Cenas Educacionais

Thematic Dossier
TEACHING, EDUCATION AND TECHNOLOGY

e-ISSN: 2595-4881

COMPUTATIONAL MODEL FOR REGISTERING INFODEMIA ABOUT COVID-19 IN SOCIAL NETWORK APPLICATIONS

MODELO COMPUTACIONAL PARA EL REGISTRO DE INFODEMIA SOBRE COVID-19 EN APLICACIONES DE REDES SOCIALES

MODELO COMPUTACIONAL PARA REGISTRO DE INFODEMIA SOBRE A COVID-19 EM APLICATIVOS DE REDES SOCIAIS

Emily Lorane da Silva Cerqueira [1]
Mariana Farias da Silva [2]
Vinicius Gama Nascimento [3]
Fernanda Vasques Ferreira [4]
Leandro Brito Santos [5]

Manuscript received on: January 15, 2023.
Approved on: April 07, 2023.
Published on: May 07, 2023.

Abstract
The COVID-19 pandemic generated a large volume of information that was quickly spread and disseminated in society. The infodemic is caused by the ability of social networks to increase the flow of information sharing, making it easier to spread false news and more difficult to differ from true news. Thus, it is necessary to use a computational model to classify information, aimed at detecting fake news. Therefore, an application was idealized which allows the registration of information in text or image format, which are classified by an algorithm and presented in a panel, thus facilitating the verification of the veracity of this information.

Keywords: Infodemic; Fake News; Computational model.

Resumen
La pandemia del COVID-19 generó un gran volumen de información que fue rápidamente difundida en la sociedad. La infodemia es causada por la capacidad de las redes sociales para aumentar el flujo de intercambio de información, lo que facilita la difusión de noticias falsas

1 Graduating in Electrical Engineering from the Federal University of Western Bahia.
ORCID: https://orcid.org/0000-0001-9627-7984 Contact: emilycerqueiraeng@gmail.com
2 Graduating in Electrical Engineering from the Federal University of Western Bahia.
ORCID: https://orcid.org/0000-0002-8270-5041 Contact: engmarifarias@gmail.com
3 Graduating in Electrical Engineering from the Federal University of Western Bahia.
ORCID: https://orcid.org/0000-0002-8960-8148 Contact: vinicius.n4670@ufob.edu.br
4Doctorate in Communication from the University of Brasilia, with a Postdoctoral degree in Computational Modeling and Industrial Technology from National Industrial Apprenticeship Service. Professor at the Postgraduate Program in Communication at the Federal University of Mato Grosso. Professor at the University of Brasilia.
ORCID: https://orcid.org/0000-0003-4242-0057 Contact: fernanda.jornalista82@gmail.com
5 Doctorate in Computational Modeling and Industrial Technology by the National Service of Industrial Learning. Professor at the Federal University of Western Bahia.
ORCID: https://orcid.org/0000-0003-0132-4712 Contact: lbsantos@ufrb.edu.br

e-ISSN: 2595-4881

y dificulta la diferenciación de las noticias verdaderas. Por lo tanto, es necesario utilizar un modelo computacional para clasificar la información, con el objetivo de detectar noticias falsas. Por ello, se idealizó una aplicación que permita el registro de información en formato de texto o imagen, las cuales son clasificadas por un algoritmo y presentadas en un panel, facilitando así la verificación de la veracidad de esta información.

Palabras clave: Infodemia; Fake News; Modelo computacional.

Resumo
A pandemia da COVID-19 gerou um grande volume de informações que foram rapidamente espalhadas e difundidas na sociedade. A infodemia é ocasionada devido a capacidade das redes sociais em elevar o fluxo de compartilhamento de informações, tornando cada vez mais fácil a disseminação de notícias falsas e mais difícil de diferi-las das notícias verídicas. Dessa forma, se faz necessário a utilização de um modelo computacional para classificar informações, voltado à detecção de *fake news*. Sendo assim, idealizou-se uma aplicação a qual permite o cadastro de informações no formato de texto ou imagem, as quais são classificadas por um algoritmo e apresentadas em um painel, facilitando assim a verificação da veracidade dessas informações.

Palavras-chave: Infodemia; *Fake news*; Modelo computacional.

Introduction

In March 2020, the World Health Organization (WHO) declared a state of pandemic for the infection caused by the new coronavirus - SARS CoV-2. The impacts of the disease caused by the virus were felt worldwide. The disease has claimed 3.71 million victims and was responsible for the infection of 151 million people. In Brazil, there was no effective control of Covid-19, presenting a tragic scenario that, on April 30, 2021, totaled numbers such as: 401 thousand deaths and 14.6 million infected. The data are from April 30, 2021, considered the most lethal month of the pandemic in Brazil.

