tempo, ensinar o PC é denominada *Scratch*.

O *Scratch* é um *software* desenvolvido por um grupo de pesquisadores do *Massachusetts Institute of Technology* (MIT), nos Estados Unidos. Criado, principalmente, para crianças, pois trabalha, de forma lúdica, a introdução à lógica de programação e ao Pensamento Computacional.

O *Scratch* apresenta um “guia curricular”, que traz um quadro de referência a respeito do aprendizado das habilidades do PC para mediar conceitos, práticas e perspectivas computacionais.

Pensadores computacionais, de acordo com o guia do *Scratch* são “indivíduos que podem recorrer a conceitos computacionais, práticas e perspectivas em todos os aspectos das suas vidas, em várias disciplinas e contextos” (COMPUTAÇÃO CRIATIVA, 2011).

Acredita-se, assim, que o incentivo à aprendizagem de lógica de programação por meio do *Scratch*, pode abrir novos caminhos, além de possibilitar mudanças de paradigmas culturais e tecnológicos.

De acordo com Blikstein (2008), “o mundo atual exige muito mais do que ler, escrever, adição e subtração”, que são habilidades básicas a serem aprendidas no ensino regular. O PC trabalha alguns pilares no ensino-aprendizagem, servindo “para aumentar nossa produtividade, inventividade e criatividade”. Aptidões muito exigidas e valorizadas no cenário atual.

Neste artigo, inicialmente, aborda-se o Pensamento Computacional, descrevendo sua definição, contexto histórico, etapas que auxiliam na resolução de problemas, contextualizando as habilidades relacionadas ao PC abordadas pela Sociedade Brasileira de Computação (SBC) e apresentando atividades que auxiliam o seu desenvolvimento. Em seguida são detalhados os conceitos, as práticas e as perspectivas do Pensamento Computacional trabalhados com a utilização do *software Scratch*. A última seção trata da Licenciatura em Computação e o Pensamento Computacional aborda a relação do curso com o ensino-aprendizagem do PC.

2 | PENSAMENTO COMPUTACIONAL

Pensamento Computacional consiste em processos, com base em conceitos computacionais, que auxiliam na compreensão e resolução de problemas, sendo eles executados por pessoas ou computadores.

A ideia de Pensamento Computacional foi mencionada primeiramente pelo matemático Seymour Papert, em 1980, em seu livro “Logo: computadores e educação” (PAPERT, 1985). No livro, ele idealizava um mundo onde a tecnologia seria essencial para os afazeres diários, dando grande importância para desenvolver habilidades computacionais desde criança, tendo como base a inserção de tal conceito na educação desde cedo. Uma das questões de interesse do autor estava direcionada para “como as pessoas pensam e como aprendem a pensar” (PAPERT, 1985, p. 24) e em “como o computador poderia contribuir para os processos mentais não somente como instrumento, mas, essencialmente, de maneira conceitual, influenciando o pensamento das pessoas, mesmo quando estas estiverem fisicamente distantes dele” (PAPERT, 1985, p. 16). O matemático enfocava a utilização do computador, juntamente com a utilização da linguagem de programação LOGO, para obter respostas dos questionamentos.

A linguagem computacional LOGO, que foi criada em 1967 pelo grupo LOGO (composto por professores e alunos do MIT), objetiva fazer a criança se comunicar com o computador. O ambiente LOGO permite que a criança, através de comandos de computador, controle uma tartaruga, que desenha traços na tela, conforme o código informado. “Essa tartaruga serve ao único propósito de ser fácil de programar e boa para se pensar” (PAPERT, 1985).

Com a publicação do artigo “*Computational Thinking: it represents a universally applicable attitude and skill set everyone, not just computer scientists, would be eager to learn and use*” (WING, 2006) o conceito de PC ganhou destaque na comunidade da ciência da computação. Wing (2006) defende que este auxilia na vida de qualquer pessoa, desde crianças, declarando-o como habilidade fundamental para todos. Uma vez, que ajuda na resolução de variados problemas, inclusive os problemas do dia a dia, além de relatar a importância e influência do PC em várias disciplinas, como estatística e biologia, tudo isso baseando-se nos conceitos de ciência da computação.

Pensamento computacional envolve a resolução de problemas, projeção de sistemas, e compreensão do comportamento humano, através da extração de conceitos fundamentais da ciência da computação. O pensamento computacional inclui uma série de ferramentas mentais que refletem a vastidão do campo da ciência da computação (WING, 2006, p. 2).

A eficiência na resolução de problemas relativas ao PC é dada por alguns fundamentos teóricos da ciência da computação (Quadro 1). Esses, divididos em etapas, objetivam transformar um problema grande e complexo em problemas menores, de tal forma que possam ser compreendidos, facilitando a visualização da resolução, tornando-o tratável.

Etapas	Descrição
Decomposição	Decompor o problema difícil em partes menores, tornando assim, tais partes, em problemas mais simples que a pessoa saiba resolver
Reconhecimento de padrões	Agrupar as partes do problema por similaridade, identificando algum padrão entre elas.
Abstração	Filtrar o que é realmente importante para resolver o problema e focar nisso, ignorando detalhes que não são essenciais.
Algoritmo	Criar uma sequência de passos, baseada nos resultados de todas as etapas anteriores, para a resolução do problema.

Quadro 1 – Etapas para resolução de problemas pensando computacionalmente.

Fonte: Elaborado pela Autora, 2020.

De acordo com Silva, Souza e Morais (2016) e Cavalcante et al. (2017), as organizações americanas *Computer Science Teachers Association* (CSTA), *International Society for Technology in 325 Education* (ISTE) e *National Science Foundation* (NSF) desenvolveram algumas ferramentas e diretrizes, a *Computational Thinking Task Force*, para auxiliar no ensino de PC nas escolas do ensino primário e secundário (equivalente a ensino fundamental e médio, no Brasil). Tais diretrizes contemplam nove etapas, algumas

já mencionadas na Quadro 1, do ensino sobre o PC. São elas: coleta de dados, análise de dados, representação de dados, decomposição do problema, abstração, algoritmos e procedimentos, automação, paralelização e simulação.

A Sociedade Brasileira de Computação, em 2017, disponibilizou um documento denominado “Referenciais de Formação em Computação: Educação Básica”. Este documento foi criado por uma comissão composta por vários professores, uma vez que a SBC “entende que é fundamental e estratégico para o Brasil que conteúdos de Computação sejam ministrados na Educação Básica” (RAAB; RIBEIRO, 2017, p. 1).

Esse documento trabalha com os três eixos principais da área de computação:

- **Pensamento Computacional:** “capacidade de sistematizar, representar, analisar e resolver problemas” (RAAB; RIBEIRO, 2017, p. 3).
- **Mundo Digital:** composto por componentes físicos, que são as máquinas (computadores e celulares, por exemplo), e virtuais compostos por dados e programas, como por exemplo a internet.
- **Cultura Digital:** interação do homem com as tecnologias do mundo digital e sua influência no cotidiano.

Em cada eixo são exploradas competências que podem ser trabalhadas da Educação Infantil ao Ensino Médio, como mostra o Quadro 2:

Eixo	Competências
Pensamento Computacional	<p>Abstração: transformar as informações em processos algorítmicos.</p> <p>Automação: descrever soluções de forma que a máquina entenda. Através de algoritmos.</p> <p>Análise: verificar as possibilidades de solução de problemas, bem como correção de erros e avaliação de eficiência.</p>
Mundo Digital	<p>Codificação: entender como as informações podem ser descritas e armazenadas.</p> <p>Processamento: perceber como o computador processa a informação.</p> <p>Distribuição: compreender como é feita a comunicação entre dispositivos digitais e como os dados chegam, de forma segura e íntegra.</p>
Cultura Digital	<p>Computação e Sociedade: compreender o impacto dos avanços do uso da tecnologia na sociedade.</p> <p>Fluência Tecnológica: utilizar ferramentas tecnológicas para seus diversos fins.</p> <p>Ética Digital: analisar as questões éticas e morais causadas pelo mundo digital.</p>

Quadro 2 – Competências desenvolvidas por cada eixo da computação - documento da SBC.

Fonte: Adaptado de Raab e Ribeiro (2017).

O documento destaca habilidades de cada eixo em todas as etapas da educação

básica¹:

Etapa	Habilidades
Educação Infantil	<ul style="list-style-type: none">• Compreender uma situação problema criando e identificando sequências de passos de uma tarefa para sua solução.• Representar os passos de uma tarefa através da notação pictórica, de forma organizada e relacional.• Criar passos para solução de problemas relacionados ao movimento do corpo e trajetórias espaciais.
Ensino Fundamental Anos Iniciais (1.º ao 5.º ano)	<ul style="list-style-type: none">• Representar em experiências concretas as principais abstrações para descrever dados: registros, listas e grafos.• Identificar as principais abstrações para construir processos: escolha, composição e repetição, simulando e definindo algoritmos simples que representem situações do cotidiano infantil• Utilizar linguagem lúdica visual para representar algoritmos.• Compreender a técnica de decompor um problema para solucioná-lo.
Ensino Fundamental Anos Finais (6.º ao 9º ano)	<ul style="list-style-type: none">• Utilizar linguagens visuais e língua nativa para representar dados e processos.• Formalizar os conceitos de dados estruturados (registros, listas, grafos).• Empregar o conceito de recursão, para a compreensão mais profunda da técnica de solução através de decomposição de problemas.• Construir soluções de problemas usando a técnica de generalização, permitindo o reuso de soluções de problemas em outros contextos, aperfeiçoando e articulando saberes escolares.• Relacionar um algoritmo descrito em uma linguagem visual com sua representação em uma linguagem de programação.
Ensino Médio	<ul style="list-style-type: none">• Elaborar projetos integrados às áreas de conhecimento curriculares, em equipes, solucionando problemas, usando computadores, celulares, e outras máquinas processadoras de instruções.• Compreender a técnica de solução de problemas através de transformações: comparar problemas para reusar soluções.• Analisar algoritmos quanto ao seu custo (tempo, espaço, energia, ...) para poder justificar a adequação das soluções a requisitos e escolhas entre diferentes soluções.• Argumentar sobre a correção de algoritmos, permitindo justificar que uma solução de fato resolve o problema proposto.• Reconhecer o conceito de meta-programação como uma forma de generalização.• Entender os limites da Computação para diferenciar o que pode ou não ser mecanizado, buscando uma compreensão mais ampla dos processos mentais envolvidos na resolução de problemas.

Quadro 3 – Habilidades relacionadas ao PC nas etapas da educação - documento da SBC.

Fonte: Adaptado de Raab e Ribeiro (2017).

¹ Neste artigo são destacadas apenas as habilidades relacionadas ao eixo PC, divididas pelas etapas educacionais, descritas no Quadro 3.

Para desenvolver as habilidades do PC, algumas atividades práticas podem ser implementadas para auxiliar no aprendizado, tais como se apresentam no Quadro 4.

Prática	Descrição
Computação desplugada	São atividades propostas que auxiliam o aluno a pensar computacionalmente, mas sem a utilização do computador.
Jogos digitais	Os recursos de jogos são utilizados com finalidade de ajudar na aprendizagem sobre PC.
Linguagens de programação	As linguagens de programação de alto nível, como C e JAVA, trabalham fortemente com a habilidade de PC
Linguagens de programação visual	São linguagens que trabalham com a utilização de blocos gráficos, permitindo clicar e arrastar, facilitando a programação. Uma linguagem muito utilizada para trabalhar PC é o <i>Scratch</i> .
Robótica	Emprega objetos robóticos para auxiliar no entendimento do PC.

Quadro 4 – Atividades para desenvolvimento do PC.

Fonte: Adaptado de Zanetti; Borges e Ricarte (2016).

3 | SCRATCH

O *Scratch* é uma linguagem de programação visual que foi desenvolvida pelo grupo de pesquisa *Lifelong Kindergarten Group* (LLK), pertencente ao MIT, criado em 2007. Essa ferramenta foi desenvolvida, para crianças entre os 8 e os 16 anos, bem como para auxiliar pessoas que estão começando a programar. O *Scratch* foi desenvolvido baseado na linguagem LOGO, criada por Seymour Papert, e é uma ferramenta gratuita e de código aberto.

Essa linguagem de programação tem caráter educativo, apoiando no aprendizado de conceitos de matemática e computação. Ela foi desenvolvida baseada no conceito *drag-and-drop* (arrastar e soltar) com comandos em forma de blocos que se encaixam, lembrando o brinquedo Lego.

O *Scratch* possui uma interface gráfica composta por quatro ambientes: os botões de programação; a área de código (onde os comandos são inseridos); a tela de animação; os objetos e o palco. Essa ferramenta utiliza os comandos, botões de programação, em forma de blocos. Esses blocos são arrastados para a área de código e possuem um formato que pode ser comparado a um quebra-cabeça, ou seja, os blocos se encaixam e, dessa forma, os códigos são criados.

Os comandos são divididos em categorias e, cada uma delas, se relacionam às ações que serão dadas para que o programa funcione. Tais ações podem ser atribuídas tanto para o objeto quanto para o cenário, bastando selecionar um dos dois no canto inferior direito da tela do *Scratch*. Por fim, o resultado do código desenvolvido pode ser visto pela

tela de animação que retrata as ações descritas pelo programa. A Figura 1 apresenta a tela do *Scratch* de forma geral, identificando os quatro ambientes citados.

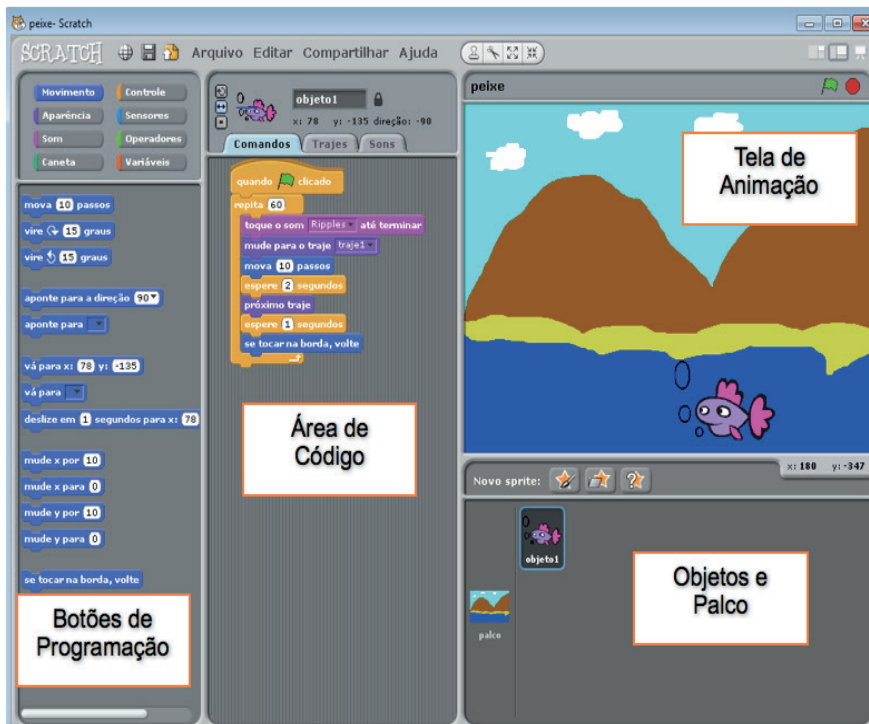


Figura 1 – Tela *Scratch*

Fonte: Adaptado de Brandt (2012).

O *Scratch* apoia o desenvolvimento de competências de aprendizagem do século XXI. Conforme descrito por Rusk, Resnick e Maloney (2006), ao trabalhar nove tipos de habilidades de aprendizagem divididas em três grupos-chave:

- **Grupo 1 – Habilidades em informação e comunicação:** ao ter contato com o *Scratch* os alunos aprendem a manipular mídias, como imagens, animações, textos e áudios. Com isso ficam familiarizados com os meios de comunicação, podendo criar seu próprio meio de comunicação.
- **Grupo 2 – Habilidades de pensamento e solução de problemas:** ao longo do desenvolvimento de projetos (construção de códigos) com o *Scratch* os alunos aprimoram o pensamento crítico e o raciocínio lógico,

- **Grupo 3 – Habilidades interpessoais:** o *Scratch* incentiva o trabalho em grupo e a reutilização de código para seu aprimoramento. Todos os projetos desenvolvidos podem ser compartilhados pela internet, serem acessados por outros estudantes, e assim, trabalharem e comentarem colaborativamente o código disponível do projeto. Tais propostas auxiliam na motivação, superação de desafios e frustrações, além de provocar discussões entre as pessoas que participam da comunidade *Scratch*.

3.1 O SCRATCH E O PENSAMENTO COMPUTACIONAL

Em 2011, foi criado pelo MIT um guia de atividades (COMPUTAÇÃO CRIATIVA, 2011), com programa de aulas, destinado a educadores que queriam explorar conceitos, práticas e perspectivas do PC utilizando o *Scratch* como ferramenta da computação criativa. Esse guia, esboça um projeto curricular, descrevendo habilidades ligadas ao PC e, de maneira geral, alia os interesses dos jovens com a computação, tendo como base a criatividade e imaginação, apoiando também o desenvolvimento desses jovens como pensadores computacionais.

De acordo com Brennan e Resnick (2012), após vários anos de pesquisa e estudo das atividades desenvolvidas nas comunidades que programam com *Scratch*, eles desenvolveram três dimensões-chave do pensamento computacional:

- **Conceitos computacionais:** são os conceitos empregados no programa.
- **Práticas computacionais:** na prática os desenvolvedores se envolvem com os conceitos computacionais, depuram o programa ou o corrigem, por exemplo.
- **Perspectivas computacionais:** consiste em como os desenvolvedores se sentem com eles mesmos e com o ambiente ao redor.

O Guia Computação Criativa (2011) apresenta uma lista de competências e de habilidades ligadas aos conceitos, práticas e perspectivas computacionais, o Quadro 5, resume as três dimensões citadas:

Conceitos Computacionais	Descrição
Sequência	Identificar uma série de etapas de uma tarefa
Ciclos	Executar a mesma sequência várias vezes
Execução em paralelo	Fazer as ações decorrerem ao mesmo tempo
Eventos	Fazer um acontecimento causar outro acontecimento
Condições	Tomar decisões com base em condições
Operadores	Expressar operações matemáticas e lógicas
Dados	Armazenar, recuperar e atualizar valores

Práticas Computacionais	Descrição
Ação iterativa e incremental	Desenvolver um pouco, depois verificar se funciona e, em seguida, desenvolver um pouco mais
Teste e depuração	Certificar-se de que tudo funciona e encontrar e corrigir erros
Reutilização e reformulação	Fazer algo utilizando o que outros – ou a própria pessoa – já fizeram
Abstração e modulação	Construir algo grande unindo conjuntos de partes mais pequenas
Perspectivas Computacionais	Descrição
Expressar	Perceber que a computação é um meio de criação. “Eu posso criar.”
Conectar	Reconhecer a vantagem de criar com e para outros “Eu posso ter novas ideias quando tenho acesso a outros.”
Questionar	Sentir que se pode fazer perguntas sobre o mundo “Eu posso (utilizar a computação para) suscitar questões que façam sentido (com entes computacionais) para o mundo.”

Quadro 5 – Conceitos, práticas e perspectivas do pensamento computacional

Fonte: Computação Criativa (2011).

3.1.1 CONCEITOS COMPUTACIONAIS

Os conceitos computacionais abrangem sete definições muito usadas nos projetos *Scratch*. Cada conceito será descrito e exemplificado conforme alíneas “a” a “g”.

a) Sequência

Na programação as instruções do código nada mais são que uma sequência de passos que são executados. Cada passo tem uma função e um comportamento específico.

b) Ciclos

Em algumas situações uma ação precisará ser executada várias vezes e os ciclos auxiliam nesta tarefa, cujo código que se repetirá, ficará dentro de um comando chamado “repita” no *Scratch*.

c) Execução em paralelo

No *Scratch* permite a execução em paralelo, que nada mais é do que a execução de várias ações ao mesmo tempo.

d) Eventos

Os eventos são comandos que causam acontecimentos de outros acontecimentos. Eles são importantes componentes de interação dentro do *Scratch*.

e) Condições

Os comandos condicionais executam determinadas ações de acordo com condições pré-estabelecidas.

f) Operadores

Os comandos operadores permitem executar ações de operações lógicas e matemáticas, além de trabalhar com manipulação de *strings*, ao qual se trabalha com concatenação e contagem de caracteres.

g) Dados

Os dados têm a função de armazenar, recuperar e atualizar valores. No *Scratch* é possível fazer isso por meio de variáveis e listas.

3.1.2 Práticas computacionais

As práticas computacionais descrevem o processo de construção dos programas, em outras palavras, descrevem quais as práticas e estratégias são mais utilizadas ao longo do desenvolvimento dos projetos no *Scratch*.

Brennan e Resnick (2012, p.7) entendem que “as práticas computacionais se concentram no processo de pensar e aprender, além do que você está aprendendo a aprender” e destacam quatro principais grupos de práticas que serão detalhadas a seguir:

- **Ação incremental e iterativa**

Essa prática retrata que ao se ter um problema para resolver com o *Scratch* ele pode mudar e ser alterado a qualquer momento, em outras palavras, o projeto exposto é revisado ciclicamente, acrescentando ou retirando dados, até que o problema seja resolvido. Todo o processo desta prática é adaptativo.

- **Teste e depuração**

Ao se desenvolver um código é necessário ver se o mesmo está funcionando conforme o propósito do projeto, para isso são feitos testes ao longo do desenvolvimento e se no teste for detectado algum erro, o código é revisto até o problema ser encontrado, podendo assim, ser corrigido, sendo feita assim a depuração.

- **Reutilização e reformulação**

Esta prática é utilizada para agilizar o desenvolvimento, permitindo a criação de novos programas baseados nos que já existem, reformulando-os, tornando-os mais complexos e melhores. Para Brennan e Resnick (2012, p. 8) a prática reutilizar e reformular auxilia no “desenvolvimento de capacidades críticas de leitura de código e provoca questões importantes sobre propriedade e autoria”. O *Scratch* possui várias comunidades no site que disponibilizam os códigos dos projetos compartilhados para que possam ser estudados e baixados de forma gratuita.

- **Abstração e modulação**

A abstração e a modularização nada mais são do que dividir um problema grande em partes menores e juntar essas partes menores conforme as afinidades de cada parte, facilitando, a partir de então, a visualização da questão levantada, a qual poderá ser analisada por partes.

3.1.3 PERSPECTIVAS COMPUTACIONAIS

Além dos conceitos e práticas computacionais, uma outra habilidade do PC é observada em quem desenvolve programas com o *Scratch*, as perspectivas computacionais que descrevem algumas mudanças de perspectivas, como a própria denominação sugere.

Como já mencionado, ao longo deste trabalho, as perspectivas computacionais foram divididas em três, e serão descritas com mais detalhes a seguir:

a) Expressar

O *Scratch* permite que as pessoas possam se expressar tecnologicamente. “Um pensador computacional vê a computação mais do que algo para consumir. A computação é algo que eles podem usar para design e autoexpressão” (BRENNAN; RESNICK, 2012, p. 8).

b) Conectar

A partir da criação de programas com o *Scratch* é possível trabalhar as interações em redes sociais, uma vez que esta ferramenta possui vários canais de compartilhamento dos projetos criados. Brennan e Resnick (2012, p.11) apontam duas questões de interação social que foram observadas com quem cria com o *Scratch*: o valor de criar com outros e criar para outros. Ao se criar com outros (visão colaborativa) e para os outros (visão de publicação), amplia-se o alcance das possibilidades de desenvolvimento. O valor de criar “para outros”, envolve a apreciação de suas criações pelo público, se conectar com as pessoas através dos projetos desenvolvidos, inclusive através da educação.

c) Questionar

A tecnologia tem permitido a criação de sistemas cada vez mais complexos. Esses sistemas parecem, para muitas pessoas, intangíveis, uma vez que, pensar em “como funcionam” ou em “como alterá-los” parece muito distante para qualquer um.

As pessoas que aprendem a programar com o *Scratch*, começam a questionar a forma que programas e dispositivos foram desenvolvidos, mudando o pensamento, se aproximando da computação. A perspectiva computacional explora a distância que as pessoas que programam com o *Scratch* sentem entre suas habilidades e a tecnologia que as cerca.

4 | A LICENCIATURA EM COMPUTAÇÃO E O PENSAMENTO COMPUTACIONAL

Segundo o Guia do Estudante (2018) o curso de Licenciatura em Computação forma profissionais que podem atuar como professores que lecionam informática para ensino fundamental, médio, cursos profissionalizantes, dentre outros. O licenciado, além de lecionar, pode atuar como consultor na área pedagógica de cursos de informática e como desenvolvedor de *softwares*.

Muitas instituições de ensino aderem às aulas de informática no currículo, necessitando de profissionais com conhecimento tanto na área específica, quanto na área pedagógica. Daí a importância de se formar profissionais licenciados em computação.

O Curso de Licenciatura em Computação tem apoiado a aprendizagem do PC, de forma que seja inserido no ensino já a partir da infância. Alunos deste curso têm muitas vantagens em aprender conceitos do PC e são responsáveis por ministrarem vários projetos com esse foco.

Alguns projetos de extensão têm sido fomentados por alunos de Licenciatura em Computação. Alunos do Instituto Federal de Educação, Ciência e Tecnologia da Bahia (IFBA) ministraram um curso, no qual o objetivo foi ensinar o PC, focando em mulheres, estudantes da rede pública, com idade entre 11 e 18 anos, notando-se a masculinização das ciências exatas, principalmente da Computação (SILVA; FRANCO; JUNIOR, 2017).

Outro projeto foi aplicado por estudantes de Licenciatura em Computação, da Universidade Federal do Paraná (UFPR), com o intuito de ensinar programação de computadores às crianças. O curso foi ministrado em uma escola estadual da cidade de Palotina-PR e abrangendo alunos entre o 6º e o 8º ano do ensino fundamental (SOARES; CERCI; MONTE-ALTO, 2016).

Na resolução nº 5, do Ministério da Educação (MEC) de novembro 2016 (BRASIL, 2016), o Art. 4 enumera vários elementos importantes, que os profissionais formados nos cursos de Bacharelado e Licenciatura em Computação devem ser dotados. O PC entra no inciso VII, sendo citado da seguinte forma: “da capacidade de reconhecer a importância do pensamento computacional na vida cotidiana, como também sua aplicação em outros domínios e ser capaz de aplicá-lo em circunstâncias apropriadas”.

A Licenciatura em Computação é um curso-chave para fortalecer e ser multiplicador do ensino do pensamento computacional, pois tem o objetivo de capacitar profissionais para o exercício do magistério que podem atuar no ensino básico, técnico e tecnológico na área de computação e informática.

5 | CONCLUSÃO

Este trabalho se ateve à conceituação e à exposição de possibilidades de ensino-aprendizagem do Pensamento Computacional no contexto educacional, em especial, enfocando conceitos, práticas e perspectivas do PC, contextualizado e trabalhado por

meio software *Scratch*. Além disso, foi destacada a importância do PC pela Sociedade Brasileira de Computação e a inserção do PC nas diretrizes do curso de Licenciatura em Computação, principal multiplicador deste conhecimento nas escolas.

Um tema de tal relevância e de grande efervescência pode levar as pessoas a desenvolverem muitos tipos de aplicações. Quando essas aplicações se dão em contexto educacional, com finalidade de ensino-aprendizagem, cabe ao docente (seja da educação formal ou informal) o papel de mediador responsável. Nesse sentido, este artigo busca trazer uma sistematização conceitual, a fim de mostrar que há diferentes definições e níveis de relações conceituais; que há diferentes objetivos conforme as finalidades das ações educativas; que há possibilidades de aplicações e avaliação ou validação dos projetos desenvolvidos com o PC.

É necessário também que a Sociedade Brasileira de Computação, os Institutos Tecnológicos de Educação, Ciência e Tecnologia, as Universidades, os gestores da educação e organizações não governamentais, deem atenção, estabeleçam diálogo e colaborem com o fortalecimento das políticas públicas de formação de docentes a fim de valorizar e enriquecer o trabalho profissional de quem atua na área de informática e educação. Isso pode ser fomentado por meio de projetos formativos, nos quais, além dos cursos de Licenciatura em Computação, outros profissionais interessados (tais como, cursos técnicos em Informática, cursos superiores de Ciência da Computação, Análise e Desenvolvimento de Sistemas ou Engenharia da Computação), para que possam vir a ser multiplicadores do PC, do mundo e da cultura digital no país.

REFERÊNCIAS

BLIKSTEIN, P. **O pensamento computacional e a reinvenção do computador na educação**. In: PAULO BLIKSTEIN. Stanford, CA, 2008. Disponível em: http://www.blikstein.com/paulo/documents/online/ol_pensamento_computacional.html. Acesso em: 09 out. 2018.

BRANDT, I. **Scratch**. In: ESCADINHA DO SABER. [S.I.], 2012. Disponível em: <http://escadinhadosaber.blogspot.com.br/2012/06/scratch.html>. Acesso em: 26 abr. 2017.

BRASIL. Ministério da Educação. Conselho Nacional de Educação Câmara de Educação Superior. **Resolução nº 5, de 16 de novembro de 2016**. Institui as Diretrizes Curriculares Nacionais para os cursos de graduação na área da Computação, abrangendo os cursos de bacharelado em Ciência da Computação, em Sistemas de Informação, em Engenharia de Computação, em Engenharia de Software e de licenciatura em Computação, e dá outras providências. Brasília, DF, 2016. Disponível em: http://portal.mec.gov.br/index.php?option=com_docman&view=download&alias=52101-rces005-16-pdf&category_slug=novembro-2016-pdf&Itemid=30192. Acesso em: 20 jun. 2018.

BRENNAN, K.; RESNICK, M. **New frameworks for studying and assessing the development of computational thinking**. In: *Annual American Educational Research Association Meeting*, 2012, Vancouver, Canadá. AERA 2012. Washington, DC, 2012. Disponível em: http://web.media.mit.edu/~kbrennan/files/Brennan_Resnick_AERA2012_CT.pdf. Acesso em: 24 nov. 2017.

CAVALCANTE, A. F. et al. Um estudo exploratório da aplicação de pensamento computacional baseado nas perspectivas de professores do ensino médio. In: CONGRESSO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO, 6., 2017, Recife. **Anais dos Workshops do CBIE 2017**. Porto Alegre: Sociedade Brasileira de Computação, 2017. p. 992-1001. Disponível em: <<http://www.br-ie.org/pub/index.php/wcbie/article/view/7488>>. Acesso em: 20 jan. 2018.

COMPUTAÇÃO CRIATIVA: **uma introdução ao pensamento computacional baseada no conceito de design**. Tradução: EduScratch. Massachusetts: MIT, 2011. Disponível em: <<http://projectos.ese.ips.pt/cttic/wp-content/uploads/2011/10/Guia-Curricular-ScratchMIT-EduScratchLPpdf.pdf>>. Acesso em: 20 ago. 2017.

GUIA do Estudante. **Computação**. São Paulo: Grupo Abril, 2018. Disponível em: <<https://guiadoestudante.abril.com.br/profissoes/computacao/>>. Acesso em: 24 jun. 2018.

LIFELONG KINDERGARTEN GROUP MIT MEDIA LAB. *About Scratch*. [S.l.], [201-]. Disponível em: <<https://scratch.mit.edu/about/>>. Acesso em: 18 abr. 2017.

PAPERT, S. **A máquina das crianças**: repensando a escola na era da informática. Porto Alegre, Artes Médicas, 1994.

PAPERT, S. **LOGO**: computadores e educação. São Paulo, Brasiliense, 1985.

RAAB, A. L. A.; RIBEIRO, L.. (Coord.). **Referenciais de formação em computação: educação básica**. Porto Alegre: SBC, 2017. Documento aprovado pela Comissão de Educação e apresentado no CSBC 2017 durante as Assembleias do WEI e da SBC. Disponível em: <<http://www.sbc.org.br/files/ComputacaoEducacaoBasica-versaofinal-julho2017.pdf>>. Acesso em: 25 jun. 2018.

RUSK, N.; RESNICK, M.; MALONEY, J. **Learning with scratch**: 21st century learning skills. Cambridge, MA: MIT Media Laboratory, 2006. Disponível em: <<https://llk.media.mit.edu/papers/scratch-21st-century.pdf>>. Acesso em: 24 abr. 2017.

SILVA, A. P. S.; FRANCO, J. S. S.; JUNIOR, J. C. L. Desenvolvimento do pensamento computacional e discussões sobre representação feminina na computação: um estudo de caso. In: CONGRESSO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO, 6., 2017, Recife. **Anais dos Workshops do CBIE 2017**. Porto Alegre: Sociedade Brasileira de Computação, 2017. p. 1111-1120. Disponível em: <<http://www.br-ie.org/pub/index.php/wcbie/article/view/7500>>. Acesso em: 31 jul. 2018.

SILVA, V. SOUZA, A. MORAIS, D. Pensamento computacional no ensino de computação nas escolas: um relato de experiência de estágio em licenciatura em computação em escolas públicas. In: CONGRESSO REGIONAL SOBRE TECNOLOGIAS NA EDUCAÇÃO, 1., 2016, Natal. **Anais...** Natal, RN. p. 324-335. Disponível em: <http://ceur-ws.org/Vol-1667/CtrlE_2016_AC_paper_55.pdf>. Acesso em: 11 jan. 2018.

SOARES, J. P. R. S.; CERCI, R. G.; MONTE-ALTO, H. H. L.C. Clube de programação e oficinas com o Scratch: um relato de experiência. In: CONGRESSO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO, 5., 2016, Uberlândia. **Anais dos Workshops do CBIE 2016**. Porto Alegre: Sociedade Brasileira de Computação, 2016. p. 958-962. Disponível em: <<http://www.br-ie.org/pub/index.php/wie/article/view/6618>>. Acesso em: 31 jul. 2018.

WING, J. M. Pensamento Computacional: um conjunto de atitudes e habilidades que todos, não só cientistas da computação, ficaram ansiosos para aprender e usar. Tradução: Cleverson Sebastião dos Anjos. **Revista Brasileira de Ensino de Ciência e Tecnologia**, Ponta Grossa, v. 9, n.2, p. 1-10, maio/ago. 2016. Disponível em: < <https://periodicos.utfpr.edu.br/rbect/article/view/4711>>. Acesso em: 01 jan. 2018.

ZANETTI, H. A. P.; OLIVEIRA, C. L. V. Prática de ensino de programação de computadores com robótica pedagógica e aplicação de pensamento computacional. In: CONGRESSO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO, 4., 2015, Maceió. **Anais dos Workshops do CBIE 2015**. Porto Alegre: Sociedade Brasileira de Computação, 2015. p. 1236-1245. Disponível em: <<http://www.br-ie.org/pub/index.php/wcbie/article/view/6268>>. Acesso em: 22 nov. 2018.

SISTEMA DE LOCALIZAÇÃO HÍBRIDO BASEADO EM NUVEM PARA AMBIENTES INTERNOS E EXTERNOS

Data de aceite: 27/08/2020

Data de submissão: 07/08/2020

Raul de Queiroz Mendes

Universidade Federal de São Carlos
(UFSCar)

São Carlos – São Paulo

<http://lattes.cnpq.br/5042730285685241>

Roberto Santos Inoue

Universidade Federal de São Carlos
(UFSCar)

São Carlos – São Paulo

<http://lattes.cnpq.br/6221209121565990>

Tatiana de Figueiredo Pereira Alves Taveira Pazelli

Universidade Federal de São Carlos
(UFSCar)

São Carlos – São Paulo

<http://lattes.cnpq.br/2574900514697052>

Rafael Vidal Aroca

Universidade Federal de São Carlos
(UFSCar)

São Carlos – São Paulo

<http://lattes.cnpq.br/9262228584082064>

com cinemática em tempo real (RTK). O GPS RTK, no entanto, ainda tem um alto custo de implantação, depende de estações base individuais e seu desempenho é prejudicado em ambientes fechados ou mesmo externos na presença de árvores ou nuvens. Ademais, existe uma necessidade crescente de se localizar dispositivos móveis, como tablets e smartphones, ou mesmo robôs móveis em ambientes internos, onde o GPS não funciona bem. Assim, este trabalho propõe um sistema de localização híbrido capaz de usar diferentes tecnologias para fornecer a localização interna e externa de robôs ou de outros dispositivos móveis. Para a localização externa, explora-se a tecnologia GPS. Para a localização interna, utiliza-se a trilateração de sinais de Beacons Bluetooth Chiiip 01 da Animaltag e de sinais Wifi gerados por módulos ESP8266. Acrescentando-se a isso, um sistema em nuvem foi desenvolvido para processar, armazenar e visualizar dados do sistema de localização híbrido.

PALAVRAS-CHAVE: GPS, sistema de localização híbrido, Beacons Bluetooth, ESP8266, sistema em nuvem.

