



ENGENHARIA ELÉTRICA E DE COMPUTAÇÃO:

Docência, pesquisa e inovação tecnológica



Lilian Coelho de Freitas
(Organizadora)



Atena
Editora
Ano 2023



ENGENHARIA ELÉTRICA E DE COMPUTAÇÃO:

Docência, pesquisa e inovação tecnológica



Lilian Coelho de Freitas
(Organizadora)



Atena
Editora
Ano 2023

Editora chefe

Profª Drª Antonella Carvalho de Oliveira

Editora executiva

Natalia Oliveira

Assistente editorial

Flávia Roberta Barão

Bibliotecária

Janaina Ramos

Projeto gráfico

Bruno Oliveira

Camila Alves de Cremo

Luiza Alves Batista

Imagens da capa

iStock

Edição de arte

Luiza Alves Batista

2023 by Atena Editora

Copyright © Atena Editora

Copyright do texto © 2023 Os autores

Copyright da edição © 2023 Atena

Editora

Direitos para esta edição cedidos à Atena Editora pelos autores.

Open access publication by Atena Editora



Todo o conteúdo deste livro está licenciado sob uma Licença de Atribuição *Creative Commons*. Atribuição-Não-Comercial-NãoDerivativos 4.0 Internacional (CC BY-NC-ND 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.

Todos os manuscritos foram previamente submetidos à avaliação cega pelos pares, membros do Conselho Editorial desta Editora, tendo sido aprovados para a publicação com base em critérios de neutralidade e imparcialidade acadêmica.

A Atena Editora é comprometida em garantir a integridade editorial em todas as etapas do processo de publicação, evitando plágio, dados ou resultados fraudulentos e impedindo que interesses financeiros comprometam os padrões éticos da publicação. Situações suspeitas de má conduta científica serão investigadas sob o mais alto padrão de rigor acadêmico e ético.

Conselho Editorial**Ciências Exatas e da Terra e Engenharias**

Prof. Dr. Adélio Alcino Sampaio Castro Machado – Universidade do Porto

Profª Drª Alana Maria Cerqueira de Oliveira – Instituto Federal do Acre

Profª Drª Ana Grasielle Dionísio Corrêa – Universidade Presbiteriana Mackenzie

Profª Drª Ana Paula Florêncio Aires – Universidade de Trás-os-Montes e Alto Douro

Prof. Dr. Carlos Eduardo Sanches de Andrade – Universidade Federal de Goiás

Profª Drª Carmen Lúcia Voigt – Universidade Norte do Paraná

Prof. Dr. Cleiseano Emanuel da Silva Paniagua – Instituto Federal de Educação, Ciência e Tecnologia de Goiás

Prof. Dr. Douglas Gonçalves da Silva – Universidade Estadual do Sudoeste da Bahia
Prof. Dr. Eloi Rufato Junior – Universidade Tecnológica Federal do Paraná
Prof^o Dr^a Érica de Melo Azevedo – Instituto Federal do Rio de Janeiro
Prof. Dr. Fabrício Menezes Ramos – Instituto Federal do Pará
Prof^o Dr^a Glécilla Colombelli de Souza Nunes – Universidade Estadual de Maringá
Prof^o Dr^a Iara Margolis Ribeiro – Universidade Federal de Pernambuco
Prof^o Dra. Jéssica Verger Nardeli – Universidade Estadual Paulista Júlio de Mesquita Filho
Prof. Dr. Juliano Bitencourt Campos – Universidade do Extremo Sul Catarinense
Prof. Dr. Juliano Carlo Rufino de Freitas – Universidade Federal de Campina Grande
Prof^o 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
Prof. Dr. Marco Aurélio Kistemann Junior – Universidade Federal de Juiz de Fora
Prof^o Dr^a Maria José de Holanda Leite – Universidade Federal de Alagoas
Prof. Dr. Miguel Adriano Inácio – Instituto Nacional de Pesquisas Espaciais
Prof. Dr. Milson dos Santos Barbosa – Universidade Tiradentes
Prof^o Dr^a Natiéli Piovesan – Instituto Federal do Rio Grande do Norte
Prof^o Dr^a Neiva Maria de Almeida – Universidade Federal da Paraíba
Prof. Dr. Nilzo Ivo Ladwig – Universidade do Extremo Sul Catarinense
Prof^o Dr^a Priscila Tessmer Scaglioni – Universidade Federal de Pelotas
Prof^o Dr Ramiro Picoli Nippes – Universidade Estadual de Maringá
Prof^o Dr^a Regina Célia da Silva Barros Allil – Universidade Federal do Rio de Janeiro
Prof. Dr. Sidney Gonçalo de Lima – Universidade Federal do Piauí
Prof. Dr. Takeshy Tachizawa – Faculdade de Campo Limpo Paulista

Engenharia elétrica e de computação: docência, pesquisa e inovação tecnológica

Diagramação: Camila Alves de Cremo
Correção: Flávia Roberta Barão
Indexação: Amanda Kelly da Costa Veiga
Revisão: Os autores
Organizadora: Lilian Coelho de Freitas

Dados Internacionais de Catalogação na Publicação (CIP)

E57 Engenharia elétrica e de computação: docência, pesquisa e inovação tecnológica / Organizadora Lilian Coelho de Freitas. – Ponta Grossa - PR: Atena, 2023.

Formato: PDF

Requisitos de sistema: Adobe Acrobat Reader

Modo de acesso: World Wide Web

Inclui bibliografia

ISBN 978-65-258-0946-5

DOI: <https://doi.org/10.22533/at.ed.465231601>

1. Energia elétrica. 2. Computação. I. Freitas, Lilian Coelho de (Organizadora). II. Título.

CDD 623.3

Elaborado por Bibliotecária Janaina Ramos – CRB-8/9166

Atena Editora

Ponta Grossa – Paraná – Brasil

Telefone: +55 (42) 3323-5493

www.atenaeditora.com.br

contato@atenaeditora.com.br

DECLARAÇÃO DOS AUTORES

Os autores desta obra: 1. Atestam não possuir qualquer interesse comercial que constitua um conflito de interesses em relação ao artigo científico publicado; 2. Declaram que participaram ativamente da construção dos respectivos manuscritos, preferencialmente na: a) Concepção do estudo, e/ou aquisição de dados, e/ou análise e interpretação de dados; b) Elaboração do artigo ou revisão com vistas a tornar o material intelectualmente relevante; c) Aprovação final do manuscrito para submissão.; 3. Certificam que os artigos científicos publicados estão completamente isentos de dados e/ou resultados fraudulentos; 4. Confirmam a citação e a referência correta de todos os dados e de interpretações de dados de outras pesquisas; 5. Reconhecem terem informado todas as fontes de financiamento recebidas para a consecução da pesquisa; 6. Autorizam a edição da obra, que incluem os registros de ficha catalográfica, ISBN, DOI e demais indexadores, projeto visual e criação de capa, diagramação de miolo, assim como lançamento e divulgação da mesma conforme critérios da Atena Editora.

DECLARAÇÃO DA EDITORA

A Atena Editora declara, para os devidos fins de direito, que: 1. A presente publicação constitui apenas transferência temporária dos direitos autorais, direito sobre a publicação, inclusive não constitui responsabilidade solidária na criação dos manuscritos publicados, nos termos previstos na Lei sobre direitos autorais (Lei 9610/98), no art. 184 do Código Penal e no art. 927 do Código Civil; 2. Autoriza e incentiva os autores a assinarem contratos com repositórios institucionais, com fins exclusivos de divulgação da obra, desde que com o devido reconhecimento de autoria e edição e sem qualquer finalidade comercial; 3. Todos os e-book são *open access*, *desta forma* não os comercializa em seu site, sites parceiros, plataformas de *e-commerce*, ou qualquer outro meio virtual ou físico, portanto, está isenta de repasses de direitos autorais aos autores; 4. Todos os membros do conselho editorial são doutores e vinculados a instituições de ensino superior públicas, conforme recomendação da CAPES para obtenção do Qualis livro; 5. Não cede, comercializa ou autoriza a utilização dos nomes e e-mails dos autores, bem como nenhum outro dado dos mesmos, para qualquer finalidade que não o escopo da divulgação desta obra.