This whole context of public health emergency cannot be "detached" from a problem that the Pan American Health Organization (PAHO) named, also in March 2020, as "infodemic", defined as an excess of accurate or inaccurate information which, when necessary, make it difficult to find reputable sources and reliable guidance. "The word infodemic refers to a large increase in the volume of information associated with a specific subject, which can multiply exponentially in a short time" (OPAS, 2020).

e-ISSN: 2595-4881

The context of misinformation and fake news that surrounds the Covid-19 pandemic and seems to touch questions about the disease since the first alerts issued is the object of research due to the complexity that the phenomenon presents. The infodemic and the disinformation epidemic have fertile ground in societies where access to information is given, above all, via cell phones connected to the internet and social media. According to the 2021 Digital Report by We Are Social and Hootsuite, the spread of fake news in digital media is one of the main concerns of Brazilians interviewed. Although the term infodemic emerged from the Covid-19 pandemic, misinformation has deeper roots in history.

The fake news phenomenon itself is not new. However, when fake news reaches digital networks - due to hyperconnection and hyperpoliticization - it becomes a concern for researchers. In 2016, two important events marked the rise of the term fake news: the Brexit process in the United Kingdom and the election of Donald Trump as president of the United States. During the 2016 campaign, Trump began to accuse the press in that country of producing fake news, when it put him in an unfavorable situation. In 2018, Trump even created the Fake News Awards for reporters who made mistakes and made wrong predictions, as well as for the media that produced false information before and during his administration.

In 2018, the same strategy for disseminating fake news, such as propaganda, during the elections was used in Brazil by the victorious candidate in the country. Given the complexity that the phenomenon of fake news, disinformation and infodemic present as a challenge for complex societies, research and technological innovation are required to act in this context of multifaceted problems, demanding the development of practical solutions that contribute to improving the scenario, solve part of the problem and provide conditions for society to face the informational chaos generated by the infodemic.

In this way, fake news can generate irreversible consequences in political and economic scenarios. However, in situations involving human health and health emergencies such as the Covid-19 pandemic, the consequences can be disastrous, leading to preventable deaths. Henriques (2018) warns that erroneous information can lead individuals to risk-generating behaviors, either by indiscriminately using medications, for example, or by refusing technologies, such as vaccines and

e-ISSN: 2595-4881

necessary protection measures, which can lead to health to collapse. The recommendation by authorities and some health professionals of treatments without any scientific evidence, popularly known as the Covid Kit, has created a false scenario of safety and protection.

Santos (2020) clarifies that in a specific context such as the pandemic, the effects of the circulation of fake news are peculiar, since the situation of social distancing, poor knowledge about the disease and exploration of possible strategies to combat Covid-19 tend to create a scenario conducive to misinformation. In this way, we identified that the enemy of people's health and lives - fake news about Covid-19 - is disseminated thanks to automation, the use of algorithms creating "information bubbles" harmful to public health, collective and individual health.

In this sense, our work has the following guiding question: "How can we facilitate the tracking of false information about Covid-19 with the use of artificial intelligence?". Our question is based on the fact that digital networks such as Facebook, Twitter, Instagram, WhatsApp, Telegram, among others, have robust intellectual and technological capital to combat fake news.

Therefore, it is intended to contribute to the analysis of information regarding COVID-19 through the creation of a dashboard that can be accessed both in the mobile and desktop environment, facilitating the visualization of data and improving knowledge for tracking and detecting fake news. In addition, to demonstrate that with knowledge of artificial intelligence, use of Python language and processing of neural networks, from machine learning, it is possible to direct actions so that the competent authorities and organizations can pressure the large corporations that integrate digital networks to commit, involve and take responsibility for combating the phenomenon of fake news, highlighting the importance of preserving people's lives and health.

- Objectives

- Creation of the user interaction interface;
- Development of a REST API, to exchange data with the frontend;

Cenas Educacionais
e-ISSN: 2595-4881

- Modeling an AI for classifying fake news.