RESUMO: Com o surgimento do sistema de posicionamento global (GPS), diversos setores passaram a se beneficiar desta tecnologia, como as áreas de: serviços de localização, cadeias logísticas, jogos, agricultura, negócios, segurança, entre outras. O GPS é um mecanismo que retorna a localização de dispositivos em ambientes externos com uma precisão da ordem de metros e cerca de centímetros quando empregado o GPS

CLOUD-BASED HYBRID LOCATION SYSTEM FOR INDOOR AND OUTDOOR ENVIRONMENTS

ABSTRACT: With the emergence of the global positioning system (GPS), several sectors started to benefit from this technology, such as the areas of location services, logistics chains, games, agriculture, business, security, among others. GPS is a mechanism that returns the

location of devices in outdoor environments with an accuracy of the order of meters and about centimeters when using GPS with real-time kinematics (RTK). The GPS RTK, however, still has high costs, it depends on individual base stations and its performance is impaired indoors or even outdoors in the presence of trees or clouds. Moreover, there is a growing need to locate mobile devices, such as tablets and smartphones, or even indoor mobile robots, where GPS does not work well. Thus, this work proposes a hybrid location system capable of using different technologies to provide internal and external localization of robots or other mobile devices. For external localization, GPS technology is explored. For internal localization, the signals trilateration of Beacons Bluetooth Chiip 01 from Animalltag and ESP8266 Wifi modules is applied. Furthermore, a cloud system was developed to process, store and visualize data from the hybrid location system.

KEYWORDS: GPS, hybrid location system, Bluetooth Beacons, ESP8266, cloud system.

1 | INTRODUÇÃO

1.1 Sistemas de localização para ambientes externos

Devido ao seu baixo custo e facilidade de uso, o GPS é um sistema de localização amplamente presente em dispositivos móveis como tablets, telefones celulares, robôs, veículos aéreos não tripulados, e outros tipos de veículos. Contudo, mesmo em ambientes externos o GPS pode não funcionar adequadamente em certas situações e informações adicionais para localização podem ser necessárias. Um exemplo do funcionamento inadequado do GPS em ambientes externos pode ser encontrado em um sistema robótico que emprega um receptor GPS integrado com sensores inerciais para navegação na floresta amazônica (STEIN; REIS, 2007). A alta densidade da vegetação da floresta atenua os sinais enviados pelos satélites, podendo até impedir que o receptor GPS colete tais sinais. Sendo assim, o sistema inercial foi integrado através de uma técnica denominada filtro de Kalman, a fim de se solucionar os problemas do sistema de localização (STEIN; REIS, 2007).

Ademais, uma limitação do sistema GPS está relacionada com a sua precisão de posicionamento, cujo erro chega a ser da ordem de metros. Assim, surgiu o GPS RTK (em português: Posicionamento Cinemático em Tempo Real), que é uma técnica que utiliza uma estação de referência que fornece correções em tempo real para estações móveis, fazendo com que a precisão final obtida seja da ordem de centímetros (FENG et al., 2008). Esta técnica está, aos poucos, ocupando espaço em diversas áreas como em: Obras Viárias, Cadastros, Mineração, Agricultura de Precisão, Topografia, Loteamento, Locação (Construção Civil), Batimetria e Prospecção de Petróleo (RODRIGUES, 2014).

Para a robótica, o emprego da técnica de RTK pode ser observado em uma pesquisa desenvolvida por Mirisola et al. (2011), em que foi validado, experimentalmente, um veículo robótico específico para ambientes externos. Tal validação experimental foi realizada por meio do método de seguimento de trajetória. No veículo robótico em questão, foram

utilizados um GPS RTK Trimble e uma estação AgGPS-332 para aumentar a precisão de localização do robô (MIRISOLA et al., 2011).

1.2 Sistemas de localização para ambientes internos

Em geral, a localização por GPS não é adequada para ambientes internos, já que o sinal proveniente dos satélites pode ser atenuado ou extinguido até atingir o receptor em um local fechado. Sabendo-se disso, os Beacons Bluetooth estão preenchendo esta lacuna em sistemas de localização para ambientes internos (MAYCOTTE, 2015). Beacons são dispositivos de dimensões reduzidas que transmitem sinais periódicos de rádio que são captados por receptores/módulos, com tecnologia Bluetooth, por exemplo, que estão próximos a tais dispositivos (MAYCOTTE, 2015).

Como o alcance do sinal Bluetooth emitido pelos Beacons é limitado, e uma vez que o preço destes dispositivos é relativamente baixo, a disposição de vários Beacons no interior de certos locais pode ser utilizada como uma ferramenta para a localização indoor de aparelhos móveis. Chawathe (2008) desenvolveu um sistema que emprega o sinal Bluetooth de Beacons, juntamente com o método cell-based, para a determinação da localização de aparelhos móveis. Através do método cell-based, é possível encontrar a interseção dos alcances limitados dos sinais de Beacons visíveis, sem incluir medidas de distância ou de ângulo, a fim de se determinar a localização do aparelho móvel em ambientes internos (CHAWATHE, 2008).

Uma desvantagem na aplicação de sinais Bluetooth para a localização indoor de aparelhos móveis está relacionada a latência, que pode chegar a aproximadamente 20 segundos na fase de descoberta do protocolo (fase de detecção dos Beacons) (CHAWATHE, 2008). Deste modo, em uma outra pesquisa, Chawathe (2009) introduz o método de localização probe-based, cujo conceito está vinculado ao uso de um mapa de Beacons e do último local conhecido para determinar possíveis Beacons visíveis na vizinhança do local atual e, a partir disso, determinar sua visibilidade por meio de sondas diretas (CHAWATHE, 2009).

Em um artigo publicado por Raghavan et al. (2010), foi descrito um método de trilateração de sinais Bluetooth para a determinação da localização de um robô móvel em ambientes internos. Tal método consiste em determinar a posição de um objeto por meio de medidas simultâneas de distância de 3 locais conhecidos. Além disso, este método de trilateração visa determinar uma solução que minimiza o erro relativo total e não o erro absoluto (RAGHAVAN et al., 2010).

Além da tecnologia Bluetooth, a tecnologia de rede sem fio Wifi pode ser utilizada como base para localização de objetos em ambientes internos (BISWAS; VELOSO, 2010). Seguindo a mesma linha de pesquisa para localização de robôs em ambientes internos, Biswas e Veloso (2010) descreveram um sistema de localização e navegação para um robô

autônomo. Como foi relatado nesta pesquisa, tal robô é capaz de operar em ambientes internos através de um mapa de assinaturas Wifi com restrições geométricas, de um modelo de percepção do ambiente por meio da amostragem da força do sinal Wifi e de dados de odometria (BISWAS; VELOSO, 2010).

Uma outra tecnologia que vem ganhando interesse da comunidade científica para sistemas de localização em geral é a de identificação por radiofrequência (RFID). Essa tecnologia é uma forma de comunicação sem fio que utiliza ondas de rádio para identificar e rastrear objetos com tags (etiquetas de baixo custo capazes de armazenar e trocar informações), por meio de leitores (VIOLINO, 2005). Em um trabalho realizado por Azzouzi et al. (2011), foi determinada uma nova abordagem para medida de ângulo de chegada para localizar tags de RFID UHF. Nesta pesquisa, obteve-se um erro de posição de 0,21 metros em uma área de teste de 3x3 metros, considerando-se 25 pontos de teste (AZZOUZI et al., 2011).

1.3 Sistemas de localização híbridos

Sistemas de localização híbridos são aqueles que integram múltiplas tecnologias para a localização de objetos em ambientes internos e externos. A utilização de múltiplas tecnologias de localização, simultaneamente, pode ser uma técnica benéfica aos sistemas de localização internos e externos, uma vez que é possível aumentar a disponibilidade e a exatidão de tais sistemas (RODRIGUES, 2011). No entanto, há a necessidade de se estudar as melhores formas de unir as informações de vários tipos de tecnologia de localização e evitar que esta união cause prejuízos para a integridade da posição encontrada (RODRIGUES, 2011).

Em um artigo publicado por Zhu et al. (2012), foi descrito um sistema de localização para ambientes internos utilizando Wifi e Beacons Bluetooth. O sistema em questão opera da seguinte maneira: depois de solicitar a entrada ao usuário, o dispositivo móvel envia assinaturas Wifi para um servidor de mapas e também permite que os Beacons compartilhem suas posições e informações de assinaturas para preencher o mapa. Deste modo, os próximos dispositivos móveis que entrarem no local podem visualizar os Beacons e receber informações de localização no cômodo sem a necessidade de solicitações ao usuário (ZHU et al., 2012).

DiGiampaolo e Francesco (2011) desenvolveram um sistema que combina dados de odometria de um robô e leituras RFID para criar um sistema de localização global. Nesta pesquisa, tags de RFID de alta frequência (UHF) foram dispostas no teto de ambientes para serem detectadas por um robô móvel, até que este se encontrasse abaixo das tags. Tal método apresentou um erro de posição de 0,1 metros (DIGIAMPAOLO; FRANCESCO, 2011).

Existem ainda diversas outras técnicas de localização e uma vasta literatura sobre sistemas de localização usando visão de máquina e técnicas chamadas de localização e mapeamento simultâneo (SLAM). O foco deste trabalho está na integração de tecnologias de radiofrequência, como Wifi e Bluetooth, e a tecnologia GPS para implementar um sistema híbrido de localização que funcione tanto em ambiente interno quanto externo.

2 | MATERIAL E MÉTODOS

2.1 Ambiente de desenvolvimento

Inicialmente, propôs-se para este projeto, o estudo teórico e experimental de diversas tecnologias que poderiam ser aplicadas para localização de robôs e dispositivos móveis, e após este estudo, buscou-se desenvolver um sistema híbrido de localização, que integra GPS (localização *outdoor*) e outras tecnologias para localização *indoor* simultaneamente.

Para o desenvolvimento deste trabalho, foram utilizados os seguintes dispositivos e equipamentos: 3 *Animalltag Chiip 01 Beacons Bluetooth*, 3 módulos *Wifi NodeMCU (ESP8266)*, um *smartphone* com sistema operacional *Android 7.0* e que suporta *Bluetooth 4.0* e um PC com sistema operacional *Windows 10*. A Fig. 1 ilustra a estrutura completa do sistema de localização híbrido, juntamente com o sistema em nuvem.

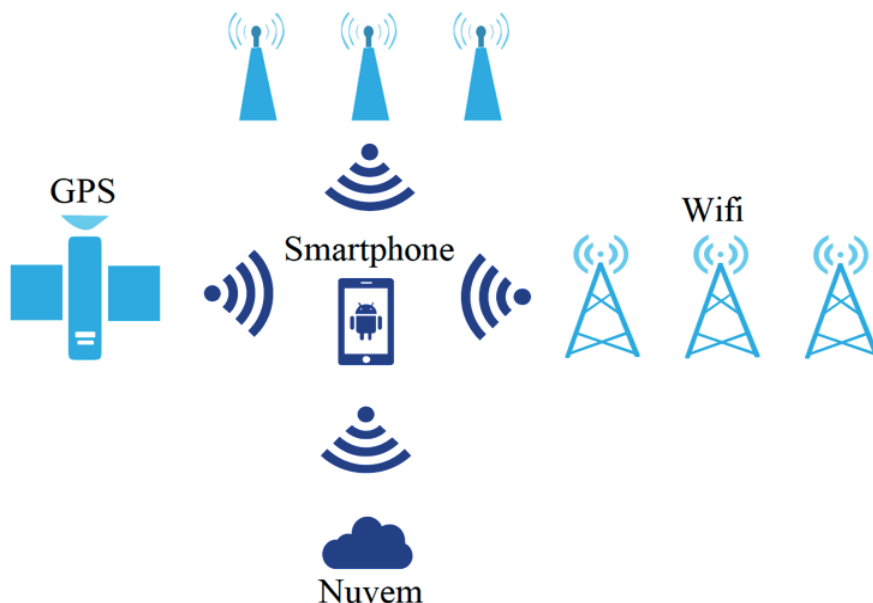


Figura 1: Esquemático que ilustra a composição do sistema de localização híbrido, juntamente com o sistema em nuvem

Fonte: Elaborada pelo autor

Na estrutura da Fig. 1, é possível notar que o sistema opera a partir da aquisição dos sinais *Bluetooth* (localização *indoor*), *Wifi* (localização *indoor*) e GPS, simultaneamente, por meio de um aplicativo desenvolvido para *smartphone*, o qual transmite estes dados para um *Web Service*, programado em linguagem PHP, na mesma frequência do sinal dos *Beacons Bluetooth* (1 Hz). Neste webservice PHP, são executados todos os cálculos de localização, assim qualquer atualização no algoritmo e técnica de localização, beneficia todos usuários, já que ele funciona em nuvem. A partir do tratamento (no servidor) dos dados coletados com o *smartphone*, estes são armazenados em um banco de dados em nuvem, o qual foi implementado utilizando um banco de dados MySQL. Após a etapa de armazenamento, os dados são plotados em um mapa e em um gráfico (em tempo real) disponíveis via web para usuários e outros sistemas.

2.2 Sistema de localização híbrido (HLS)

O desenvolvimento do aplicativo denominado de HLS (*Hybrid Location System*) para a aquisição dos sinais dos dispositivos *Wifi*, *Bluetooth* e GPS, para a transmissão de tais sinais ao *Web Service* e para a visualização das posições do *smartphone* foi feito em linguagem Java na plataforma Android Studio, uma vez que *smartphones* e outros dispositivos móveis que utilizam o sistema operacional Android, atualmente, são amplamente utilizados.

Através de tal aplicativo, são determinadas as distâncias em metros entre o dispositivo móvel e os 3 *Beacons*, com o auxílio de um algoritmo pré-desenvolvido da biblioteca disponibilizada por Young (2018). Este algoritmo foi calibrado por meio de testes em laboratório. Os *Beacons* são diferenciados de acordo com seu identificador único.

Além disso, por meio do HLS, são coletados os dados de frequência e RSSI (*Received Signal Strength Indicator* – indica a intensidade do sinal recebido) dos 3 módulos *Wifi*, os quais foram programados para funcionar em modo AP (ponto de acesso *Wifi*), de forma que fosse possível identificar uma rede específica para cada módulo e diferenciar seus valores de RSSI. Os valores de latitude e longitude do *smartphone*, quando disponíveis, também são coletados pelo HLS via receptor de GPS.

2.3 Sistema em nuvem

A partir da aquisição de todos os dados de localização interna e externa, o HLS envia os dados coletados a um servidor (a uma frequência de 1 Hz), onde são realizados os cálculos de distância entre os módulos *Wifi* e o *smartphone* através da equação de transmissão em espaço livre de Friis (FRIIS, 1946) e dos valores de frequência (em MHz) e RSSI (em dB) de tais módulos. Esta equação de transmissão em espaço livre foi calibrada por meio da Tabela II do artigo escrito por Barai, Biswas e Sau (2017).

A equação de espaço livre de Friis determina a relação de perdas entre a potência recebida por uma antena e a potência transmitida por outra antena, considerando um espaço livre entre elas (BREINBJERG; KASLIS, 2017). No entanto, esta equação pode ser rearranjada para que se possa determinar a distância entre a antena transmissora e a receptora. A equação de transmissão em espaço livre de Friis, para determinar a distância entre antenas, pode ser encontrada da seguinte forma (BREINBJERG; KASLIS, 2017; POOLE, 2018):

$$\frac{P_r}{P_t} = A_t A_r \left(\frac{\lambda}{4\pi d} \right)^2 \quad (1)$$

sendo P_r a potência recebida [W], P_t a potência transmitida [W], A_t o ganho da antena transmissora, A_r o ganho da antena receptora, d a distância da antena transmissora à receptora [m] e λ o comprimento de onda do sinal transmitido [m]. Seja P_L a perda em espaço livre definida por:

$$P_L = \left(\frac{4\pi d}{\lambda} \right)^2 \quad (2)$$

$$P_L(dB) = 10 \log \left(\frac{4\pi d f}{c} \right)^2 \quad (3)$$

sendo c a velocidade da luz no vácuo [Mm/s] e f a frequência do sinal transmitido [MHz]. Assim, a Eq. (3) pode ser rearranjada da seguinte forma:

$$P_L(dB) = 20 \log \left(\frac{4\pi d f}{c} \right) \quad (4)$$

$$P_L(dB) = 20 \log d + 20 \log f - 27,55 \quad (5)$$

$$d = 10^\kappa [m] \quad (6)$$

$$\kappa = \frac{P_L}{20} - \log f + 1,3775 \quad (7)$$

De (1) para (2), assumiu-se que $A_t = 1$; $A_r = 1$ e $P_t/P_r = P_L$. Além disso, de (2) para (3), assumiu-se que $\lambda = c/f$ [m], e, de (4) para (5), assumiu-se que $\pi = 3,1415$ e que $c = 299,7925$ [Mm/s].

No *Web Service*, também são calculadas as trilaterações em duas dimensões (x e y) dos sinais dos *Beacons Bluetooth* e dos sinais dos módulos *Wifi* a fim de se determinar a posição *indoor* do dispositivo móvel (x,y). A Fig. 2 ilustra o funcionamento do sistema de trilateração em duas dimensões aplicado neste projeto.

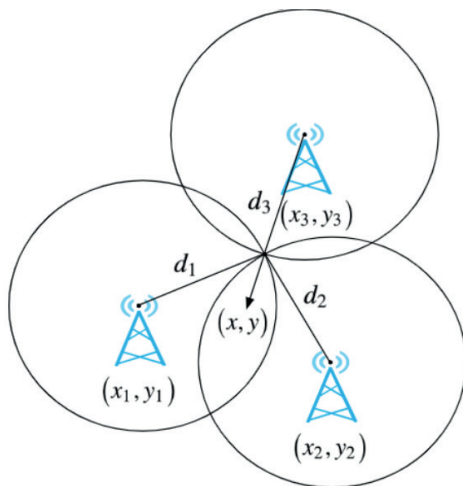


Figura 2: Sistema de trilateração em duas dimensões de sinais de radiofrequência

Fonte: Elaborada pelo autor

Os cálculos de trilateração são realizados considerando-se as três coordenadas x e y das posições fixas dos 3 dispositivos transmissores de sinais de radiofrequência ((x_1, y_1) , (x_2, y_2) e (x_3, y_3)), além das distâncias de cada um dos 3 dispositivos ao *smartphone* (d_1 , d_2 e d_3). A posição (x, y) é determinada a partir da interseção dos 3 sinais dos dispositivos transmissores. Tendo em vista a Fig. 2, pode-se encontrar as equações de trilateração da seguinte forma (WEDDING, 2013):

$$(x - x_1)^2 + (y - y_1)^2 = d_1^2 \quad (8)$$

$$(x - x_2)^2 + (y - y_2)^2 = d_2^2 \quad (9)$$

$$(x - x_3)^2 + (y - y_3)^2 = d_3^2 \quad (10)$$

$$x^2 - 2x_1 + x_1^2 + y^2 - 2y_1 + y_1^2 = d_1^2 \quad (11)$$

$$x^2 - 2x_2 + x_2^2 + y^2 - 2y_2 + y_2^2 = d_2^2 \quad (12)$$

$$x^2 - 2x_3 + x_3^2 + y^2 - 2y_3 + y_3^2 = d_3^2 \quad (13)$$

$$\alpha x + \beta y = \gamma \quad (14)$$

$$\delta x + \varepsilon y = \zeta \quad (15)$$

$$\alpha = -2x_1 + 2x_2 \quad (16)$$

$$\beta = -2y_1 + 2y_2 \quad (17)$$

$$\gamma = d_1^2 - d_2^2 - x_1^2 + x_2^2 - y_1^2 + y_2^2 \quad (18)$$

$$\delta = -2x_2 + 2x_3 \quad (19)$$

$$\varepsilon = -2y_2 + 2y_3 \quad (20)$$

$$\zeta = d_2^2 - d_3^2 - x_2^2 + x_3^2 - y_2^2 + y_3^2 \quad (21)$$

$$x = \frac{\beta\zeta - \gamma\varepsilon}{\delta\beta - \alpha\varepsilon} \quad (22)$$

$$y = \frac{\gamma\delta - \alpha\zeta}{\gamma\beta - \alpha\varepsilon} \quad (23)$$

A fim de se determinar os valores de α , β , γ , δ , ε e ζ , efetuou-se as seguintes manipulações matemáticas com as equações em questão: (11) – (12), e (12) – (13).

Após o tratamento dos dados, no servidor, para a determinação das posições interna e externa do dispositivo móvel, onde também foram implementadas as calibrações da equação de transmissão de Friis e do valor da distância dos *Beacons Bluetooth*, tais dados tratados são enviados para um banco de dados em nuvem. Os dados de posição do dispositivo móvel, armazenados no banco de dados, são plotados em um mapa que pode ser visualizado por meio do aplicativo HLS (ou via *web*), o qual acessa o servidor do mapa em questão e mostra a posição atual do *smartphone* (Fig. 3).

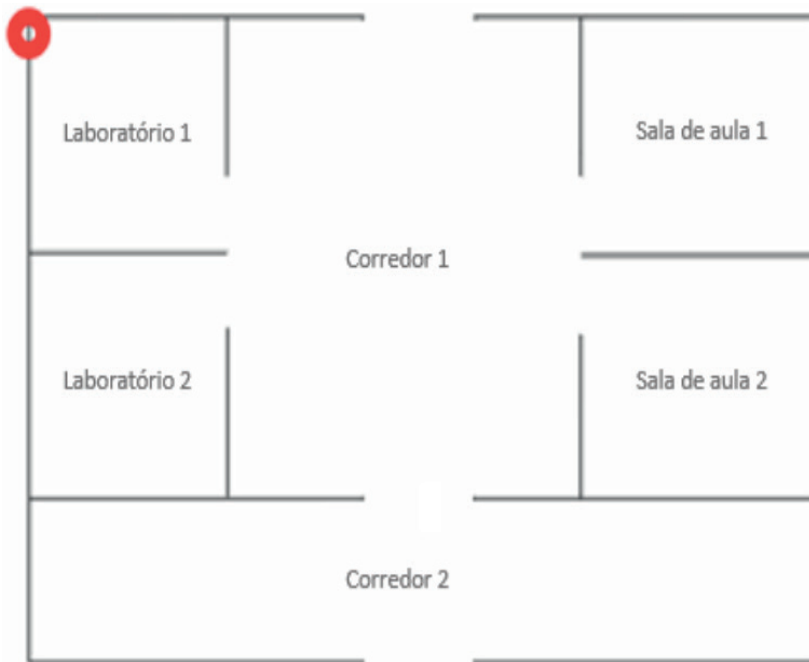


Figura 3: Mapa do Web Service apresentado na tela do aplicativo HLS que mostra a posição do smartphone

Fonte: Elaborada pelo autor

Além disso, os valores de latitude e longitude do dispositivo móvel, armazenados no banco de dados, são plotados em um gráfico em tempo real que apresenta a posição externa de tal dispositivo em função do tempo e que atualiza os dados plotados a cada 5 pontos armazenados de latitude e longitude (MENDES, 2018a).

Finalmente, como mencionado, os tratamentos e armazenamentos de dados foram realizados em nuvem neste projeto, pois o sistema se torna flexível e de fácil manutenção, atualização e otimização, beneficiando todos os usuários. A Fig. 4 ilustra as funções que são executadas em nuvem.

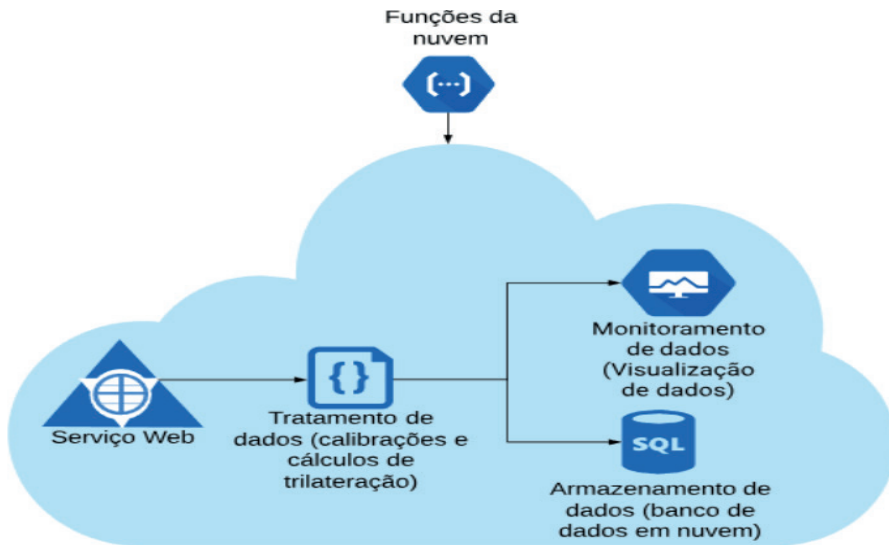


Figura 4: Esquemático da estrutura do sistema em nuvem

Fonte: Elaborada pelo autor

2.4 Calibrações

A calibração da equação de transmissão Friis (Eqs. (6) e (7)) foi realizada com o auxílio da Tabela I, onde os valores de RSSI com relação a distâncias físicas para módulos *NodeMCU* foram retirados da Tabela II do artigo escrito por Barai, Biswas e Sau (2017). Os valores de distância da coluna “Equação de Distância de Friis [m]” da Tabela I foram obtidos aplicando-se os valores de RSSI da coluna “RSSI [dB]” desta mesma tabela nas Eqs. (6) e (7) (para valores de frequência fixos dos módulos *Wifi NodeMCU*).

Distância Esperada [m]	RSSI [dB]	Equação de Distância de Friis [m]
0,3	-55	5,561
0,4	-59	8,813
0,5	-60	9,888
0,6	-62	12,449
0,7	-66	19,730
0,8	-68	24,838
0,9	-69	27,869
1	-71	35,085
2	-74	49,559

3	-79	88,130
4	-81	110,950
5	-83	139,680
6	-84	156,720
7	-85	175,840
8	-86	197,300
9	-87	221,370
10	-88	248,380

TABELA I

CALIBRAÇÃO DA EQUAÇÃO DE ESPAÇO LIVRE DE FRIIS (BARAI; BISWAS; SAU, 2017)

Na Tabela I, os valores de distância obtidos pela equação de transmissão em espaço livre de Friis foram determinados por meio dos valores de RSSI de acordo com as devidas distâncias físicas. O gráfico da Fig. 5 relaciona os valores das distâncias esperadas com os valores determinados pela equação de Friis. Nos eixos x e y, foram plotados os valores da coluna “Equação de Distância de Friis [m]” da Tabela I e da coluna “Distância Esperada [m]” desta mesma tabela respectivamente. De acordo com o gráfico da Fig. 5, pode-se concluir que a relação mencionada é linear ($R^2 = 0,9948$).

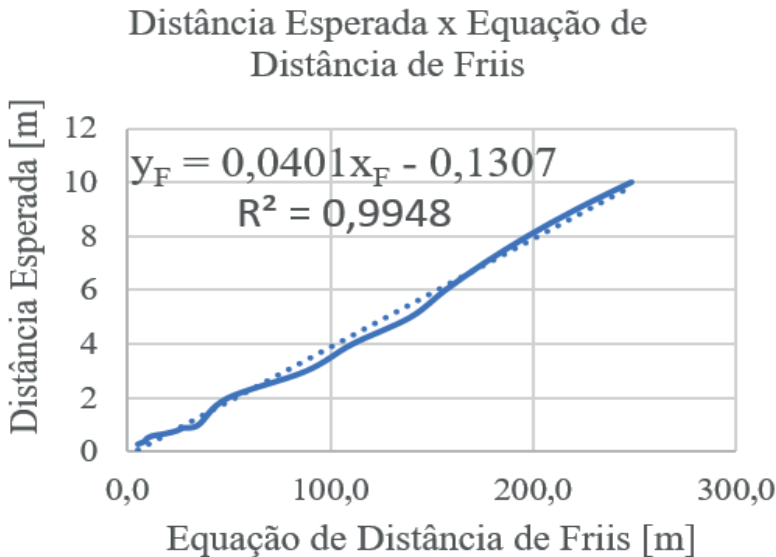


Figura 5: Gráfico utilizado para estabelecer a relação para calibração da equação de transmissão em espaço livre de Friis

Fonte: Elaborada pelo autor

Por meio da linha de tendência plotada no Gráfico da Fig. 5, obteve-se a Eq. (24), a qual foi implementada no servidor para calibrar a equação de Friis.

$$y_F = 0,0401x_F - 0,1307 \quad (24)$$

O teste de calibração do algoritmo pré-desenvolvido da biblioteca disponibilizada por Young (2018) foi realizado de acordo com o plano de testes da Fig. 6.

Operador (Raul de Queiroz Mendes)	1														
Tensão de alimentação do Beacon (pilha de Lítio Sony CR2016 3V)	1														
Smartphone	1														
Local do Beacon (coordenada: (0, 0))	1														
Amostra do Beacon	1					2					3				
Local do Smartphone (coordenadas: (0,5, 0);(1, 0);(1,5, 0);(2, 0);(2,5, 0))	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
Y (Distância do emissor ao Smartphone)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

Figura 6: Plano de testes para a calibração do algoritmo pré-desenvolvido da biblioteca de Beacons (YOUNG, 2018)

Fonte: Elaborada pelo autor

No plano da Fig. 6, é possível verificar que foram selecionados 5 valores de distâncias esperadas (0,5; 1; 1,5; 2; 2,5) entre os *Beacons* e o dispositivo móvel. Os valores de distância entre os *Beacons* e o *smartphone*, obtidos pelo algoritmo em questão, foram armazenados no banco de dados e posteriormente agrupados na Tabela II.

Distância Esperada [m]	Distância do Beacon 1 (algoritmo) [m]	Distância do Beacon 2 (algoritmo) [m]	Distância do Beacon 3 (algoritmo) [m]
0,5	3,050	3,177	2,963
1	6,019	5,376	5,192
1,5	7,548	6,850	6,556
2	8,735	7,702	7,481
2,5	9,298	8,468	8,457

TABELA II

CALIBRAÇÃO DE DISTÂNCIA PARA OS BEACONS 1, 2 E 3

A partir da Tabela II, foram plotados gráficos da Fig. 7, da Fig. 8 e da Fig. 9, relacionando as distâncias esperadas com as distâncias obtidas pelo algoritmo da biblioteca de *Beacons* (YOUNG, 2018) para os *Beacons* 1, 2 e 3 respectivamente.

Nos gráficos da Fig. 7, da Fig. 8 e da Fig. 9 foram plotadas linhas de tendência a fim de se determinar uma expressão matemática que correlaciona os valores de distância teóricos com os valores de distância experimentais entre os *Beacons* e o dispositivo móvel. Além disso, visou-se gerar um termo de correlação R^2 para cada gráfico.

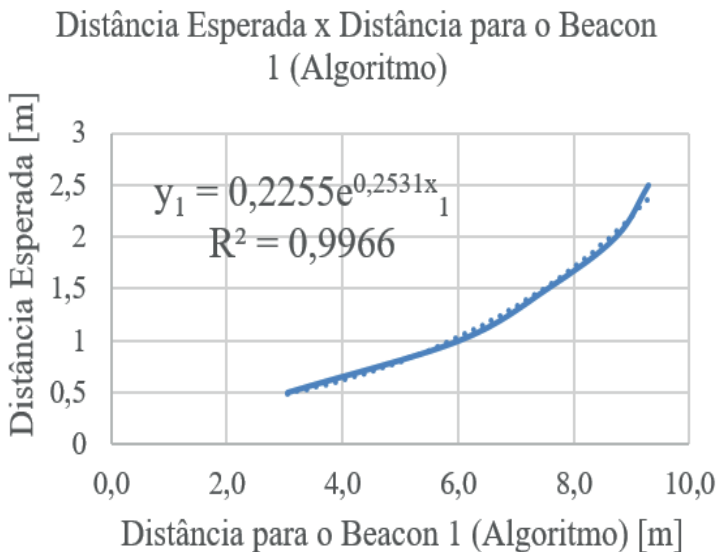


Figura 7: Gráfico que relaciona as distâncias esperadas com as distâncias obtidas para o Beacon 1 por meio do algoritmo

Fonte: Elaborada pelo autor

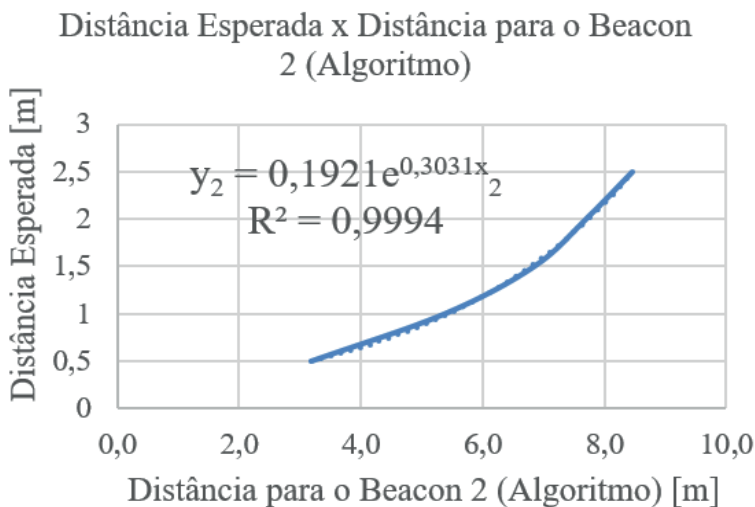


Figura 8: Gráfico que relaciona as distâncias esperadas com as distâncias obtidas para o Beacon 2 por meio do algoritmo

Fonte: Elaborada pelo autor

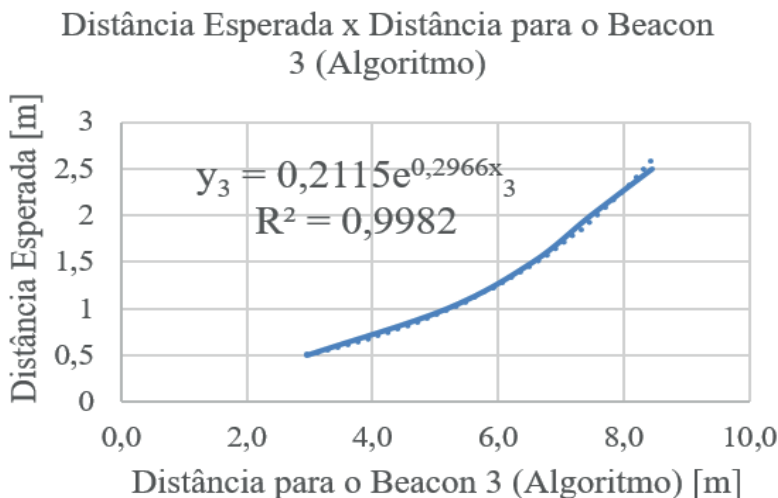


Figura 9: Gráfico que relaciona as distâncias esperadas com as distâncias obtidas para o Beacon 3 por meio do algoritmo

Fonte: Elaborada pelo autor

Com os dados plotados da Tabela II, foram geradas as Eqs. (25), (26) e (27), respectivamente aos Beacons 1, 2 e 3.

$$y_1 = 0,2255e^{0,2531x_1} \quad (25)$$

$$y_2 = 0,1921e^{0,3031x_2} \quad (26)$$

$$y_3 = 0,2115e^{0,2966x_3} \quad (27)$$

Tendo em vista as Eqs. (25), (26) e (27) e os valores de correlação R^2 , tem-se que a relação entre as distâncias esperadas e as distâncias obtidas pelo algoritmo é exponencial. Tais equações também foram implementadas no *Web Service* para calibrar o algoritmo.

3 I RESULTADOS E DISCUSSÃO

3.1 Testes

Os testes realizados para verificar a funcionalidade do sistema de localização híbrido foram executados de acordo com o plano da Fig. 10.

Operador (Raul de Queiroz Mendes)	1									
Tensão de alimentação dos Beacons/módulos Wifi (pilha CR2016 3V e USB 5V respectivamente)	1									
Smartphone	1									
Local fixo dos Beacons/módulos Wifi (coordenadas: (2, 0);(2, 3);(0,5, 1))	1									
Local do Smartphone (coordenadas: (1,3, 1,1);(1,9, 1,4);(2, 2);(2,3, 1,5);(1,5, 2))	1	2	3	4	5	6	7	8	9	10
Y (Dados de localização dos Beacons (x,y))	1	2	3	4	5	6	7	8	9	10
Y (Dados de localização dos módulos Wifi (x,y))	1	2	3	4	5	6	7	8	9	10
Y (Dados de localização do GPS (latitude,longitude))	1	2	3	4	5	6	7	8	9	10

Figura 10: Plano de testes para verificar a funcionalidade do sistema de localização híbrido

Fonte: Elaborada pelo autor

De acordo com o plano de testes da Fig. 10, os *Beacons* e os módulos *Wifi* foram posicionados em locais fixos ((2, 0);(2, 3);(0,5, 1)). A partir disso, variou-se a posição do *smartphone* em 5 pontos e, para cada rodada, coletou-se os dados de localização.

A Tabela III apresenta as posições do *smartphone* determinadas pelo método de trilateração dos sinais dos *Beacons* e dos sinais dos módulos *Wifi* NodeMCU (ESP8266), em relação as posições esperadas.

Coordenada	Posição Esperada [m]	Posição (Beacon) [m]	Posição (ESP8266) [m]	Erro de Posição [%] (Beacon)	Erro de Posição [%] (ESP8266)
x	1,3	1,386	1,214	6,636	6,653
y	1,1	1,467	0,928	33,341	15,642
x	1,9	1,914	1,961	0,741	3,216
y	1,4	1,600	1,410	14,262	0,743
x	2	1,830	2,076	8,514	3,775
y	2	1,648	1,738	17,615	13,085
x	2,3	2,378	2,259	3,373	1,787
y	1,5	1,734	1,452	15,620	3,200
x	1,5	1,484	1,599	1,036	6,600
y	2	1,904	1,738	4,787	13,085

TABELA III

POSIÇÕES DO SMARTPHONE DETERMINADAS PELOS MÉTODOS DE LOCALIZAÇÃO INDOOR EM RELAÇÃO AS POSIÇÕES ESPERADAS

A Tabela IV mostra as coordenadas geográficas do *smartphone* obtidas por meio do aplicativo HLS e as coordenadas geográficas verificadas pelo *Google Maps* (2018) (ambas com relação as posições locais do dispositivo móvel utilizadas em teste).