O e-book intitulado “Engenharia elétrica e de computação: Docência, pesquisa e inovação tecnológica” está organizado em 12 capítulos e reúne importantes trabalhos científicos desenvolvido por pesquisadores de Norte a Sul do Brasil, que atuam em renomadas instituições de ensino e pesquisa.

Cada capítulo apresenta uma experiência única, com resultados práticos, consistentes e didáticos. Dessa forma, ao ler este livro, o leitor poderá aprofundar seus conhecimentos em desenvolvimento e teste de softwares, jogos digitais, aprendizagem de máquina, automação, geração de energia, entre outros assuntos relacionados à engenharia elétrica e de computação.

Além de uma base teórica aprofundada, nota-se que os autores de cada capítulo adotaram uma linguagem pedagógica e educativa. Assim, acredito que este livro é um excelente referencial teórico, especialmente para alunos de engenharia elétrica e de computação que estejam desenvolvendo trabalhos de conclusão de curso e que buscam exemplos de aplicações práticas para os conhecimentos teóricos estudados durante o curso. Através da reprodução dos resultados apresentados, é possível por exemplo propor melhorias, apresentar soluções alternativas para os problemas propostos ou desenvolver estudos comparativos. Assim o conhecimento científico avança.

Registro meus sinceros agradecimentos aos autores deste e-book, pelas significativas contribuições e pela parceria com a Atena Editora para tornar o conhecimento científico acessível de forma gratuita.

Aos nossos leitores, desejo um ótimo estudo, repleto de *insights* criativos e inovadores.

Lilian Coelho de Freitas

| | |
|---|-----------|
| CAPÍTULO 1 | 1 |
| ESTUDO E DESENVOLVIMENTO DE UMA METODOLOGIA PARA O PROCESSO DE REVISÃO EM HOMOLOGAÇÕES DE RELEASES ANDROID | |
| Pedro Ivo Pereira Lancellotta | |
| Heryck Michael dos Santos Barbosa | |
| João Gabriel C. Santos | |
| Klirssia M. Isaac Sahdo | |
| Janisley Oliveira De Sousa | |
| Abda Myrria De Albuquerque | |
| Roger Porty Pereira Vieira | |
|  https://doi.org/10.22533/at.ed.4652316011 | |
| CAPÍTULO 2 | 11 |
| ENGENHARIA DE REQUISITOS E SUA IMPORTÂNCIA NO DESENVOLVIMENTO DE <i>SOFTWARE</i> | |
| Henderson Matsuura Sanches | |
|  https://doi.org/10.22533/at.ed.4652316012 | |
| CAPÍTULO 3 | 21 |
| ALGORITMOS NÃO SUPERVISIONADOS E <i>WEB SCRAPING</i> PARA DESCOBERTA DE CONHECIMENTO DE CONHECIMENTO EM REDES SOCIAIS | |
| Carlos Daniel de Sousa Bezerra | |
|  https://doi.org/10.22533/at.ed.4652316013 | |
| CAPÍTULO 4 | 38 |
| MODELOS MENTAIS DIFUSOS PARA TOMADA DE DECISÃO SOBRE O CRESCIMENTO POPULACIONAL EM CIDADES INTELIGENTES USANDO TÉCNICAS COGNITIVAS | |
| Márcio Mendonça | |
| Caio Ferreira Nicolau | |
| Fabio Rodrigo Milanez | |
| Vicente de Lime Gonogora | |
| Luiz Henrique Geromel | |
| Marcio Aurélio Furtado Montezuma | |
| Rodrigo Henriques Lopes da Silva | |
| Marcos Antônio de Matos Laia | |
| Marco Antônio Ferreira Finocchio | |
| Renato Augusto Pereira Lima | |
| Edson Hideki Koroishi | |
| Gilberto Mitsuo Suzuki Trancolin | |
| André Luís Shiguemoto | |
|  https://doi.org/10.22533/at.ed.4652316014 | |
| CAPÍTULO 5 | 57 |
| CUSTOMIZED EXPERIENCE: DIGITAL GAMES POSSIBILITIES BEYOND | |

THEIR MECHANICS

Paula Poiet Sampedro
 Nicholas Bruggner Grassi
 Isabela Zamboni Moschin
 Vânia Cristina Pires Nogueira Valente
 Emilene Zitkus

 <https://doi.org/10.22533/at.ed.4652316015>

CAPÍTULO 673**O USO DA AUTOMAÇÃO DIGITAL PARA AGILIZAR PROCESSOS E SUPRIMIR ERROS NA EXECUÇÃO DE ROTINAS**

Geovane Griesang
 Pedro Henrique Giehl
 Mateus Roberto Algayer

 <https://doi.org/10.22533/at.ed.4652316016>

CAPÍTULO 780**HOSPITAL INTELIGENTE: UMA SIMULAÇÃO DE MONITORAMENTO DE PACIENTES UTILIZANDO INTERNET DAS COISAS**

Júlia Borges Santos
 Vinicius da Rocha Motta
 Saymon Castro de Souza
 Ciro Xavier Maretto

 <https://doi.org/10.22533/at.ed.4652316017>

CAPÍTULO 887**DESENVOLVIMENTO DE UM APLICATIVO NO AMBIENTE *APP DESIGNER* DO *SOFTWARE* MATLAB® PARA PLANEJAMENTO DE TRAJETÓRIA DO ROBÔ PUMA 560**

Eber Delgado de Souza
 Flávio Luiz Rossini
 Luiz Fernando Pinto de Oliveira

 <https://doi.org/10.22533/at.ed.4652316018>

CAPÍTULO 9110**ANÁLISE DE MOTIVAÇÃO E SATISFAÇÃO NA INSTALAÇÃO DE PAINÉIS SOLARES FOTOVOLTAICOS POR MEIO DE MAPAS COGNITIVOS FUZZY**

Márcio Mendonça
 Angelo Feracin Neto
 Carlos Alberto Paschoalino
 Matheus Gil Bovolenta
 Emerson Ravazzi Pires da Silva
 Marcio Aurelio Furtado Montezuma
 Kazuyochi Ota Junior
 Marcos Antonio de Matos Laia
 Augusto Alberto Foggiato
 Vicente de Lima Gongora

Andre Luis Shiguemoto
Francisco de Assis Scannavino Junior
Nikolas Catib Boranelli

 <https://doi.org/10.22533/at.ed.4652316019>

CAPÍTULO 10..... 126

DESENVOLVIMENTO E AVALIAÇÃO DE UM CONTROLADOR PREDITIVO NÃO-LINEAR BASEADO EM MODELO QUASILINEAR MODIFICADO

Manoel de Oliveira Santos Sobrinho

Adhemar de Barros Fontes

 <https://doi.org/10.22533/at.ed.46523160110>

CAPÍTULO 11 140

IMPLEMENTAÇÃO DE ATERRAMENTO EM UMA RESIDÊNCIA COM DR PARA ELIMINAR O CHOQUE ELÉTRICO

Eliandro Marquetti

Elielton Christiano de Oliveira Metz

Luciana Paro Scarin Freitas

 <https://doi.org/10.22533/at.ed.46523160111>

CAPÍTULO 12..... 156

PANORAMA DAS FONTES TÉRMICAS PARA GERAÇÃO DE ENERGIA ELÉTRICA NO BRASIL

Bruno Knevez Hammerschmitt

Felipe Cirolini Lucchese

Marcelo Bruno Capeletti

Renato Grethe Negri

Leonardo Nogueira Fontoura da Silva

André Ross Borniatti

Fernando Guilherme Kaehler Guarda

Alzenira da Rosa Abaide

 <https://doi.org/10.22533/at.ed.46523160112>

SOBRE A ORGANIZADORA 171

ÍNDICE REMISSIVO..... 172

ALGORITMOS NÃO SUPERVISIONADOS E WEB SCRAPING PARA DESCOBERTA DE CONHECIMENTO DE CONHECIMENTO EM REDES SOCIAIS

Data de aceite: 02/01/2023

Carlos Daniel de Sousa Bezerra

A Respectiva Publicação é resultado do Trabalho de Conclusão de Curso apresentado ao Curso à Distância de Especialização em Inteligência Artificial e Aprendizado de Máquina da Pontifícia Universidade Católica de Minas Gerais como requisito parcial à obtenção do título de especialista.