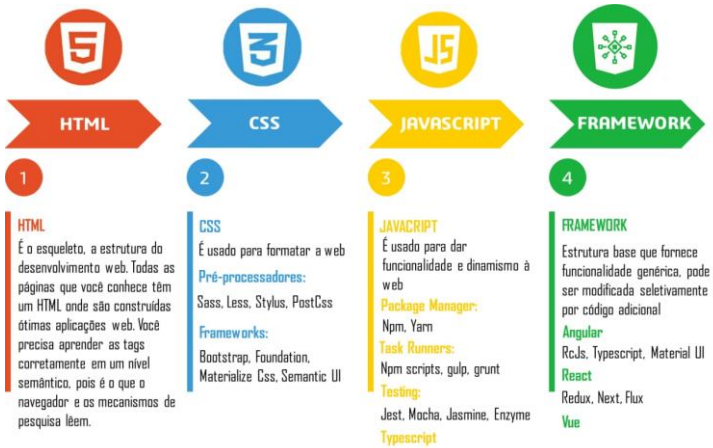
Theoretical Basis

For the creation of the idealized application, it is necessary to master the tools that allow the implementation, in this way the main concepts and technologies used for the computational modeling of the system of tracking, classification and display of fake news will be presented. The architecture of web systems are divided into the presentation layer (frontend) and the business rule layer (backend), the resources of libraries and frameworks structure these two layers.

- Frontend

The presentation layer is based on HTML - HyperText Markup Language, CSS - Cascading Style Sheets and JavaScript. According to Flanagan (2013), HTML is used to specify the content of web pages; CSS, to specify the presentation of these pages; and JavaScript, to specify their behavior. Although these technologies form the basis of the frontend, there are libraries and frameworks that facilitate web development, such as React, Bootstrap, Vue.Js, among others. In Figure 01 it is possible to observe a Roadmap (script) of the main technologies applied to frontend development.

Figure 01. Main technologies applied to frontend development

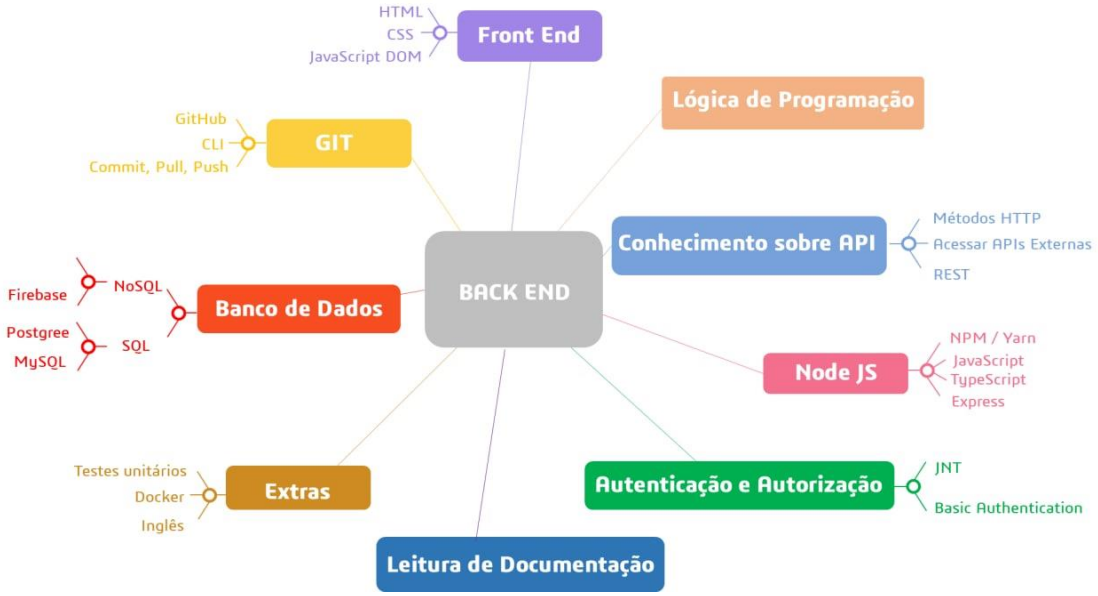


Source: Authors (2022)

- Backend

The backend is the whole system that structures an application, so that's where the business rules are, Figure 02 shows the main parts of this system. Although JavaScript is mostly used on the user side (client-side), it can also be used on the server side, so Node.js is used for backend development. Node.js is an open source server environment that enables the execution of applications written in JavaScript and can act, in Internet applications, as a server-side programming language (OLIVEIRA & ZANETTI 2020). For applications that unify the frontend and the backend, the EJS framework (Embedded JavaScript templating) can be used, which provides an easy way to transport data from the server side to the client side.