Posição local do smartphone (x, y) [m]	Coordenada Geográfica Obtida (Latitude, Longitude) [°]	Coordenada Geográfica Esperada (Latitude, Longitude) [°]	Erro de Posição GPS [m]
(1,3, 1,1)	(-21,980833, -47,878049)	(-21,980714, -47,877901)	21,102563
(1,9, 1,4)	(-21,980832, -47,878048)	(-21,980706, -47,877904)	21,261992
(2, 2)	(-21,980830, -47,878050)	(-21,980703, -47,877912)	20,839963
(2,3, 1,5)	(-21,980830, -47,878051)	(-21,980699, -47,877904)	21,879638
(1,5, 2)	(-21,980828, -47,878053)	(-21,980711, -47,877912)	20,359537

TABELA IV

RELAÇÃO ENTRE AS COORDENAS GEOGRÁFICAS OBTIDAS PELO SMARTPHONE E AS COORDENADAS GEOGRÁFICAS VERIFICADAS PELO GOOGLE MAPS

Através da Eq. (28), pode-se determinar o erro relativo entre as coordenadas da posição do *smartphone* esperadas e aquelas obtidas pelos dois métodos utilizados para determinação de posição interna (BARAI; BISWAS; SAU, 2017).

$$Erro\ Relativo\ [\%] = \frac{\left| \rho_{esperado} - \rho_{obtido} \right|}{\rho_{esperado}} \times 100 \quad (28)$$

De acordo com a Tabela III, os erros máximo e mínimo de coordenada de posição do dispositivo móvel, gerados por meio dos *Beacons*, foram iguais a 33,341% e 0,741% respectivamente, e, aqueles gerados por meio dos módulos *Wifi*, foram iguais a 15,642% e 0,743% respectivamente. Os erros máximos de posição indoor podem ser explicados pela presença de não linearidades no ambiente onde foram executados os testes de validação do sistema em questão, gerando interferências na recepção do sinal pelo dispositivo móvel (apesar dos *Beacons* e dos módulos *Wifi* terem sido calibrados).

Tendo em vista tais erros relativos máximos e mínimos de coordenada de posição interna, conclui-se que o sistema híbrido aplicado para localização em ambientes internos é confiável.

A coluna de erro de posição GPS da Tabela IV relaciona as coordenadas geográficas obtidas em teste e as coordenadas geográficas verificadas pelo *Google Maps* (2018). Tais erros foram calculados a partir das Eqs. (29), (30) e (31) (BENI, 2009).

$$DLA = \left| lat_{esperada} - lat_{obtida} \right| \times 60 \times 1852 \quad (29)$$

sendo DLA a diferença de latitude [m] e lat o valor da latitude [°]. Seja DLO a diferença de longitude em metros e long o valor da longitude [°], têm-se:

$$DLO = \left| long_{esperada} - long_{obtida} \right| \times 60 \times 1852 \quad (30)$$

$$Erro\ (GPS) = \sqrt{(DLA)^2 + (DLO)^2} [m] \quad (31)$$

As constantes 60 e 1852 das Eqs. (29) e (30) foram utilizadas para converter a unidade grau [°] em milha náutica [NM] e a unidade milha náutica [NM] em metros [m] respectivamente. Além disso, o método utilizado nas Eqs. (29), (30) e (31) foi validado comparando-se os dados da coluna “Erro de Posição GPS [m]” da Tabela IV com os valores

de posição GPS em metros obtidos por meio do site disponibilizado por Williams (2003). Neste site, foram inseridos os dados das coordenadas geográficas obtida e esperada da Tabela IV, bem como o modelo da Terra WGS84/NAD83/GRS80 a fim de se determinar os valores de posição GPS em metros.

Por meio da comparação dos valores dos erros de posição GPS em metros da Tabela IV com os dados gerados pelo site disponibilizado por Williams (2003), obteve-se um erro máximo de 4,575% (erro da ordem de centímetros), comprovando que o método utilizado em questão é válido para a aplicação de robótica móvel. Os valores dos erros de posição GPS da Tabela IV são da ordem de dezenas de metros e, devido a isso, conclui-se que o sistema GPS não é confiável para a determinação da localização de dispositivos móveis em ambientes internos, restringindo assim sua utilização para ambientes externos.

Finalmente, a Fig. 11 mostra a tela inicial do aplicativo HLS com um ícone (*START RANGING*) para ser clicado a fim de se iniciar o processo de monitoramento do dispositivo móvel.

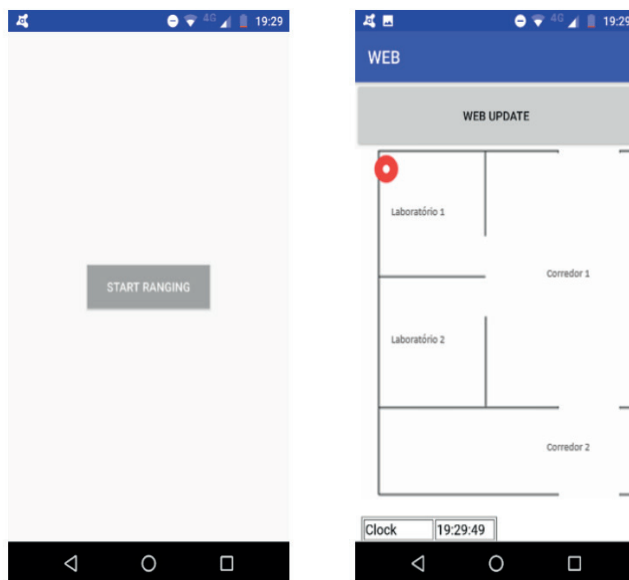


Figura 11. Tela inicial e tela que mostra o mapa para monitoramento do *smartphone*. do aplicativo HLS.

A Fig. 11 também mostra a tela do aplicativo que acessa o servidor em nuvem e monitora a posição *indoor* do dispositivo móvel através de um mapa, enquanto a posição *outdoor* do dispositivo é monitorada por meio de um outro servidor que apresenta um gráfico em tempo real (MENDES, 2018a).

4 | CONCLUSÕES

O sistema de localização híbrido desenvolvido neste projeto apresentou-se como um sistema aplicável, funcional e com uma considerável precisão para a determinação da posição de robôs e dispositivos móveis em ambientes internos e externos, tendo em vista os erros das coordenadas de posição obtidos por meio dos testes executados. Além disso, a utilização de sistemas em nuvem para processamento, armazenamento e visualização de dados mostrou-se viável, prática e eficaz, considerando-se todas as manipulações de dados que foram realizadas em tempo real durante o funcionamento do sistema de localização híbrido. O sistema de localização desenvolvido neste projeto pode ser disseminado e aplicado em diversos outros dispositivos móveis e em locais distintos, uma vez que o aplicativo HLS apresenta uma interface simples e fácil de se utilizar, e sua codificação está aberta e publicada no Github (MENDES, 2018b; MENDES, 2018c). Como próximos passos, novas tecnologias de localização *indoor* e *outdoor*, para dispositivos móveis, serão avaliadas, testadas e, possivelmente, incorporadas no sistema vigente a fim de se aumentar sua precisão e sua eficiência. Acrescentando-se a isso, trabalhos já estão sendo realizados para melhorar a precisão dos dispositivos implementados.

REFERÊNCIAS

- AZZOUZI, Salah et al. New measurement results for the localization of uhf rfid transponders using an angle of arrival (aoa) approach. In: **2011 IEEE International Conference on RFID**. IEEE, 2011. p. 91-97.
- BARAI, Suvankar; BISWAS, Debajyoti; SAU, Buddhadeb. Estimate distance measurement using NodeMCU ESP8266 based on RSSI technique. In: **2017 IEEE Conference on Antenna Measurements & Applications (CAMA)**. IEEE, 2017. p. 170-173.
- BENI, E. *Calculando distâncias e direções utilizando Coordenadas Geográficas*. Piloto Policial, set. 2009. Disponível em: <https://www.pilotopolicial.com.br/calculando-distancias-e-direcoes-utilizando-coordenadas-geograficas/>. Acesso em: 9 mai. 2018.
- BISWAS, Joydeep; VELOSO, Manuela. Wifi localization and navigation for autonomous indoor mobile robots. In: **2010 IEEE international conference on robotics and automation**. IEEE, 2010. p. 4379-4384.
- BREINBJERG, Olav; KASLIS, Kyriakos. On the accuracy of Friis' transmission formula at short range. In: **2017 XXXIInd General Assembly and Scientific Symposium of the International Union of Radio Science (URSI GASS)**. IEEE, 2017. p. 1-2.
- CHAWATHE, Sudarshan S. Beacon placement for indoor localization using bluetooth. In: **2008 11th International IEEE Conference on Intelligent Transportation Systems**. IEEE, 2008. p. 980-985.
- CHAWATHE, Sudarshan S. Low-latency indoor localization using bluetooth beacons. In: **2009 12th International IEEE Conference on Intelligent Transportation Systems**. IEEE, 2009. p. 1-7.

DIGIAMPAOLO, Emidio; MARTINELLI, Francesco. A passive UHF-RFID system for the localization of an indoor autonomous vehicle. **IEEE Transactions on Industrial Electronics**, v. 59, n. 10, p. 3961-3970, 2011.

FENG, Yanming et al. GPS RTK performance characteristics and analysis. **Positioning**, v. 1, n. 13, 2008.

FRIIS, Harald T. A note on a simple transmission formula. **Proceedings of the IRE**, v. 34, n. 5, p. 254-256, 1946.

GOOGLE MAPS. **UFSCar - Universidade Federal de São Carlos**, São Carlos - SP. 2018. Disponível em: <https://www.google.com.br/maps/place/UFSCar++Universidade+Federal+de+S%C3%A3o+Carlos/@-21.9841396,-47.8824341,17z/data=!3m1!4m5!3m4!1s0x94b870d88899b96c5:0x26db4c677a5af1d4!8m2!3d-21.9841446!4d-47.8802454>. Acesso em: 9 mai. 2018.

MAYCOTTE, H. O. **Beacon Technology: The Where, What, Who, How and Why**. Forbes, set. 2015. Disponível em: <https://www.forbes.com/sites/homaycotte/2015/09/01/Beacon-technology-the-what-who-how-why-and-where/#289bbd881aaf>. Acesso em: 18 jan. 2018.

MENDES, R. Q. **Coordenadas GPS x Tempo**. Robótica UFSCar, jun. 2018. Disponível em: <http://www.robotica.ufscar.br/~raul/mapa/html/graph.html>. Acesso em: 2 dez. 2018.

MENDES, R. Q. **RauldeQueirozMendes/Cloud-System**. Github, jun. 2018. Disponível em: <https://github.com/RauldeQueirozMendes/Cloud-System>. Acesso em: 2 dez. 2018.

MENDES, R. Q. **RauldeQueirozMendes/Web-Service---Beacon---Wifi---GPS**. Github, jun. 2018. Disponível em: <https://github.com/RauldeQueirozMendes/Web-Service---Beacon---Wifi---GPS>. Acesso em: 29 jun. 2018.

MIRISOLA, L. et al. Validação experimental de um veículo robótico terrestre para ambientes externos. **Anais do X Simpósio Brasileiro de Automação Inteligente**, p. 1322-1327, 2011.

POOLE, I. **Free Space Path Loss: Details, Formula, Calculator**. radio-electronics.com. Adrio Communications Ltd. Disponível em: <https://www.electronics-notes.com/articles/antennas-propagation/propagation-overview/free-space-path-loss.php>. Acesso em: 25 abr. 2018.

RAGHAVAN, Aswin N. et al. Accurate mobile robot localization in indoor environments using bluetooth. In: **2010 IEEE international conference on robotics and automation**. IEEE, 2010. p. 4391-4396.

RODRIGUES, D. **Como Utilizar o seu GPS/GNSS RTK**. MundoGEO, maio. 2014. Disponível em: <http://mundogeo.com/blog/2014/05/30/artigo-como-utilizar-o-seu-gpsgnss-rtk/>. Acesso em: 17 jan. 2017.

RODRIGUES, Moises Lisboa. Localização em ambientes internos utilizando múltiplas tecnologias sem fio. 2011.

STEIN, Procópio Silveira; REIS, N. R. S. D. Sistema integrado de navegação para o Robô Ambiental Híbrido na Floresta Amazônica. **XIII Simpósio Brasileiro de Sensoriamento Remoto**, p. 21-26, 2007.

VIOLINO, Bob. What is RFID. **RFID Journal**, p. 1-3, 2005.

WEDDING, J. **Find X location using 3 known (X,Y) location using trilateration**. StackExchange MATHEMATICS, ago. 2014. Disponível em: <https://math.stackexchange.com/questions/884807/find-x-location-using-3-known-x-y-location-using-trilateration>. Acesso em: 6 mar. 2018.

WILLIAMS, E. **Great Circle Calculator**. edwilliams.org, jun. 2003. Disponível em: <http://edwilliams.org/gccalc.htm>. Acesso em: 2 dez. 2018.

YOUNG, G. D. **AltBeacon/android-beacon-library-reference**. Github, nov. 2018. Disponível em: <https://github.com/AltBeacon/android-beacon-library-reference>. Acesso em: 2 dez. 2018.

ZHU, Jindan et al. Improving crowd-sourced wi-fi localization systems using bluetooth beacons. In: **2012 9th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON)**. IEEE, 2012. p. 290-298.

SUMARIZAÇÃO AUTOMÁTICA DE ARTIGOS DE NOTÍCIAS EM PORTUGUÊS USANDO PROGRAMAÇÃO LINEAR INTEIRA E REGRESSÃO

Data de aceite: 27/08/2020

Data de submissão: 05/06/2020

Hilário Tomaz Alves de Oliveira

Instituto Federal do Espírito Santo – Campus Serra
Serra - Espírito Santo
<http://lattes.cnpq.br/8980213630090119>

Laerth Bruno de Brito Gomes

Centro Universitário de João Pessoa – Unipê
João Pessoa – Paraíba

RESUMO: Sumarização Automática de Textos (SAT) tem por objetivo desenvolver métodos capazes de gerar resumos contendo as informações mais relevantes a partir de um ou mais documentos. Sistemas de SAT podem auxiliar no processo de identificar informações de interesse de maneira mais eficiente. Neste trabalho, apresentamos um sistema de sumarização multidocumento de artigos de notícias escritos em português do Brasil. A solução proposta inicialmente cria múltiplos resumos candidatos usando uma abordagem baseada em conceitos com programação linear inteira. Posteriormente, um modelo de regressão é aplicado para estimar e selecionar o resumo candidato mais informativo. Experimentos realizados no corpus CSTNews demonstram que o sistema desenvolvido obteve resultados promissores comparados com outros sistemas da literatura, com base nas medidas de avaliação do ROUGE.

PALAVRAS-CHAVE: Sumarização

automática de textos; Programação linear inteira; Regressão.

AUTOMATIC SUMMARIZATION OF NEWS ARTICLES IN PORTUGUESE USING INTEGER LINEAR PROGRAMMING AND REGRESSION

ABSTRACT: Automatic Text Summarization (ATS) aims to develop methods capable of generating summaries containing the most relevant information from one or more documents. ATS systems can assist in the process of identifying information of interest more efficiently. In this paper, we present a multi-document summarization system of news articles written in Brazilian Portuguese. The proposed solution initially creates multiple summary candidates using a concept-based approach with integer linear programming. Afterward, a regression model is applied to estimate and select the most informative candidate summary. Experiments carried out in the CSTNews corpus demonstrate that the developed system obtained encouraging results compared to other systems in the literature, based on the ROUGE evaluation measures.

KEYWORDS: Automatic text summarization; Integer linear programming; Regression.

1 | INTRODUÇÃO

Ao longo das últimas décadas, com o desenvolvimento tecnológico e o advento da era da informação, as pessoas passaram a ter a possibilidade de criar, compartilhar e

acessar informações rapidamente. Com isso, uma quantidade sem precedentes de novas informações é criada diariamente na forma de artigos de notícias, postagens em redes sociais, artigos científicos, entre outros. Essa vasta quantidade de dados também causa uma sobrecarga de informações, já que essa abundância excede a capacidade humana de processamento. Esse cenário tem se tornado um problema para as pessoas.

A Sumarização Automática de Textos (SAT) pode desempenhar um papel fundamental para mitigar os efeitos da sobrecarga de informação (MANI, 2001). O objetivo da SAT é criar sistemas computacionais capazes de gerar um resumo contendo as informações mais relevantes a partir de um ou mais documentos textuais (NENKOVA e MCKEOWN, 2012). Dessa forma, as pessoas podem ler o resumo que é uma representação condensada das fontes de entrada e, posteriormente, decidir se os documentos são relevantes para o seu interesse.

A SAT é uma tarefa bem conhecida pela comunidade de Processamento de Linguagem Natural (PLN) e tem demandado muita pesquisa ao longo dos anos. Em especial, um maior interesse tem sido dado nos últimos anos a tarefa de Sumarização Multidocumento (SMD) que consiste na criação de um único resumo a partir de vários documentos relacionados ao mesmo assunto. Muitos sistemas de SMD têm sido desenvolvidos adotando uma abordagem extrativa que identifica e seleciona o subconjunto de frases mais relevantes dos documentos de entrada para criação do resumo, evitando a inclusão de informações redundantes.

A maioria dos sistemas de SMD extrativos tem focado em estimar um escore de relevância para frases ou n-gramas pertencentes aos documentos a serem sintetizados. Para isso, uma grande variedade de técnicas supervisionadas e não supervisionadas tem sido explorada. Contudo, poucos trabalhos têm investigado a viabilidade de estimar a relevância de um resumo já construído (OLIVEIRA et al., 2017). Dessa forma, seria possível selecionar o resumo mais informativo a partir de um conjunto de resumos candidatos.

Neste trabalho, apresentamos um sistema para sumarização multidocumento de artigos de notícias escritos em português do Brasil. Inicialmente, diversos resumos candidatos são criados usando um método de sumarização baseado em conceitos via Programação Linear Inteira (PLI). Em uma segunda etapa, um modelo de regressão é usado para estimar a cobertura de informações relevantes (informatividade) dos resumos candidatos. Dez atributos baseados em indicadores de relevância, medidas de similaridade e divergência são extraídos a partir da análise do resumo e da coleção de documentos de entrada para construção do modelo. Por fim, o resumo estimado como mais informativo é selecionado.

Para avaliar a efetividade do sistema desenvolvido, experimentos foram executados utilizando o corpus CSTNews (CARDOSO et al. 2011) que é comumente utilizado para avaliar sistemas de sumarização de artigos de notícias escritos em português do Brasil. Como medida de informatividade dos resumos, analisamos as medidas de cobertura,

precisão e medida-F geradas a partir da ferramenta do *Recall-Oriented Understudy for Gisting Evaluation* (ROUGE) (LIN, 2004). Os resultados experimentais demonstram que a estratégia de gerar diversos resumos candidatos e, posteriormente, aplicar um modelo de regressão para estimar a informatividade dos candidatos é viável e resultou na criação de resumos mais informativos.

2 | TRABALHOS RELACIONADOS

Diversos trabalhos investigando o desenvolvimento de sistemas de sumarização multidocumento extrativos podem ser encontrados na literatura, especialmente para documentos escritos em inglês. Um dos primeiros sistemas de SMD para o português do Brasil foi o GistSumm (PARDO, 2005). Nesse sistema, cada sentença dos documentos de entrada recebe um escore de relevância que pode ser gerado com base na frequência de suas palavras ou usando um método baseado no tradicional *Term Frequency - Inverse Document Frequency* (TF-IDF). A frase com escore de relevância mais alto é selecionada como principal. Posteriormente, o sistema seleciona as frases que contêm pelo menos uma palavra em comum com a frase principal e têm uma relevância superior a um determinado limiar, que é a média de todos os escores das frases.

Castro Jorge e Pardo (2010) introduziram uma abordagem para SMD que explora relações semânticas e métodos de seleção de conteúdo com base no modelo da Teoria da Estrutura de Documentos Cruzados, do inglês *Cross-document Structure Theory* (CST). O sistema proposto, chamado *CSTSumm (CSTbased SUMMARizer)*, aborda alguns dos principais problemas da tarefa de sumarização de múltiplos documentos, como redundância. Cardoso e Pardo (2016) desenvolveram uma pesquisa semelhante para criar métodos de sumarização multidocumento explorando modelos de CST e da Teoria da Estrutura Retórica, do inglês *Rhetorical Structure Theory* (RST). Ambos os modelos foram adotados para identificar informações relevantes nos documentos a serem sintetizados.

Sodré e Oliveira (2019) apresentaram um sistema supervisionado para sumarização de múltiplos documentos que combina diversos indicadores de relevância baseados em frequência, posição, centralidade, entre outros. Um modelo de regressão foi desenvolvido para estimar a relevância das frases. Posteriormente, aplicando um método de seleção guloso, as frases mais relevantes são utilizadas para geração dos resumos desde que elas não tenham uma similaridade maior que uma dado limiar com sentenças já selecionadas.

Recentemente, abordagens baseadas em conceito usando Programação Linear Inteira (PLI) têm alcançado um desempenho do estado da arte para tarefas de SMD (OLIVEIRA et al., 2018; GOMES e OLIVEIRA, 2019), especialmente para documentos escritos em inglês. Tais abordagens representam o processo de sumarização como um problema de otimização de máxima cobertura, cujo objetivo é extrair o subconjunto de frases que possuem o maior número de fragmentos textuais relevantes (chamados de

conceitos), respeitando o tamanho máximo do resumo desejado.

Gomes e Oliveira (2019) desenvolveram um sistema de sumarização multidocumento baseado em conceitos usando PLI para textos de notícias em português. O sistema proposto utiliza bigramas como conceitos e aplica métodos baseados na quantidade de documentos (frequência dos documentos) e na posição das frases que mencionam o bigrama para estimar a sua relevância.

Todos os trabalhos supracitados usam um conjunto de parâmetros e técnicas predefinido para mensurar a relevâncias das frases ou n-gramas e, posteriormente, construir o resumo selecionando as sentenças consideradas mais importantes. Conforme apontado por trabalhos recentes (OLIVEIRA et al., 2017), usar um único método de sumarização é uma significativa limitação, já que, em geral, ele não consegue gerar resumos informativos para todos os documentos, mesmo quando eles pertencem ao mesmo domínio. O sistema desenvolvido neste trabalho busca mitigar essa limitação explorando a estratégia de gerar múltiplos resumos candidatos e, em uma segunda etapa, identificar e selecionar o resumo mais informativo utilizando um modelo de regressão.

3 | SISTEMA DE SUMARIZAÇÃO USANDO PLI E REGRESSÃO

O sistema desenvolvido neste trabalho é composto por três etapas principais, conforme apresentado na Figura 1. Dada uma coleção de artigos de notícias, inicialmente os documentos são processados para estruturá-los em um formato adequado. Posteriormente, diversos resumos candidatos são gerados aplicando uma extensão do método de sumarização baseado em conceitos usando PLI desenvolvido por Gomes e Oliveira (2019). Na última etapa, um modelo de regressão é usado para estimar a informatividade de cada resumo candidato, e aquele com maior relevância estimada é selecionado como o resumo mais representativo.

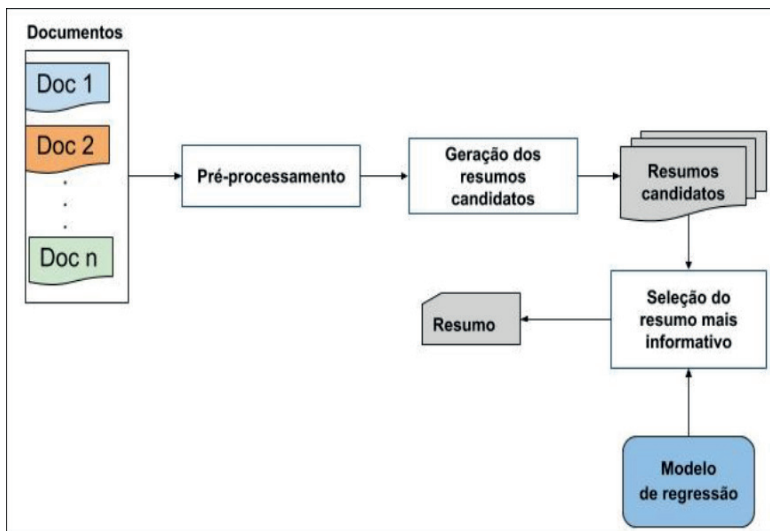


Figura 1. Visão geral das etapas do sistema proposto.

Os documentos de entrada estão escritos em linguagem natural, ou seja, estão em um formato não estruturado. Esse tipo de formato requer estruturação prévia para que eles possam ser usados para a aplicação dos algoritmos de sumarização. Portanto, na primeira etapa é realizado o pré-processamento dos documentos usando técnicas de PLN. As seguintes técnicas foram utilizadas: divisão de frases, tokenização, remoção de *stopwords* e *stemming*. Todas as técnicas foram implementadas usando as ferramentas *spaCy* e *Natural Language Toolkit* (NLTK), ambas na linguagem de programação Python.

Após a realização da etapa de pré-processamento, cada documento de entrada é dividido em uma ou mais frases, com cada frase possuindo uma ou mais palavras, e cada palavra possui associado o seu radical identificado pelo processo de stemming e uma marcação indicando se ela é uma stopwords.

As etapas de geração dos resumos candidatos e da seleção do resumo mais informativo são explicadas nas subseções a seguir.

3.1 GERAÇÃO DOS RESUMOS CANDIDATOS

O objetivo desta etapa é gerar um conjunto de resumos candidatos que possam ser possíveis representações das informações mais relevantes dos documentos de entrada. Para isso, o sistema de sumarização multidocumento baseado em conceitos usando PLI desenvolvido por Gomes e Oliveira (2019) foi estendido. Esse sistema foi escolhido por ter obtido resultados competitivos em comparação com outros trabalhos na literatura, além de permitir a exploração de diferentes configurações visando gerar um conjunto diversificado de resumos candidatos.

A abordagem de sumarização baseada em conceitos usada neste trabalho considera

o processo de sumarização como um problema de máxima cobertura, cujo objetivo é selecionar o subconjunto de frases que maximize a seleção de elementos textuais (conceitos) considerados relevantes, respeitando o tamanho máximo do resumo desejado. Para isso, duas questões que precisam ser definidas é que elemento textual será usado para representar a noção de um conceito e como mensurar a sua relevância. A abordagem desenvolvida é realizada em três subetapas, descritas a seguir:

Extração dos conceitos. Nesta primeira subetapa, n-gramas são extraídos para cada frase dos documentos de entrada, como formas de representação de conceitos para serem usados no processo de sumarização. Com o objetivo de gerar diversos resumos candidatos, adotamos cinco formas de representações separadamente, sendo que cada uma delas é responsável pela geração de um resumo. As formas de representação utilizadas neste trabalho são: unigramas, bigramas, trigramas, quadrigramas e uma combinação de todas as anteriores em uma só representação.

O processo de construção dos n-gramas é feito a partir das palavras de cada sentença, sendo que cada palavra é representada pelo seu radical extraído processo de *stemming*. Como adotado em trabalhos anteriores (OLIVEIRA et al., 2018), conceitos formados somente por *stopwords* são removidos. Ao final desta subetapa, cada frase possui cinco listas de conceitos, sendo uma para cada forma de representação.

Ponderação dos conceitos. Após a extração dos conceitos, é necessário realizar o processo de ponderação visando gerar um escore (peso) que represente a sua relevância. Para isso, utilizamos o método não supervisionado de ponderação proposto por Oliveira et al. (2018) que é computado conforme apresentado na Equação 1. Nesse método, um conceito é considerado relevante se ele for mencionado em muitos dos documentos a serem sumarizados e se ele for mencionado pela primeira vez nas frases iniciais desses documentos.

$$\text{Peso}(c_i) = \text{FreqDoc}(c_i) \times \text{SentPos}(c_i) \quad (1)$$

$$\text{SentPos}(c_i) = \sum_{d_i \in D} 1 - \frac{I_{S_{c_i}}}{|S_{d_i}|} \quad (2)$$

Nessa equação, **Peso (c_i)** é o escore de relevância gerado para o conceito **c_i**; **FreqDoc(c_i)** é o método de frequência dos documentos que retorna o total de documentos **d_i** da coleção de documentos de entrada **D** que mencionam o conceito **c_i** no seu texto; **|S_{d_i}|** é o total de frases do documento **d_i**; **I_{S_{c_i}}** é o índice da primeira frase que contém o conceito **c_i** em cada documento **d_i ∈ D**, com a contagem das frases iniciando com zero.

Ao final desta subetapa, cada conceito presente nas cinco listas de conceitos possui um escore que representa a sua relevância.

Geração do resumo. Esta última subetapa é responsável pela seleção das frases para compor os resumos candidatos. Frases com conteúdo menos do que dez palavras são removidas por serem consideradas muito curtas para serem inseridas no resumo. Após essa filtragem, as frases remanescentes são usadas para construir o modelo de PLI baseado em conceitos, conforme descrito na Equação 3.

$$\text{Maximize } \sum_i w_i \times c_i \quad (3a)$$

$$\text{s.t. } \sum_j l_j \times s_j \leq L \quad (3b)$$

$$s_j \times Occ_{ij} \leq c_i \forall i, j \quad (3c)$$

$$\sum_j s_j \times Occ_{ij} \geq c_i \forall i, j \quad (3d)$$

$$c_i, s_j, Occ_{ij} \in \{0, 1\} \forall i, j \quad (3e)$$

Nessa equação, w_i representa o peso de um conceito c_i computado conforme apresentado na Equação 1. A variável Occ_{ij} indica a ocorrência de um conceito c_i na frase s_j . A variável l_j representa a quantidade de palavras da sentença s_j e L é a quantidade máxima de palavras que o resumo a ser gerado pode ter. As Equações 3c e 3d são restrições que garantem a consistência do modelo, assegurando que, se uma frase for selecionada para a solução do modelo, todos os seus conceitos também devem ser selecionados, e um conceito só é selecionado se estiver presente em pelo menos uma frase selecionada. Um valor binário é atribuído a cada sentença, sendo o valor 1 atribuído a frases selecionadas para compor o resumo e 0, caso contrário. Por fim, as frases são ordenadas usando um índice gerado na etapa de pré-processamento, que indica a posição da frase no documento ao qual ela pertence.

O modelo baseado em conceitos é executado cinco vezes para gerar cinco resumos para cada uma das formas de representação de conceitos adotada neste trabalho. Ao final desta etapa, um conjunto com cinco resumos candidatos é gerado para a coleção de documentos de entrada. Esses candidatos representam diferentes possíveis resumos contendo as informações mais relevantes dos documentos a serem resumidos.

3.2 SELEÇÃO DO RESUMO MAIS INFORMATIVO

Nesta etapa, um algoritmo de regressão é aplicado para estimar a cobertura de informações relevantes (informatividade) de cada um dos resumos candidatos gerados na etapa anterior. Por ser uma tarefa de aprendizado supervisionado, a criação desse modelo de regressão requer a disponibilidade de um conjunto de treinamento contendo resumos

anotados com algum escore que represente a sua informatividade. Neste trabalho, nós avaliamos a utilização das medidas de cobertura, precisão e medida-F, como possíveis valores para representar a informatividade dos resumos candidatos. Essas medidas foram escolhidas por serem comumente usadas na literatura para avaliar sistemas de sumarização. Dessa forma, o objetivo do modelo desenvolvido é gerar um escore que represente uma estimativa da relevância de um dado resumo.

Identificar atributos para a construção do modelo de regressão é uma tarefa essencial para o seu desempenho. Por isso, selecionamos um conjunto com dez atributos que apresentaram bons resultado, conforme reportado em (OLIVEIRA et al., 2017) para sumarização de artigos de notícias escritos em inglês. Esses atributos são baseados em indicadores de relevância, como frequência e posição, sendo extraídos em nível de resumo, frase e conceitos. Alguns dos atributos adotados são gerados aplicando medidas de similaridade e divergência entre a lista de pesos de todos os conceitos presentes nos documentos de entrada e no resumo. Os dez atributos utilizados para construção do modelo são:

- **Palavras únicas:** computado pelo quociente entre a quantidade de palavras distintas no resumo, em relação ao total de palavras do resumo.
- **Palavras nos títulos dos documentos:** representa a proporção do total de palavras do resumo que estão nos títulos dos documentos, com *stopwords* sendo desconsideradas.
- **Palavras nas primeiras frases:** calculado pelo quociente entre as palavras do resumo que estão na primeira frase de cada um dos documentos de entrada, com *stopwords* sendo removidas.
- **Quantidade de frases:** é o total de frases do resumo.
- **Posição das frases:** computado pela média da medida de posição das sentenças (Equação 2) presentes no resumo.
- **Sobreposição dos pesos:** essa medida representa a sobreposição de pesos dos conceitos presentes no resumo e nos documentos de entrada, sendo computada conforme apresentado na Equação 4.

$$\text{Sobreposicao}(D, R) = \sum_{c_i \in D} \text{Min}(D_{c_i}, R_{c_i}) \quad (4)$$

Nessa equação D é a lista de conceitos e seus pesos dos documentos de entrada, R é a lista de pesos dos conceitos presentes no resumo, D_{c_i} é o peso do conceito c_i em D , e R_{c_i} é o peso do conceito c_i em R .

- **Similaridade com outros resumos candidatos:** A medida de similaridade do cosseno é computada entre o resumo candidato e os outros resumos candidatos gerados.
- **Medidas de divergência:** As medidas de divergência Kullback-Leibler (KL) (KULLBACK, 1959), Jensen-Shannon (JS) (LIN, 2006) e Coeficiente de Divergência (SHIRKHORSHIDI; AGHABOZORGI; WAH, 2015) são computadas entre a lista de pesos dos conceitos presentes no resumo candidato e nos documentos de entrada.

Ao final desta fase, o resumo candidato com a maior escore de relevância estimado pelo modelo de regressão é selecionado como o resumo contendo mais informações relevantes dos documentos de entrada.

4 | EXPERIMENTOS

Nesta seção são apresentados e discutidos os experimentos realizados para avaliar o desempenho do sistema de SAT desenvolvido neste trabalho. Um primeiro experimento foi executado para avaliar cinco algoritmos de regressão para estimar a informatividade dos resumos candidatos. Por fim, um segundo experimento comparou o desempenho obtido pelo sistema desenvolvido em comparação com outros trabalhos da literatura.

Neste trabalho, utilizamos o corpus do CSTNews (CARDOSO et al. 2011) que foi criado para auxiliar pesquisas na área de SAT de artigos de notícias escritos em português do Brasil. Esse corpus possui cinquenta grupos de notícias, e em cada grupo existem aproximadamente três documentos de fontes diferentes abordando o mesmo assunto. Cada grupo de notícias possui cinco resumos extrativos criados para serem utilizados na avaliação de sistemas de sumarização multidocumento.

Para avaliar os resumos criados foram adotadas as medidas de cobertura, precisão e medida-F geradas a partir da ferramenta do ROUGE (LIN, 2004), usando uma variação chamada de ROUGE-1 (R-1). Essa variação computa a sobreposição de unigramas entre os resumos gerados automaticamente e os resumos de referência. Como cada grupo de documentos do corpus CSTNews possui mais de um resumo de referência, os valores das medidas de avaliação são gerados computando a média entre a comparação do resumo gerado automaticamente com todos os resumos de referência. O R-1 foi computado utilizando a ferramenta *py-rouge*, aplicando *stemming* e sem remoção de *stopwords*.

Os resumos gerados pelo sistema desenvolvido neste trabalho possuem no máximo de 110 palavras. Esse limiar foi escolhido para gerar resumos com tamanhos compatíveis com outros sistemas da literatura que geram resumos com aproximadamente de 100 palavras.

4.1 AVALIAÇÃO DOS ALGORITMOS DE REGRESSÃO

Neste primeiro experimento, avaliamos a aplicação de cinco algoritmos de aprendizado de máquina para realizar a estimação da informatividade dos resumos candidatos. Os seguintes algoritmos disponíveis na ferramenta *Scikit-learn* foram avaliados: Regressão Bayesiana, Regressão Linear, Regressão Ridge, Máquina de vetores de suporte do inglês *Support Vector Regression* (SVM) com kernel linear (SVR-Linear). Esses algoritmos foram usados com suas configurações padrões, ou seja, nenhuma calibração de parâmetros foi realizada.

Para avaliar o desempenho dos algoritmos utilizamos a metodologia de validação cruzada com k conjuntos (*k-fold cross validation*). O parâmetro k foi definido com o valor cinquenta ($k=50$) que é total de grupos de artigos de notícias do corpus CSTNews. Dessa forma, para cada grupo de documentos, o processo de geração do resumo é executado de acordo com as duas etapas a seguir:

- **Treinamento:** Nesta etapa, 49 grupos de artigos são usados para treinar o modelo de regressão. Como cada grupo possui 5 resumos candidatos, um total de 245 resumos são usados para treinamento.
- **Teste:** O grupo de documentos não selecionado na etapa de treinamento é usado para testar o modelo de regressão criado na etapa.