RESUMO: Os algoritmos denominados *word embedding (WB)*, são técnicas computacionais capazes de realizar a representação de palavras como vetores densos e conseqüentemente transformação de dados textuais em numéricos para processamento computacional eficaz, permitindo a execução de cálculos matemáticos importantes no campo de mineração de textos e análise de sentimentos, bem como o uso de técnicas de aprendizagem de máquinas e inteligência artificial (IA). Este trabalho, originalmente um trabalho de conclusão de curso, atua na linha de desenvolvimento de “**Inteligência Artificial na Análise de Dados da Web e das Redes Sociais**” e baseia-se na seguinte hipótese: se é possível hibridizar

as técnicas de *WS* e *word2vec*, bem como aplicação de aprendizagem de máquina para obter conhecimento não obvio em redes sociais, mais especificamente no *Instagram*. A metodologia e software proposto possui aplicações comerciais, principalmente na área de jornalismo de dados e administração de redes sociais.

PALAVRAS-CHAVE: Exploração de Dados, Processamento de Linguagem Natural, *word2vec*.

ABSTRACT: Algorithms called *word embedding (WB)* are computational techniques capable of performing the representation of words as dense vectors and consequently transforming textual data into numerical ones for efficient computational processing, allowing the execution of important mathematical calculations in the field of text mining and analysis. of feelings, as well as the use of machine learning techniques and artificial intelligence (AI). This work, originally a course completion work, operates in the line of development of “**Artificial Intelligence in Web and Social Networks Data Analysis**” and is based on the following hypothesis: if it is possible to hybridize *WS* and *word2vec* techniques, as well as applying machine learning to obtain

non-obvious knowledge on social networks, more specifically on Instagram. The proposed methodology and software has commercial applications, mainly in the field of data journalism and social media management.

KEYWORDS: Data Exploration, Natural Language Processing, word2vec.

1 | INTRODUÇÃO

1.1 Contextualização

1.1.1 Redes sociais e exploração de dados

De acordo com Junior (2008) a partir da 3ª revolução industrial, com advento da *internet* e comunicação de dados, grandes volumes de informações digitalizadas passaram a surgir no cotidiano da sociedade global. A atual implementação de conceitos de *indústria 4.0* passou a aumentar o fluxo de dados nas organizações, assim um dos principais desafios desta década é lidar com esta massa crescente de dados.

Desta forma a vida da sociedade tende, cada vez mais, a ser pautada em dados e as redes sociais contribuem significativamente com o crescente volume massivo de informações. A extração e compreensão destas informações passa a ser uma estratégia competitiva entre as grandes organizações, seja para obtenção de dados valiosos e consequente vantagem comercial, quanto para auxílio nas tomadas de decisões internas, bem como a análise do *feedback* dos usuários. Consequentemente surgem novas áreas de estudo em computação, como a descoberta de conhecimento (*Knowledge Discovery in Database – KDD*), mineração de texto na *web* e recuperação da informação em redes sociais agregadas com as tradicionais técnicas de processamento de linguagem natural (PNL).

A *internet* é a mais popular fonte de informação do mundo e é utilizada para instruir e esclarecer pessoas sobre assuntos diversos. Por exemplo, segundo uma pesquisa eleitoral realizada no ano de 2018, pelo Instituto Paraná, 42,5% dos entrevistados afirmaram que a *internet* é o principal meio de informação sobre as eleições daquela época, superando meios de comunicação como a televisão e jornais impressos. “As mídias sociais conferem nova perspectiva ao termo cidadania, com um público que lê, interpreta e passa adiante as mensagens recebidas. As redes provocaram uma mudança na forma de se comunicar e, ao alterar a comunicação, as modificaram igualmente as relações sociais” (CAGGIANO, 2014,p.239). Dessa forma, minerar dados na *web* se torna importante em diversas análises e áreas de conhecimento, principalmente para análise de redes sociais.

Uma das técnicas mais conhecidas para minerar textos na *Internet* é a *raspagem de dados (web scraping - WS)*. Segundo Guimarães (2018) define-se a técnica *WS* como um programa computacional capaz de obter informações e dados em sites da *web* de forma ágil e automática, emulando o comportamento da navegação humana.

Como supracitado, o *WS* é uma ferramenta poderosa e ampla, por tanto é inevitável que ela possa ser utilizada para fins antiéticos, uma vez que a legislação no Brasil ainda é incerta e nebulosa no campo jurídico. Segundo Michel (2019), a proteção dos direitos autorais é uma das preocupações legais desta prática, entretanto a proteção de direitos autorais engloba principalmente trabalhos criativos, não incluindo estatísticas ou fatos, itens estes que a maioria dos *scapers* (programa do *WS*) procuram executar em sua busca. Ou seja, projetar um *software* baseado em *WS* que colete informações estatísticas ou dados factuais sobre determinado site não é violação de direitos autorais. Além disso os *softwares* que não coletam dados privados dos usuários, não desrespeitam as políticas de privacidade da página *web* (*robot.txt*), utilize os dados que estejam dispostos de forma pública e não sobrecarregue a página com números exagerados de requisições são vistos como práticas legais e éticas de coleta de dados (COELHO e BEZERRA 2020).

1.1.2 Representação de Palavras

Os algoritmos denominados *word embedding* (*WB*), são técnicas computacionais capazes de realizar a representação de palavras como vetores densos e conseqüentemente transformação de dados textuais em numéricos para processamento computacional eficaz, permitindo a execução de cálculos matemáticos importantes no campo de mineração de textos e análise de sentimentos, bem como o uso de técnicas de aprendizagem de máquinas e inteligência artificial (*IA*).

Destaca-se na área de *word embedding* o algoritmo de aprendizagem não supervisionado *word2vec*, proposto por Mikolov et.al (2013). Segundo Li (2019), esta técnica vem se tornando a mais popular para vetorização utilizando redes neurais artificiais (*RNA*). O principal objetivo do *word2vec* objetivo é extrair conhecimento semântico em documentos textuais. Algumas das aplicações destes algoritmos estão na área de análise de sentimentos, sistemas de recomendação e mineração de textos.

O *word2vec* pode ser dividido em duas técnicas inteligentes: O Saco de Palavras Contínua (*Continuous Bag of Words - CBOW*) e o *Skip Gram*. Ambos os algoritmos utilizam *RNA* e possuem a capacidade de prever palavras similares dentro de uma janela espacial pré sintonizada. A diferença entre eles é a arquitetura da rede neural, onde o *Skip Gram* é capaz de prever o contexto de uma determinada palavra, já o *CBOW* prediz a palavra em função de um determinado contexto. A Figura 1 exemplifica espaço vetorial da representação de palavras proposta pelo *word2vec*, bem como a arquitetura de predição *CBOW* associada a uma rede neural artificial para classificação de uma palavra em função do contexto.

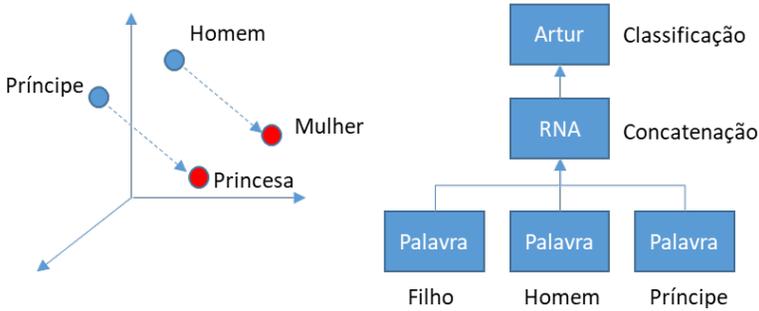


Figura 1- Fluxo de Operação do Word2vec. Autoria Própria

Este trabalho de conclusão de curso atua na linha de desenvolvimento de **“Inteligência Artificial na Análise de Dados da Web e das Redes Sociais”** e baseia-se na seguinte hipótese: se é possível hibridizar as técnicas de *WS* e *word2vec*, bem como aplicação de aprendizagem de máquina para obter conhecimento não obvio em redes sociais, mais especificamente no *Instagram*. A metodologia e *software* proposto possui aplicações comerciais, principalmente na área de jornalismo de dados e administração de redes sociais. Tem-se como objetivo principal a criação de uma metodologia para implementação de sistema inteligente que explore estes dados em textos de comentários dispostos por usuários desta rede social.

