

# 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 Internazionale 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<sup>ª</sup> Dr<sup>ª</sup> Anelise Levay Murari – Universidade Federal de Pelotas  
Prof. Dr. Benedito Rodrigues da Silva Neto – Universidade Federal de Goiás  
Prof<sup>ª</sup> Dr<sup>ª</sup> 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<sup>ª</sup> Dr<sup>ª</sup> Eleuza Rodrigues Machado – Faculdade Anhanguera de Brasília  
Prof<sup>ª</sup> Dr<sup>ª</sup> Elane Schwinden Prudêncio – Universidade Federal de Santa Catarina  
Prof<sup>ª</sup> Dr<sup>ª</sup> Eysler Gonçalves Maia Brasil – Universidade da Integração Internacional da Lusofonia Afro-Brasileira  
Prof. Dr. Ferlando Lima Santos – Universidade Federal do Recôncavo da Bahia  
Prof<sup>ª</sup> Dr<sup>ª</sup> 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<sup>ª</sup> Dr<sup>ª</sup> 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<sup>ª</sup> Dr<sup>ª</sup> Magnólia de Araújo Campos – Universidade Federal de Campina Grande  
Prof. Dr. Marcus Fernando da Silva Praxedes – Universidade Federal do Recôncavo da Bahia  
Prof<sup>ª</sup> Dr<sup>ª</sup> Mylena Andréa Oliveira Torres – Universidade Ceuma  
Prof<sup>ª</sup> Dr<sup>ª</sup> 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<sup>ª</sup> Dr<sup>ª</sup> Regiane Luz Carvalho – Centro Universitário das Faculdades Associadas de Ensino  
Prof<sup>ª</sup> Dr<sup>ª</sup> Renata Mendes de Freitas – Universidade Federal de Juiz de Fora  
Prof<sup>ª</sup> Dr<sup>ª</sup> Vanessa Lima Gonçalves – Universidade Estadual de Ponta Grossa  
Prof<sup>ª</sup> Dr<sup>ª</sup> 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<sup>ª</sup> Dr<sup>ª</sup> 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<sup>ª</sup> Dra. Jéssica Verger Nardeli – Universidade Estadual Paulista Júlio de Mesquita Filho  
Prof. Dr. Juliano Carlo Rufino de Freitas – Universidade Federal de Campina Grande  
Prof<sup>ª</sup> Dr<sup>ª</sup> 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 Andreza 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. Lilian 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](http://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<sub>2C</sub> 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 esta 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<sub>2C</sub> 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**

<b>CAPÍTULO 6</b> .....	<b>50</b>
MODELO PARA DETERMINAR PERFIS DE DESEMPENHO ACADÊMICO NA UNNE COM MINERAÇÃO DE DADOS EDUCACIONAIS Julio César Acosta David Luis La Red Martínez <b>DOI 10.22533/at.ed.6132004096</b>	
<b>CAPÍTULO 7</b> .....	<b>59</b>
O USO DO CONTROLADOR LÓGICO PROGRAMÁVEL (CLP) Viviane Alencar Marques Araújo do Nascimento <b>DOI 10.22533/at.ed.6132004097</b>	
<b>CAPÍTULO 8</b> .....	<b>72</b>
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 <b>DOI 10.22533/at.ed.6132004098</b>	
<b>CAPÍTULO 9</b> .....	<b>86</b>
QENADE: APLICATIVO MÓVEL PARA PREPARAÇÃO DE ESTUDANTES PARA O ENADE Helder Guimarães Aragão <b>DOI 10.22533/at.ed.6132004099</b>	
<b>CAPÍTULO 10</b> .....	<b>93</b>
SÍNTESE DOS CONCEITOS DO PENSAMENTO COMPUTACIONAL VOLTADA PARA EDUCAÇÃO BÁSICA BRASILEIRA Nayara Poliana Massa <b>DOI 10.22533/at.ed.61320040910</b>	
<b>CAPÍTULO 11</b> .....	<b>109</b>
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 <b>DOI 10.22533/at.ed.61320040911</b>	

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



<b>SOBRE O ORGANIZADOR.....</b>	<b>210</b>
<b>ÍNDICE REMISSIVO.....</b>	<b>211</b>

## 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<sup>1</sup>).

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 Energia<sup>2</sup>. 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

<sup>2</sup> 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.

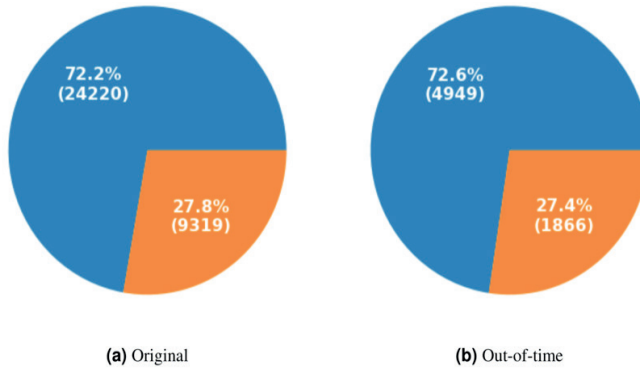


Figure 1: Fraud distribution in original (a) and out-of-time (b) datasets. Regular and irregular situations are displayed in blue and orange, respectively. The figures in bracket represent the true numbers of investigated customers for each category.

### 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.

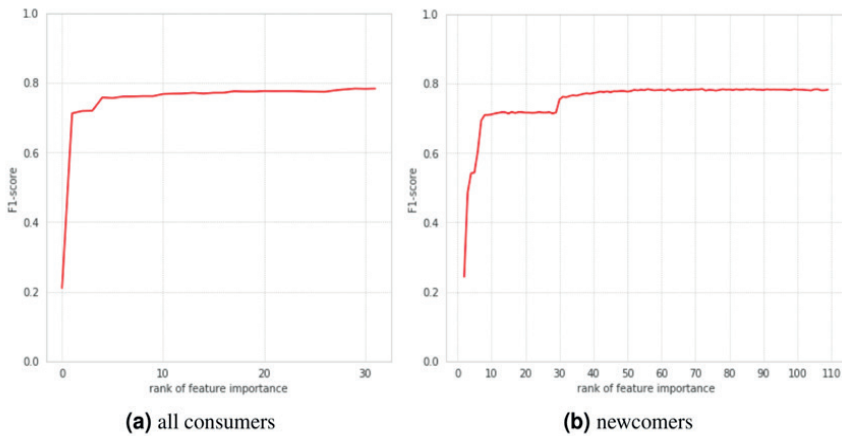


Figure 2: Refinement of predictive models in order to keep only the most relevant features for predictive efficiency for (a) all costumers and (b) costumers with no previous inspection history (newcomers).

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.

## 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](http://www.atenaeditora.com.br) 

[contato@atenaeditora.com.br](mailto:contato@atenaeditora.com.br) 

[@atenaeditora](https://www.instagram.com/atenaeditora) 

[www.facebook.com/atenaeditora.com.br](https://www.facebook.com/atenaeditora.com.br) 

 **Atena**  
Editora

Ano 2020



# TECNOLOGIAS, MÉTODOS E TEORIAS NA ENGENHARIA DE COMPUTAÇÃO

[www.atenaeditora.com.br](http://www.atenaeditora.com.br) 

[contato@atenaeditora.com.br](mailto:contato@atenaeditora.com.br) 

[@atenaeditora](https://www.instagram.com/atenaeditora) 

[www.facebook.com/atenaeditora.com.br](https://www.facebook.com/atenaeditora.com.br) 

 **Atena**  
Editora

Ano 2020