Na Tabela 1 são apresentados os resultados deste experimento considerando as medidas de cobertura, precisão e medida-F do R-1 dos resumos selecionados como mais informativos. Essas três medidas também foram avaliadas como possíveis escores de informatividade dos resumos (atributos alvos).

Algoritmos	Atributo Alvo	Cobertura	Precisão	Medida-F
<i>R. Bayesiana</i>	<i>Cobertura</i>	64,64 (9,35)	62,34 (8,93)	63,47
	<i>Precisão</i>	64,66 (9,36)	62,29 (8,89)	63,45
	<i>Medida-F</i>	64,57 (9,39)	62,29 (8,95)	63,41
<i>R. Linear</i>	<i>Cobertura</i>	64,75 (9,56)	62,40 (8,83)	63,55
	<i>Precisão</i>	64,59 (9,40)	62,24 (8,91)	63,40
	<i>Medida-F</i>	64,57 (9,39)	62,29 (8,95)	63,41
<i>R. Ridge</i>	<i>Cobertura</i>	64,73 (9,40)	62,36 (8,94)	63,52
	<i>Precisão</i>	63,67 (9,76)	61,29 (9,42)	62,46
	<i>Medida-F</i>	64,49 (9,51)	62,17 (9,07)	63,31
<i>SVM-Linear</i>	<i>Cobertura</i>	64,35 (9,40)	61,96 (8,99)	63,13
	<i>Precisão</i>	63,92 (9,70)	61,61 (9,28)	62,74
	<i>Medida-F</i>	64,34 (9,42)	61,95 (8,95)	63,12

Tabela 1. Resultados da avaliação (%) e desvio padrão (entre parênteses) dos algoritmos de regressão.

Em geral, os algoritmos apresentaram resultados similares, sendo que o algoritmo de Regressão linear usando a medida de cobertura do R-1 apresentou o melhor desempenho nas três medidas de avaliação. O resultado obtido na medida de cobertura demonstra que esse algoritmo conseguiu gerar resumos que possuem 64,75% das informações presentes nos resumos de referência. É importante ressaltar que o algoritmo de Regressão linear obteve 62,40% na medida de precisão e 63,55% na medida-F. Esses resultados indicam que os resumos gerados por esse algoritmo apresentam um bom equilíbrio entre a quantidade de informações presentes nos resumos de referência e aquelas que o algoritmo selecionou.

Em geral, adotar a medida de cobertura do R-1 como escore de informatividade dos resumos foi a que levou os algoritmos de regressão a selecionarem melhor os resumos informativos. Tal resultado indica que essa medida é mais adequada para refletir informatividade dos resumos.

4.2 COMPARAÇÃO COM OUTROS TRABALHOS

Neste segundo experimento é realizado uma comparação dos resultados obtidos pelo sistema desenvolvido usando o algoritmo de Regressão linear com outros cinco sistemas identificados na literatura, sendo eles: GistSumm (PARDO, 2005), CSTSumm (CASTRO JORGE e PARDO, 2010), RC-4 (CARDOSO e PARDO, 2016), SentReg (SODRÉ e OLIVEIRA, 2019) e ILP (GOMES e OLIVEIRA, 2019). Além desses sistemas, identificamos o sistema de Oráculo que representa o limite máximo que pode ser obtido pelo sistema proposto neste trabalho, selecionando sempre dentre os cinco resumos candidatos de cada grupo de documentos, aquele que possui o maior valor real da medida de cobertura do R-1.

Na Tabela 2 são apresentados os resultados obtidos usando as medidas de cobertura, precisão e medida-F do R-1. O melhor sistema (com exceção do Oráculo) em cada medida de avaliação é destacado em negrito.

Sistemas	Cobertura	Precisão	Medida-F
CSTSumm	56,21 (11,34)	56,40 (10,17)	56,31
GistSumm	56,23 (12,34)	54,87 (10,76)	55,54
RC-4	57,66 (8,44)	58,33 (10,61)	57,99
SentReg	62,23 (9,28)	59,43 (8,08)	60,80
ILP	63,08 (8,56)	59,23 (7,66)	61,10
RegILP	64,75 (9,56)	62,40 (8,83)	63,55
<i>Oracle</i>	<i>66,30 (8,94)</i>	<i>63,79 (8,15)</i>	<i>65,02</i>

Tabela 2. Resultados da comparação (%) e desvio padrão (entre parênteses) com outros trabalhos da literatura.

O sistema desenvolvido neste trabalho obteve desempenho superior aos demais sistemas considerados em todas as medidas de avaliação do R-1. O sistema SentReg apresenta o segundo melhor resultado na medida de precisão, enquanto que o sistema ILP alcançou o segundo melhor desempenho nas medidas de cobertura e medida-F. É importante ressaltar que o sistema ILP utiliza uma abordagem similar ao sistema desenvolvido neste trabalho, mas sem a geração de múltiplos resumos candidatos e o módulo de seleção usando regressão. Dessa forma, é possível concluir que a estratégia de sumarização baseada em duas etapas adotada neste trabalho é viável e resultou na geração de resumos mais informativos do que os outros sistemas avaliados.

Apesar dos resultados encorajadores, ainda existe muito espaço para melhorias, como pode ser observado comparando o resultado do sistema proposto com o sistema de Oráculo identificado. Tais resultados evidenciam que a etapa de geração dos resumos candidatos foi capaz de gerar resumos informativos para os documentos de entrada, mas esses não foram identificados na etapa de seleção. Tal problema ocorreu devido a erros de estimação cometido pelo algoritmo de regressão na seleção do resumo mais informativo.

5 | CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS

Neste trabalho foi apresentado um sistema de sumarização automático de artigos de notícias escritas em português. A solução proposta é baseada em uma estratégia de sumarização em duas etapas. Em uma primeira etapa, uma abordagem baseada em conceitos usando programação linear inteira é aplicada para gerar um conjunto de resumos candidatos. Posteriormente, um modelo de regressão foi construído para estimar a informatividade dos resumos candidatos visando selecionar o resumo mais informativo. Para isso, dez atributos baseados em medidas de similaridade, divergência e tradicionais indicadores de relevância foram utilizados.

Para avaliar o desempenho do sistema desenvolvido, experimentos executados usando o corpus CSTNews (CARDOSO et al., 2011) e as medidas de avaliação do ROUGE. Os resultados experimentais demonstraram que o algoritmo de Regressão linear obteve o melhor desempenho em comparação com outros algoritmos para a tarefa de estimar um escore de informatividade dos resumos candidatos. Além disso, o sistema proposto obteve resultados superiores em comparação com outros trabalhos da literatura.

REFERÊNCIAS

CARDOSO, P. C.; PARDO, T. A. S. **Multi-document summarization using semantic discourse models**. *Procesamiento del Lenguaje Natural*, (56): p. 57–64, 2016.

CASTRO JORGE, M. L. d. R.; PARDO, T. A. S. **Experiments with cst-based multidocument summarization**. In: *Workshop on Graph-based Methods for Natural Language Processing*, Stroudsburg, PA, USA, p. 74–82. 2010.

CARDOSO, P.; MAZIERO, E.; JORGE, M.; SENO, E.; FELIPPO, A. D.; RINO, NUNES, L. M.; PARDO, T. A. S. **CstNews - a discourse-annotated corpus for single and multi-document summarization of news texts in Brazilian Portuguese**. In: 3^a RST Brazilian Meeting. p. 88–105. 2011.

GOMES, Laerth; OLIVEIRA, Hilário. **A Multi-document Summarization System for News Articles in Portuguese using Integer Linear Programming**. In: Encontro Nacional de Inteligência Artificial e Computacional (ENIAC), 2019, Salvador. Anais do XVI ENIAC. Porto Alegre: Sociedade Brasileira de Computação, p. 622-633. 2019.

KULLBACK, S. **Information theory and statistic**. Wiley, 1959.

LIN, C.-Y. **Rouge**: A package for automatic evaluation of summaries. In: Annual Meeting of the Association for Computational Linguistics. Barcelona, Spain. Association for Computational Linguistics, 2004. p. 74–81.

LIN, J. **Divergence measures based on the shannon entropy**. IEEE Trans. Inf. Theor., IEEE Press, Piscataway, NJ, USA, v. 37, n. 1, p. 145–151, 2006.

MANI, I. **Summarization evaluation: An overview**. In: Third Second Workshop Meeting on Evaluation of Chinese & Japanese Text Retrieval and Text Summarization, NTCIR-2, Tokyo, Japan, 2001.

NENKOVA, A.; McKeown, K. **A survey of text summarization techniques**. In Mining Text Data. Springer. p. 43-76. 2012.

OLIVEIRA, H.; LINS, R. D.; LIMA, R.; FREITAS, F. SIMSKE, S. J. **A regression-based approach using integer linear programming for single-document summarization**. In: 2017 IEEE 29th International Conference on Tools with Artificial Intelligence (ICTAI). Boston, USA. p. 270-277. 2017.

OLIVEIRA, H.; LINS, R. D.; LIMA, R.; FREITAS, F.; SIMSKE, S. J. **A concept-based ILP approach for multi-document summarization exploring centrality and position**. In: 7th Brazilian Conference on Intelligent Systems (BRACIS). São Paulo, Brazil. p. 37–42, 2018.

PARDO, T. A. S. **Gistsumm-gist summarizer**: Extensões e novas funcionalidades. Serie de Relatórios do NILC. 2005.

SHIRKHORSHIDI, A. S.; AGHABOZORGI, S.; WAH, T. Y. **A comparison study on similarity and dissimilarity measures in clustering continuous data**. PLOS ONE, Public Library of Science, v. 10, n. 12, p. 1–20, 2015.

SODRÉ, Lucas; OLIVEIRA, Hilário. **Avaliando Algoritmos de Regressão para Sumarização Automática de Textos em Português do Brasil**. In: Encontro Nacional de Inteligência Artificial e Computacional (ENIAC). Anais do XVI ENIAC. Porto Alegre: Sociedade Brasileira de Computação, p. 634-645. 2019.

TÉCNICAS DE APRENDIZADO DE MÁQUINA APLICADAS NA PREVISÃO DE VULNERABILIDADES QUANTO AO USO DA INTERNET PELO PÚBLICO INFANTOJUVENIL

Data de aceite: 27/08/2020

Data de submissão: 12/06/2020

Franciele Cristina Espanhol Ferreira Alves

Universidade Federal de Uberlândia- UFU,
Faculdade de Computação-FACOM
Monte Carmelo- MG
<http://lattes.cnpq.br/2882868713508417>

Fernanda Maria da Cunha Santos

Universidade Federal de Uberlândia-UFU,
Faculdade de Computação- FACOM
Monte Carmelo- MG
<http://lattes.cnpq.br/6802596562404346>

RESUMO: A Internet oferece a conexão de bilhões de pessoas em todo o mundo por meio de várias formas de interações e organizações sociais. As redes sociais têm se tornado uma das principais formas de comunicação, que proporciona boas experiências e, contraditoriamente, preocupações para pais e professores de crianças e adolescentes. As redes sociais estão sendo utilizadas pelo público infantojuvenil demasiadamente, possibilitando gerar situações de riscos e vulnerabilidades à eles. Diante destes fatores, o objetivo deste trabalho foi empregar duas técnicas de Aprendizado de Máquinas para prever se parte do público infantojuvenil da cidade de Monte Carmelo estar suscetível à algum risco ou situação constrangedora nas redes sociais. A base de dados em estudo é constituída de características pessoais que qualificam as formas de uso deste público em relação ao seu acesso à redes sociais da Internet. As técnicas de Aprendizado de

Máquina escolhidas foram os algoritmos de Árvore de Decisão e Naive Bayes, ambos definidos no WEKA. Os resultados alcançados pelas implementações apresentaram valores satisfatórios com acurácia acima de 90%.

PALAVRAS-CHAVE: Árvore de Decisão; Infantojuvenil; Naive Bayes; Redes sociais.

MACHINE LEARNING TECHNIQUES APPLIED IN PREDICTING VULNERABILITIES REGARDING THE USE OF THE INTERNET BY CHILDREN AND YOUTH

ABSTRACT: The Internet connects billions of people worldwide through various forms of interactions and social organizations. Social networks have become one of the main forms of communication, which provides good experiences and, contradictorily, concerns for parents and teachers of children and adolescents. Social networks are being used by children and youth audiences too much, making it possible to generate situations of risks and vulnerabilities to them. In view of these factors, the objective of this work was to employ two machine learning techniques to predict whether part of the children's population in the city of Monte Carmelo is susceptible to some risk or embarrassing situation on social networks. The database under study consists of personal characteristics that qualify the ways of using this public access to social networks on the Internet. The machine learning techniques chosen were the Decision Tree and Naive Bayes algorithms, both defined in WEKA. The results achieved by the

implementations presented satisfactory values with accuracy above 90%.

KEYWORDS: Decision Tree; Children; Naive Bayes; Social Networks.

1 | INTRODUÇÃO

A sociedade vive em constante modificação. A Internet e as tecnologias estão cada vez mais presentes na vida das pessoas, conseqüentemente, crianças começam cada vez mais cedo a lidar com elas (RODRIGUES, 2015). O uso imprudente das inovações tecnológicas pode colocar em risco a segurança e a saúde de seus usuários. Nesse contexto, os pais e responsáveis surgem como orientadores e mediadores da relação que os seus filhos devem estabelecer com as diferentes tecnologias.

A utilização da Internet, principalmente, o uso das redes sociais pelo público infantojuvenil é um tema que vem gerando grande preocupação entre pais e educadores. Pessoas mal-intencionadas estabelecem contato com as crianças por meio de conversas em salas de bate-papo, jogos compartilhados, mensagens instantâneas e e-mail. Em um mundo globalmente conectado e com uma sociedade rica em informação, é preciso haver o discernimento e equilíbrio das informações expostas, identificando comportamentos em ambientes digitais e promovendo tomadas de decisões críticas (BERSON; BERSON, 2005).

A busca por softwares que realizem filtros em redes sociais não é suficientemente eficaz para gerir o comportamento dos jovens perante o mundo digital, sendo preciso a conscientização e atenção dos usuários da falta de segurança na Internet (FLEMING et al., 2006). Para tentar suprir a falta de segurança na Internet e visando o monitoramento do comportamento de crianças e adolescentes quando estão no mundo virtual, alguns autores estão utilizando os algoritmos de Aprendizagem de Máquina (AM) para tentar detectar ameaças e vulnerabilidades (JÚNIOR et al., 2015; SANTIN P. L.; FREITAS, 2012; SANTOS; GUEDES, 2019)

Nas últimas décadas, o AM (do inglês *Machine Learning*) tornou-se um dos pilares da tecnologia da informação, destacando algoritmos e técnicas computacionais que são capazes de aprender a partir de dados existentes. O AM tem uma gama extremamente ampla de aplicações. Por exemplo, tem grande contribuição em processamento de linguagem natural, motores de busca, diagnósticos médicos, bioinformática, reconhecimento de fala, reconhecimento de escrita, visão computacional e locomoção de robôs e sistemas de previsão (AMORIM; BARONE; MANSUR, 2008).

Devido ao uso intenso, tanto os sites quanto os aplicativos das redes sociais reúnem materiais suficientes para subsidiar previsões e classificar padrões que caracterizam os seus usuários em larga escala (FREITAS et al., 2008). As técnicas de AM podem utilizar esta vasta quantidade de dados para aprender e gerar conclusões diante os questionamentos e tomadas de decisões. Ademais, prevendo ou classificando um padrão a partir do modo de uso dos usuários nas redes sociais, poder-se-ia alertar pais e responsáveis por um perigo

eminente à seu filho, que fica exposto a um mundo até então desconhecido e formado por distintos tipos de personalidades (PEREIRA, 2009).

Por conseguinte, objetivou para este estudo, o uso de técnicas de Aprendizagem de Máquina para prever se o público infantojuvenil está suscetível à algum risco ou situação constrangedora na Internet. Esta previsão será feita ao analisar uma base de dados constituída de características que qualificam as formas de uso deste público em relação ao seu acesso a redes sociais da Internet.

2 | REFERENCIAL TEÓRICO

2.1 Aprendizagem de Máquina

Para Mitchell (1997), AM é uma área de pesquisa da Inteligência Artificial que tem a finalidade de estudar métodos capazes de extrair conceitos com base em um modelo de informações. Um sistema de aprendizado é um programa de computador que possui capacidade de tomar decisões com base em experiências acumuladas por meio de soluções bem-sucedidas de problemas anteriores. Em outras palavras, o sistema de aprendizado tem a capacidade de tomar decisão por já ter tido experiência acumulada em casos de sucesso (MONARD; BARANAUSKAS, 2003b).

Existem duas formas de adquirir conhecimento pelos algoritmos propostos pelas técnicas de AM: supervisionado e não-supervisionado. Lorena e Carvalho (2003) relata que no aprendizado supervisionado tem-se a representação de um professor externo, onde o algoritmo de aprendizagem de máquina é treinado a partir de conjuntos de exemplos rotulados. Uma característica de todos os paradigmas de algoritmos supervisionados é que o conceito a ser aprendido deve estar relacionado com exemplos já observados, e cada exemplo deve estar rotulado com a classe à qual pertence. Já no algoritmo não-supervisionado, não há instâncias rotuladas da função a ser aprendida, e o algoritmo faz a análise de um modelo e tenta agrupá-los de alguma forma.

Existem técnicas de AM que são implementados em sistemas computacionais que atuam na classificação e previsão de dados tanto em trabalhos científicos quanto em atividades rotineiras em diferentes áreas de atuação. Essas técnicas são: redes neurais artificiais, algoritmos genéticos, árvores de decisão, Support Vector Machine (SVMs), sistemas fuzzy, classificadores Bayesianos e outras (LORENA; CARVALHO, 2003). Um sistema computacional que possa atuar em situações de incertezas deve ser capaz de atribuir níveis de confiabilidade para todas as sentenças em sua base de conhecimento, e ainda, estabelecer relações entre as sentenças. Lidar com falta de informação significa lidar com incertezas, e tais situações podem ser representadas por grafos que retratam as relações causais entre eventos. As redes bayesianas oferecem uma abordagem para o raciocínio probabilístico que engloba teoria de grafos para o estabelecimento das relações entre sentenças e ainda, teoria de probabilidades, para a atribuição de níveis de confiabilidade (MARQUES R. L.; DUTRA, 2002).

De forma semelhante, uma árvore de decisão pode ser construída a partir de um determinado conjunto de atributos, além de existir, exponencialmente, várias árvores que representam um mesmo conjunto (TAN et al., 2009). As definições das técnicas de árvore de decisão e de aprendizado Bayesiano serão descritas sucintamente nas próximas subseções.

2.1.1 Árvore de Decisão

Árvore de decisão é uma estrutura de dados recursiva, onde um nó folha corresponde a uma classe (um rótulo) ou um nó de decisão (nó interno), que contém um teste sobre algum atributo. Em cada nó interno possui exatamente uma aresta chegando e duas ou mais saindo para uma subárvore. Cada subárvore tem a mesma estrutura que a árvore (MONARD; BARANAUSKAS, 2003b).

A árvore é representada como um conjunto de regras. Cada regra tem início na raiz da árvore e caminha até uma de suas folhas. Aplica-se a condição de teste do nó raiz ao registro e segue a ramificação apropriada baseada no resultado do teste. Isto direcionará a um nó interno, que executará um novo teste resultando numa aresta direcionada à um nó folha ou à outro nó interno. Se for encaminhado a um nó folha, o rótulo da classe associado ao nó folha é atribuída ao registro. A Figura 1 mostra o caminho na árvore de decisão que prevê o rótulo da classe de um flamingo.

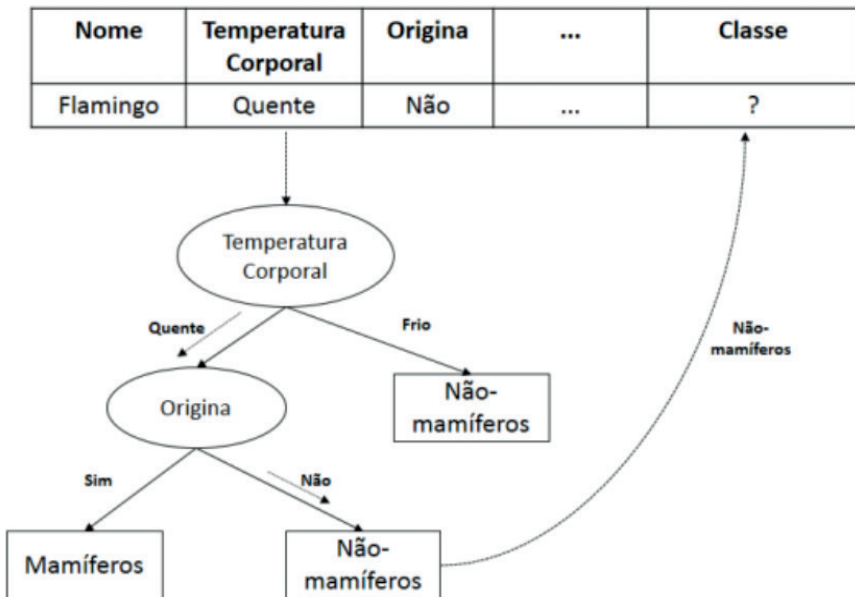


Figura 1 – Árvore de decisão para o problema de classificação de mamíferos.

Fonte: Retirado do livro Tan et al. (2009, p. 179).

Estas árvores são criadas por um conjunto de dados para treinamento e, posteriormente, outros exemplos, que são o conjunto de dados de teste, são classificados pela mesma árvore. Para a construção destas árvores são usados algoritmos como o ID3, ASSISTANT e C4.5.

2.1.2 Aprendizado Bayesiano

Em algumas situações, o rótulo da classe de um registro da base de dados de teste não pode ser previsto com certeza, devido a alguns atributos com ruídos ou da presença de fatores de confusão que afetam a classificação deste registro (TAN et al., 2009). Nestes casos, modelos estatísticos podem encontrar uma boa aproximação do conceito induzido. Dentre os métodos estatísticos, pode-se citar o Aprendizado Bayesiano, que usa uma maneira probabilística e o conhecimento prévio do problema, combinado com os exemplos de treinamento, para saber qual a porcentagem final de uma hipótese (MONARD; BARANAUSKAS, 2003a).

Uma vez definida a topologia da rede, basta especificar as probabilidades dos nós que participam em dependências diretas, e utilizar estas para computar as demais probabilidades que se deseje. A probabilidade condicional $P(a | b)=x$, pode ser interpretada como: “dado o evento b, a probabilidade do evento a é x”. Por exemplo, $P(\text{Cárie} | \text{Dor})=0.8$, indica que caso um paciente esteja com dor de dente e nenhuma outra informação esteja disponível, então, a probabilidade do paciente ter uma cárie é 0.8. Portanto, a regra de Bayes é definida por:

$$P(b | a) = \frac{P(a | b)P(b)}{P(a)}$$

3 | METODOLOGIA

A base de dados utilizada pelo presente estudo foi originada por informações obtidas por crianças e adolescentes que residem na cidade de Monte Carmelo, no Estado de Minas Gerais. A página do site QEDU¹ traz informações sobre o número de estudantes por série na educação escolar. As séries pesquisadas foram do 5º ao 9º ano que compreende alunos de 9 a 14 anos, que é o foco da pesquisa. Segundo o site QEDU, o município de Monte Carmelo possui 3.182 alunos matriculados nessa faixa etária. A pesquisa foi feita por uma amostra que totalizaram 498 pessoas, que corresponde a 16% dos alunos de escolas da rede pública e da rede privada, sendo indivíduos do sexo masculino e feminino.

¹ https://qedu.org.br/cidade/2246-montecarmelo/censoescolar?year=2018dependence=0localization=0education_stage=0item=

Após a coleta dos dados nas cinco escolas, foi realizado a análise das informações contidas nas respostas dos questionários, organizando e tabulando os dados em planilhas eletrônicas. O software utilizado para organizar as respostas foi o PSPP², um software livre direcionado à cálculos estatísticos, o qual foi escolhido para demonstrar os dados de estatística descritiva. Assim, foi possível identificar o perfil dos alunos de acordo com a sua idade, escolaridade, quais redes sociais que utilizam, a frequência de uso nas redes sociais e outras informações.

Analisando a base de dados, gerou-se gráficos para realçar e facilitar sua interpretação. A Figura 2 exibe a relação entre a idade dos entrevistados e a rede social que mais utilizam. Dentre os 240 entrevistados com idade entre 9 a 10 anos, 73,5% utilizam o YouTube, e em segundo lugar, com 11,3%, o Facebook. Isso pode levantar a questão de que o conteúdo que as crianças e os adolescentes querem assistir, sem horário predeterminado, justifica a alta audiência do YouTube e a substituição dos programas de televisão por esta rede social.

Em contrapartida, a rede social menos utilizada por essa faixa de idade é o Twitter, com 0,5%, abaixo do Instagram, com 6,9%. Pré-adolescentes com idade entre 11 e 12 anos, totalizaram 128 alunos pesquisados, o que corresponde a 25,8% do total da amostra analisada. Estes alunos também utilizam com mais frequência o canal do YouTube, com um total de 47,7%, seguido pelo Facebook com 21,9%. A faixa etária dos 13 aos 14 anos, totaliza 163 alunos e corresponde a 32,8% do conjunto total. Estes adolescentes utilizam em primeiro lugar o Facebook, com 39,3%, seguido do canal YouTube com 27% e em seguida a rede social Instagram com 15,3% dos entrevistados.

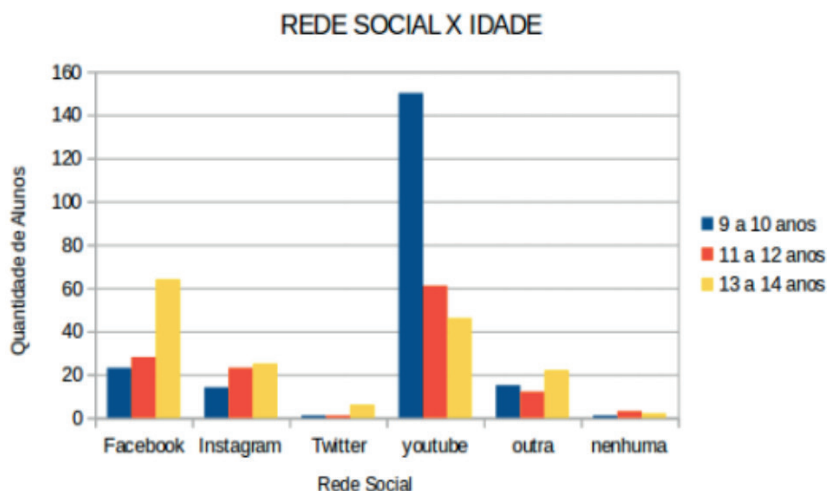


Figura 2 – Relação entre a idade dos entrevistados e as redes sociais que mais utilizam.

Ao analisar a quantidade de alunos que possuem Internet em casa, dentre os alunos de 9 aos 14 anos, 88,76% tem acesso à rede mundial de computadores. Esses dados demonstram que as crianças de hoje têm crescido em um ambiente com maior infraestrutura tecnológica e com acesso a informação do que a que seus pais vivenciaram.

Segundo os dados apresentados na Figura 3, o meio eletrônico para o acesso as redes sociais pelas crianças de 9 a 10 anos são: 60,9% usam smartphone; 40,6% usam computador e smartphone; e 37,1% apenas o computador. O restante não possui acesso. De acordo com a faixa etária de 11 a 12 anos, 31,6% utilizam smartphone, 24,3% utilizam computador e smartphone e, 15,6% utilizam somente o computador. No que diz respeito aos entrevistados de 13 a 14 anos, 37,9% utiliza o smartphone, 27,7% utilizam ambos, computador e smartphone e, 23,4% utilizam computador.

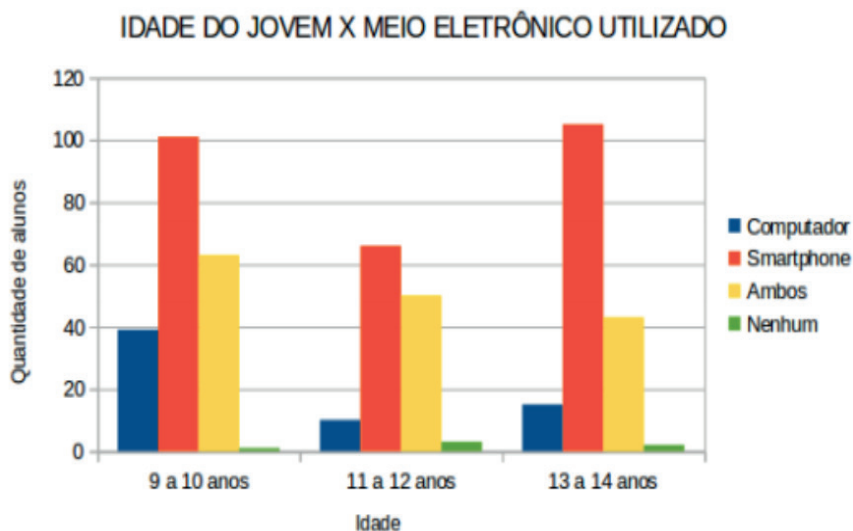


Figura 3 – Meio eletrônico mais utilizado pelos entrevistados.

A Figura 4 exibe os resultados alcançados ao questionar os entrevistados se já passaram por alguma situação constrangedora nas redes sociais.

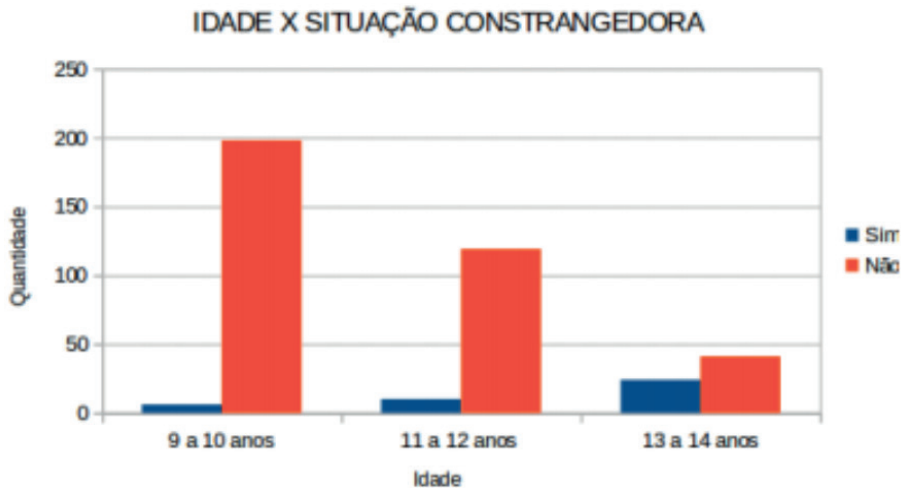


Figura 4 – Relação entre idade e se passaram por alguma situação constrangedora nas redes sociais.

Além dos dados já apresentados, foram analisados mais alguns gerados pela entrevista com os alunos da cidade de Monte Carmelo, para atender os objetivos deste estudo. São eles: frequência de acesso, tempo de interação contínua, controle de acesso feito pelos pais. Todos os atributos destacados compõe a base de dados, a qual foi analisada pelas técnicas de AM que identificaram se os entrevistados são susceptíveis ou não à alguma situação constrangedora pelas redes sociais.

As técnicas de AM testadas na base de dados foram Árvore de Decisão e Rede Bayesiana. Estas duas técnicas foram definidas utilizando a ferramenta computacional WEKA³.

O WEKA possui uma coleção de algoritmos de Aprendizagem de Máquina escrito na linguagem de programação Java, e que foi criado por pesquisadores da Universidade de Waikato, Nova Zelândia. Esta ferramenta consiste de vários programas, o que denominaremos de bibliotecas, dos principais algoritmos e técnicas de AM disseminados, tais como redes bayesianas, máquina de vetores de suporte, árvores de decisão, redes neurais artificiais, entre outras.

Na base de dados formatada, cada linha é uma instância, formada por atributos que estão organizados em setes colunas. A última coluna é a classe ou o rótulo que se deseja alcançar com as técnicas de AM, ou seja, se o público infantojuvenil já passou ou não por alguma situação constrangedora ao usar as redes sociais.

A escolha pelo algoritmo de árvore de decisão se deu pelo fato de ser um método de aprendizado de máquina supervisionado não-paramétrico (que significa que não assume nada sobre os dados de antemão), muito utilizados em tarefas de classificação

3 <https://www.cs.waikato.ac.nz/ml/weka/>

e regressão. A árvore de decisão é de fácil implementação e também não requer nenhum conhecimento estatístico para sua interpretação. São instáveis, pois pequenas alterações produzem novas árvores.

Por outro lado, a opção pelo classificador Bayesiano deve-se pelo algoritmo ser muito rápido, desconsidera a correlação entre as variáveis e possui um desempenho relativamente maior do que outros classificadores com um pequeno número de dados de teste para fazer classificações com uma boa precisão. O classificador Bayesiano utilizado neste estudo foi Naive Bayes, o qual é um modelo simples e que se destaca pelos sucessos obtidos na aplicação em diversos problemas, mesmo comparado à classificadores mais complexos (FRIEDMAN; GEIGER; GOLDSZMIDT, 1997).

As técnicas de AM propostas foram treinadas e avaliadas sob a base de dados configurada por uma validação cruzada com 70% para o conjunto de treinamento e 30% para o conjunto de teste. Posteriormente, os resultados gerados pelos algoritmos serão exibidos usando a matriz de confusão, que ilustra a precisão da solução para um problema de classificação, e pelo cálculo da acurácia, que consiste na proporção de predições corretas, sem levar em consideração o que é positivo e o que é negativo. A fórmula da acurácia é definida por:

$$Acuracia = \frac{Total\ de\ Acertos}{Total\ de\ dados\ no\ conjunto}$$

$$Acuracia = \frac{(VP + VN)}{(P + N)}$$

Onde VP corresponde ao verdadeiro positivo; VN corresponde ao verdadeiro negativo; P são os positivos e N negativos.

4 | RESULTADOS E DISCUSSÃO

A finalidade do presente estudo foi a aplicação de técnicas de AM para classificar a vulnerabilidade de crianças e adolescentes nas redes sociais da Internet quanto à sua forma de uso da mesma. Assim sendo, definiu-se os algoritmos de árvore de decisão e de Naive Bayes pelas bibliotecas do WEKA aplicados sob a base de dados gerada com a pesquisa realizada à alguns alunos da cidade de Monte Carmelo. Os resultados serão analisados a seguir.

O algoritmo utilizado pela árvore de decisão foi o J48, classificando um total de 458 instâncias corretas da base de dados, o que corresponde a 91.96%. No entanto, as instâncias classificadas incorretas foram 40, totalizando 8.03%. A matriz de confusão destes resultados está exibida na Tabela 1.

Não passou por Constrangimento	Passou por Constrangimento	
458	0	Não Passou por Constrangimento
40	0	Passou por Constrangimento

Tabela 1 – Matriz de confusão dos resultados gerados pela árvore de decisão.

Por conseguinte, foi analisado o algoritmo Naive Bayes. O total de instâncias classificadas corretas foram 453 que corresponde a 90.9639% de acertos. Já as instâncias classificadas como incorretas foram 45, totalizando 9.0361% de erros. Esses resultados podem ser vistos na matriz de confusão apresentada na Tabela 2.

Não passou por Constrangimento	Passou por Constrangimento	
449	9	Não passou por Constrangimento
36	4	Passou por Constrangimento

Tabela 2 – Matriz de confusão dos resultados gerados pelo algoritmo Naive Bayes.

A acurácia obtida pela árvore de decisão foi de 91.95%. Enquanto que a acurácia obtida pelo algoritmo Naive Bayes foi de 90.96%. Ou seja, a precisão de ambos se aproximam, mas os resultados alcançados pelo algoritmo Naive Bayes foi mais eficiente ao detectar uma pequena quantidade de alunos que já passaram por uma situação constrangedora. Contudo, é importante destacar que a quantidade de instâncias da base de dados foi pequena, o que limitou ambos os algoritmos de AM durante as fases de treinamento e teste.

5 | CONSIDERAÇÕES FINAIS

Com a crescente irradiação da Internet e sua intensa utilização pelos mais variados públicos da sociedade surgem conflitos e problemas sociais a serem enfrentadas. A exposição das crianças e adolescentes nas redes sociais podem colocá-las em risco, que vão desde passar por uma situação constrangedora até se tornarem alvo ao alcance de pedófilos.

A pesquisa realizada em textos referenciais mostrou que há um grande número de crianças e adolescentes cada vez mais inseridos na tecnologia digital. A adoção do uso de redes sociais por parte desses jovens, aliada à facilidade do acesso aos dispositivos

eletrônicos, têm tornado um cenário preocupante no que diz respeito a informações, exposição e amizades que são trocadas quando estão online. As políticas de segurança das redes sociais têm feito sua parte ao restringir acessos aos menores de treze anos. Entretanto, o que se pode notar que, mesmo em número menor, há a reincidência de possuírem contas falsas nesses sites. Uma vez cadastrados, passam a utilizar as redes sociais de forma irresponsável.

O estudo desenvolvido neste artigo procurou demonstrar que pelas técnicas de AM é possível identificar e classificar se o público infantojuvenil passou ou não por situações constrangedoras, mediante a análise da forma de uso das redes sociais. Desta forma, os resultados alcançados neste trabalho são um subsídio ao desenvolvimento de estudos mais minuciosos que enfatizam metodologias capazes de identificar possíveis vulnerabilidades na Internet, priorizando ações de medidas protetivas, principalmente à crianças e adolescentes.