1.1.3 Aplicação em Ambiente Comercial

1.2 O problema proposto

Propõe-se neste trabalho a hibridização de quatro técnicas computacionais para o desenvolvimento metodológico de exploração de dados e análise de sentimento em redes sociais:

- *WS*: técnica responsável para a obtenção dos dados referentes aos comentários na rede social *Instagram*;
- *word2vec*: Técnica não supervisionada para exploração semântica de palavras e vetorização numérica para aplicação de aprendizagem de máquina;
- *doc2vec*: Variação da técnica *word2vec*, porém busca realizar a representação numérica de um documento por inteiro.
- *k-means*: Técnica de Aprendizado de máquina para classificação dos comentários obtidos na rede social.

Observa-se que os algoritmos de inteligência artificial são respectivamente: o ***word2vec***, ***doc2vec*** e o **classificador *k-means***, pois eles utilizam de uma estrutura não supervisionada (ou semi supervisionada como citada por alguns autores) para obter

aprendizado sobre os dados dispostos. O *WS* é uma metodologia complementar para a extração de dados na *web*.

A linguagem de programação utilizada para implementação do *software* proposto é o *Python*. A Tabela 1 dispõe os frameworks recorridos para auxílio da construção dos algoritmos.

| Framework | Finalidade |
|-------------------|---|
| <i>Gensim</i> | Implementação dos algoritmos <i>word2vec</i> e <i>doc2vec</i> |
| <i>nltk</i> | Processamento de linguagem natural |
| <i>Pandas</i> | Visualização dos Dados Obtidos |
| <i>Requests</i> | Realização do <i>WS</i> no Instagram |
| <i>sklearn</i> | Implementação do classificador <i>kmeans</i> |
| <i>matplotlib</i> | Visualização em forma de gráfico dos dados e resultados |

Tabela 1- Frameworks Utilizados

Utiliza-se no *WS* de dados contas do *Instagram* de órgãos governamentais para atendimento ao público, onde o feedback da população passa a ser importante na relação usuário/administrador, podendo ser utilizado, inclusive, para o compreensão da satisfação da população ao atendimento de um determinado órgão ou governo.

Para a execução dos testes do algoritmo proposto, o perfil selecionado para a coleta de dados referentes aos comentários dos usuários foi o ***Instagram do Ministério da Cidadania***, órgão governamental para atendimento à população e prestação de serviços. Atualmente, em função do período de crise sanitária em que vivemos (COVID-19) é importante compreender o nível de satisfação dos usuários e necessidades da população principalmente em relação a prestação de serviços públicos. Os administradores da página devem compreender também de que maneira devem realizar o feedback, ou responder, estes usuários e a descoberta de conhecimento, ou conteúdo, referente aos comentários pode auxiliar neste quesito.

A coleta e análise de dados proposta possui o seguinte objetivo geral: Obter conhecimento não obvio e útil nos comentários dos usuários, isto é, obter informações e padrões a partir de dados processados, estruturados e organizados intencionalmente para tal fim. O conhecimento não obvio é obtido por características consideradas como 'julgamentos, ou seja questionamentos aprofundados sobre a base de dados em busca de conhecimento novo e não observável inicialmente.

Desta forma, as seguintes características são exploradas nos dados obtidos:

Frequência da Ocorrência de Palavras, Nuvem de Palavras e Similaridade Vetorial. Ao final, cria-se um modelo inteligente capaz de obter as três características quando submetido à um determinado comentário. A Figura 2 ilustra o fluxo do modelo proposto.

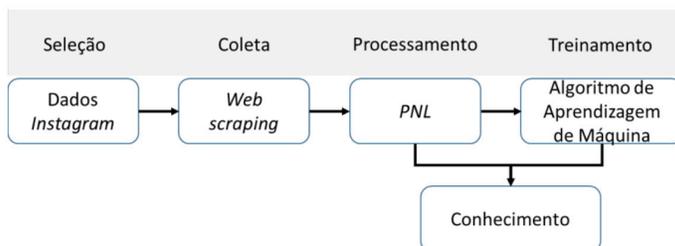


Figura 2- Etapas do Modelo Proposta. Autoria Própria

São coletados dados referentes a 20 publicações no *Instagram* ministério da cidadania (@**mincidadania**), que ocorreram entre os dias 03 a 17 de Agosto de 2020.

2 | COLETA DE DADOS

Nesta etapa, verifica-se a aplicação dos dois primeiros blocos da Figura 2, ou seja, a seleção dos dados e o *WS*. Para coleta dos dados no *Instagram* utiliza-se de uma interface de aplicação (API), ou seja, um padrão de comunicação entre servidores e *softwares* mesmo que escritos em linguagem diferentes. Portanto, realiza-se uma requisição a API que fornece uma resposta em formato JSON (*Java script object notation*). Ressalta-se que para coleta de dados utilizou-se da API pública, e mais antiga, do *Instagram* ('/?__a=1') e desta forma alguns recursos são limitados, como por exemplo o número de comentários por publicação, que são reduzidos em relação ao número de comentários originais.

Para realizar a requisição a API utiliza-se do *framework* da linguagem *Python requests*. Além desta ferramenta permitir o *download* de dados obtidos por uma requisição ela possibilita ainda a decodificação de JSON para linguagem *Python* dispondo os dados obtidos sob a forma de dicionário. Como os principais dados coletados para o problema em questão são os comentários dos usuários, necessita-se realizar o acesso a chave específica do dicionário criado.

A Tabela 2 dispõe as 37 chaves com o conteúdo obtido na requisição. A chave "***edge_media_to_parent_comment***" contém os comentários dos usuários em cada publicação explorada.

| Dicionário da Requisição | | |
|---------------------------------------|----------------------------------|--|
| Chave | Chave | Chave |
| <i>accessibility_caption</i> | <i>fact check overall rating</i> | <i>edge media to sponsor user</i> |
| <i>caption is edited</i> | <i>gating info</i> | <i>edge media to tagged user</i> |
| <i>commenting_disabled_for_viewer</i> | <i>has_ranked_comments</i> | <i>edge_related_profiles</i> |
| <i>comments_disabled</i> | <i>id</i> | <i>edge_web_media_to_related_media</i> |
| <i>dimensions</i> | <i>is_ad</i> | <i>fact check information</i> |
| <i>display_resources</i> | <i>is video</i> | <i>viewer in photo of you</i> |
| <i>display url</i> | <i>location</i> | <i>taken at timestamp</i> |
| <i>edge_media_preview_comment</i> | <i>media_overlay_info</i> | <i>tracking_token</i> |
| <i>edge media preview like</i> | <i>media_preview</i> | <i>viewer can reshare</i> |
| <i>edge media to caption</i> | <i>owner</i> | <i>viewer has liked</i> |
| <i>edge_media_to_hoisted_comment</i> | <i>sensitivity_friction_info</i> | <i>viewer_has_saved</i> |
| <i>edge_media_to_parent_comment</i> | <i>shortcode</i> | <i>viewer_has_saved_to_collection</i> |

Tabela 2 - Parâmetros do Dicionário Resultante da Requisição

Inicialmente os comentários de uma respectiva publicação são obtidos com o acesso a chave do dicionário da requisição, conforme supracitado. Em seguida este comentário é salvo em uma variável do tipo lista. O fluxo se repete para cada publicação. A Figura abaixo ilustra o trecho da codificação do software desenvolvido para obter a coleta dos dados. A variável *shortcode* é uma identificação do *Instagram* para cada publicação em uma determinada página.

```

###
#Coletando no site do Ministério da Cidadania
i = 1
comentariosfinal = []
shortcode = ['CD_7zSCJA91', 'CD_kkhUJ8Cx', 'CD9Nk59pMtg', 'CD4dw9gpR6x', 'CD1bIAEpGe1',
            'CDzH0ZEpyOP', 'CDy0rFNpIT6', 'CDwyzPsJqaU',
            'CDwHMRxJO8T', 'CDtgOqSpago', 'CDq7tsEpabq', 'CDoUbjMJa6n',
            'CDmketmJqIU', 'CDjQyJlJmIz', 'CDhLASfJ6Df', 'CDgx5wapAvR',
            'CDepxjHJlGt', 'CDb8gampkKp', 'CDWhsjoJuYf', 'CDRr-SZp8mt' ]
for shortlist in shortcode:
    url2 = 'https://www.instagram.com/p/' + shortlist + '/?_a=1'
    r2 = requests.get(url2)
    dados2 = r2.json()
    comentarios = dados2['graphql']['shortcode_media']['edge_media_to_parent_comment']['edges']
    for comentario in comentarios:
        usuario = 'usuario'
        texto = comentario['node']['text']
        comentariosfinal.append(texto)
        print(usuario, texto, '\n')
        i+=1
print (i)