O estudo mostra grande desafio aos pais e responsáveis à fiscalização do conteúdo encontrado no mundo virtual, pois é um ambiente em constante transformação. É preciso a implementação de políticas públicas de forma a prevenir e conscientizar a todos os desenvolvedores de conteúdo digital, sobre a segurança das crianças e adolescentes.

REFERÊNCIAS

AMORIM, M. J.; BARONE, D.; MANSUR, A. U. **Técnicas de aprendizado de máquina aplicadas na previsão de evasão acadêmica**. In: Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE). [S.l.: s.n.], 2008. v. 1, n. 1, p. 666–674.

BERSON, I. R.; BERSON, M. J. **Challenging online behaviors of youth: Findings from a comparative analysis of young people in the united states and new zealand**. Social Science Computer Review, Sage Publications Sage CA: Thousand Oaks, CA, v. 23, n. 1, p. 29–38, 2005.

FLEMING, M. J. et al. **Safety in cyberspace: Adolescents' safety and exposure online**. Youth & Society, Sage Publications Sage CA: Thousand Oaks, CA, v. 38, n. 2, p. 135–154, 2006.

FREITAS, C. M. D. S. et al. **Extração de conhecimento e análise visual de redes sociais**. In: Anais do XXVIII Congresso da Sociedade Brasileira de Computação. [S.l.: s.n.], 2008. p. 106–120.

FRIEDMAN, N.; GEIGER, D.; GOLDSZMIDT, M. **Bayesian network classifiers**. Machine Learning, v. 29, n. 2-3, p. 131–163, 1997.

JÚNIOR, M. S. R. F. et al. **Análise comportamental para proteção da criança nas redes sociais por meio de mineração de interações e sistemas multiagentes**. In: Anais do XII Encontro Nacional de Inteligência Artificial e Computacional. [S.l.: s.n.], 2015.

LORENA, A. C.; CARVALHO, A. C. P. d. L. F. **Introdução às máquinas de vetores suporte (support vector machines)**. [S.l.]: ICMC-USP, 2003.

MARQUES R. L.; DUTRA, I. **Redes bayesianas: o que são, para que servem, algoritmos e exemplos de aplicações**. Coppe Sistemas–Universidade Federal do Rio de Janeiro, p. 1–22, 2002.

MITCHELL, T. M. **Does machine learning really work?** AI magazine, v. 18, n. 3, p. 11–11, 1997.

MONARD, M. C.; BARANAUSKAS, J. A. **Conceitos sobre aprendizado de máquina**. In: **Sistemas Inteligentes Fundamentos e Aplicações**. 1. ed. Barueri-SP: Manole Ltda, 2003. p. 89–114. ISBN 85-204-168.

_____. **Indução de regras e árvores de decisão**. Sistemas Inteligentes. Rezende, SO Editora Manole Ltda, p. 115–140, 2003.

PEREIRA, S. E. F. N. **Redes sociais de adolescentes em contexto de vulnerabilidade social e sua relação com os riscos de envolvimento com o tráfico de drogas**. Tese (Doutorado) — Universidade de Brasília, 2009.

RODRIGUES, L. M. **As crianças e os media: análise de discursos dirigidos aos pais**. Tese (Doutorado), 2015.

SANTIN P. L. ; FREITAS, C. O. A. . P. E. . S. A. **Modelagem de aliciamento de menores em mensagens instantâneas de texto**. In: Anais do XII Simpósio Brasileiro de Segurança da Informação e de Sistemas Computacionais. [S.l.: s.n.], 2012.

SANTOS, L.; GUEDES, G. **Identificação de predadores sexuais brasileiros por meio de análise de conversas realizadas na internet**. In: Anais do VIII Brazilian Workshop on Social Network Analysis and Mining. Porto Alegre, RS, Brasil: SBC, 2019. p. 143–154. ISSN 2595-6094.

TAN, P. et al. **Introdução ao datamining: mineração de dados**. Ciencia Moderna, 2009. ISBN 9788573937619.

TECNOLOGIAS DISRUPTIVAS NO CONTEXTO DA TRANSFORMAÇÃO DIGITAL

Data de aceite: 27/08/2020

Data de submissão: 25/06/2020

Rejane Maria da Costa Figueiredo

Universidade de Brasília (UnB), Faculdade do Gama (FGA), *Information Technology - Research and Application Center (ITRAC)* – Brasília, DF
<http://lattes.cnpq.br/2187680174312042>

Leonardo Sagmeister de Melo

Universidade de Brasília (UnB), Faculdade do Gama (FGA) – Brasília, DF
<http://lattes.cnpq.br/0283236652451520>

John Lenon Cardoso Gardenghi

Universidade de Brasília (UnB), Faculdade do Gama (FGA), *Information Technology - Research and Application Center (ITRAC)* – Brasília, DF
<http://lattes.cnpq.br/1723982878095429>

Ricardo Ajax Dias Kosloski

Universidade de Brasília (UnB), Faculdade do Gama (FGA), *Information Technology - Research and Application Center (ITRAC)* – Brasília, DF
<http://lattes.cnpq.br/8309011123228244>

e, mais recentemente, a realidade digital e o *blockchain*. Este trabalho teve como objetivo identificar as principais tecnologias que estão sendo utilizadas no contexto de Transformação Digital no cenário mundial. Foram empregadas as técnicas mapeamento científico e revisão sistemática. Observa-se que as Tecnologias Disruptivas podem ser introduzidas por meio da aplicação de estratégias de Transformação Digital, levando em consideração as particularidades de cada caso.

PALAVRAS-CHAVE: Transformação Digital, Estratégia Digital, Tecnologias Digitais, Tecnologias Disruptivas.

DISRUPTIVE TECHNOLOGIES IN THE CONTEXT OF DIGITAL TRANSFORMATION

ABSTRACT: With the technological progress, new ways of doing ordinary tasks emerge daily in various contexts, including the organizational one. To maintain their competitiveness in this changing scenario, companies must change the way they perform their activities, undergoing a process of Digital Transformation. During the past few years, several technologies have been developed, such as Cloud Computing and the Internet of Things, and more recently Digital Reality and Blockchain, which caused great impact on the market. This study aimed to identify the main technologies used in the context of Digital Transformation. Disruptive Technology application models were found in the literature, and similarities were observed between the

RESUMO: Com o constante desenvolvimento tecnológico, novas formas de realizar tarefas cotidianas vêm surgindo em diversos contextos. Para manterem sua competitividade, organizações têm buscado a Transformação Digital em um cenário em que diversas tecnologias têm causado grandes impactos no mercado, tais como a computação em nuvem, a *Internet of Things*

model found and the case of a Brazilian company, as well as the case of the digitization plan created by the Brazilian government. Disruptive Technologies can be introduced using Digital Transformation Strategies, taking the particular needs of each case into account.

KEYWORDS: Digital Transformation, Digital Strategy, Digital Technologies, Disruptive Technologies.

1 | INTRODUÇÃO

Os avanços tecnológicos obtidos nos últimos anos vêm transformando o modo como diversos processos são realizados. É possível observar mudanças na área da saúde (VOLLENBROEK-HUTTEN *et al.*, 2015), no mundo do entretenimento (VIAL, 2019) e na estruturação e execução de serviços e processos empresariais (DELOITTE INSIGHTS, 2019; OESTERREICH *et al.*, 2016; VON LEIPZIG *et al.*, 2017) e governamentais (DELOITTE INSIGHTS, 2018; FIGUEIREDO *et al.*, 2019).

Quando surge uma tecnologia mais barata, simples e/ou conveniente, essa tecnologia causa uma ruptura na estrutura existente e recebe o nome de *Tecnologia Disruptiva*. Existem dois níveis de disrupção: um localizado, em que a disrupção se mantém dentro dos limites de um mercado ou indústria específicos; e um mais amplo, no qual a disrupção se espalha, abrangendo diversos mercados (SCHUELKE-LEECH, 2018). Tecnologias Disruptivas possuem um potencial de expandir nichos emergentes de mercado. Isso leva a uma alteração de paradigmas, que frequentemente muda o cenário do mercado de tal forma que empresas ou indústrias acabam por se deparar com a necessidade de se adequar ao novo cenário, remodelando-se em função dele: uma grande quantidade de empresas *pré-digitais* sente agora a necessidade de realizar um processo de *Transformação Digital* para se manter competitiva.

Um processo de Transformação Digital requer a mudança de diversos aspectos e áreas-chave dentro de uma organização (CHANIAS *et al.*, 2019; OESTERREICH *et al.*, 2016; VON LEIPZIG *et al.*, 2017). Não apenas isso – uma empresa pré-digital também deve passar por uma mudança de cultura para se digitalizar de forma efetiva. Muitos dos conceitos e modelos utilizados na Transformação Digital de uma empresa vêm de metodologias ágeis e da cultura de *startups* (CHANIAS *et al.*, 2019). Entre as mudanças a serem realizadas, vale ressaltar a da visão da estratégia de sistemas de informação da empresa. Com a evolução digital vivenciada nos últimos anos, a noção de que a estratégia de negócios e a estratégia de sistemas de informação de uma empresa são coisas distintas e separadas vem perdendo força. Agora, entende-se que não há uma fronteira bem definida entre ambas, e não há como gerir uma empresa sem levar em consideração sua parte tecnológica. Toda essa mudança de postura de uma organização exige uma *estratégia para a transformação digital*, orientada a negócios e inspirada em tecnologia (CHANIAS *et al.*, 2019).

Dado o contexto, o objetivo deste trabalho é *investigar que Tecnologias Disruptivas têm sido empregadas no contexto da Transformação Digital*.

O presente trabalho está organizado em 5 seções. Na Seção 2, são apresentados os conceitos de Transformação Digital, Estratégia de Transformação Digital e Tecnologias Disruptivas. Na Seção 3, expõe-se o Mapeamento Científico e a Revisão Sistemática da Literatura. Na Seção 4, são apresentados os resultados. Na Seção 5, são abordadas as conclusões.

2 I CONCEITOS E TRABALHOS CORRELATOS

Nesta seção, são apresentados os conceitos de Transformação Digital, Estratégias de Transformação Digital e Tecnologias Disruptivas.

2.1 Transformação Digital

A visão de que as organizações e os Sistemas de Informação são duas coisas separadas, cada uma com sua estratégia específica que deve ser alinhada à da outra, está ultrapassada. As tecnologias digitais vêm se desenvolvendo e se enraizando tanto ao redor das organizações como um todo que, em seu relatório tecnológico mais recente, a empresa de consultoria Deloitte (2019) afirmou que “[...] toda empresa é uma empresa de tecnologia, e todo funcionário é um tecnólogo”.

Esta fusão entre empresas e processos, de um lado, e suas tecnologias, de outro, pode ainda ser observada pela terminologia emergente do contexto de Transformação Digital. É importante, por exemplo, fazer a diferenciação entre os termos “digitalização” e “digitização”. Segundo o Gartner Institute (2015), “digitalização” refere-se ao ato de utilizar tecnologias digitais para mudar um modelo, gerando assim novos valores, ao passo que “digitização” seria um conceito mais específico, haja vista referir-se ao ato de usar tecnologias para efetivamente transformar algo que antes era analógico (um processo, por exemplo) em algo digital.

De fato, a tecnologia digital se faz presente nas organizações modernas de forma ubíqua. Empresas como a Uber, a Amazon e a Netflix são exemplos de como a estrutura de uma empresa, assim como seus processos e modelos de negócio, podem ser modelados do zero ao redor de tecnologias digitais (DELOITTE INSIGHTS, 2019). Tais empresas, ditas *nascidas digitais*, passaram a representar um risco àquelas que, até então, não haviam sentido pressão alguma para se modernizarem, as ditas empresas *pré-digitais* (CHANIAS et al., 2019).

Segundo uma pesquisa realizada pela SAP (2017), 84% das empresas pré-digitais com presença no mercado mundial veem a *Transformação Digital* como fator crítico para sua sobrevivência pelos próximos anos. Apesar dessa necessidade evidente, elaborar, implantar e manter um processo de Transformação Digital em uma organização não é uma tarefa fácil (CHANIAS et al., 2019; DELOITTE INSIGHTS, 2019; SAP SE, 2017). Entre

as empresas entrevistadas no mesmo estudo da SAP, apenas 3% haviam conseguido executar com sucesso seus esforços de Transformação Digital planejados até então.

2.2 Estratégia de Transformação Digital

Para executar um processo de Transformação Digital com sucesso, uma empresa deve desenvolver uma *Estratégia de Transformação Digital* (ETD) (CHANIAS et al., 2019). É difícil definir com exatidão como essa estratégia deve ser elaborada e implementada, principalmente ao considerar que cada organização possui características próprias, devendo, portanto, ser tratada de forma diferente (CHANIAS et al., 2019; DELOITTE INSIGHTS, 2019; OESTERREICH; TEUTEBERG, 2016; VIAL, 2019).

Chanias et al. (2019) propõem um modelo de formulação de Estratégias de Transformação Digital que busca identificar atividades e fatores-chave para a sua implementação em uma organização. O modelo proposto, ilustrado na Figura 1, é dividido em quatro áreas:

- *Contextos organizacionais*: referem-se ao estado da organização, sua estrutura, sua cultura, seus processos, seus funcionários e o que mais define a organização.
- *Práticas de ETD*: auxiliam a elaboração e implementação da ETD.
- *ETD realizada*: é a estratégia efetivamente aplicada e executada pela organização. O resultado observado é utilizado como insumo para a tomada de decisão em novos episódios de formação de Transformação Digital.
- *Episódios de formação de Estratégia Digital*: são todos os momentos em que uma decisão estratégica é tomada. Podem ser reuniões formais, *workshops*, palestras e até conversas informais.



Figura 1 – Modelo de definição de ETD.

Fonte: adaptada de CHANIAS et al. (2019).

Em um estudo de caso, Chantias et al. (2019) identificaram uma sequência de fases executadas durante os esforços de Transformação Digital, que são ilustradas na Figura 2. Vale destacar nesse fluxo de fases que o reconhecimento da necessidade por transformação deve partir da própria organização, sem a influência externa. A preparação do palco consiste na criação de estruturas de apoio para o processo de Transformação Digital e disseminação da mesma visão em todos os níveis da organização. A ETD é elaborada de forma recursiva, com planejamentos e passos formais a serem executados para que o estado final desejado seja alcançado. A implementação da ETD é contínua e passa por constantes melhorias, por isso, é vista como um processo recursivo que itera entre os três últimos passos do fluxo ilustrado.



Figura 2 – Fases de implementação da ETD.

Fonte: adaptada de Chantias et al. (2019).

2.3 Tecnologias Disruptivas

Ao longo dos últimos anos, um grande volume de tecnologias vem surgindo em diversas áreas (DELOITTE INSIGHTS, 2019). Como consequência, a variedade de aplicações de uma tecnologia para modernizar e transformar uma organização digitalmente é vasta.

Neste trabalho, considera-se que Tecnologia Disruptiva é aquela cuja inserção no mercado é capaz de alterar ou romper algum paradigma estabelecido dentro de organizações, seja em suas operações internas, seja quanto à produção ou entrega de seu produto (SCHUELKE-LEECH, 2018).

3 | DESENVOLVIMENTO

O desenvolvimento consiste na aplicação de duas técnicas: Mapeamento Científico e Revisão Sistemática de Literatura (RSL).

3.1 Mapeamento Científico

A pesquisa realizada seguiu o fluxo de trabalho proposto por Zupic e Čater (2015), dividido em cinco etapas: desenho da pesquisa, compilação dos dados bibliométricos, análise, visualização e interpretação.

- *Desenho da Pesquisa:* divide-se em duas atividades: (i) definir a pergunta de pesquisa e (ii) selecionar o método bibliométrico apropriado. No presente trabalho, adotamos os métodos de *co-word*, análise de citação e análise geográfica. O método de *co-word* foi selecionado para identificar as principais tecnologias disruptivas mencionadas na literatura; o método de análise geográfica foi selecionado para analisar o volume de publicações originárias de diferentes países, com a intenção de observar o nível de envolvimento do Brasil no tema; e o método de análise de citação foi utilizado como insumo para a RSL.
- *Compilação dos Dados Bibliométricos:* refere-se à seleção de uma base de pesquisa apropriada e à filtragem e exportação dos dados bibliométricos a partir dela. Neste trabalho, foi utilizada a base Scopus, por seu volume de artigos científicos revisados e por sua aceitação no meio acadêmico. Dessa base, obtêm-se os dados por uma busca utilizando-se a *string* definida na RSL, que, por sua vez, são filtrados e exportados. Esta filtragem se deu de forma seletiva, ao definir quais dados seriam úteis para a execução do Mapeamento Científico.
- *Análise:* essa etapa é composta pela seleção da ferramenta de estudo bibliométrico a ser utilizada de acordo com as necessidades do trabalho, pela limpeza dos dados coletados e pela identificação dos subgrupos. Essa etapa é responsável por gerar o conjunto final de dados a ser utilizado para criar as visualizações na etapa seguinte.
- *Visualização:* são gerados os gráficos que permitem abstrair os *insights* desejados dos dados coletados. De posse dos gráficos finais, os pesquisadores podem analisá-los e interpretá-los, tirando conclusões. Nesta etapa, foi utilizado o *software* Vos Viewer para a geração de gráficos.
- *Interpretação:* é o momento no qual as visualizações obtidas são analisadas de forma a gerar os *insights* e resultados da pesquisa.

3.2 Revisão Sistemática de Literatura

A RSL foi executada de acordo com as etapas exibidas na Figura 3.

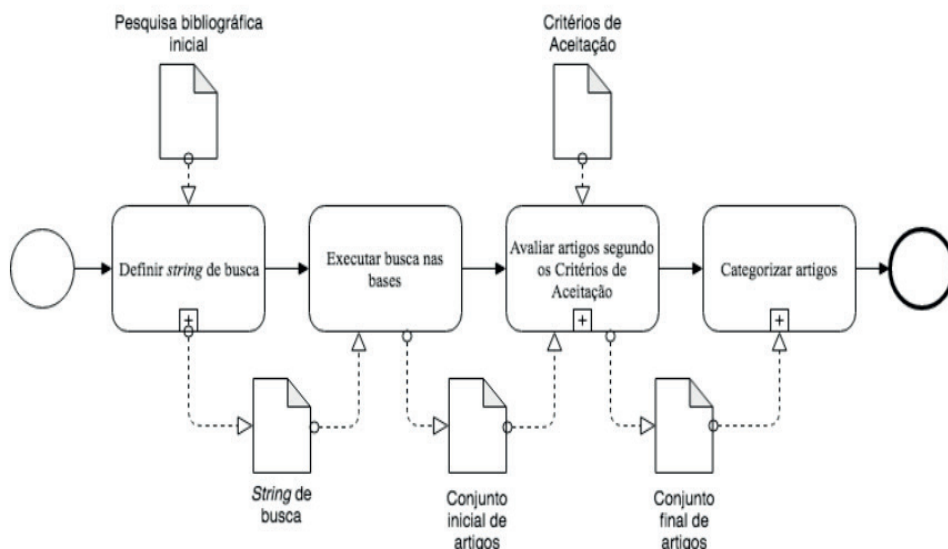


Figura 3 – Processo da Revisão Sistemática de Literatura

Fonte: os autores.

Definir string de busca: a *string* de busca utilizada foi definida a partir das informações advindas da pesquisa bibliográfica inicial e aprimorada por meio da realização de três rodadas de teste na base da Scopus. A *string* final adotada foi: (“*digital transformation*” OR *strategy* OR *digitization*) AND (“*disruptive technologies*”) AND (*application* OR *utilization* OR *use*). Para melhorar a relevância dos artigos retornados pela Scopus, foram utilizados os seguintes filtros na busca: (LIMIT-TO (SRCTYPE, “j”) OR LIMIT-TO (SRCTYPE, “p”) OR LIMIT-TO (SRCTYPE, “d”)) AND (LIMIT-TO (DOCTYPE, “ar”) OR LIMIT-TO (DOCTYPE, “cp”) OR LIMIT-TO (DOCTYPE, “re”)) AND (LIMIT-TO (LANGUAGE, “English”)) AND (LIMIT-TO (EXACTKEYWORD, “Disruptive Technology”)) AND (LIMIT-TO (SUBJAREA, “COMP”) OR LIMIT-TO (SUBJAREA, “ENGI”) OR LIMIT-TO (SUBJAREA, “BUSI”)).

- *Executar busca nas bases:* a *string* de busca é aplicada nas bases científicas selecionadas. A busca pela *string* definida retornou 157 artigos. Os artigos encontrados como resultado dessa atividade são avaliados na próxima etapa.
- *Avaliar artigos segundo os critérios de aceitação:* são definidos critérios de aceitação para avaliar os artigos encontrados. Artigos foram excluídos com base

nos seguintes critérios: (i) artigo não estar em inglês; (ii) texto completo estar indisponível; (iii) artigo não dar exemplos de *Tecnologias Disruptivas*; (iv) artigo não apresentar o contexto de Transformação Digital.

- *Categorizar artigos*: é o resultado do presente trabalho, que consiste em categorizar os artigos obtidos na etapa anterior.

4 | RESULTADOS

Após a definição da *string* de busca e sua aplicação na base Scopus, 157 artigos foram encontrados. A Figura 4 representa a divisão dos artigos retornados por autor. O autor com o maior número de contribuições é Zaidan. Esta informação pode ser útil para formar uma base de conhecimento sobre os autores mais envolvidos em determinada linha de pesquisa.

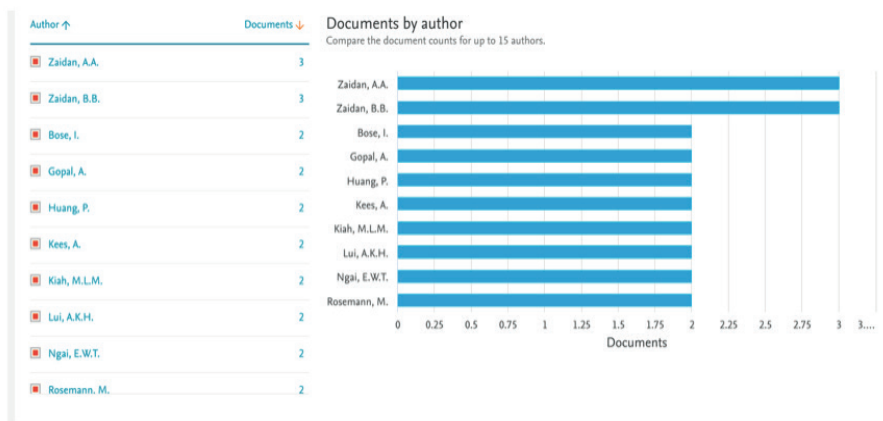


Figura 4 – Documentos por autor

Fonte: busca na base Scopus.

A Figura 5 representa a evolução da quantidade de publicações recuperadas na pesquisa ao longo do tempo. Podemos observar que o tema vem ganhando relevância com o passar dos anos, com um crescimento a partir de 2012. O declínio apresentado entre 2018 e 2019 não representa necessariamente uma queda no volume geral de publicações, mas é resultado do fato de grande parte dos artigos produzidos neste período ainda não terem sido avaliados ou publicados. Se for levada em consideração a tendência apresentada no gráfico, o tema de Tecnologias Disruptivas no contexto de Transformação Digital é um campo fértil para futuros trabalhos.

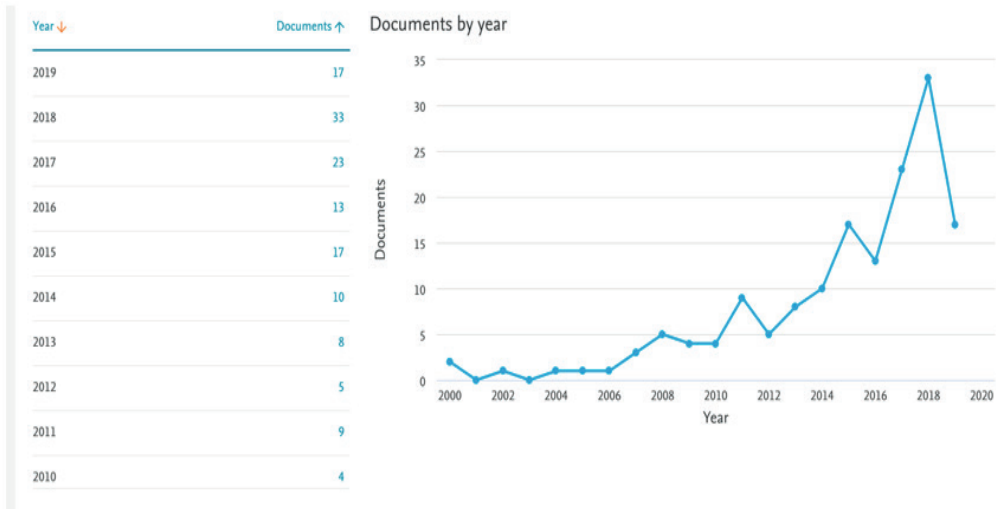


Figura 5 – Volume de publicações ao longo do tempo

Fonte: busca na base Scopus.

A Figura 6 representa o volume de publicações por país. Esta informação pode ser especialmente interessante em trabalhos nos quais o contexto tem papel importante na análise dos dados.

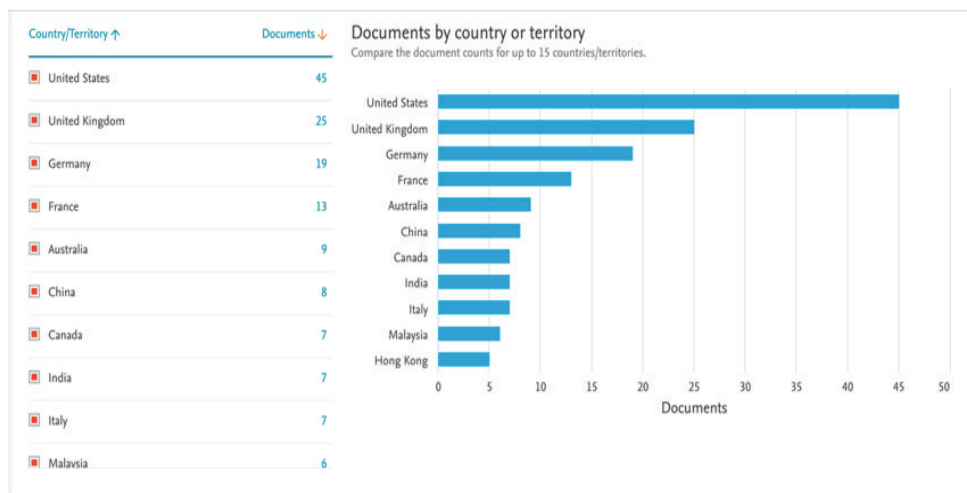


Figura 6 – Volume de publicações por país

Fonte: busca na base Scopus

Os dados obtidos pela busca na base Scopus foram analisados utilizando-se a ferramenta Vos Viewer. Para iniciar essa análise, apresentamos, na Figura 7, o volume de citações de cada autor, bem como o período de maior relevância do mesmo por meio da cor dos nós.

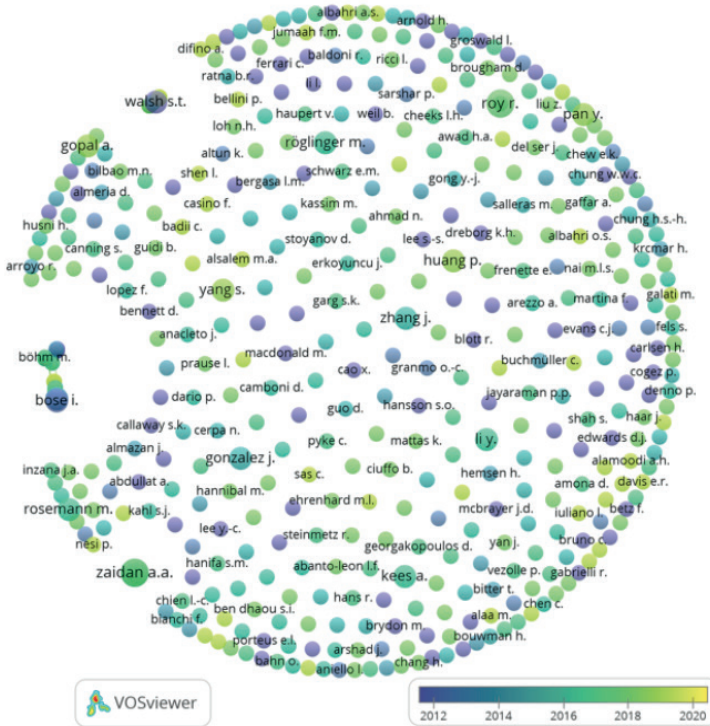


Figura 7 – Gráfico de citações por autor por ano

Fonte: os autores.

O gráfico da Figura 7 está em concordância com os dados vistos na base da Scopus, uma vez que o autor apresentado como de maior relevância (Zaidan) aparece em ambos com destaque. É importante verificar que, no gráfico gerado pelo Vos Viewer, outros nomes também ganham destaque, tais como Böhm M., Bose i. e Wzals S.T. Esta informação é importante porque, para avaliar a importância de um trabalho científico, a quantidade de citações é uma informação de peso. Frequentemente, um grande volume de citações indica que determinado artigo é bem-aceito em meio à comunidade acadêmica. Existem exceções, tais como em casos de citações negativas; porém, há técnicas que permitem evitar interferências causadas por este tipo de ocorrência, tais como a utilização de mais de um método bibliométrico para análise. As informações obtidas a partir deste gráfico servem

de insumo para a condução e avaliação de uma RSL e podem ser utilizadas para auxiliar na seleção de artigos de controle de forma sistemática, reduzindo o risco de os resultados serem distorcidos por viés.

A Figura 8 representa o gráfico de co-ocorrência de palavras entre os artigos (buscadas em seus metadados). Por meio dele, é possível observar como está estruturada a área de conhecimento em questão.

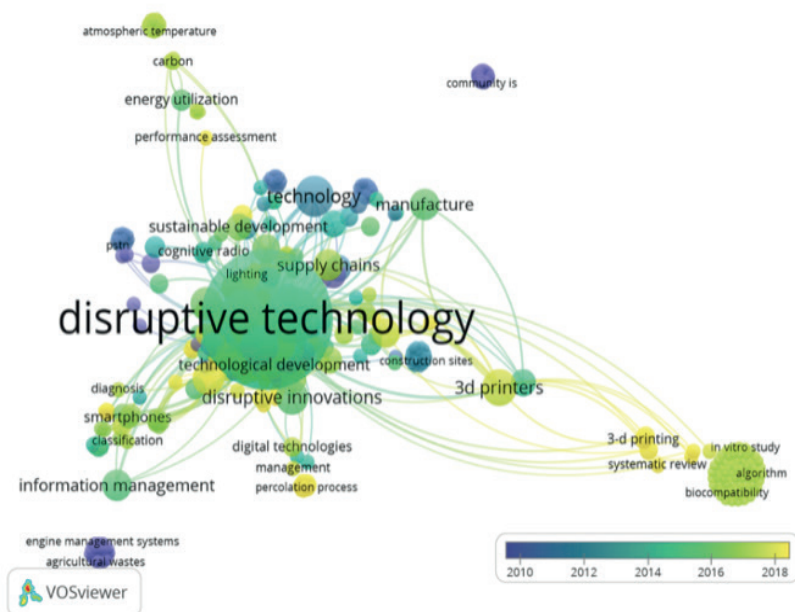


Figura 8 – Gráfico de co-ocorrência

Fonte: os autores.

No gráfico, fica evidente a recorrência de termos como “3d printers” e “manufacture”, o que indica que um conjunto dos artigos encontrados trata do uso de Tecnologias Disruptivas (por exemplo, a impressão 3d) no contexto de manufaturas. Também se pode esperar encontrar artigos sobre desenvolvimento sustentável e energia. A expressão “digital technologies” (tecnologias digitais) encontra-se fortemente ligada ao termo “management” (gestão), o que indica que grande parte dos artigos retornados que tratam de tecnologias disruptivas digitais foram escritos no contexto de gestão. Há ainda forte ligação entre termos como “diagnosis” e “smartphones”, o que pode indicar um tipo específico de utilização de tecnologia (por exemplo, artigos sobre aplicativos médicos).

A partir dos 157 artigos recuperados na Scopus, inicia-se a fase de RSL, começando-se por um processo de seleção que resultou em 18 artigos. Esse processo é ilustrado na

Figura 9, e os artigos selecionados são apresentados no Quadro 1.

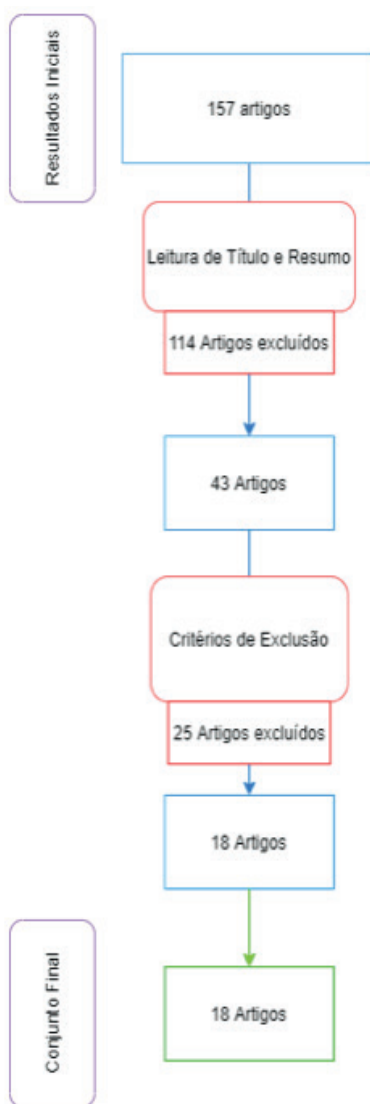


Figura 9 – Fluxo da Revisão Sistemática de Literatura

Fonte: os autores.

Autores	Tecnologia	Contexto
(CASINO <i>et al.</i> , 2019)	Blockchain	Gestão
(DAVIS <i>et al.</i> , 2019)	Impressão 3d	Tomada de decisão
(KOK <i>et al.</i> , 2018)	Impressão 3d	Física de materiais
(GARBUIO e LIN, 2019)	IA; IoT	Gestão
(MAESA <i>et al.</i> , 2018)	Blockchain	Acadêmico
(LO e CAMPOS, 2018)	IoT	Marketing
(SABERI <i>et al.</i> , 2019)	Blockchain	Gestão de suprimentos
(OBERLÄNDER <i>et al.</i> , 2018)	IoT	Gestão
(SHAH <i>et al.</i> , 2017)	Impressão 3d	Gestão de suprimentos
(BUSACHI <i>et al.</i> , 2016)	Impressão 3d	Construção naval
(CALIGNANO <i>et al.</i> , 2019)	Impressão 3d	Biomedicina
(WOODHEAD <i>et al.</i> , 2018)	IoT	Construção civil
(LUI <i>et al.</i> , 2016)	RFID	Gestão
(CALLAWAY e HAMILTON, 2008)	Internet Banking	Gestão
(ROY e ROY, 2019)	IoT; Big Data	Sistemas de gestão
(BADII <i>et al.</i> , 2019)	IoT; IoTE	Construção civil
(MA e MCGROARTY, 2017)	HFT	Finanças
(NIEUWENHUIS <i>et al.</i> , 2018)	Comp. em Nuvem	Gestão

Quadro 1 – Artigos selecionados para revisão

Fonte: os autores.

No Quadro 2, apresentamos uma breve descrição das tecnologias encontradas nos artigos selecionados para revisão.

Tecnologia	Descrição
Blockchain	protocolo que se baseia em três princípios: descentralização, consenso e criptografia. Utilizado para garantir a segurança e confiabilidade de transações.
Big Data	técnicas para extrair informações relevantes de grandes conjuntos de dados.
Comp. em Nuvem	sistemas computacionais e de armazenamento de dados disponíveis sob demanda sem a necessidade de instalação ou configuração por parte do utilizador.
HFT	<i>high-frequency trading</i> , ou negociações de alta frequência, são computadores de algoritmos capazes de executar várias transações por frações de segundo baseados em análises de condições do mercado.
IA	inteligência artificial, técnicas computacionais que procuram simular a inteligência humana.
IoT	<i>internet of everything</i> é o termo que se usa para descrever a situação, vivida atualmente, em que dispositivos de todos os tipos possuem alguma tecnologia acoplada (sensor, processamento, entre outros).

IoT	<i>internet of things</i> , ou internet das coisas, refere-se à capacidade de interconexão entre dispositivos cotidianos (celular, relógio, eletrodomésticos, etc).
Impressão 3D	é a tecnologia capaz de gerar protótipos tridimensionais rapidamente.
Internet banking	tecnologias utilizadas para transações bancárias pela internet.
RFID	<i>radio-frequency identification</i> , ou identificação por radiofrequência, é uma tecnologia de identificação por sinais de rádio, comunicando-se remotamente com etiquetas RFID.

Quadro 2 – Descrição das tecnologias encontradas nos artigos selecionados para revisão

Fonte: os autores.

No Quadro 3, apresenta-se uma classificação dos artigos resultantes da aplicação da RSL. Os artigos selecionados fornecem exemplos de utilizações de Tecnologias Disruptivas dentro do contexto de Transformação Digital.

Tecnologia	Contextos
Impressão 3d	Tomada de decisão; física de materiais; gestão de linha de suprimento; construção naval; biomedicina
IoT	Construção civil; sistemas de gestão; marketing; gestão
Blockchain	Gestão de linha de suprimento; gestão; estudos acadêmicos
IoE	Construção civil
Computação em Nuvem	Gestão
High Frequency Trading	Finanças
Big Data	Sistemas de gestão
IA	Gestão
RFID	Gestão

Quadro 3 – Classificação dos artigos da revisão

Fonte: os autores.

Entre os artigos selecionados, a tecnologia mais presente foi a Internet of Things (IoT), tema de 6 artigos. Logo depois, veio a impressão 3D, com 5 artigos, e por fim Blockchain, com 3 artigos. Outros temas, como RFID, IA, IoE e computação em nuvem, foram mencionados por apenas um artigo.

A partir dos artigos selecionados, é possível observar como as Tecnologias Disruptivas podem ser usadas para diferentes fins mesmo em um contexto específico como a Transformação Digital. O artigo de Busachi *et al.* (2016), por exemplo, trata a utilização de impressão 3D no contexto de construção naval pelo Ministério da Defesa da Inglaterra. Woodhead *et al.* (2018) trazem a aplicação de IoT na área de construção civil,

desenvolvendo-se sobre o tema de cidades inteligentes. Lo e Campos (2018) exemplificam o uso de IoT para estratégias de marketing empresarial.

Ao comparar os resultados da RSL com os obtidos pelo Mapeamento Científico, é possível, por exemplo, que as conclusões realizadas a partir do gráfico de co-ocorrência estavam corretas, uma vez que os padrões esperados foram de fato observados, tais como a relação entre Tecnologias Disruptivas (impressão 3d) e manufatura. Foi também observada a ligação entre *smartphones* e o contexto médico.

Entretanto, os nomes dos principais autores identificados no Mapeamento Científico não apareceram nos artigos selecionados. Isso pode indicar que os artigos escritos por esses autores não estão inseridos no escopo específico deste trabalho.

5 | CONCLUSÃO

O presente trabalho aplicou o mapeamento científico e a RSL para identificar de que formas as Tecnologias Disruptivas vêm sendo empregadas no contexto de Transformação Digital. Além disso, foram relacionados às Tecnologias Disruptivas encontradas seus contextos de aplicação e seu uso.

Trabalhos futuros podem explorar mais a fundo técnicas de RSL e mapeamento científico para expandir a cobertura de artigos encontrados e determinar ainda mais tecnologias que são temas atuais de pesquisa, abrindo outras possibilidades para pesquisas e para a aplicação dessas tecnologias em processos de Transformação Digital, além de estudar aplicações de tecnologias de acordo com a localização geográfica, e determinar tendências e formas de aplicação típicas de locais determinados.

REFERÊNCIAS

BADII, C. et al. Sii-Mobility: An IoT/IoE Architecture to Enhance Smart City Mobility and Transportation Services. **Sensors**, 19, 2019.

BUSACHI, A. et al. Defining Next-Generation Additive Manufacturing Applications for the Ministry of Defence (MoD). **Procedia CIRP**, 55, 2016. 302-307.

CALIGNANO, F. et al. Design of Additively Manufactured Structures for Biomedical Applications: A Review of the Additive Manufacturing Processes Applied to the Biomedical Sector. **Journal of Healthcare Engineering**, 2019.

CALLAWAY, S. K.; HAMILTON, R. D. Managing disruptive technology - internet banking ventures for traditional banks. **International Journal of Innovation and Technology Management**, 5, 2008. 55-80.

CASINO, F.; DASAKLIS, T. K.; PATSAKIS, C. A systematic literature review of blockchain-based applications: Current status, classification and open issues. **Telematics and Informatics**, 36, 2019. 55-81.

- CHANIAS, S.; MYERS, M. D.; HESS, T. Digital transformation strategy making in pre-digital organizations: The case of a financial services provider. **The Journal of Strategic Information Systems**, 28, 2019. 17-33.
- DAVIS, E. R. et al. An Analytics Framework for Structuring 3D Printing Deployment Decisions. **2019 Annual Reliability and Maintainability Symposium (RAMS)**, 2019. 1-6.
- DELOITTE INSIGHTS. **Digital reality in government: How AR and VR can enhance**. Londres. 2018.
- DELOITTE INSIGHTS. **Leading the social enterprise: Reinvent with a human focus**. Londres. 2019.
- FIGUEIREDO, R. M. C. et al. **Governo digital brasileiro: relatório técnico**. Universidade de Brasília. Brasília. 2019.
- GARBUIO, M.; LIN, N. Artificial Intelligence as a Growth Engine for Health Care Startups: Emerging Business Models. **California Management Review**, 61, 2019. 59-83.
- GARTNER. **Government portals are evolving to enable digital government**. Stanford. 2015.
- KOK, Y. et al. Anisotropy and heterogeneity of microstructure and mechanical properties in metal additive manufacturing: A critical review. **Materials & Design**, 139, 2018. 565-586.
- LO, F.-Y.; CAMPOS, N. Blending Internet-of-Things (IoT) solutions into relationship marketing strategies. **Technological Forecasting and Social Change**, 137, 2018. 10-18.
- LUI, A. K. H.; NGAI, E. W. T.; LO, C. K. Y. Disruptive information technology innovations and the cost of equity capital: The moderating effect of CEO incentives and institutional pressures. **Information & Management**, 53, 2016. 345-354.
- MA, T.; MCGROARTY, F. Social Machines: how recent technological advances have aided financialisation. **Journal of Information Technology**, 32, 2017. 234-250.
- MAESA, D. D. F. et al. BITKER: A P2P Kernel Client for Bitcoin. **2018 International Conference on High Performance Computing & Simulation (HPCS)**, 2018. 130-137.
- NIEUWENHUIS, L. J. M.; EHRENHARD, M. L.; PRAUSE, L. The shift to Cloud Computing: The impact of disruptive technology on the enterprise software business ecosystem. **Technological Forecasting and Social Change**, 129, 2018. 308-313.
- OBERLÄNDER, A. M. et al. Conceptualizing business-to-thing interactions – A sociomaterial perspective on the Internet of Things. **European Journal of Information Systems**, 27, 2018. 486-502.
- OESTERREICH, T. D.; TEUTEBERG, F. Understanding the implications of digitisation and automation in the context of Industry 4.0: A triangulation approach and elements of a research agenda for the construction industry. **Computers in Industry**, 83, 2016. 121-139.
- ROY, M.; ROY, A. Nexus of Internet of Things (IoT) and Big Data: Roadmap for Smart Management Systems (SMgS). **IEEE Engineering Management Review**, 47, 2019. 53-65.

SABERI, S. et al. Blockchain technology and its relationships to sustainable supply chain management. **International Journal of Production Research**, 57, 2019. 2117-2135.

SAP. SAP Study Reveals Four Key Traits of a Digital Transformation Leader, 2017. Disponível em: <<https://news.sap.com/2017/07/sap-study-reveals-four-key-traits-digital-transformation-leader/>>. Acesso em: 19 jun. 2020.

SCHUELKE-LEECH, B. A. A model for understanding the orders of magnitude of disruptive technologies. **Technological Forecasting and Social Change**, 129, 2018. 261-274.

SHAH, S. et al. Contribution of Additive Manufacturing Systems to Supply Chain. **2017 International Conference on Industrial Engineering, Management Science and Application (ICIMSA)**, 2017. 1-5.

VIAL, G. Understanding digital transformation: A review and a research agenda. **The Journal of Strategic Information Systems**, 28, 2019. 118-144.

VOLLENBROEK-HUTTEN, M. et al. **From Telemedicine Technology to Telemedicine Services**. Proceedings of the 3rd 2015 Workshop on ICTs for Improving Patients Rehabilitation Research Techniques. Lisboa, Portugal: ACM. 2015. p. 113–117.

VON LEIPZIG, T. et al. Initialising customer-orientated digital transformation in enterprises. **Procedia Manufacturing**, 8, 2017. 517-524.

WOODHEAD, R.; STEPHENSON, P.; MORREY, D. Digital construction: From point solutions to IoT ecosystem. **Automation in Construction**, 93, 2018. 35-46.

ZHUKOV, D. et al. Managing social networks: Applying the percolation theory methodology to understand individuals' attitudes and moods. **Technological Forecasting and Social Change**, 129, 2018. 297-307.

ZUPIC, I.; ČATER, T. Bibliometric Methods in Management and Organization. **Organizational Research Methods**, 18, 2015. 429-472.

UM SISTEMA ESTATÍSTICO PARA APOIO AO ACOMPANHAMENTO DE DESEMPENHO ACADÊMICO

Data de aceite: 27/08/2020

Guilherme Álvaro Rodrigues Maia Esmeraldo

Instituto Federal de Educação, Ciência e Tecnologia do Ceará
Crato – Ceará
<http://lattes.cnpq.br/7082611091690747>

Francisco Wilcley Lacerda de Lima

Instituto Federal de Educação, Ciência e Tecnologia do Ceará
Crato – Ceará
<http://lattes.cnpq.br/0161365616738465>

Rennan Rodrigues Isídio Teles

Instituto Federal de Educação, Ciência e Tecnologia do Ceará
Crato – Ceará
<http://lattes.cnpq.br/8502809472144286>

Francisca Alves de Souza

Instituto Federal de Educação, Ciência e Tecnologia do Ceará
Crato – Ceará
<http://lattes.cnpq.br/9410013773982707>

Cícero Carlos Felix de Oliveira

Instituto Federal de Educação, Ciência e Tecnologia do Ceará
Crato – Ceará
<http://lattes.cnpq.br/0479940625032580>

RESUMO: Uma das atribuições do setor pedagógico no ensino médio é o acompanhamento do desempenho acadêmico dos discentes a cada bimestre. Para tanto, se faz necessário agilidade e clareza na análise dos resultados emitidos pelo corpo docente

em cada disciplina, para que seja possível, quando necessário, intervir com instrumentos eficazes de recuperação da aprendizagem dos alunos na disciplina que não houve o êxito desejado. Portanto, esse artigo descreve os elementos utilizados na construção de um sistema computacional, sem custo financeiro para a instituição e de fácil compreensão para o usuário, que utiliza os conhecimentos estatísticos para realizar a descrição, a apresentação e análise dos dados coletados. **PALAVRAS-CHAVE:** Acompanhamento Pedagógico. Ferramenta computacional. Sistema Web. Análise Estatística.

A STATISTICAL SYSTEM TO SUPPORT THE MONITORING OF ACADEMIC PERFORMANCE

ABSTRACT: One of the attributions of the pedagogical sector in high school is the monitoring of the academic performance of students every two months. Therefore, it is necessary to have agility and clarity in the analysis of the results emitted by the faculty in each discipline, so that it is possible, when necessary, to intervene with effective instruments to recover students' learning in the discipline that did not have the desired success. Therefore, this paper describes the elements used to build a computer system, with no financial cost for the institution and easy to understand for the users, who uses the statistical knowledge to perform the description, presentation and analysis of the collected data.

KEYWORDS: Pedagogical Accompaniment. Computational tool. Web System. Statistical Analysis.

1 | INTRODUÇÃO

O acompanhamento do desempenho acadêmico, na grande maioria das instituições de ensino brasileiras, vem sendo realizado através de avaliações quantitativas por disciplina, e seus resultados têm sido utilizados como parâmetro base para o cálculo do rendimento escolar. No ensino médio, por exemplo, essa avaliação ocorre a cada bimestre, de forma que pode-se obter parâmetros periódicos e parciais de rendimento, os quais podem ser utilizados como referencial para que o estudante possa, em tempo, estabelecer ações para recuperar, manter ou aumentar seu desempenho no decorrer de um ano.

As instituições de ensino frequentemente têm buscado otimizar os processos de ensino-aprendizagem visando aumentar os desempenhos de seus estudantes, contudo muitos são os fatores que colaboram para a retenção escolar, como a Hiperatividade, Transtorno do Deficit de Atenção, Transtornos do Espectro Autista, Dislexia, *Bullying*, problemas domésticos, renda familiar, tempo de estudo, formação dos pais e índice de violência (FEITOSA; ESMERALDO; DE OLIVEIRA, 2013) (AMÂNCIO-VIEIRA et al., 2015). Avaliar cada um desses aspectos torna-se uma tarefa que pode demandar tempo, recursos humanos e financeiros excessivos, muitas vezes indisponíveis para os setores pedagógicos das instituições de ensino, e o resultado pode não ser suficientemente eficiente ao ponto de se evitar uma retenção (FEITOSA; ESMERALDO; DE OLIVEIRA, 2013). Nesse sentido, a análise quantitativa torna-se um instrumento essencial e factível de resultados no acompanhamento de desempenho acadêmico.

A análise estatística é uma importante ferramenta para a pesquisa em diversos campos do saber, como na economia, engenharias, fisiologia, etc. (CASTRUCCI, 2005). A estatística consiste de um conjunto de métodos quantitativos que tratam de elementos de pesquisa e inclui uma série de etapas de análise, como coleta, exploração, descrição e interpretação de dados numéricos, para, ao final, suportar a tomada de decisões (SILVESTRE, 2007) (CARVALHO e CAMPOS, 2008). Dependendo das técnicas estatísticas a serem aplicadas e dos tamanhos das bases de dados a serem analisadas, serão necessários cálculos cada vez mais complexos para se completar cada uma das fases do método estatístico.

Na literatura científica, há diversos tipos de ferramentas estatísticas, onde sua aplicabilidade e opção são condicionados a uma série de fatores, como: custo, procedimento de instalação, necessidade de treinamento, pacotes complementares, modelo de interação, suporte ao usuário, geração de gráficos atraentes, entre outros (OZGUR; KLECKNER; LI, 2015). No entanto, mesmo com as facilidades tecnológicas que elas oferecem, deve-se considerar que, para sua utilização efetiva, o usuário necessita ter conhecimentos

de estatística e, em muitos casos, de lógica de programação. Esses fatores podem tornar impraticável o uso dessas ferramentas por usuários leigos (PRADHANANGA, Y.; KARANDE, S; KARANDE, 2016).

Diante desse contexto, o presente trabalho visa contribuir no auxílio à análise do desempenho acadêmico dos estudantes do Instituto Federal do Ceará (IFCE) *campus* Crato, através de uma nova solução de análises estatísticas. Objetiva-se que a ferramenta proposta possa ser utilizada pelo setor pedagógico para descrição, apresentação e análise de dados quantitativos escolares, como, por exemplo, notas, número de faltas e número de estudantes evadidos, de forma a estabelecer ações preventivas de evasão e/ou retenção.

2 | REVISÃO TEÓRICA

Na literatura científica, é possível encontrar variados trabalhos que utilizam métodos estatísticos para análise e diagnóstico de diferentes indicadores educacionais.

Um dos exemplos é o trabalho em (FEITOSA; ESMERALDO; DE OLIVEIRA, 2013), o qual apresenta uma abordagem de uso de modelos de regressão múltipla, que considera as notas obtidas nas avaliações bimestrais, para estimar o desempenho acadêmico de estudantes do ensino médio. A abordagem proposta objetiva oferecer um instrumento para apoio à tomada de decisões para a redução da retenção escolar.

Assim como no trabalho anterior, a pesquisa em (AMÂNCIO-VIEIRA et al., 2015) utiliza modelos de regressão múltipla para caracterizar o rendimento escolar. No entanto, utilizou-se uma gama maior de variáveis candidatas ao modelo, distribuídas entre três categorias. São elas: 1) variáveis técnicas, que envolveram, por exemplo, número de alunos, número de professores graduados e pós-graduados e relação aluno-professor; 2) variáveis de custo pedagógico, sociais e administrativos; e 3) variáveis de desempenho acadêmico, como nota no Ideb e Brasil/Saeb. Os resultados comprovaram que a atuação docente e os fatores orçamentários possuem grande impacto nos resultados acadêmicos.

Já o trabalho em (DA SILVA, 2013) realiza modelagem estatística com o objetivo de identificar as principais variáveis que aumentam a correlação com a evasão em comparação com a média dos alunos do ensino superior da instituição. O estudo mostrou que a instituição analisada pode utilizar os resultados obtidos para tomar medidas administrativas para sanar dificuldades pedagógicas/financeiras e com isso reduzir os índices de evasão.

Por fim, o estudo em (CAPOVILLA; DIAS, 2007) buscou identificar o desenvolvimento das estratégias de leitura logográfica, alfabética e ortográfica em crianças da 1a. a 4a. Séries do ensino fundamental. A abordagem buscou correlacionar, através de cálculos do coeficiente de Pearson, o desempenho do uso dessas técnicas de leitura às notas escolares obtidas pelos estudantes, a fim de delinear medidas de intervenção para tratamento ou prevenção de dificuldades de leitura.

Nesse cerne, percebe-se que a estatística é um instrumento muito importante para compreensão dos diferentes contextos no mundo acadêmico e que pode ser utilizada para estabelecer indicadores para otimizar os resultados e aumentar a qualidade dos processos educacionais. O trabalho proposto neste artigo consiste de uma nova ferramenta que busca simplificar a realização de análises estatísticas, potencializando sua utilização na caracterização de indicadores para apoio pedagógico.

3 | MATERIAIS E MÉTODOS

Na metodologia adotada para o desenvolvimento da solução aqui proposta, inicialmente, realizou-se um levantamento bibliográfico sobre as tecnologias de desenvolvimento de sistemas Web, seguido por discussões em equipe para elencar os requisitos iniciais de interface, arquitetura de software e principais funções.

Em seguida, após a codificação das funções em módulos e da interface do sistema proposto, realizou-se novo estudo visando aumentar o desempenho da camada de análises estatísticas e de armazenamento de dados. Do estudo, foram adicionadas novas bibliotecas de software para processamento de alto desempenho, além da substituição do sistema de gerenciamento de bancos de dados por outro com abordagem NoSQL, a qual tem se mostrado mais eficiente em relação à relacional, quando se trata de processamento paralelo distribuído (MONIRUZZAMAN; HOSSAIN, 2013) e de bancos de dados maiores (ALMEIDA, BERNARDINO; FURTADO, 2015). Ressalta-se que, nesta etapa, realizou-se diversos experimentos, que incluíram o uso de bases de dados com diferentes tamanhos e diferentes cálculos estatísticos. Além disso, a interface do sistema proposto sofreu diversos refinamentos para suportar os princípios de usabilidade, objetividade, amigabilidade, acessibilidade e responsividade (CYBIS; BETIOL; FAUST, 2015). A arquitetura de software resultante do sistema proposto pode ser vista na Figura 1.

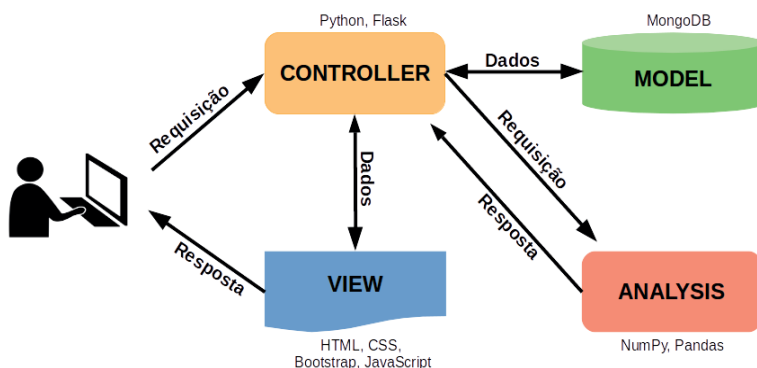


Figura 1 - Diagrama de arquitetura de software.

Fonte: Elaborado pelo autor (2018).

A arquitetura de software proposta, ilustrada na Figura 1, é baseada no padrão MVC (*Model View Controller*) (GAMMA et al., 2015), e está dividida em quatro camadas: 1) **Model**: envolve o gerenciamento das bases de dados que serão analisadas e seu armazenamento em banco de dados; 2) **View**: disponibiliza uma interface para apresentação e interação com o sistema; 3) **Controller**: inclui a lógica de negócio do sistema e permite a comunicação entre as demais camadas; e 4) **Analysis**: inclui as rotinas para análises de alto desempenho e geração de gráficos estatísticos.

Para a codificação da arquitetura proposta, em cada camada utilizou-se as seguintes tecnologias: *Model*) sistema de gerenciamento de bancos de dados orientado a documentos MongoDB; *View*) HTML5, CSS, JavaScript e o framework Bootstrap; *Controller*) linguagem de programação Python e framework Web Flask; e *Analysis*) Pandas e Numpy.

As subseções a seguir trazem descrições de cada uma dessas tecnologias, de acordo com a respectiva camada.

3.1 Model

O MongoDB é um banco de dados escalável e flexível com abordagem NoSQL (Not Only SQL). A ferramenta armazena dados em documentos semelhantes a JSON, o que significa que os campos podem variar de documento para documento e a estrutura de dados pode ser alterada ao longo do tempo. O modelo de documento é mapeado para os objetos no código, facilitando a manipulação dos dados. Consultas *ad hoc*, indexação e agregação em tempo real fornecem formas poderosas de acessar e analisar seus dados. É um banco de dados distribuído em seu núcleo, de modo que a alta disponibilidade, o dimensionamento horizontal e a distribuição geográfica são integrados e fáceis de usar. Além do mais, é gratuito e de código aberto, publicado sob a licença GNU Affero General Public License.

3.2 View

HTML5 é a versão mais recente do HTML. Ela introduz tags simplificadas, nova semântica e elementos de mídia, e se baseia em um conjunto de bibliotecas JavaScript que permite aplicativos web. A nova versão introduz elementos que adicionam nova semântica a suas páginas, trazendo ao desenvolvedor mais opções para a criação da estrutura da página Web.

CSS, que é uma abreviação para o termo em inglês *Cascading Style Sheet*, ou traduzido para o português como “folhas de estilo em cascata”, consiste em “(..) um mecanismo simples para adicionar estilos (por exemplo: fontes, cores, espaçamentos) aos documentos web.” (SILVA, 2012. p.24).

JavaScript é a linguagem de programação dinâmica da Web. A ampla maioria

dos sites modernos utiliza JavaScript e todos os navegadores modernos, presentes em computadores de mesa, consoles de jogos, tablets e smartphones, incluem interpretadores JavaScript, tornando-a a linguagem de programação mais onipresente da história. JavaScript faz parte da tríade de tecnologias que todos os desenvolvedores Web devem conhecer: HTML, para especificar o conteúdo de páginas Web; CSS para especificar a apresentação dessas páginas e JavaScript, para especificar o comportamento delas.

Bootstrap¹ é um kit de ferramentas de código aberto para desenvolvimento com HTML, CSS e JS. Através dele é possível criar de maneira rápida interfaces responsivas, amigáveis, acessíveis e objetivas. O framework utiliza um sistema de “grids responsivos” que permitem que página se adapte ao tamanho da resolução do equipamento utilizado no acesso. O Bootstrap está licenciado sob a licença MIT.

3.3 Controller

Python é uma linguagem de altíssimo nível orientada a objetos, de tipagem dinâmica e forte, interpretada e interativa. O Python tem uma sintaxe clara e concisa que favorece a legibilidade do código-fonte, tornando a linguagem mais produtiva. Mantida pela *Python Software Foundation* (PSF), é um software de código aberto (com licença compatível com a *General Public License - GPL*, porém menos restritiva, permitindo que o Python seja inclusive incorporado em produtos proprietários). (BORGES, 2014. p.13)

Flask² é um micro-framework (*framework* minimalista) desenvolvido em Python e baseado nas tecnologias:

- **Werkzeug**: consiste de uma biblioteca para desenvolvimento de aplicações WSGI (*Web Server Gateway Interface*), ou, em outras palavras, inclui uma especificação universal de como deve ser a interface (ou modelo de interação) entre uma aplicação desenvolvida em Python e um servidor Web. Para tanto, Werkzeug define um padrão para interceptar requisições Web, e lidar com as respectivas respostas, além de suportar o controle de *cache*, *cookies*, roteamento de *urls* e também conta com uma poderosa ferramenta de depuração de código.
- **Jinja2**: É um *template engine* escrito em Python, onde, através dele, é possível gerar dinamicamente (renderizar) páginas Web estáticas a partir de marcações em um template.

Flask é mantido por uma comunidade bastante ativa, seu código é baseado em diferentes padrões de projeto de software, possuindo alta qualidade e legibilidade, e permite que desenvolvedores tenham liberdade em estruturar, de formas diferentes, novos

¹ BOOTSTRAP. Disponível em: <<https://getbootstrap.com/>>. Acesso em: 28 mai. 2018.

² FLASK. Disponível em: <<http://flask.pocoo.org/>>. Acesso em: 28 mai. 2018.

aplicativos em desenvolvimento. Flask inclui ainda os *Blueprints* - mecanismos criados para centralizar o registro de novas operações providas de uma extensão -, os quais tornam possível o reaproveitamento simplificado de extensões predefinidas para a adição de novos recursos ao projeto.

3.4 Analysis

Pandas³ é uma biblioteca de código aberto e fornece estruturas de dados e ferramentas de análise de dados de alto desempenho para a linguagem de programação Python. O Pandas é um projeto promovido pela NumFOCUS (*Numerical Foundation for Open Code and Useable Science*), que é uma fundação sem fins lucrativos que busca promover diversas tecnologias de desenvolvimento de software para aplicações científicas.

NumPy⁴ é a biblioteca fundamental para computação científica com a linguagem de programação Python. Ela inclui uma estrutura de dados de array N-dimensional, a qual permite representar dados em memória em baixo nível e, com isso, manipulá-los de forma mais eficiente, ferramentas para integração de código com C/C++ e Fortran, funções sofisticadas para análise de dados, como, por exemplo, para álgebra linear e transformada de Fourier. Além de seu uso científico, o NumPy também pode ser usado como um contêiner multidimensional eficiente de dados genéricos. Desta forma, tipos de dados arbitrários podem ser facilmente convertidos, permitindo que o NumPy integre-se de forma simplificada e eficiente a uma ampla variedade de bancos de dados.

4 | RESULTADOS E DISCUSSÃO

O sistema Web proposto atualmente inclui os seguintes módulos funcionais:

- **Login/Cadastro:** é o subsistema responsável por gerenciar os usuários do sistema. Ele permite a criação e edição das informações do usuário, autenticação, autorização bem como acesso aos recursos do sistema proposto;
- **Dashboard:** consiste na área de trabalho do usuário. Este subsistema é o ponto central que orienta às atividades de cadastro de novas bases de dados, realização de análises estatísticas e visualização de resultados, através da geração de relatórios;
- **Análises Estatísticas:** módulo de suporte à edição das bases de dados cadastradas e à realização de análises estatísticas sobre elas. Neste módulo, pode-se ainda criar perfis de análises, onde o usuário do sistema pode selecionar determinados cálculos estatísticos, que poderão ser executados conjuntamente em instantes predeterminados;

3 PANDAS. Disponível em: <<https://pandas.pydata.org/>>. Acesso em: 28 mai. 2018.

4 NUMPY. Disponível em: <<http://www.numpy.org/>>. Acesso em: 28 mai. 2018.

- **Relatórios:** lista os relatórios gerados a partir de análises estatísticas previamente realizadas, permitindo visualizá-los e/ou exportá-los a qualquer momento.

A Figura 2 mostra o *dashboard* de um determinado usuário, onde é possível visualizar um total de 6 bases de dados cadastradas e a disponibilidade de relatórios de 10 análises previamente realizadas.

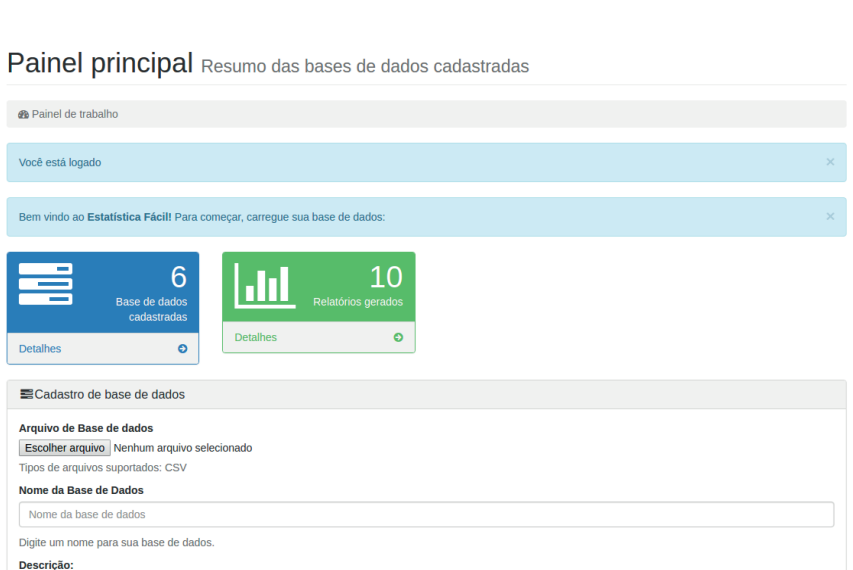


Figura 2. Área de trabalho de um determinado usuário.

Fonte: Elaborado pelo autor (2018).

Para realização de uma análise estatística, a partir do *dashboard*, são necessários apenas três passos básicos:

- 1. Cadastro de uma base de dados:** esta etapa consiste em selecionar um arquivo em formato CSV⁵, o qual inclui a base dados a ser analisada, definir um nome e uma descrição para a base e, por fim, realizar o upload dos dados para o sistema Web proposto. A Figura 2 mostra os controles para cadastro de uma nova base de dados;
- 2. Análise de dados:** após o cadastro, a base de dados fica disponível para manipulação e/ou análise estatística. A análise consiste em selecionar os tipos de cálculos e gráficos estatísticos que serão gerados a partir da base dados. Na Figura

⁵ CSV, ou *Comma Separated Values*, é um formato de arquivo de dados estruturados em tabela e que pode ser gerado pela maioria dos aplicativos modernos de planilha eletrônica.

3, pode-se observar parte da base de dados intitulada “notas de todos os alunos”, a qual inclui os registros de notas dos 4 bimestres letivos de vários alunos. Ainda na figura, verifica-se que foram selecionados os cálculos de média, mediana, moda, somatório, variância, desvio padrão, desvio absoluto, amplitude e a geração de histogramas.

notas de todos os alunos Selecione a análise estatística desejada:

Média
 Mediana
 Moda
 Somatório
 Variância
 Desvio Padrão
 Desvio Absoluto
 Amplitude
 Histograma
Gerar Análise

☰ Dados da Base de Dados

NOME	n1	n2	n3	n4	
ALUNO 0	6.5	1.8	3.0	6.0	<input type="checkbox"/>
ALUNO 1	2.5	0.0	0.0	0.0	<input type="checkbox"/>
ALUNO 2	5.0	5.5	7.0	6.0	<input type="checkbox"/>
ALUNO 3	5.0	6.0	3.0	3.0	<input type="checkbox"/>
ALUNO 4	7.0	6.5	6.0	7.0	<input type="checkbox"/>
ALUNO 5	0.0	0.0	0.0	0.0	<input type="checkbox"/>
ALUNO 6	0.0	0.0	0.0	0.0	<input type="checkbox"/>
ALUNO 7	4.0	0.0	0.0	0.0	<input type="checkbox"/>
ALUNO 8	0.0	0.0	0.0	0.0	<input type="checkbox"/>
ALUNO 9	6.5	6.0	4.5	5.8	<input type="checkbox"/>

Figura 3. Apresentação de uma base de dados cadastrada no sistema Web proposto.

Fonte: Elaborado pelo autor (2018).

3. Geração de Relatório: por fim, após uma análise de dados, o relatório torna-se disponível no *dashboard* e pode ser visualizado sempre que desejado. Na Figura 4, pode-se observar o relatório da análise estatística configurada na Figura 3. Nesse relatório, é possível observar os resultados dos cálculos solicitados para cada bimestre, bem como a presença dos gráficos histograma.

Resultado de análise: notas de todos os alunos

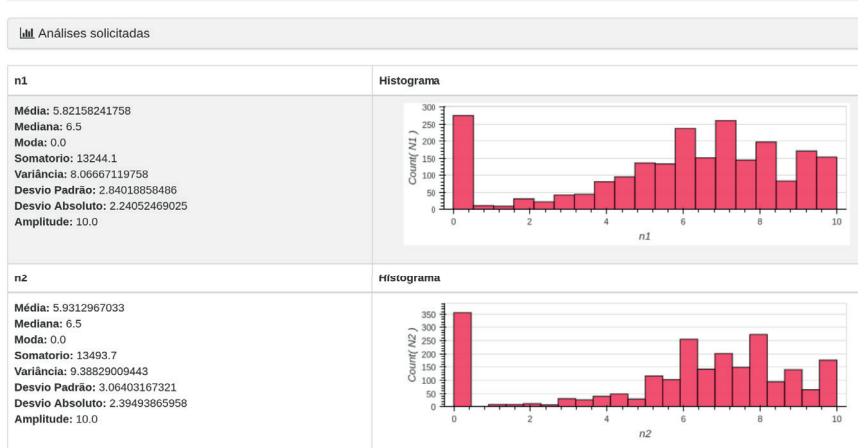


Figura 4. Exemplo de relatório de análise estatística.

Fonte: Elaborado pelo autor (2018).

Para avaliar o sistema Web proposto, elaborou-se um estudo de caso, onde analisou-se as notas da disciplina de Português dos estudantes do ensino médio do IFCE *campus* Crato, nos 3 primeiros bimestres letivos, com o objetivo de oferecer suporte ao Setor Pedagógico, no acompanhamento do desempenho individual, por turmas e por nível de escolaridade, e com isso, quando necessário, buscar mecanismos para reduzir os índices de retenção.

No estudo, calculou-se as estatísticas descritivas média, desvio padrão, nota máxima, nota mínima, mediana, 25° e 75° percentis, bem como plotou-se o histograma das notas em quatro cenários: 1 estudante, 31 estudantes de uma turma do 1° ano, 615 estudantes de todas as turmas do 1° ano e 2275 estudantes de todas as turmas dos 1°, 2° e 3° anos. A Figura 5 apresenta um gráfico de *Speedup* com os resultados de desempenho para o cálculo das estatísticas descritivas e geração de histogramas para cada um desses cenários.

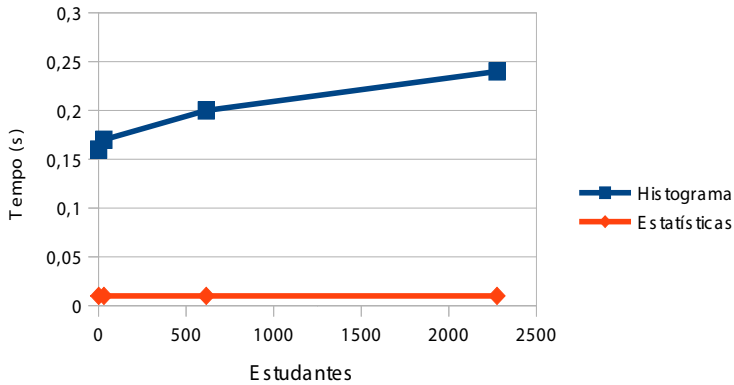


Figura 5. Gráficos de *Speedup* para cálculos e para geração de gráficos de estatísticos.

Fonte: Elaborado pelo autor (2018).

As curvas no gráfico na Figura 5 mostram que, tanto para o cálculo das estatísticas e quanto para geração dos histogramas, nos 4 cenários, os tempos ficaram abaixo de 0,25 segundos. Na figura, percebe-se ainda que os comportamentos das curvas tendem a ser lineares, mostrando assim um grande potencial de escalabilidade do tamanho das bases de dados que podem ser manipuladas pelo sistema de Web proposto.

5 | CONCLUSÕES

Esse trabalho apresentou uma proposta de sistema para apoio às análises estatísticas na Web. Em seu projeto, buscou-se aplicar um conjunto de técnicas de projetos de interfaces visando suportar usabilidade, objetividade, amigabilidade, acessibilidade e responsividade, bem como de computação de alto desempenho para permitir análises estatísticas mais eficientes.

Para avaliá-lo, utilizou-se um estudo de caso de análise de notas de alunos do ensino médio do IFCE *campus* Crato e os resultados mostraram que o sistema proposto dispõe de uma interface simplificada para suportar as tarefas de cadastro de bases de dados, análise estatística e geração de relatórios, bem como apresentou alto desempenho e escalabilidade.

A infraestrutura do sistema de análises proposto atualmente está em fase de testes. Em paralelo, o novo ciclo de desenvolvimento inclui adicionar o suporte de análises estatísticas e geração de gráficos mais complexos. Além disso, será necessária a inclusão de um módulo de agendamento de tarefas ao sistema proposto, uma vez que objetiva-se compartilhar seus serviços entre grupos e núcleos de pesquisa da região do Cariri.

Como trabalhos futuros, cogita-se a inclusão de Big Data e Inteligência Artificial,

visando a recomendação de tipos de análises estatísticas de acordo com diferentes perfis (estrutura da base de dados cadastrada, tipo de usuário, linha de pesquisa cadastrada, histórico de uso do sistema, entre outros.). Outros esforços concentram-se em otimizações de desempenho, como a distribuição de cálculos estatísticos entre diferentes nós computacionais e gerenciamento mais eficiente de consumo de memória.

REFERÊNCIAS

ALMEIDA, R.; BERNARDINO, J.; FURTADO, P. **Testing SQL and NoSQL approaches for big data warehouse systems**. International Journal of Business Process Integration and Management, 7(4), 2015. pp.322-334.