```

Figura 3 - Trecho da Codificação Para Extração dos Dados

Após o preenchimento por completo da lista, os dados são salvos em uma planilha no formato *CSV* para facilitar a visualização e futuro carregamento dos dados no mesmo

programa.

```
###  
#Criando a Planilha  
arquivo = open('mincida_fil.csv', 'w', encoding='utf8', newline='')  
planilha = csv.DictWriter(arquivo, fieldnames=['Source', 'Target'], )  
planilha.writeheader()  
for comentario in novo_c:  
    # com = comentario['node']['text']  
    planilha.writerow({'Source': 'usuario', 'Target': comentario})  
arquivo.close()
```

Figura 4 - Trecho da codificação para criação da planilha

3 I PROCESSAMENTO/TRATAMENTO DE DADOS

A etapa de processamento e tratamento consiste em operar sobre os dados obtidos para que estes possam estar aptos a serem utilizados por técnicas de aprendizagem de máquina. Nesta etapa o PNL é fundamental para desempenhar este papel. Além disso, a visualização dos dados, com o auxílio da biblioteca *pandas*, permite identificar atributos que devem ser removidos, como por exemplo os *emojis* utilizados pelos usuários.

Do PNL, utiliza-se as seguintes técnicas: *tokenização* e *stopwords*. A tokenização consiste em separar, por exemplo um comentário unificando em forma de “string”, em palavras separadas, ou seja, segmentando-as em *strings* separadas. Os algoritmos *word2vec* e *doc2vec* utilizam *tokens* (palavras separadas) para desempenhar seus cálculos e por isto a importância da aplicação dessa etapa.

As *stopwords* são palavras que não trazem conteúdo semântico para um texto. Desta forma a sua filtragem permite a retirada destas expressões de comentários aumentando o processamento computacional do algoritmo.

A Figura 4 apresenta uma amostra dos 306 comentários obtidos nas 20 publicações analisadas. Ressalta-se a limitação da API em obter todos os comentários originais. Por motivos de privacidade o nome do perfil dos usuários foi suprimido. Observa-se a presença de diversos *emojis* (vide detalhe das setas em vermelho) nos dados iniciais obtidos, sem o processamento e tratamento necessário.

| Usuário | Comentário |
|---------|--|
| usuario | Eu morro no sítio e não tenho nem uma caixa d'água já coloquei meu esposo colocou o nome é ainda não veio as cisternas |
| usuario | Tira o povo dessa avaliação de bosta🤔🤔🤔🤔🤔🤔 @dataprev @mincidadania @jairmessiasbolsonaro @onyxlorenzoni 🤔🤔🤔🤔🤔🤔@datena.ofc |
| usuario | Tá mais e o pessoal da Avaliação? Quando é a Estimativa? |
| usuario | 🤔🤔 |
| usuario | Cheio de golpista por aí tempo todofico com pena de quem não tem esclarecimento ,cai fácil 😊 |
| usuario | Quando sai a quinta parcela do auxílio |
| usuario | 🤔🤔🤔🤔 |
| usuario | Avaliação eterno #ficamosprtras |
| usuario | 🤔🤔🤔🤔 |
| usuario | Tira nois dessa avaliação 🤔🤔🤔🤔@jairmessiasbolsonaro @mincidadania @dataprev @caixa @datenareal @datena.ofc |
| usuario | @mincidadania quando será dado uma posição para liberação do auxílio pra quem teve a 3° e 4° parcela do auxílio bloqueada (em reavaliação) ? |
| usuario | Quem tá Avaliação 4 meses |
| usuario | Avaliação de pessoas desde abril ninguém fala ! |
| usuario | Impacto desastroso na vida de milhares de cidadãos desempregados e que estão passando fome pois ficaram para trás com seus saldos bloqueados indevidamente , desde maio, pelo INEFICIENTE MINISTÉRIO DA CIDADANIA. #AuxilioEmergencial só no nome . Os recursos não lhes pertencem, PAGA O POVO que estão HÁ 4 MESES em reavaliação e PASSANDO FOME. |

Figura 5- Exemplo de Comentários obtidos no Instagram sem Tratamento

| Usuário | Comentário |
|---------|---|
| usuário | ['morro', 'sítio', 'caixa', 'd'água', 'coloquei', 'esposo', 'colocou', 'nome', 'ainda', 'veio', 'cisternas'] |
| usuário | ['tira', 'povo', 'dessa', 'avaliação', 'bosta', '@dataprev', '@mincidadania', '@jairmessiasbolsonaro', '@onyxlorenzoni', '@datena.ofc'] |
| usuário | ['tá', 'pessoal', 'avaliação?', 'é', 'estimativa?'] |
| usuário | ['anos', 'esperando', 'cisterna'] |
| usuário | ['cheio', 'golpista', 'aí', 'tempo', 'todo', '.....fico', 'pena', 'esclarecimento', 'cai', 'fácil'] |
| usuário | ['saí', 'quinta', 'parcela', 'auxílio'] |
| usuário | ['avaliação', 'eterno', '#ficamosprtras'] |
| usuário | ['tira', 'nois', 'dessa', 'avaliação', '@jairmessiasbolsonaro', '@mincidadania', '@dataprev', '@caixa', '@datenareal', '@datena.ofc'] |
| usuário | ['https://www.in.gov.br/en/web/dou/-/resolucao-n-972-de-11-de-agosto-de-2020-271713731'] |
| usuário | ['recebi', 'hj?', '????'] |
| usuário | ['@mincidadania', 'dado', 'posição', 'liberação', 'auxílio', 'pra', '3°', '4°', 'parcela', 'auxílio', 'bloqueada', '(em', 'reavaliação)'] |
| usuário | ['tá', 'avaliação', '4', 'meses'] |
| usuário | ['avaliação', 'pessoas', 'desde', 'abril', 'ninguém', 'fala'] |
| usuário | ['impacto', 'desastroso', 'vida', 'milhares', 'cidadãos', 'desempregados', 'passando', 'fome', 'pois', 'ficaram', 'trás', 'saldos', 'bloqueados', 'indevidamente', 'desde', 'maio', 'ineficiente', 'ministério', 'cidadania', '#auxilioemergencial', 'nome', 'recursos', 'pertencem', 'paga', 'povo', '4', 'meses', 'reavaliação', 'passando', 'fome.'] |

Figura 6- Exemplo de Comentários obtidos no Instagram com Tratamento

Na Figura 5 é possível observar os mesmos comentários, entretanto com a remoção de *emojis*, *stopwords* e sob a forma de *tokens*. Após o tratamento os dados foram reduzidos para 277 comentários, pois 29 comentários eram formados exclusivamente por *emojis*.

4 | CRIAÇÃO DE MODELOS DE MACHINE LEARNING

4.1 Algoritmos Word2vec e Doc2vec

O modelo de *machine learning* utilizado para extrair conhecimento dos dados obtidos no perfil do ministério da cidadania é o *word2vec*. Este algoritmo faz parte do processo de *embedding*. Este modelo baseia-se em uma rede neural artificial, ou seja, um modelo

computacional do neurônio biológico humano, capaz de realizar cálculos complexos e aproximar funções universais. A rede neural artificial utilizada no processo de *embedding* é relativamente simples, duas camadas rasas (*shallow*) e pode ser ilustrada pela Figura 5.