AMÂNCIO-VIEIRA, S. F.; BORINELLI, B.; DE NEGREIROS, L. F.; DALMAS, J. C. **A Relação entre custo direto e desempenho escolar: uma análise multivariada nas escolas de ensino fundamental de Londrina/Pr**. Educação em Revista, 31(1), 2015. pp.169-194.

BORGES, L. E. **Python para Desenvolvedores**. Novatec, São Paulo 2014.

CAPOVILLA, A. G. S.; DIAS, N. M. **Desenvolvimento de estratégias de leitura no ensino fundamental e correlação com nota escolar**. Psicol. rev. (Belo Horizonte), Belo Horizonte , v. 13, n. 2, p. 363-382, dez. 2007. Disponível em: <http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1677-11682007000200010&lng=pt&nrm=iso>. Acesso em: 21 maio 2018.

CASTRUCCI, P. L. **Modelos Computacionais para Gestão: Princípios e Aplicações**. Editora Manole, 2005.

CYBIS, W. A.; BETIOL, A. H.; FAUST, E. **Ergonomia e Usabilidade: Conhecimentos, Métodos e Aplicações**. Novatec, 2015.

DA SILVA, G. P. **Análise de evasão no ensino superior: uma proposta de diagnóstico de seus determinantes**. Avaliação: Revista da Avaliação da Educação Superior, 18(2), 2013.

ESMERALDO, G.; ESMERALDO, D.; ALVES, M.; FERRAZ, J. **Uma Solução de Datacenter para Suporte às Pesquisas na Região do Cariri**. In: Revista Conexões, 9(1), 2015.

FEITOSA, R. G. F.; ESMERALDO, G. A.; DE OLIVEIRA, C. C. F. **Estimativa de Desempenho Acadêmico a partir de Métodos Estatísticos Preditivos: Um Estudo de Caso com Alunos do 1º Ano do Ensino Médio do Instituto Federal do Ceará, campus Crato**. Revista Conexões - Ciência e Tecnologia, 7(1), 2013.

FLANAGAN, D. **JavaScript - O guia definitivo, 6ª edição**. Editora Bookman, 2013.

FREEMAN, E. **Use a cabeça!: programação em HTML 5: Desenvolvendo aplicativos para web com JavaScript**. Eric Freeman, Elisabeth Robson - Rio de Janeiro: Alta Books, 2014.

GAMMA, E.; HELM, R.; JOHNSON, R. E.; VLISSIDES, J. **Design Patterns: Elements of Reusable Object-oriented Software**. Pearson Education, 2015.

MONIRUZZAMAN, A. B. M.; HOSSAIN, S. A. **Nosql database: New era of databases for big data analytics-classification, characteristics and comparison.** International Journal of Database Theory and Application, Vol. 6, No. 4, 2013.

OZGUR, C.; KLECKNER, M.; LI, Y. **Selection of Statistical Software for Solving Big Data Problems: A Guide for Businesses, Students, and Universities.** *SAGE Open*, 5(2), 2015.

PRADHANANGA, Y.; KARANDE, S; KARANDE, C. **High performance analytics of bigdata with dynamic and optimized hadoop cluster.** In: IEEE International Conference on Advanced Communication Control and Computing Technologies, 2016. pp. 715-720.

SILVESTRE, A. L. **Análise de Dados e Estatística Descritiva.** Editora Escolar, 2007.

LUTZ, M. **Learning Python: Powerful Object-Oriented Programming.** 5. ed. United States of America: O'Reilly Media, Inc., 2013. 1600 p.

R Core Team. **R: A Language and Environment for Statistical Computing.** Vienna, Austria, 2017. Disponível em: <<https://www.R-project.org/>>.

SILVA, M. S. **CSS3: desenvolva aplicações web profissionais com uso dos poderosos recursos de estilização das CSS3.** Novatec Editora, São Paulo, 2012.

TAJRA, S. F. **Internet na educação: o professor na era digital.** 1. ed. São Paulo: Érica, 2002.

UMA DISCUSSÃO ACERCA DA INTERNET: DESAFIOS PARA CONFIABILIDADE DA INFORMAÇÃO

Data de aceite: 27/08/2020

Data de submissão: 03/06/2020

Breno Meirelles Costa Brito Passos

Universidade Federal do Sul da Bahia, IHAC
- Campus Paulo Freire
Teixeira de Freitas – Bahia
<http://lattes.cnpq.br/8260973788557348>

Eli Shuab Carvalho Lima

Universidade Federal do Sul da Bahia, IHAC
- Campus Paulo Freire
Teixeira de Freitas – Bahia
<http://lattes.cnpq.br/1219208893127917>
<https://orcid.org/0000-0001-6036-6010>

Bruno Soares Galdino

Universidade Federal do Sul da Bahia, IHAC
- Campus Paulo Freire
Teixeira de Freitas – Bahia
<http://lattes.cnpq.br/9534781388416069>

Livia Santos Lima Lemos

Universidade Federal do Sul da Bahia, IHAC
- Campus Paulo Freire
Teixeira de Freitas – Bahia
<http://lattes.cnpq.br/4155634257662178>

RESUMO: A internet possibilita agilidade demandada pela vida pós-moderna, através da difusão rápida e armazenamento de informações. Entretanto, um dos desafios do ciberespaço é a fidedignidade dos dados, ocasionados pela desinformação e *fake news*. Diante disso, o presente trabalho objetiva discutir acerca da confiabilidade das informações disseminadas na internet, para entender os riscos e a importância da

avaliação dos conteúdos encontrados no ambiente virtual. Aplicou-se, assim, uma pesquisa bibliográfica qualitativa e documental através da perspectiva de vários autores para esboçar aspectos da comunidade virtual, descentralização do ciberespaço, produção colaborativa da *Wikipédia*, *Fake News* e golpes na rede. Permitiu-se evidenciar alguns critérios para avaliar as informações, como os propostos por Tomaél *et al.* (2001), que expõe a relevância da averiguação da atualidade das fontes, avaliação da clareza do conteúdo e pesquisa em outros espaços virtuais para uma mesma informação. Deste modo, os usuários de internet não devem ser meros espectadores, mas sim agentes capazes de discernir com base em pesquisa criteriosa a veracidade dos dados que recebem e compartilham no ambiente virtual.

PALAVRAS-CHAVE: Ciberespaço, credibilidade, dados, informações falsas, *fake news*.

A DISCUSSION ABOUT INTERNET: CHALLENGES FOR CREDIBILITY OF INFORMATION

ABSTRACT: Internet enables the agility demanded by postmodern life, through the speed dissemination and storage of information. However, one of the challenges of cyberspace is the trustworthiness of data, caused by misinformation and fake news. Therefore, this paper aims to discuss the reliability of information disseminated on the internet, in order to understand the risks and the importance of evaluating the content

found in virtual environment. Thus, a qualitative and documentary bibliographic research was applied through the perspective of several authors to outline aspects of the virtual community, decentralization of cyberspace, collaborative production of Wikipedia, Fake News and scams on the network. It was possible to highlight some criteria to evaluate the information, such as those proposed by Tomaél *et al.* (2001), which exposes the relevance of checking the currentness of sources, evaluating the clarity of the content and researching other virtual spaces for the same information. Thus, internet users should not be mere spectators, but agents capable of discerning, based on careful research, the veracity of the data they receive and share in virtual environment.

KEYWORDS: Cyberspace, credibility, data, false information, fake news

1 | INTRODUÇÃO

A rede de telecomunicação *on-line* caracteriza por ser um processo evolutivo com impactos sociais, econômicos, políticos e institucionais (TOMAÉL *et al.*, 2001). Neste contexto, a internet representa uma verdadeira revolução na geração, recuperação, dispersão e armazenamento de conteúdos, seguindo os paradigmas estabelecidos pelo mundo pós-moderno, que demandam agilidade dos indivíduos que nele estão inseridos. Porém, a velocidade e quantidade exagerada dos dados no espaço virtual requer tempo, que muitos usuários não utilizam, para verificar a procedência e veracidade da informação (BRISOLA; BEZERRA, 2018).

Neste sentido, a capacidade para criar, recriar e replicar conteúdos no ambiente virtual mostra-se como um empecilho para maior segurança no que tange à confiabilidade. Algumas pessoas estão sujeitas a compartilhar e tomar por verdade o que se encontra *on-line*, sem verificação própria e prévia, e, em virtude do caráter descentralizado da internet, os dados que são tidos como verídicos são compartilhados, de modo tal que os receptores absorvam tudo quanto recebem (TOMAÉL *et al.*, 2001).

Diante da popularização da internet, objetivou-se discutir acerca da confiabilidade das informações difundidas no ciberespaço, através de pesquisa bibliográfica e documental, para entender a importância da avaliação dos conteúdos encontrados na rede e os riscos da desinformação e informação falsa. Aprofundando conhecimento sobre os conceitos que abarcam a internet e a credibilidade das notícias, investigando-os e examinando-os sob diferentes óticas de autores.

2 | A IDEIA DA COMUNIDADE VIRTUAL

Machado e Tijiboy (2005) expõe a discussão da construção virtual de uma comunidade, destacando a possibilidade da criação de diversos 'eus' - público e privado. A ideia de teia ou rede proposta por algumas plataformas presentes na internet é bastante significativa, em que há interligação dos indivíduos, tornando-os canais para difusão de conteúdos, informações e ações. Na rede, há a circulação simultânea de inúmeras

informações, o que a confere poder na construção de conhecimento de modo cooperativo. Em vista disso, o computador torna-se uma ferramenta social e de educação.

As relações sociais virtuais podem se dar de dois níveis, os quais abarcam os níveis de confiança humana: mútua - na qual a negociação das ações, palavras e atitudes são realizadas concomitantemente aos desenvolvimentos delas -, e relativa - em que há assuntos pré-programados. Essa relação forja valores tanto de comportamento quanto de interação virtual entre os usuários, fazendo com que determinadas informações tenham preferência em relação a outras, sendo elas encontradas e compartilhadas entre os interlocutores. Essa seleção silenciosa de informações faz com que o capital social seja rico em determinado tema e pobre em outro (RECUERO, 2007).

Para exemplificar a ideia de rede, há o Facebook, criado em 2004 por Mark Zuckerberg. Esta mídia social pode ser utilizada tanto na integração social quanto na formação de um ambiente de aprendizagem e divulgação de informações e conteúdos, em decorrência dos compartilhamentos livres, espaço para debate e exposição coletiva de conhecimento, o que proporciona conotação imediata de ensino, sendo ferramenta educacional inovadora (ANDRADE; AZEVEDO; DÉDA, 2012). Todavia, em virtude da liberdade que essa plataforma oferece, a questão de confiabilidade é fundamental, pois que muitas pessoas tomam por verdade tudo o que a eles chegam.

3 | DIFUSÃO DE FAKE NEWS NA INTERNET

O termo *fake news* (notícias falsas, em português), apesar de mais antigo, ganhou popularidade após uma coletiva de imprensa do 45º presidente dos Estados Unidos, Donald Trump. Embora o significado desse termo seja impreciso, pode-se adotar a designação de que são materiais disseminados virtualmente que, assemelhando-se a notícias verdadeiras, subvertem a verdade, abandonam a relação com a realidade e geram impactos na sociedade. Seis são as possíveis definições para o termo: “sátira, paródia, fabricação, manipulação, publicidade e relações públicas e propaganda” (CARVALHO, 2019).

As *fake news* são instrumentos empregados para disseminação de ideias e são capazes de interferir no debate político, encontrando nas redes sociais campo propício para sua propagação, como é o caso do aplicativo *Whatsapp*, podendo ter sido utilizado com essa finalidade nas eleições brasileiras de 2018, o que pode explicar o motivo da instalação da CPMI das *Fake News*, no Congresso Nacional, em 2019. A força delas pode advir da ideia de comunidade virtual, acima exposta, atrelada à velocidade de propagação das informações e ao amplo alcance delas. Além disso, a facilidade e a liberdade que os usuários possuem em criar, curtir, comentar e compartilhar os conteúdos no meio virtual, bem como o apelo emocional e o excesso deles, podem, igualmente, dar-lhes força, criando desafios a serem solucionados (BARBOSA; SANTI, 2019).

Segundo Brisola e Bezerra (2018) são dois os motivos para produção e disseminação das *fake news*: o primeiro trata da rentabilidade proporcionada pelas notícias falsas, enquanto o segundo é ideológico. Neste cenário, os algoritmos e *bots* - programas virtuais que possibilitem a realização de tarefas - podem funcionar como difusores de *fake news* na bolha informacional.

Diante de cenários de polarizações, algumas pessoas são motivadas, também, em disseminar informações falsas por acreditar em determinadas teses (BRISOLA; BEZERRA, 2018). Notícias falsas podem tanto trazer vantagens políticas e econômicas, bem como instaurar o medo e o caos entre os usuários das mídias digitais, principalmente na área da saúde (DE SOUSA JUNIOR *et al.*, 2020). Neste sentido, um exemplo dos impactos da *fake news* na saúde é o atual cenário da pandemia de COVID-19, no qual mensagens falsas acerca da SARS-COV-2 estão espalhando desinformação no ciberespaço afetando a vida individual e coletiva das pessoas. Neste contexto, a *fake news* ganha espaço nas redes sociais, pois conjuga a velocidade da disseminação, quantidade de informações e a ausência da verificação da credibilidade da notícia.

4 | CONSIDERAÇÕES SOBRE A PRODUÇÃO COLABORATIVA ATRAVÉS DA WIKIPÉDIA

A *Wikipédia* – desenvolvida por Jimmy Wales e Larry Sanger e lançada em 2001 - vem sendo muito utilizado no mundo inteiro para pesquisas e suas maiores inovações são a liberdade de construção e a cultura heterogênea (GOMES, 2006). Esse site, de disponibilidade *online*, possibilita a participação coletiva e colaborativa das pessoas na elaboração dos textos disponibilizados, de modo que não está limitado à autoria única, mas é passível de modificação contínua, o que gera incertezas quanto a sua confiabilidade.

A *Wikipédia* chama a atenção dos usuários pela sua análise particular e pessoal, dinamicidade, facilidade e comodidade, além de ser um trabalho colaborativo e redutor de tempo. Isso evidencia que o poder de busca não está senão pela facilidade de disponibilização da maior gama de resultados na maior velocidade possível.

De acordo com Vieira e Christofolletti (2013), um estudo publicado pela revista *Nature* em 2005 apontou que a *Wikipédia* é tão confiável quanto a tradicional *Enciclopédia Britânica* - enciclopédia de língua inglesa cujo público alvo é leitores adultos e cultos e sua publicação cabe à editora privada *Encyclopædia Britannica, Inc.* Isso permite inferir que as desconfianças que se tem quanto à utilização dessa ferramenta de busca não condizem com a realidade, mas que, pelo contrário, esse site apresenta nível de credibilidade e confiabilidade. Ademais, a *Enciclopédia Livre, Wikipédia*, se preocupa em impedir a vandalização de verbetes e artigos de sua plataforma *online*, através de normas, diretrizes e princípios que asseguram que os colaboradores sigam os propósitos das políticas de recomendações (BRAZ, SOUZA, 2014).

Os *Cinco Pilares* são uma das diretrizes estabelecidas pela *Wikipédia* para manter conteúdos livres de achismo, convivência amistosa entre os editores, garantir a essência da enciclopédia, permitir edições do conteúdo e alterações nas suas diretrizes, exceto, *Os Cinco Pilares*. Em casos de infrações que comprometam a confiabilidade das informações, existem penalidades que variam entre a limitação na edição de conteúdo ou, em situações mais graves, bloqueio do protocolo de internet (BRAZ, DE SOUZA, 2014).

Todas as diretrizes, normas, medidas e ferramentas adotadas pela *Wikipédia* para produção de conteúdo colaborativamente, não são adotadas em muitas outras empresas com plataformas *online*, a exemplo o *Facebook*, por possuir objetivos diferentes, o que compromete a qualidade da informação que nelas circulam. O propósito das redes sociais é aumentar a permanência de pessoas conectadas e descobrir as informações dos usuários para fins publicitários, enquanto que o jornalismo gera lucro irrisório e atritos que essas empresas buscam evitar (FRIAS FILHO, 2018).

5 | DESCENTRALIZAÇÃO DA INTERNET

Com a democratização da internet, circular por ela tornou-se um risco, já que está cada vez mais difícil saber discernir o que é de fato verdade, em virtude da facilidade com que os usuários difundem qualquer tipo de informação. Em 2016, um levantamento realizado pelo *Buzzfeed Brasil* - empresa de mídias sociais fundada em 2006 por Jonah Peretti em Nova Iorque nos Estados Unidos da América - mostrou que as interações e engajamentos por partes dos usuários sobre a Operação Lava-Jato nas 10 notícias falsas foram superiores às 10 notícias verdadeiras, no mesmo período. Isso demonstra a importância da adoção de postura mais crítica pelo receptor das informações.

É na internet que a relação entre os indivíduos e as fontes de informação se confronta, objetivando o estabelecimento de relações de confiança e credibilidade entre ambos. A credibilidade tem sido questionada pela tradição jornalística, em parte, com o uso de fontes oficiais. Conforme Traquina (2005), “as fontes são quem são porque estão diretamente ligadas a setores decisivos da atividade” humana, de modo que elas sejam importante instrumento de conexão entre setores de influência na sociedade e a sociedade em geral. Desse modo, a introdução de fontes duvidosas se caracteriza como problema.

A legitimidade que carece uma parte das fontes no ambiente virtual pode ser explicado pela descentralização do ciberespaço, o que gera a multiplicação das fontes sem tradição. Desse modo, a desarticulação do modelo clássico do jornalismo – causada pela arquitetura descentralizada do ciberespaço – faz com que o exercício da profissão exija novos critérios que garantam a confiabilidade do sistema de apuração no mundo digital. Informações de qualidade e com credibilidade surgem, lado a lado, a informações falsas - fato acentuado no ciberespaço - onde a informação deve ser tratada pelos jornalistas com os mesmos critérios aplicados às fontes tradicionais. Neste aspecto, a edição e a filtragem

da informação tornam-se ainda mais importantes na *web*, em que o usuário pode publicar o que quer que seja, dando-lhe aspecto substancial (BASTOS, 2000).

6 | A COMERCIALIZAÇÃO NA REDE

A rede tem mostrado eficácia na comercialização de bens. Miranda e Arruda (2004) colocam que as empresas modernas estão em um processo de encontrar respostas para questionamentos que dizem respeito às “expectativas, demandas, hábitos e comportamento de compra”, e, por isso, estudos que visem descrever o perfil do consumidor *online* são relevantes, a fim de fortalecer esse mercado. Porém, o ambiente virtual está sujeito à ocorrência de golpes e fraudes de venda e compra de produtos, que podem traumatizar as vítimas e trazer-lhes prejuízos, o que diminui a credibilidade do comércio virtual.

O surgimento, na década de 90, dos primeiros sites de comércio virtual em território nacional ocasionou desconfiança, mesmo entre os usuários (MARTINS; SILVEIRA, 2012). No Brasil, grande parte das lojas virtuais tem seu fechamento atrelado à ocorrência de fraudes que causam prejuízos, já que a compra virtual com cartão de crédito não tem verificações de irregularidades, de acordo com *Serasa Experian*, empresa nacional que visa analisar e informar assuntos relacionados aos créditos e negócios (E-COMMERCE BRASIL, 2015).

7 | VERIFICAÇÃO DE INFORMAÇÕES NA REDE

Diante da proliferação de informações falsas e golpes na rede de telecomunicação *online*, ocasionados, em parte, pela excessiva confiança dos usuários de internet, é preciso que eles desenvolvam atitudes de verificação das informações. Para isso, alguns requisitos são utilizados: i) as impressões do usuário quanto a plataforma; ii) a facilidade de navegação oferecida pela página; iii) a existência de imagens, sons e músicas; iv) a organização das informações; v) o nível de formalidade dos autores; e, vi) a clareza nas informações (NIELSEN, 1996; McLACHLAN, 1999, *apud* TOMAÉL et al., 2001). Ademais, há a análise de atualidade das informações, da autoridade da autoria e da instituição responsável pela publicação (TOMAÉL et al., 2001).

Flanagin e Metzger (2007, *apud* FERREIRA; GRADIM, 2015) propõe quatro possibilidades básicas de verificação da informação: a checagem de atualidade, a inteligibilidade do texto, o tipo de ideia que é transmitida pelo argumento do emissor da mensagem e realização de pesquisa em outras fontes. Com isso, expõe-se o papel ativo que os receptores de informações devem ter no cenário virtual, diferenciando-se da ideia de que eles sejam meros espectadores, fazendo-se necessária a atenção deles ao fato de que é preciso ter consciência e curiosidade, bem como buscar o desenvolvimento de habilidades e práticas que possibilitem análise de confiabilidade, creditando veracidade àquilo que recebem (LÓPEZ HIDALGO; FERNÁNDEZ BARRERO, 2012).

Concomitante a essas habilidades e práticas para avaliação de referência em plataformas *online*, os autores Tomaél *et al.* (2001) propõem critérios de confiabilidade para avaliar as fontes de informação na internet conforme Tabela 1. De acordo com os autores supracitados, os critérios não são inertes, pois a internet e os aparatos tecnológicos e informáticos são dinâmicos. Porém, assume-se que a adoção destes critérios permite reduzir as incertezas em relação às fontes nas redes digitais, pois assegura maior qualidade da informação.

Critérios	Objetivos
Identificação	Identificar e detalhar os dados do responsável pelo <i>site</i> .
Consistência	Avaliar a integridade e desenvolvimento das informações.
Confiabilidade	Avaliar a autoria das informações.
Ajustamento da fonte	Checar a linguagem usada e a concordância com os objetivos.
Links	Avaliar endereços de <i>sites</i> e outras fontes de informações.
Uso simples	Examinar a simplicidade para exploração do documento.
Lay-out	Analisar as mídias empregadas.
Limitações	Avaliar ocasiões que podem ser obstáculo para uso de fontes de informação.
Assistência	Acessar componentes de auxílio aos usuários.
Observações diversas	Checar os auxílios proporcionados aos deficientes e as possibilidades de acesso a outras línguas.

Tabela 1. Critérios de credibilidade das fontes de informação no ciberespaço.

Fonte: elaborado pelos autores com base em Tomaél *et al.* (2001)

A adesão dos critérios de avaliação pelos usuários de internet é uma tarefa custosa. Para tornar as pessoas aptas em aplicar estratégias e técnicas no tratamento da informação, é necessário que, através da educação, sejam desenvolvidas competências, habilidades e valores para lidar com o fenômeno informacional (BRISOLA; BEZERRA, 2018). É preciso educar e sensibilizar as pessoas para responsabilidade e os perigos existentes na *web*. A alfabetização digital permite escapar com maior segurança da desinformação e *fake news* na internet (BRANCO, 2017).

8 | CONCLUSÃO

A internet é um ambiente propício para a construção e o compartilhamento de conhecimentos e informações em virtude da sua velocidade e flexibilidade. No entanto, a cautela e atenção para com aquilo que se consome na rede devem ser continuamente presentes e os usuários devem estar sempre de sobreaviso quanto à possibilidade de

informações irreais. Nota-se que uma parte significativa do que se está disponível no meio virtual pode levar o leitor a enganar-se advindos de *fake news*.

Um dos fatores que influenciam para a circulação de desinformações na internet é a sua própria natureza, pois esta permite que qualquer indivíduo, sem restrição e/ou correção, exponha o que quer que seja. Portanto, cabe ao usuário desenvolver práticas e mecanismos que o façam capaz de discernir as informações que circulam na internet.

REFERÊNCIAS

ANDRADE, Polyana Bittencourt; AZEVEDO, Denio Santos; DÉDA, Talita de Azevedo. Práticas de Ensino e Redes Sociais na Internet: um estudo de caso do Facebook como ambiente de aprendizagem. **Simpósio Educação e Comunicação-infoinclusão: possibilidades de ensinar e aprender**, v. 3, p. 301-316, 2012.

ARAGÃO, Alexandre. **Notícias falsas da Lava Jato foram mais compartilhadas que verdadeiras**. Disponível em: <<https://www.buzzfeed.com/br/alexandrearagao/noticias-falsas-lava-jato-facebook>>. Acesso em: 22 de maio de 2020.

BASTOS, Helder. Jornalismo electrónico. Internet e Reconfiguração de Práticas nas Redações, Coimbra. **Minerva**, 2000.

BARBOSA, Maria Luciene Sampaio; SANTI, Vilso Junior. A INTENCIONALIDADE NAS NOTÍCIAS FALSAS: A NOTA DE REPÚDIO COMO ESTRATÉGIA DE DEFESA DO JORNALISMO NA ERA DAS FAKES NEWS. **Aturá-Revista Pan-Amazônica de Comunicação**, v. 3, n. 3, p. 93-109, 2019.

BRANCO, Sergio. **Fake News e os caminhos para fora da bolha**. Rio de Janeiro, 2017. Disponível em: <<https://itsrio.org/wp-content/uploads/2017/08/sergiofakenews.pdf/>>. Acesso em 29 de maio de 2020.

BRANSKI, Regina Meyer. Localização de informações na internet: características e formas de funcionamento dos mecanismos de busca. **Transinformação**, v. 12, n. 1, p. 11-19, 2000.

BRISOLA, Anna; BEZERRA, Arthur Coelho. DESINFORMAÇÃO E CIRCULAÇÃO DE “FAKE NEWS”: DISTINÇÕES, DIAGNÓSTICO E REAÇÃO. In: **XIX ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO (XIX ENANCIB)**. 2018.

CARVALHO, Rafiza Luziani Varão Ribeiro. Notícias falsas ou propaganda?: Uma análise do estado da arte do conceito fake news. **Questões Transversais**, v. 7, n. 13, 2019.

COELHO, Elisa Quaresma; COELHO, Augusto Quaresma; CARDOSO, José Eduardo Dias. Informações médicas na internet afetam a relação médico-paciente?. **Revista Bioética**, v. 21, n. 1, p. 142-149, 2013.

DE FIGUEIRÊDO BRAZ, Sandrine Cristina; DE SOUZA, Edivanio Duarte. Os Desafios da Confiabilidade da Informação na Produção Colaborativa de Conteúdos: análises na Wikipédia, a Enciclopédia Livre. **Ciência da Informação em Revista**, v. 1, n. 3, p. 19-31, 2014.

DE SOUSA JÚNIOR, João Henriques et al. Da Desinformação ao Caos: uma análise das Fake News frente à pandemia do Coronavírus (COVID-19) no Brasil. **Cadernos de Prospecção**, v. 13, n. 2 COVID-19, p. 331, 2020.

FERREIRA, Luciana Gomes; GRADIM, Anabela. Qualidade e credibilidade para além do Jornalismo. A informação local nas mídias sociais. **Revista Mídia e Cotidiano**, v. 7, n. 7, p. 34-49, 2015.

FLANAGIN, Andrew J.; METZGER, Miriam J. The role of site features, user attributes, and information verification behaviors on the perceived credibility of web-based information. **New media & society**, v. 9, n. 2, p. 319-342, 2007.

FRIAS FILHO, Otavio. O que é falso sobre fake news. **Revista Usp**, n. 116, p. 39-44, 2018.

GONDIM, Ana Paula Soares; WEYNE, Davi Pontes; FERREIRA, Bruno Sousa Pinto. Qualidade das informações de saúde e medicamentos nos sítios brasileiros. **Einstein (São Paulo)**, v. 10, n. 3, p. 335-341, 2012.

GUIZZARDI FILHO, Osvaldo; CONTI, Vivaldo Luiz. Produção e disseminação de informações socioeconômicas. **Transinformação**, v. 13, n. 2, p. 43-54, 2001.

LÓPEZ HIDALGO, Antonio; FERNÁNDEZ BARRERO, María Ángeles. Notícias falsas, incorretas e incompletas: os desafios dos jornalistas em busca da retificação voluntária. A experiência espanhola. **Brazilian Journal of Research**, 8 (2), 212-236, 2012.

MACHADO, Joice Mengue Ribeiro; TIJIBOY, Ana Vilma. Redes Sociais Virtuais: um espaço para efetivação da aprendizagem cooperativa. **RENOTE-Revista Novas Tecnologias na Educação**, v. 3, n. 1, 2005.

_____. **Maioria das fraudes no e-commerce acontece na madrugada de quinta-feira.** Disponível em: <<https://www.e-commercebrasil.com.br/noticias/maioria-das-fraudes-no-e-commerce-acontece-na-madrugada-de-quinta-feira/>>. Acesso em: 23 de maio de 2020.

MARTINS, Ana Cristina Cazorla; SILVEIRA, Juliane Vanessa. A Influência das Redes Sociais no Comportamento do Consumidor Online. **Centro Universitário Franciscano (UNIFRA). Santa Maria, RS**, 2012.

MIRANDA, Cláudia Maria Carvalho; DE OLIVEIRA ARRUDA, Danielle Miranda. E-produtos e variáveis comportamentais determinantes de compra no varejo virtual: um estudo com consumidores brasileiros. **Revista Eletrônica de Administração**, v. 10, n. 1, 2004.

RECUERO, Raquel. Considerações sobre a difusão de informações em redes sociais na internet. **Intercom Sul**, 2007.

RECUERO, Raquel. Redes sociais na internet, difusão de informação e jornalismo: elementos para discussão. **Metamorfoses jornalísticas**, v. 2, p. 1-269, 2009.

TERRA, Ana Lúcia; SÁ, Salvina. A credibilidade da informação na web: estudo de práticas informacionais em contexto escolar. **VI Encontro Ibérico EDICIC 2013: Globalização, ciência, informação: atas**, 2013.

TOMAÉL, Maria Inês et al. Avaliação de fontes de informação na Internet: critérios de qualidade. **Informação & Sociedade**, v. 11, n. 2, 2001.

TRAQUINA, Nelson. Porque as notícias são como são. **Florianópolis: Insular**, v. 3, 2005.

VIEIRA, Marli Vick; CHRISTOFOLETTI, Rogério. Confiabilidade no uso da Wikipédia como fonte de pesquisa escolar. **Revista on line de Política e Gestão Educacional**, n. 15, 2013.

UMA ESTRATÉGIA PARA NAVEGAÇÃO DE ROBÔS DE SERVIÇO SEMIAUTÔNOMOS USANDO INFORMAÇÃO LOCAL E PLANEJADORES PROBABILÍSTICOS

Data de aceite: 27/08/2020

Data de submissão: 03/08/2020

Elias José de Rezende Freitas

Instituto Federal de Educação, Ciência e
Tecnologia de Minas Gerais
Itabirito – Minas Gerais
<http://lattes.cnpq.br/4716860363006932>

Guilherme Augusto Silva Pereira

West Virginia University
Morgantown – West Virginia – EUA
<http://lattes.cnpq.br/1726523803257092>

RESUMO: Este trabalho propõe uma estratégia para a navegação de robôs semiautônomos baseada apenas em informações locais, obtidas pelos sensores instalados no robô e um planejador probabilístico que gera caminhos a serem seguidos localmente por ele, garantindo assim o desvio de obstáculos. A missão que o robô deve cumprir é definida pelo usuário por meio de uma sequência de comandos simples, como siga em frente e vire à direita, sendo esse comando codificado por campos vetoriais artificiais, que são utilizados pelo funcional de custo a ser otimizado pelo planejador. Os resultados obtidos demonstram que os robôs foram capazes de executar a missão definida pelo usuário, desviando eficientemente e de forma planejada de obstáculos estáticos e, para algumas situações bem definidas, dinâmicos, incluindo pessoas se locomovendo no ambiente durante a missão.

PALAVRAS-CHAVE: Robôs de serviço; Navegação; Desvio de obstáculos;

Planejadores probabilísticos.

SEMIAUTONOMISTS USING LOCAL INFORMATION AND PROBABILISTIC PLANNERS

ABSTRACT: This paper proposes a strategy for navigation of semi-autonomous service robots based on local information obtained by the sensors installed on the robot and on a probabilistic planner, which calculates paths to be followed locally by the robot, thus ensuring obstacle avoidance. The robot's mission is defined by a user through a sequence of simple commands, such as “go ahead” and “turn right”. These commands are encoded by artificial vector fields, which are used by the cost functional to be optimized by the planner. The results show that the robots were able to perform the user-defined mission efficiently, avoiding static obstacles and, for some well-defined situations, also dynamic obstacles, including people moving in the environment during the mission.

KEYWORDS: Service robots; Navigation; Obstacle avoidance; Probabilistic planners.

1 | INTRODUÇÃO

O desenvolvimento e a utilização de robôs têm crescido cada vez mais no mercado mundial. Eles já fazem parte da sociedade, estão presentes nas indústrias, nas ruas e dentro das casas (Dayoub, Morris e Corke, 2015). Dentre os vários robôs, existem aqueles que são destinados para fins

não-industriais, chamados de robôs de serviço.

Muitos desses robôs são ainda semiautônomos, por dependerem ou interagirem de alguma maneira com o ser humano. Alguns exemplos de robôs de serviço semiautônomos são: as cadeiras de rodas inteligentes (Simpson, LoPresti e Cooper, 2008), os robôs que devem seguir pessoas (Müller et al., 2008), os guias de museus (Dayoub, Morris e Corke, 2015) e robôs exploradores espaciais (Goldberg, Maimone e Matthies, 2002). Esses robôs se diferem dos demais robôs em pelo menos três aspectos: (i) eles não precisam conhecer o mapa global do ambiente em que se deslocam; (ii) eles não precisam conhecer o objetivo ou alvo final; e (iii) eles não precisam se localizar globalmente no ambiente. Contudo, esses robôs ainda devem ser capazes de cumprir a missão desejada pelo usuário de maneira segura, não colidindo com obstáculos (objetos, outros robôs, pessoas). O grande desafio é garantir isso em um ambiente desconhecido e dinâmico.

Para motivar este trabalho, considere a situação em que um robô, por exemplo, uma cadeira de rodas inteligente, entra em um prédio pela primeira vez, tendo que levar um usuário até uma determinada sala. Nessa situação, não se tem informações a priori do local (ambiente desconhecido), existem pessoas se movimentando (ambiente dinâmico) e, pelo fato de se ter um ambiente interno, não há sinal de GPS para facilitar a localização do robô. Assim, o conhecimento do usuário do local ou uma indicação fornecida por outra pessoa (um porteiro, por exemplo) deveria fornecer as informações necessárias para que robô autonomamente chegasse até a sala. Seria interessante que essas informações fossem fornecidas de maneira similar às orientações comumente dadas pelas pessoas, como: siga o corredor até o final, vire à direita e entre pela terceira porta. Repare que uma pessoa teria facilidade para seguir tais comandos, mesmo em um ambiente desconhecido e sem nenhuma localização global.

Assim, este trabalho visa apresentar e avaliar uma estratégia de navegação que permita que o robô haja como um ser humano e seja capaz de seguir instruções simples, mesmo sem que ele tenha uma localização global e o mapa do ambiente. Para tanto, o robô se baseará nos dados de detecção, por exemplo, de corredores, de interseções e de portas, obtidos com seus sensores.

A principal contribuição deste trabalho é então o desenvolvimento de uma estratégia de navegação a ser aplicada, principalmente, aos robôs de serviço semiautônomos.

Essa estratégia destaca-se das demais pelas seguintes características: (i) Mapeamento local: não há necessidade do conhecimento do mapa do ambiente; (ii) Replanejamento a cada intervalo de tempo de planejamento: com base no mapa local fornecido e a cada intervalo de planejamento, um planejador de movimento irá fornecer um caminho completo a ser seguido localmente pelo robô, garantindo que o robô desvie de obstáculos existentes no ambiente e cumpra a tarefa estabelecida por um usuário de forma planejada; (iii) Tarefas codificadas por meio de um campo vetorial artificial: o campo vetorial especificado para uma tarefa, por exemplo, “siga pelo corredor”, é incorporado diretamente

ao planejador de movimento, permitindo adicionar um comportamento ou uma regra social à navegação do robô; (iv) Criação de referenciais de curta duração: a cada intervalo de planejamento um novo referencial é criado e válido apenas nesse curto intervalo de tempo, de forma que apenas a odometria relativa ao novo referencial seja suficiente para que o robô siga o caminho fornecido pelo planejador, evitando as incertezas da odometria.

A próxima seção apresenta uma revisão bibliográfica sobre estratégias de navegação. A Seção 3 é dedicada a descrever a metodologia proposta em detalhes. A Seção 4 mostra a configuração experimental usada para validar a estratégia deste trabalho e a Seção 5 apresenta os resultados experimentais que ilustram o trabalho, utilizando robôs reais. Finalmente, conclusões e trabalhos futuros são discutidos na Seção 6.

2 | REVISÃO BIBLIOGRÁFICA

Para que um robô possa se locomover em um ambiente é necessário combinar métodos de localização e/ou de mapeamento e/ou de planejamento de movimento. Existem uma grande variedade de métodos (Choset et al., 2005) e a maneira na qual eles são combinados define uma determinada estratégia de navegação capaz de executar a missão designada ao robô e capaz de atender às restrições impostas tanto pelo ambiente, seja ele conhecido, parcialmente conhecido ou desconhecido, quanto pelo robô (Alves et al., 2011). Com o intuito de facilitar a compreensão dessas combinações, pode-se agrupar as estratégias de navegação em pelo menos três grupos: (i) Navegação local reativa, (ii) Navegação global com o conhecimento a priori do mapa do ambiente e (iii) Navegação local com mapeamento.

A navegação local reativa é baseada em uma arquitetura reativa de hardware e de software, introduzida no final da década de 80, que busca permitir uma resposta mais rápida às mudanças do ambiente (Brooks 1986). Dessa forma, o robô reage instantaneamente à presença de obstáculos recém detectados sem se preocupar com as consequências de sua reação. Os principais benefícios que facilitam a sua utilização em robôs semiautônomos, principalmente, daqueles que são comandados diretamente por seres humanos (Chik et al., 2016) (Baklouti, Amor e Jallouli, 2017), é o fato de serem muito eficientes e não necessitarem de localização global nem de modelos complexos para os obstáculos e para o ambiente. No entanto, uma desvantagem é que nessa estratégia não é levado em consideração o estado futuro do robô, já que não há um planejamento de movimento, podendo levar o robô a condições indesejadas, fazendo-o parar constantemente ou seguir trajetórias muito longas.

Já a navegação global com o conhecimento a priori de um mapa tem sido a mais utilizada, pois a maioria dos trabalhos assumem que o ambiente em que o robô será inserido é pré-definido, sendo possível obter um mapa prévio desse ambiente. Nessa estratégia, métodos de planejamento de movimento, como os métodos probabilísticos (Karaman et al., 2011), são utilizados para deliberar o movimento, de tal forma que os obstáculos estáticos

considerados no modelo serão evitados no cumprimento de sua missão. Exemplo desse tipo de estratégia pode ser encontrado em robôs que navegam em museus (Burgard et al. 1999) (Thrun et al. 1999), em escritórios (Marder-Eppstein et al., 2010) e em prédios (Araujo, Caminhas e Pereira, 2015). Para a obtenção do mapa desses locais, alguns trabalhos utilizam, em uma etapa anterior à navegação, as técnicas de Localização e Mapeamento Simultâneos (SLAM) (Dayoub, Morris e Corke, 2015) (Schwesinger et al., 2016) (Araujo, Caminhas e Pereira, 2015). Note que nem sempre é possível e simples para alguns robôs semiautônomos realizar esse procedimento, como, por exemplo, uma cadeira de roda inteligente se locomovendo em um ambiente interno desconhecido (Baklouti, Amor e Jallouli, 2017).

Por último, a navegação local com mapeamento é uma estratégia que a partir de um mapa obtido em tempo de execução permite que um planejador de movimento forneça caminhos a serem seguidos localmente e que são válidos no campo de visão dos sensores. Essa estratégia é encontrada, principalmente, em robôs exploradores (Stückler et al., 2015), sendo também considerada neste trabalho. O mapeamento dessa estratégia pode ser sem memória, se utilizar diretamente os dados dos sensores, ou com memória se durante a navegação o mapa é construído, usando, por exemplo, técnicas de SLAM. Esse tipo de estratégia combina tanto características reativas, já que não possui um mapa prévio do ambiente, quanto características deliberativas, por realizar um constante replanejamento local, levando em consideração obstáculos em um horizonte futuro.

3 | METODOLOGIA

A metodologia proposta neste trabalho foi estruturada conforme a arquitetura apresentada na Figura 1, considerando que a missão definida pelo usuário pode ser dividida em uma sequência de comandos simples, como virar à esquerda, virar à direita, seguir em frente e retornar.

Em síntese, esses comandos simples podem ser codificados por meio de um campo vetorial que será inserido no cálculo de um funcional de custo utilizado por um planejador de movimento.

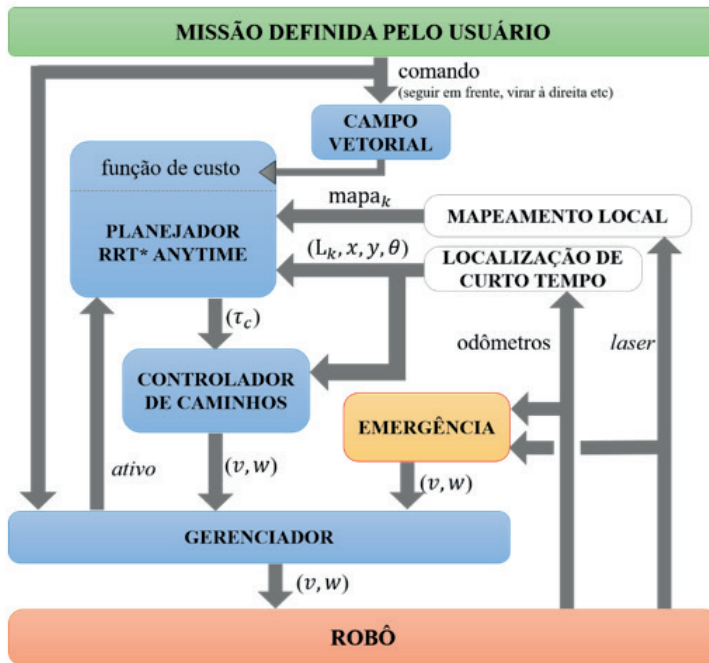


Figura 1 – Arquitetura de navegação proposta neste trabalho.

O caminho τ_c fornecido pelo planejador é utilizado por um controlador de caminho que calculará as devidas velocidades (v, ω) que o robô deve alcançar para seguir o caminho desejado. O Gerenciador da missão é o responsável por enviar as velocidades calculadas para o robô, além de ser capaz de executar alguns dos comandos simples mencionados anteriormente, e de parar o robô em caso de emergência.

Note que o planejador de movimento se baseia em um mapa local (mapa_k), obtido em tempo de execução a partir dos dados de um sensor à laser, e na localização do robô (x, y, θ) , ambos representados no mesmo referencial L_k , válido por um curto intervalo de tempo pré-determinado. Sendo assim, pode-se classificar essa estratégia de navegação como uma navegação local com mapeamento sem memória.

De forma a clarear a sequência de execução da estratégia deste trabalho, é apresentado o Algoritmo 1. Considera-se que o tempo possa ser discretizado em intervalos t_k , para $k = 1, 2, \dots$, onde o período $\Delta t = t_k - t_{k-1}$ é correspondente ao tempo de cálculo do planejador escolhido (tempo de planejamento t_s). Note que uma sequência de referenciais locais L_k será gerada na execução desse algoritmo, sendo que a cada período de planejamento um referencial local é usado como referencial fixo (linha 6 do Algoritmo 1).

Algoritmo 1: Estratégia de Navegação.

```
1  $k = 0$  ;  $v = 0.0$  ;  $\omega = 0.0$  ;  $\mathbf{v}_r = (v, \omega)$  ;  $\tau_0 \leftarrow \emptyset$  ;  $t_s =$  tempo de
   planejamento desejado ;
2  $L_0 =$  GeraReferencial();
3  $\text{mapa}_0 =$  ObtemMapa( $L_0$ );
4  $k \leftarrow k + 1$ ;
5 Enquanto ativo faça
6    $L_k =$  GeraReferencial();
7    $\tau_c =$  TransformaCaminho( $\tau_{k-1}, L_k$ );
8    $\text{mapa}_k =$  TransformaMapa( $\text{mapa}_{k-1}, L_k$ );
9    $\mathbf{o}_c =$  Integra( $\tau_c, \mathbf{p}_r, t_s, \mathbf{v}_r$ );
10  Faça em paralelo enquanto  $t \leq t_s$ :
11     $\tau_s =$  Planeja( $\text{mapa}_k, \mathbf{o}_c, t_s$ );
12     $\mathbf{v}_r =$  SegueCaminho( $\tau_c$ );
13  fim-paralelo
14   $\tau_k =$  ConcatenaCaminho( $\tau_c, \tau_s, \mathbf{p}_r, \mathbf{o}_c$ );
15   $\text{mapa}_k =$  ObtemMapa( $L_k$ );
16   $k \leftarrow k + 1$ ;
17 fim-enquanto
```

O laço de execução é realizado enquanto um sinal de ativo enviado pelo módulo gerenciador for verdadeiro. Os primeiros procedimentos do laço são gerar um novo referencial $\{L_k\}$ e em sequência transformar o caminho fornecido pelo planejador (τ_{k-1}) e o mapa local gerado a partir dos dados dos sensores (mapa_{k-1}), ambos obtidos em relação ao referencial $\{L_{k-1}\}$, para o novo referencial $\{L_k\}$ (linhas 7 e 8).

Após essa transformação, cujo tempo de cálculo é muito pequeno em relação a t_s e pode ser desprezado, uma nova instância do planejador é lançada (linha 11). Como parâmetros, esse planejador utiliza o mapa local que acaba de ser transformado para $\{L_k\}$ (linha 8) e o ponto de origem do novo caminho, \mathbf{o}_c . Pode-se estranhar a ausência de um ponto alvo como entrada do planejador. De fato, como um robô semiautônomo nem sempre conhece seu destino, caso esse ponto seja necessário para o planejador em questão, ele pode ser escolhido arbitrariamente e sem perda de generalidade como um ponto à frente do robô e no limite do campo de visão de seus sensores.

O ponto de origem, \mathbf{o}_c , é obtido integrando-se (ou simulando-se) o caminho anterior (calculado em relação a $\{L_{k-1}\}$ e transformado para $\{L_k\}$), durante um tempo t_s a partir da posição real do robô, \mathbf{p}_r , representada no referencial atual (linha 9). Para essa integração é considerada a velocidade média do robô, \mathbf{v}_r . A ideia de se escolher um ponto do caminho anterior como sendo o início do próximo caminho é manter a continuidade do caminho do robô. Com essa escolha, também se evita situações de caminhos pertencentes a classes homotópicas diferentes.

Observe que essa estratégia define um trecho do caminho anterior (entre a posição atual do robô \mathbf{p}_r e \mathbf{o}_c) a ser seguido enquanto o novo caminho é calculado. Esse trecho do caminho é chamado na literatura de caminho comprometido (*committed path*) (Karaman et al. 2011).

Enquanto é solucionado o novo problema de planejamento (linha 11), durante o intervalo t_s , um Controlador de caminhos fornece os devidos comandos de velocidade para manter o robô seguindo o caminho comprometido (linha 12). Repare que esse caminho já está no referencial $\{L_k\}$, não havendo necessidade do Controlador de caminhos transformá-lo, utilizando apenas como realimentação a posição e orientação do robô em relação ao referencial atual ($\{L_k\}$). Passado o tempo previsto de planejamento t_s , quando o planejador retorna o novo caminho, o caminho comprometido ainda não seguido pelo robô é concatenado com o caminho recém calculado (linha 14), um novo mapa com as últimas informações dos sensores é obtido (linha 15) e o processo se repete fazendo a transformação do novo caminho e do novo mapa para o recém criado referencial $\{L_{k+1}\}$.

É interessante ressaltar que, para que todo esse processo funcione adequadamente, sempre é necessário conhecer a transformação entre o referencial anterior e o referencial atual. Como t_s é, normalmente, um tempo muito pequeno (na ordem de 1s a 2s), uma boa odometria é geralmente suficiente para calcular as transformações de coordenadas necessárias com baixa incerteza, como poderá ser visto na Seção 5.

4 | IMPLEMENTAÇÃO

A estratégia proposta foi implementada em C++ e em Python, utilizando o ROS (Quigley et al., 2009) juntamente com a biblioteca OMPL (Şucan, Moll e Kavraki, 2012). A simulação foi realizada no StageROS e testada em dois robôs semiautônomos, vistos na Figura 2: (i) robô de serviço MARIA (*Manipulator Robot for Interaction and Assistance*) e (ii) uma cadeira de rodas em desenvolvimento no laboratório CORO/UFMG.

Os experimentos realizados exploram a aplicação na qual o robô navega em um ambiente desconhecido, não possuindo o mapa do ambiente. O cenário escolhido foi um prédio constituído de longos corredores de largura de 2,35 m, os quais havia alguns pilares que o estreitavam, continha alguns móveis e/ou objetos de decoração, mas que era essencialmente vazio, exceto pela presença de pessoas. Para testar o funcionamento da estratégia deste trabalho, foram também acrescentados outros obstáculos.

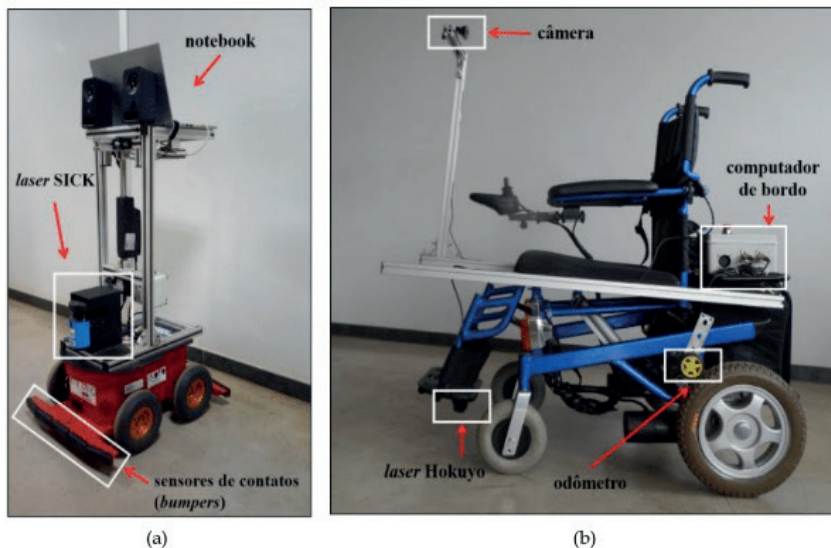


Figura 2 – Robôs de serviço semiautônomos utilizados nos experimentos reais. (a) Robô de serviço MARIA. (b) Cadeira de rodas inteligente em desenvolvimento no laboratório CORO.

5 | RESULTADOS

Nesta seção são apresentados os resultados de alguns dos experimentos realizados com o robô de serviço MARIA e com uma cadeira de rodas inteligente, sendo utilizado os mesmos parâmetros das simulações descritas em (Pereira e Freitas, 2020).

Primeiramente, o seguinte experimento foi realizado utilizando o robô MARIA: um usuário definiu a seguinte sequência de comandos: siga, vire à direita. Esses comandos foram repetidos onze vezes, de tal maneira que o robô completasse três voltas no ambiente.

A Figura 3(a) apresenta a odometria coletada durante a execução desse experimento, cujo vídeo completo pode ser encontrado em: https://youtu.be/loS_TxmtQMI. O robô foi capaz de desviar dos obstáculos existentes no ambiente, incluindo alguns cones que foram acrescentados para dificultar o cumprimento da missão.

Após a execução do experimento, observou-se que o robô terminou sua missão muito próximo do ponto de partida, o que não pôde ser observado pela odometria do robô, que apresenta um erro de integração razoavelmente grande. Mesmo com esse erro, o robô obteve sucesso na sua missão, evidenciando a importância dos referenciais de curta duração propostos na metodologia. Nesse experimento, cada referencial foi válido por apenas 1,0s, durante o qual a odometria apresenta resultados bons.

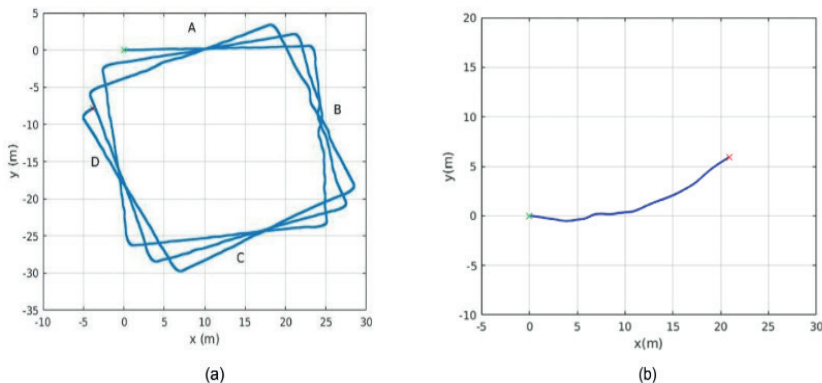


Figura 3 – Odometria coletada dos robôs: (a) MARIA durante a execução de três voltas (aprox. 300m) em um ambiente composto por quatro corredores identificados por A,B,C,D. No corredor B foram colocados obstáculos fixos (cones) no caminho do robô. (b) Cadeira de rodas inteligente ao percorrer o corredor A (aprox. 25m). Os robôs iniciam sua trajetória no ponto em verde (corredor A), finalizando no ponto em vermelho.

De forma muito similar, o segundo experimento foi realizado com a cadeira de rodas inteligente, sendo que o comando fornecido pelo usuário foi apenas siga. Dessa forma, a cadeira percorreu o corredor inteiro, mantendo-se próxima a parede à sua direita, até encontrar a primeira interseção com outro corredor. Durante esse trajeto, ela desviou com sucesso dos obstáculos estáticos presentes no ambiente. A Figura 3(b) mostra os dados da odometria durante a missão. Observe que o erro acumulado da odometria é muito maior que o do robô MARIA, chegando a quase 5m no eixo y ao fim do corredor. Novamente, mesmo com esse erro, a missão não foi comprometida devido aos referenciais de curta duração.

Para estimar como a incerteza da localização fornecida pela odometria desses robôs variou durante os experimentos, tomou-se como valor médio a própria odometria do robô e propagou-se as incertezas associadas às velocidades linear e angular dos robôs, obtidas experimentalmente, utilizando apenas a etapa de previsão do Filtro de Kalman Estendido (EKF).

Nas Figuras 4(a) e 5(a) são mostradas elipses para dois desvios padrão, o que representa um grau de confiança de aproximadamente 95 %. Observe que as áreas das elipses crescem a partir da posição inicial do robô, significando que há um aumento gradativo da incerteza da localização do robô. Esse aumento ao passar do tempo pode ser visualizado nas Figuras 4(b) e Figura 5(b). Apesar dessa incerteza sempre crescente, para a metodologia proposta, apenas a incerteza durante o tempo de planejamento t_s é, de fato, percebida pelo sistema. Como a única localização usada é relativa a um referencial criado recentemente, o erro acumulado é limitado a um valor muito pequeno, conforme mostrado nas Figuras 4(c) e 5(c), em que praticamente não se nota o crescimento das áreas das elipses e de maneira similar nas Figuras 4(d) e 5(d).

Com o intuito de também avaliar o comportamento do sistema na presença de obstáculos dinâmicos, uma pessoa moveu-se em direção ao robô, enquanto esse executava o comando siga. O robô MARIA foi configurado para uma $v_{\max} = 0,9$ m/s e o tempo de planejamento foi reduzido para $t_s = 0,45$ s. A redução do tempo de planejamento foi necessária para garantir o desvio de uma pessoa andando em uma velocidade normal de aproximadamente 1,0m/s.

Também esse experimento foi realizado utilizando a cadeira de rodas, porém dada a dificuldade de se controlar a velocidade dessa cadeira com um tempo de resposta pequeno, seja pela sua dinâmica seja pelo fato de não ser possível acionar diretamente os motores da mesma, foi configurada uma velocidade menor, $v_{\max} = 0,3$ m/s, e a pessoa caminhou com uma velocidade em torno de 0,25 m/s. Repetiu-se dez vezes esse experimento e em todos os testes os robôs foram capazes de desviar da pessoa. O vídeo completo pode ser visto em <https://youtu.be/DONxBJJ5OT8> (robô MARIA) e em <https://youtu.be/DilcPul18kc> (cadeira de rodas inteligente).

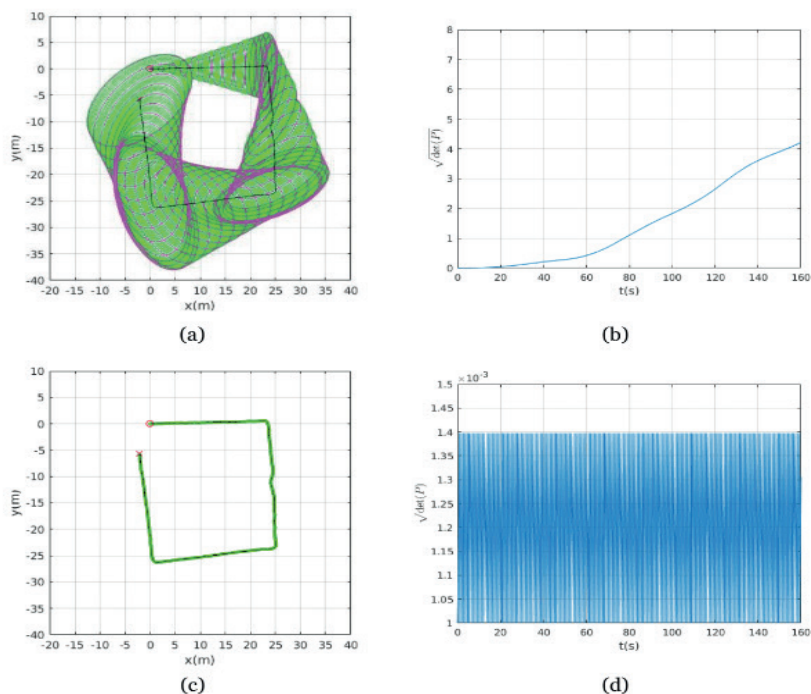


Figura 4 – Estimativa da covariância da odometria durante a primeira volta da missão, utilizando o robô MARIA. Em (a) é apresentada a estimativa da incerteza da posição do robô, representada pelas elipses, usando dois desvios padrão, ao longo dos dados da odometria em preto. Em (b) é apresentado a variação da estimativa total da incerteza de localização em relação ao tempo. Em (c) e (d) é ilustrado como a estratégia de navegação proposta neste trabalho percebe a incerteza de localização para um tempo de planejamento = 1,0s.

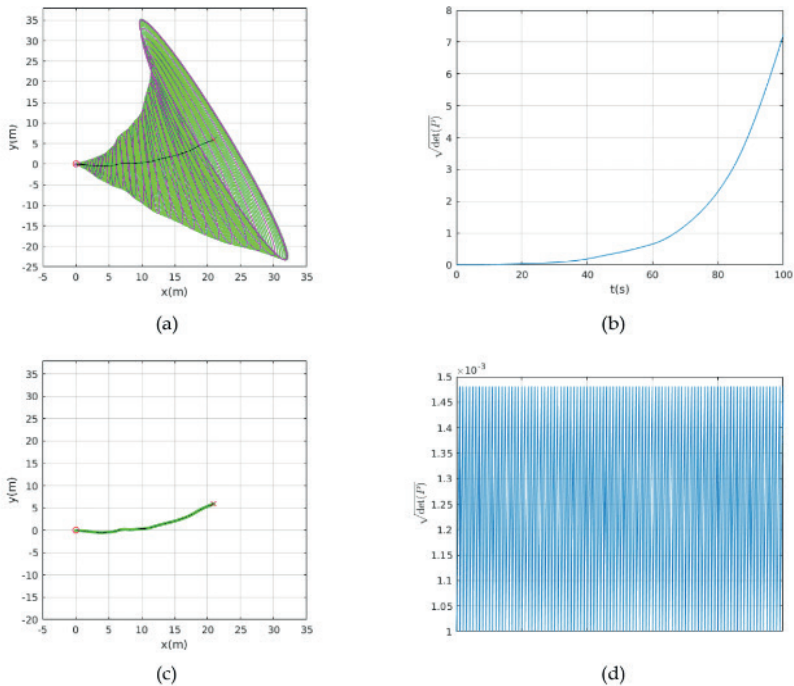


Figura 5 – Estimativa da covariância da odometria durante a navegação da cadeira de rodas inteligente por um corredor. Em (a) é apresentada a estimativa da incerteza da posição da cadeira de rodas, representada pelas elipses, usando dois desvios padrão, ao longo dos dados da odometria em preto. Em (b) é apresentado a variação da estimativa total da incerteza de localização em relação ao tempo. Em (c) e (d) é ilustrado como a estratégia de navegação proposta neste trabalho percebe a incerteza de localização para um tempo de planejamento $t_s = 1,0$ s.

6 | CONCLUSÃO E TRABALHOS FUTUROS

Neste trabalho propôs-se uma estratégia de navegação para robôs de serviço semiautônomos baseada em um planejador probabilístico que não requer a localização global do robô. Essa estratégia é destinada, principalmente, aos robôs que devem navegar com segurança, mas não possuem o mapa do ambiente nem um sistema de localização global. A ideia básica da estratégia é manter um referencial local válido somente durante o tempo necessário para que o robô siga parte do último caminho fornecido pelo planejador enquanto é computado um novo caminho que considere possíveis mudanças no espaço de trabalho do robô. Após o novo caminho ser calculado, um novo referencial é definido e o caminho é transformado para o novo referencial. O princípio da estratégia é válido já que esses dois referenciais, o anterior e o atual, estão muito próximos e, portanto, é possível definir uma transformação de coordenadas entre eles usando apenas a odometria.

Como o nosso método foi projetado para ser usado em robôs semiautônomos, foi

proposta a codificação de comandos simples do usuário por meio de campos vetoriais. Isso permitiu o uso de trabalhos anteriores que integram planejadores ótimos e campos de vetoriais para determinar o planejador de movimento utilizado neste trabalho, como apresentado em (Pereira, Choudhury e Scherer, 2016). Por meio dos experimentos realizados, verificou-se que o robô foi capaz de cumprir a missão dada pelo usuário, seguindo o mais próximo possível o campo vetorial e evitando obstáculos desconhecidos de forma planejada.

Analisando o algoritmo proposto, observou-se que um robô que executa essa estratégia é capaz de evitar obstáculos estáticos e, dependendo da relação entre a velocidade do robô, a velocidade do obstáculo e o tempo de planejamento, é também capaz de evitar obstáculos dinâmicos, incluindo pessoas, como mostrado em nossos experimentos. Portanto, trabalhos futuros pretendem aumentar a eficiência dos planejadores, por meio, por exemplo, da implementação de algumas funções em GPU ou em outro hardware dedicado. A previsão do movimento do obstáculo também é algo que poderia aprimorar este trabalho.

Uma desvantagem de nossa abordagem, que não foi crítica para nossos experimentos, mas que seria relevante se o robô pudesse se mover mais rápido, é a ausência de restrições diferenciais e/ou restrições dinâmicas no problema de otimização. Isso significa que o planejador atual pode encontrar um caminho que não pode ser seguido pelo robô, a menos que ele reduza drasticamente sua velocidade ou, até mesmo, tenha que parar. Além disso, sem restrições dinâmicas, a metodologia pode ser inviável para o uso em cadeiras de rodas inteligentes, um dos robôs semiautônomos mais comuns, já que o usuário poderia sofrer acelerações bruscas durante o movimento do robô. Uma possível solução seria utilizar um planejador de trajetória ao invés de um planejador de caminhos ou, como uma aproximação, utilizar os caminhos de Dubins para estender a árvore gerada pelo RRT*, como mostrado em (Karaman e Frazzoli, 2010).

REFERÊNCIAS

ALVES, S. F. R. et al. **Conceptual bases of robot navigation modeling, control and applications**. In: BARRERA, A. (Ed.). *Advances in Robot Navigation*. Rijeka: InTech, 2011. cap. 1.

ARAUJO, A. R.; CAMINHAS, D. D.; PEREIRA, G. A. **An architecture for navigation of service robots in human-populated office-like environments**. *IFAC-PapersOnLine*, v. 48, n. 19, p. 189 – 194, 2015. ISSN 2405-8963.

BAKLOUTI, E.; AMOR, N. B.; JALLOULI, M. **Reactive control architecture for mobile robot autonomous navigation**. *Robotics and Autonomous Systems*, v.89,p.9–14, 2017.

BROOKS, R. **A robust layered control system for a mobile robot**. *IEEE Journal on Robotics and Automation*, IEEE, v. 2, n. 1, p. 14–23, 1986.

BURGARD, W. et al. **Experiences with an interactive museum tour-guide robot**. *Artificial*

Intelligence, v. 114, n. 1-2, p. 3–55, 1999.

CHIK, S. et al. **A review of social-aware navigation frameworks for service robot in dynamic human environments.** *Journal of Telecommunication, Electronic and Computer Engineering*, v. 8, n. 11, p. 41–50, 2016. ISSN 2180-1843.

CHOSSET, H. et al. **Principles of Robot Motion: Theory, Algorithms, and Implementations** (*Intelligent Robotics and Autonomous Agents series*). Cambridge: The MIT Press, 2005. ISBN 978-0262033275.

CORKE, P. **Robotics, Vision and Control.** Berlin: Springer Berlin Heidelberg, 2011. ISBN 978-3642201431. Disponível em: <<https://doi.org/10.1007/978-3-642-20144-8>>.

DAYOUB, F.; MORRIS, T.; CORKE, P. **Rubbing shoulders with mobile service robots.** *IEEE Access*, v. 3, p. 333–342, 2015.

OLDBERG, S.; MAIMONE, M.; MATTHIES, L. **Stereo vision and rover navigation software for planetary exploration.** In: IEEE. *Proceedings of the IEEE Aerospace Conference*. 2002. v. 5, p. 2025–2036.

KARAMAN, S.; FRAZZOLI, E. **Optimal kinodynamic motion planning using incremental sampling-based methods.** In: IEEE. *Decision and Control (CDC), 2010 49th IEEE Conference on*. [S.l.], 2010. p. 7681–7687.

KARAMAN, S.; FRAZZOLI, E. **Sampling-based algorithms for optimal motion planning.** *The International Journal of Robotics Research*, Sage Publications, v. 30, n. 7, p. 846–894, 2011.

KARAMAN, S. et al. **Anytime motion planning using the RRT*.** In: *Proceedings of the IEEE International Conference on Robotics and Automation*. IEEE, 2011. p. 1478–1483. 2011.

MARDER-EPPSTEIN, E. et al. **The office marathon: Robust navigation in an indoor office environment.** In: *Proceedings of the IEEE International Conference on Robotics and Automation*. IEEE, 2010. p. 300–307.

MÜLLER, J. et al. **Socially inspired motion planning for mobile robots in populated environments.** In: *Proceedings of the International Conference on Cognitive Systems*. Springer, 2008. p. 85–90.

PEREIRA, G. A.; FREITAS, E. J. **Navigation of semi-autonomous service robots using local information and anytime motion planners.** *Robotica*, Cambridge University Press, p. 1–19.

PEREIRA, G. A. S.; CHOUDHURY, S.; SCHERER, S. **Kinodynamic Motion Planning on Vector Fields using RRT*.** Pittsburgh, PA, July 2016.

QUIGLEY, M. et al. **ROS: an open-source robot operating system.** In: *ICRA Workshop on Open Source Software*. [S.l.: s.n.], 2009.

SCHWESINGER, D. et al. **A smart wheelchair ecosystem for autonomous navigation in urban environments.** *Autonomous Robots*, Springer Nature, v. 41, n. 3, p. 519–538, 2016.

SIMPSON, R. C.; LOPRESTI, E. F.; COOPER, R. A. **How many people would benefit from a smart wheelchair?** *Journal of rehabilitation research and development*, REHABILITATION RESEARCH & DEVELOPMENT SERVICE, v. 45, n. 1, p. 53, 2008.

STÜCKLER, J. et al. NimbRo explorer: **Semiautonomous exploration and mobile manipulation in rough terrain.** *Journal of Field Robotics*, v. 33, n. 4, p. 411–430, 2015.

ŞUCAN, I. A.; MOLL, M.; KAVRAKI, L. E. **The Open Motion Planning Library.** *IEEE Robotics & Automation Magazine*, v. 19, n. 4, p. 72–82, 2012.

THRUN, S. et al. **MINERVA: a second-generation museum tour guide robot.** In: *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 1999. v. 3, p. 1999–2005.

SOBRE O ORGANIZADOR

ERNANE ROSA MARTINS - Doutorado em andamento em Ciência da Informação com ênfase em Sistemas, Tecnologias e Gestão da Informação, na Universidade Fernando Pessoa, em Porto/Portugal. Mestre em Engenharia de Produção e Sistemas, possui Pós-Graduação em Tecnologia em Gestão da Informação, Graduação em Ciência da Computação e Graduação em Sistemas de Informação. Professor de Informática no Instituto Federal de Educação, Ciência e Tecnologia de Goiás - IFG (Câmpus Luziânia) ministrando disciplinas nas áreas de Engenharia de Software, Desenvolvimento de Sistemas, Linguagens de Programação, Banco de Dados e Gestão em Tecnologia da Informação. Pesquisador do Núcleo de Inovação, Tecnologia e Educação (NITE), certificado pelo IFG no CNPq. ORCID: <https://orcid.org/0000-0002-1543-1108>. Personal homepage: <https://ernanemartins.wordpress.com/>

ÍNDICE REMISSIVO

A

Análise de negócios 72, 73, 74, 79

Análise estatística 173, 174, 180, 181, 182, 183

Aplicativo 4, 86, 87, 88, 89, 91, 114, 117, 118, 125, 127, 128, 188

Aprendizado de máquinas 20, 144

Árvore de decisão 27, 144, 147, 151, 152, 153

Automação 59, 60, 62, 63, 64, 65, 70, 71, 97, 129

B

Bag-of-features 20, 22, 23, 24, 25, 26, 28, 29

Beacons Bluetooth 109, 110, 111, 112, 113, 114, 116, 117

Bluetooth 109, 110, 111, 112, 113, 114, 116, 117, 128, 129, 130

C

Ciberespaço 186, 187, 189, 190, 192

Ciência da computação 8, 70, 93, 94, 96, 106, 210

Controlador Lógico Programável 59, 60, 61, 63, 70, 71

D

Data warehouse 50, 51, 54, 184

Desenvolvimento de software 72, 73, 74, 77, 79, 81, 82, 83, 84, 179

Dispositivos móveis 1, 4, 6, 88, 92, 109, 110, 112, 113, 114, 127, 128

E

ENADE 86, 87, 89, 90, 91, 92

Engenharia de requisitos 72, 73, 74, 79

Extreme programming 75, 77, 82, 84

F

Fake news 186, 187, 188, 189, 192, 193, 194

H

Hardware 60, 64, 198, 207

I

Inteligência artificial 143, 146, 154, 168, 183

Internet 18, 88, 94, 97, 101, 144, 145, 146, 150, 152, 153, 154, 155, 156, 168, 169, 170, 171, 185, 186, 187, 188, 190, 191, 192, 193, 194, 195

K

Kanban 73, 75, 77, 78, 82, 83, 84

k-means 24, 27

L

Ladder 59, 60, 61, 64, 66, 67, 68, 70

M

Manutenção 62, 69, 84, 118

Memória 60, 64, 65, 66, 179, 184, 199, 200

Metodologias ágeis 72, 73, 74, 78, 79, 157

Mineração de dados 8, 50, 155

MultiLayer perceptron 27, 28

P

Pensamento computacional 93, 94, 95, 96, 97, 101, 102, 105, 106, 107, 108

Programação 5, 59, 60, 62, 64, 66, 67, 70, 71, 77, 93, 94, 95, 98, 99, 102, 105, 107, 108, 131, 132, 133, 135, 142, 151, 175, 177, 178, 179, 184, 210

Programação linear 131, 132, 133, 142

R

Redes sociais 104, 132, 144, 145, 146, 149, 150, 151, 152, 153, 154, 155, 188, 189, 190, 193, 194

Região periocular 1, 2, 4, 5, 7

Regressão 27, 31, 131, 132, 133, 134, 137, 138, 139, 140, 141, 142, 143, 152, 175

Robôs 1, 109, 110, 111, 113, 128, 145, 196, 197, 198, 199, 202, 203, 204, 205, 206, 207

S

Scratch 93, 94, 95, 99, 100, 101, 102, 103, 104, 106, 107

Scrum 73, 75, 76, 77, 84

Semiautônomos 196, 197, 198, 199, 202, 203, 206, 207

Servidor 1, 4, 5, 6, 112, 114, 117, 121, 127, 178

Sistema de localização híbrido 109, 113, 114, 124, 128

Sistema em nuvem 109, 113, 114, 119

Sistema web 173

Smartphone 113, 114, 116, 117, 118, 121, 125, 126, 127, 150

Software 44, 45, 46, 57, 58, 60, 64, 66, 72, 73, 74, 75, 77, 78, 79, 80, 81, 82, 83, 84, 85, 93, 94, 95, 106, 149, 161, 171, 176, 177, 178, 179, 184, 185, 198, 208, 210

Sumarização 131, 132, 133, 134, 135, 136, 138, 139, 142, 143

T

Tecnologia 1, 42, 59, 62, 63, 70, 71, 72, 87, 88, 94, 95, 97, 104, 105, 106, 108, 109, 111, 112, 113, 145, 153, 157, 158, 160, 166, 168, 169, 173, 184, 196, 210

Tecnologias digitais 156, 158, 166

Tecnologias disruptivas 156, 157, 158, 160, 161, 163, 166, 169, 170

Transformação digital 156, 157, 158, 159, 160, 163, 169, 170

V

Variância local 1, 2, 3, 5, 6

Visão computacional 5, 20, 22, 23, 28, 29, 145

W

Web service 114, 116, 118, 124

X

XGBoost 30, 31, 33, 34, 35, 36, 38, 39

TECNOLOGIAS, MÉTODOS E TEORIAS NA ENGENHARIA DE COMPUTAÇÃO

www.atenaeditora.com.br 

contato@atenaeditora.com.br 

[@atenaeditora](https://www.instagram.com/atenaeditora) 

www.facebook.com/atenaeditora.com.br 

Atena
Editora

Ano 2020

TECNOLOGIAS, MÉTODOS E TEORIAS NA ENGENHARIA DE COMPUTAÇÃO

www.atenaeditora.com.br 

contato@atenaeditora.com.br 

[@atenaeditora](https://www.instagram.com/atenaeditora) 

www.facebook.com/atenaeditora.com.br 

 **Atena**
Editora

Ano 2020