A arquitetura de *embedding* ilustrada é a *Skip-Gram*. A rede neural recebe em sua entrada vetores de palavras sob a forma de *tokens*, codificados em *hot-hot vector*, ou seja, um vetor esparso com codificação binária (1 ou 0). Este vetor é multiplicado por um conjunto de pesos da matriz de projeção P (referentes a camada oculta da rede). Os valores são concatenados na matriz de saída M onde aplica-se a função de ativação *softmax*. Deseja-se maximizar a probabilidade de ocorrência de uma palavra w a partir das palavras de contexto (palavras próximas a ela). Os pesos da camada de projeção P representam matematicamente os vetores de *embedding* e são estes elementos que a rede neural deve aprender e atualizar durante o processo de treinamento.

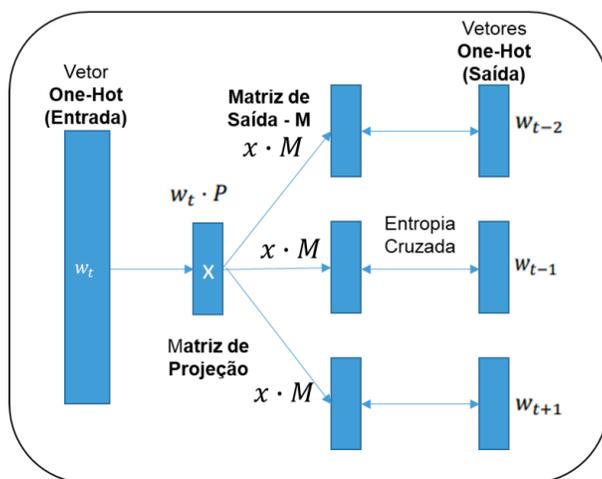


Figura 7 - Rede Neural Artificial em Arquitetura Skip-Gram. Autoria Própria

O modelo *doc2vec* é uma variação do *word2vec* e possui o mesmo princípio de funcionamento. A diferença é que este algoritmo é especializado em obter vetores para parágrafos, frases e documentos maiores. Desta forma acrescenta-se na camada de entrada da rede neural um vetor referente ao número (*tag*) do parágrafo para que ele atue como uma memória do tópico do parágrafo. Portanto, devido a sua maior aplicabilidade no problema em questão, para a obtenção dos vetores referente aos comentários dos usuários na rede social *Instagram* utiliza-se o modelo *doc2vec*.

4.2 Framework GENSIM

O algoritmo *doc2vec* é implementado em linguagem *Python* utilizando o *framework*

GEMSIM. A Figura abaixo apresenta o trecho da codificação utilizada na programação. O primeiro bloco apresenta a importação da biblioteca e um laço (**for**) que atua sobre a planilha de dados criando a lista de *tokens* que será repassada ao modelo. No segundo bloco o modelo *doc2vec* é importado do módulo GEMSIM. A parametrização do algoritmo envolve os seguintes valores: *vector size*, *window*, *min_count* e *epochs*. O *vector size* representa o tamanho do *embedding* a ser desenvolvido. Já o parâmetro *window* determina o contexto, isto é tamanho da janela, ou faixa, em que se considera um palavra-chave e as que a cercam. O *min_count* é o número mínimo de repetições de um termo para considerá-lo no modelo. O número de épocas de treinamento da rede neural é dado pela variável *epochs*.

```
###
#Testando o Modelo Doc2vec
import gensim
token_train = [] #Gerando Tokens para Treinamento
for i in range(0,len(plan)):
    if i in plan.index:
        token_c = plan["Target"][i]
        token_train.append(token_c)
###
from gensim.models.doc2vec import Doc2Vec, TaggedDocument
from gensim.models import doc2vec
model = gensim.models.doc2vec.Doc2Vec(vector_size=100, window=5, min_count=5,
                                     epochs=100)
sentences = [TaggedDocument(d, [i]) for i, d in enumerate(token_train)]
model.build_vocab(sentences)
```

Figura 8 - Trecho da codificação para determinação do modelo doc2vec

5 | ANÁLISE E EXPLORAÇÃO DOS DADOS

A etapa de análise e exploração de dados tem como objetivo demonstrar os dados obtidos de forma clara e visual, através de gráficos e demais representações, ao usuário do programa. Com as informações dispostas, estes usuários têm condições de inferir sobre os resultados da análise e tirar suas conclusões. Para explorar os dados utiliza-se das seguintes representações:

- Nuvem de Palavras
- Distribuição de Frequências
- Semelhança do Cosseno entre Vetores

Utiliza-se do gráfico de distribuição de frequências entre as palavras mais utilizadas nos comentários, bem como representação visual da nuvem de palavras.

Através da aplicação dos algoritmos *word2vec* e *doc2vec* é possível obter a probabilidade da similaridade entre palavras do texto e assim extrair o contexto destes comentários, sendo possível também, a compreensão mesmo que de forma não sofisticada

e não automatizada, dos sentimentos destes usuários, correlacionando adjetivos e advérbios similares a mesma.

Foram realizadas três baterias de testes para a melhor determinação modelo *doc2vec*, onde alterou-se os hiperparâmetros do modelo, como: Número de camadas da rede (*workers*), Tamanho do vetor a ser gerado no *embedding* (*vector_size*), tamanho da janela (*window*) e épocas de treinamento (*epochs*). Também foi alterado as topologias do modelo (*skip gram* ou *CBOW*).

| Bateria de Testes para Treinamento do Modelo <i>Doc2vec</i> | | |
|---|---|-----------|
| Nº da Bateria | Configuração | Modelo |
| 1ª | vector_size = 50, workers = 2, epochs = 50, window = 3, min count = 1 | Skip Gram |
| 2ª | vector_size = 100, workers = 5, epochs = 100, window = 10, min count = 5 | Skip Gram |
| 3ª | vector_size = 50, workers = 2, epochs = 50, window = 3, min count = 1 | CBOW |
| 4ª | vector_size = 100, workers = 5, epochs = 100, window = 10, min count = 5 | CBOW |

6 | APRESENTAÇÃO DOS RESULTADOS

6.1 Distribuição de Frequências

Os resultados referentes à análise e exploração dos dados obtidos por *WS* são apresentados nas Figura 9, Figura 10, Figura 11 e Figura 12. Foram utilizados, respectivamente, gráficos que dispõem a frequência de ocorrência das palavras nos comentários e a nuvem de palavras. A Figura 9 apresenta a distribuição de frequências dos 20 principais termos presentes nos comentários. Entretanto observa-se que alguns caracteres considerados “*stopwords*” não foram filtrados com eficiência (caracteres “q” e “d”). Além disso observou-se que a palavra auxílio apareceu duas vezes no gráfico, com e sem acento, devido as diferentes formas de escrever dos usuários que interagem com o perfil o *Instagram*.

Na Figura 10 a técnica de processamento de linguagem natural “*Stemming*” (redução da palavra ao seu radical) foi aplicada, resolvendo o problema de variação da escrita da palavra em função da acentuação. Observa-se, por exemplo, que a palavra “avaliação” foi reduzida para “avali”. As variações “avali”, “auxili” e “receb” foram as três com maiores frequências de ocorrência, respectivamente 25, 22 e 20 repetições.

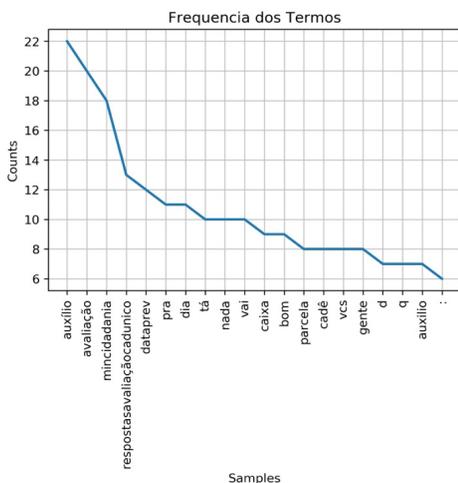


Figura 9- Frequência de Palavras (20 Termos)

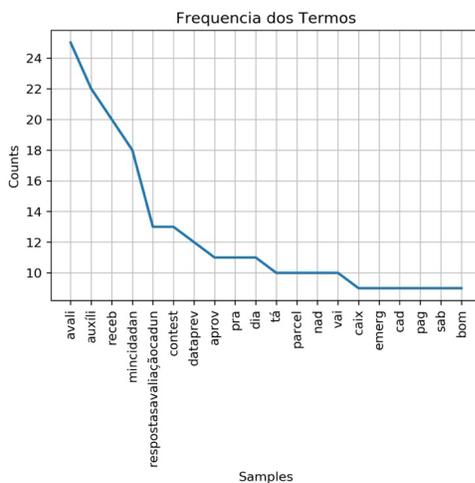


Figura 10- Frequência de Palavras com aplicação de Steam (20 Termos)

6.2 Nuvem de Palavras

A nuvem de palavras é uma representação visual dos principais termos obtidos nos comentários dos usuários. A partir desta representação pode-se inferir palavras chaves e os assuntos mais comentados. A importância de cada palavra é representada pelo tamanho da fonte e cor. A Figura 11 apresenta a nuvem de palavras dos dados sem tratamento de *Stemming* e a Figura 12 os dados com o tratamento de *Stemming*.

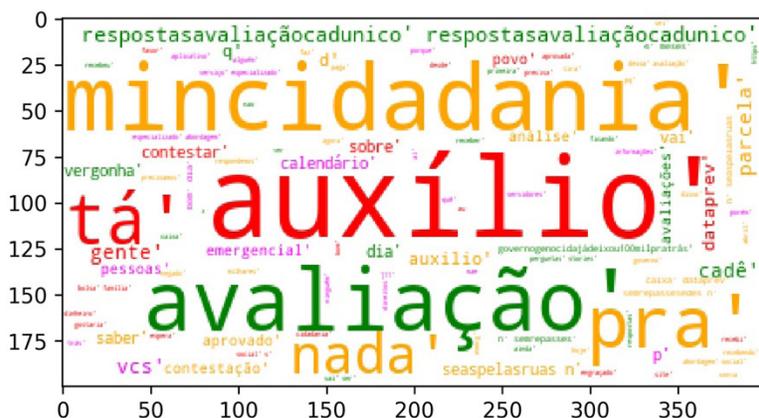


Figura 11- Nuvem de Palavras (20 Termos)

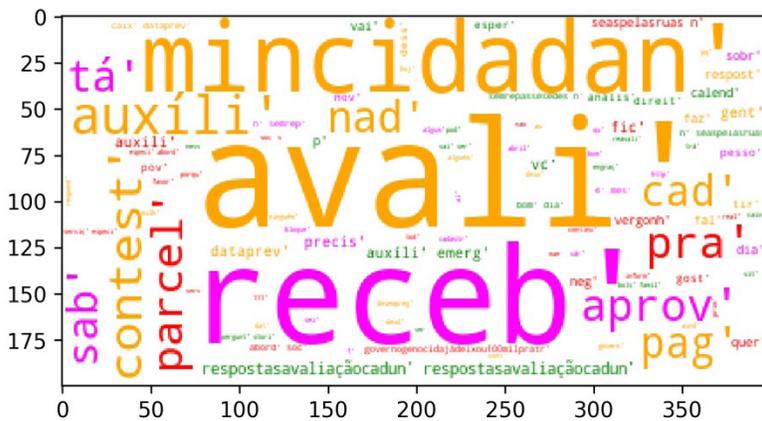


Figura 12 - Nuvem de Palavras com Steam (20 Termos)

De acordo com os dados obtidos e dispostos nas Figuras supracitadas deste tópico, observa-se que grande parte dos comentários nas publicações analisadas são referentes a questionamentos sobre o auxílio emergencial. De acordo com os termos “avaliação”, “auxílio” e “recebimento” possivelmente diversas pessoas no Brasil estão com problemas no recebimento destas parcelas em função da demora na avaliação, ou triagem, realizada pelo governo.

A aplicação do algoritmo *doc2vec* permite a análise de similaridade e analogia entre os comentários. Desta forma, é possível deduzir e compreender com maior precisão a necessidade destes usuários que estão interagindo com a rede social. Como já explicado no **Tópico 4** esta similaridade ocorre em função da similaridade do cosseno entre os vetores gerados durante o processo de *embedding*. Desta forma os comentários não foram tratados como simples “strings” e sim como vetores no espaço vetorial.

6.3 Similaridade e Semelhança do Cosseno

As Tabela 3 a Tabela 6 apresentam os resultados referentes as baterias de testes de similaridade do modelo *doc2vec* treinado com diferentes parâmetros das palavras mais similares com termo de maior frequência dentre os comentários, a palavra avaliação.

| Bateria 1 |
|--|
| % de Similaridade : Palavra "Avaliação" |
| ja', 0.3834 |
| saber', 0.3510 |
| próprio', 0.3502 |
| voltar', 0.3471 |
| brasileiro', 0.3370 |
| reverter', 0.3350 |
| mes', 0.3267 |
| repassé', 0.3260 |
| julho', 0.3249 |
| 2020', 0.3201 |

Tabela 3 – Resultados da Bateria 1

| Bateria 2 |
|--|
| % de Similaridade : Palavra "Avaliação" |
| repassé', 0.3075 |
| cadê', 0.2485 |
| direitos', 0.2424 |
| fica', 0.21457 |
| bonito', 0.1969 |
| federal', 0.1938 |
| parte', 0.1927 |
| recebeu', 0.1895 |
| todos', 0.1890 |
| governo', 0.1876 |

Tabela 4 - Resultados da Bateria 2

| Bateria 3 |
|--|
| % de Similaridade : Palavra "Avaliação" |
| ja', 0.3834 |
| saber', 0.3510 |
| próprio', 0.3502 |
| voltar', 0.3471 |
| brasileiro', 0.3370 |
| reverter', 0.3350 |
| mes', 0.3267 |
| repassé', 0.3260 |
| julho', 0.3249 |
| 2020', 0.3201 |

Tabela 5- Resultados da Bateria 3

| Bateria 4 |
|--|
| % de Similaridade : Palavra "Avaliação" |
| repassé', 0.3075 |
| cadê', 0.2485 |
| direitos', 0.2424 |
| fica', 0.21457 |
| bonito', 0.1969 |
| federal', 0.1938 |
| parte', 0.1927 |
| recebeu', 0.1895 |
| todos', 0.1889 |
| governo', 0.1876 |

Tabela 6 - Resultados da Bateria 4

Alterando a topologia do algoritmo *doc2vec*, baseado em *CBOW* ou *Skip-Gram* não houve mudanças no cálculo de similaridade realizadas pelo *framework* GENSIM. As mudanças significativas foram com relação aos demais parâmetros. Qualitativamente, observa-se que o modelo com *vector_size*, *workers* e *window* superior apresentaram palavras que melhores se ajustam ao contexto de "Avaliação", como "repassé" "cadê" e "direitos". Desta forma, deduz-se que os usuários estão, em sua maioria, questionando sobre a avaliação do repasse do auxílio emergencial.

A Figura 13 apresenta os pontos cartesianos bi dimensionais referentes aos vetores inferidos para as palavras mais similares. É possível verificar, por exemplo, a proximidade da palavra "avaliação" com "repassé".

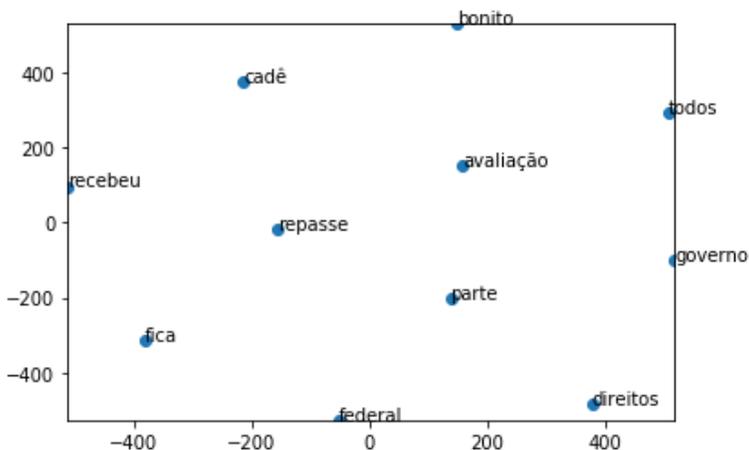


Figura 13- Distribuição das palavras mais similares

Como o *doc2vec* (ou *word2vec*) são técnicas não supervisionadas utilizou-se de dados não rotulados para extração de conhecimento. Diferentemente da aprendizagem supervisionada onde é possível utilizar métricas de avaliação objetivas, na aprendizagem não supervisionada estas métricas são subjetivas, como por exemplo a avaliação qualitativa supracitada. A tentativa de separação por classes e avaliação por *clusters* também é uma técnica avaliativa subjetiva deste tipo de aprendizagem e para este propósito o método não supervisionado *k-means* foi utilizado.

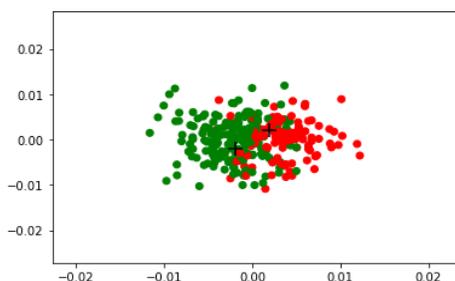


Figura 14- k-means com 100 iterações

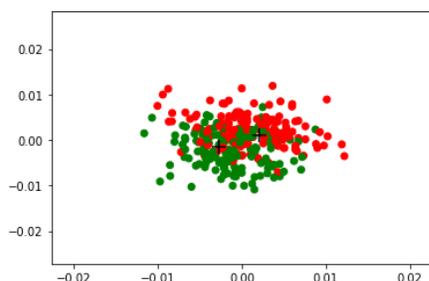


Figura 15- k-means com 200 iterações

Foi verificado, conforme ilustrado na Figura 14 e Figura 15 que o modelo de *embedding* proposto gerado pelo *doc2vec* não é completamente separável em classes. Observa-se que os dois *clusters* gerados possuem alta compactação, entretanto baixo *between* (distância entre os dois *cores*). Desta forma os comentários estão centrados, em sua maioria, em um único assunto, o auxílio emergencial da Caixa. Os pontos mais afastados mais afastados do *core*, podem provavelmente ser *outliers* (ruídos), ou seja,

comentários que não agregam valor semântico dentro desta temática.

71 CONCLUSÃO

O método proposto para exploração de dados em redes sociais possui recursos úteis e modernos no campo de inteligência artificial. A técnica de embedding utilizada permite extrair informações que a olho nu, sem os artifícios propostos, de- 23 mandariam tempo maior ao administrador ou gestor da página da rede social. Utilizou-se como exemplo prático da metodologia, a exploração de dados no Instagram do Ministério da Cidadania, entretanto qualquer página pública do Instagram poderia ser empregada. Os resultados permitiram inferir informações importantes em relação ao assunto tratado e comportamento dos usuários.

Conforme tentativa de *clusterização*, os comentários estão centrados, em sua maioria, em um único assunto: auxílio emergencial da Caixa, onde os usuários demonstram descontentamento com a lentidão no serviço de análise e triagem para distribuição deste auxílio. Os pontos, coordenadas, mais afastadas do core (centro do cluster), são provavelmente outliers (ruídos), ou seja, comentários que não agregam valor semântico dentro desta temática. Conclui-se, portanto, que a técnica é eficaz e atingiu os objetivos propostos. Como estudos futuros recomenda-se o treinamento do modelo com comentários de outras redes sociais, por exemplo de outros ministérios, a fim de verificar a sua capacidade intrínseca de classificação. Recomenda-se ainda acrescentar ao modelo a capacidade de analisar sentimentos, seja por meio de um treinamento supervisionado através de uma base de léxicos, ou por outra metodologia.

REFERÊNCIAS

CASTANHO, Maria Augusta Ferreira. O processo eleitoral na era da internet: As novas tecnologias e o exercício da cidadania. Faculdade de Direito da Universidade de São Paulo, Tese de Doutorado, 2014.

GUIMARÃES, Jonathan Wagner. Elaboração e construção de um protótipo mínimo viável para o Tingoram: Um sistema de mineração de dados web baseado em georreferenciamento para sugestão semi automatizada de doação de alimentos. Universidade Federal de Ouro Preto: Trabalho de Conclusão de Curso. João Monlevade, 2018.

JUNIOR, João Ribeiro Carrilho. Desenvolvimento de uma metodologia para mineração de textos. Pontifícia Universidade Católica do Rio: Dissertação de Mestrado, 2008. 24 LI, Zhi. A Beginner's Guide to Word Embedding with Gensim Word2Vec Model. Disponível em: . Acesso em: 18 de Agosto de 2020.

MIKOLOV, Tomas. Distributed Representations of Sentences and Documents. Proceedings of the 31 st International Conference on Machine Learning, Beijing, China, 2014. MITCHELL, Ryan. Web Scraping com Python. São Paulo: Novatec Editora, 2019. SHPERBER, Gidi. A gentle introduction to Doc2Vec. Disponível em: Acesso em: 18 de Agosto de 2020.

A

Algoritmo doc2vec 30, 34, 35

Aterramento 140, 141, 142, 145, 146, 149, 150, 151, 152, 153, 154

C

Choque elétrico 140, 141, 142, 143, 144, 149, 152, 155

Cidades inteligentes 38, 39, 40, 41, 42, 43, 45, 48, 53

Clusterização 37

Controle preditivo não-Linear 127

D

Design 40, 56, 57, 58, 59, 61, 64, 65, 71, 72, 138

Digital games 57, 58, 68, 69, 70

Dispositivo residual 140

Dispositivos móveis 1

E

Energia eólica 111, 114, 117, 157

Energias não renováveis 157

Energia solar 111, 112, 113, 114, 116, 117, 118, 123, 124, 125

Energias renováveis 114, 157, 158, 167

Energia térmica 157, 158, 159

Engenharia de requisitos 11, 12, 13, 16, 17, 20

F

Fontes térmicas 156, 157, 158, 159, 160, 163, 167, 168

Fuzzy cognitive maps 39, 40, 49, 53, 54, 55, 56, 112, 125

G

Game customization 58

Garantia de qualidade 1, 8, 14

H

Homologação de releases Android 1

Hospital inteligente 80, 82, 85

I

Inserção automática 73

Interligação de programas 73

ISO/IEC 11, 12, 13, 14, 15, 16, 20

L

Layout de inclusão facilitada 73

M

Mapas cognitivos fuzzy 39, 110, 111, 118

Matlab 87, 88, 95, 108, 109

Modelos bilineares 126, 127, 128

P

Painéis fotovoltaicos 111, 113, 117, 122

Processos 2, 8, 11, 12, 13, 14, 17, 18, 20, 47, 73, 75, 88, 89, 90, 127, 137, 159, 161, 162, 164, 166

Puma 560 87, 88, 89, 90, 91, 92, 94, 96, 97, 99, 100, 107, 108

Q

Qualidade de software 1, 4, 17, 20

R

Robô 49, 87, 88, 89, 90, 91, 92, 93, 97, 99, 101, 102, 104, 107, 108, 109

Robótica 87, 88, 89, 90, 93, 107, 108, 109, 120

S

Satisfação do cliente 111, 122

Sistemas inteligentes de computação 39

Software 1, 2, 4, 5, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 27, 41, 54, 61, 73, 74, 76, 77, 78, 83, 87, 89, 90, 95, 96, 99, 107, 118, 122

T

Teste de software 1

U

UML 11, 12, 18, 19, 20

User experience (UE) 58, 59, 61, 62, 64, 70, 72

V

Virtual things 80

W

Web of things 80, 81, 83, 86

Web scraping 21, 22, 37

ENGENHARIA ELÉTRICA E DE COMPUTAÇÃO:

Docência, pesquisa e inovação tecnológica

-  www.atenaeditora.com.br
-  contato@atenaeditora.com.br
-  [@atenaeditora](https://www.instagram.com/atenaeditora)
-  www.facebook.com/atenaeditora.com.br

ENGENHARIA ELÉTRICA E DE COMPUTAÇÃO:

Docência, pesquisa e inovação tecnológica

-  www.atenaeditora.com.br
-  contato@atenaeditora.com.br
-  [@atenaeditora](https://www.instagram.com/atenaeditora)
-  www.facebook.com/atenaeditora.com.br