Figure 02. Main knowledge for backend development



Source: Authors (2022)

- React

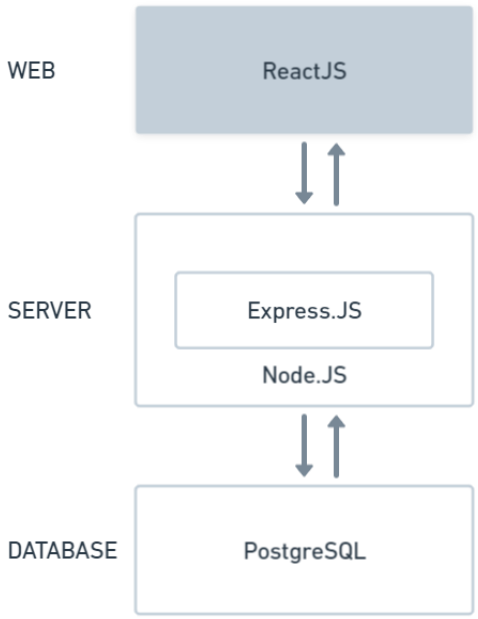
React Js is a JavaScript library for creating user interfaces, the main advantage of react is the possibility of code reuse, since the structure is based on components (REACT 2022).

e-ISSN: 2595-4881

- Express

Express is a minimal and flexible Node.js web application framework that provides a robust set of features for web and mobile applications (EXPRESS, 2021). With express it is possible to create a Node.js server that interacts with both the database and the frontend as shown in Figure 03.

Figure 03. Application structure using React and express



Source: Authors (2022)

- Machine learning

The machine learning process (GOODFELLOW, BENGIO, COURVILLE, 2016) can be carried out through two fundamental components. The first is the data set, which refers to the system to be studied, in this case a set of news texts classified as true or false, which will be divided between 80% as a training set and 20% as a training set validation. There is already a corpus in Portuguese with about 7,300 texts (MONTEIRO, et al., 2018). The second component is the model itself, which consists of a neural network to which a set of parameters will be adjusted.

e-ISSN: 2595-4881

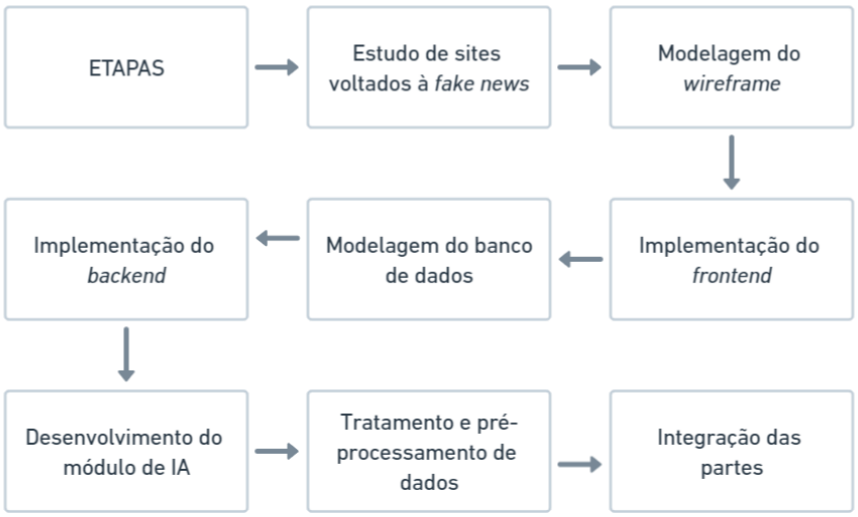
For the use of a training algorithm, this text must undergo pre-processing, in which terms that do not contribute to training are filtered, such as definite or indefinite articles, putting all letters in lower case, among others. After initial filtering, the text must go through a "tokenization" process in which words are associated with sets of numbers. For the initial processing of texts and to produce an embedded set of words in which the similarity between them can be measured numerically, Keras is used. Keras is an open source neural network library written in Python. It is capable of running on top of TensorFlow, Microsoft Cognitive Toolkit, R, Theano, or PlaidML, and designed to allow rapid experimentation with deep neural networks.

Metodology

In the methodological path, there were cycles of "theoretical foundations" with participation in courses and implementation of computational models to test the knowledge acquired in each resource throughout the training. Soon after, there was the implementation of the simulations. Thus, Figure 04 shows the steps followed for implementation.

In addition, Visual Studio Code - VSCode was used as the development environment and later Replit, as it allows that, while the server is running the application, anyone with the access link can view the project that is in the cloud storage. It was decided to make an application with separate frontend, backend and news classification module, in order to separate the complexities, idealizing that the classification module be carried out by artificial intelligence or manually.

Figure 04. Flowchart of steps



Source: Authors (2022)

Modelagem Computacional

- Web Development

At first, the skeleton of the site (wireframe) was modeled according to the application's requirements, to assist in the creation of the pages. The interface was divided into 4 parts: landing page, infodemic information presentation screen, infodemic registration screen, and dashboard. After developing the wireframe of the entire application, development of both the frontend and the backend began.

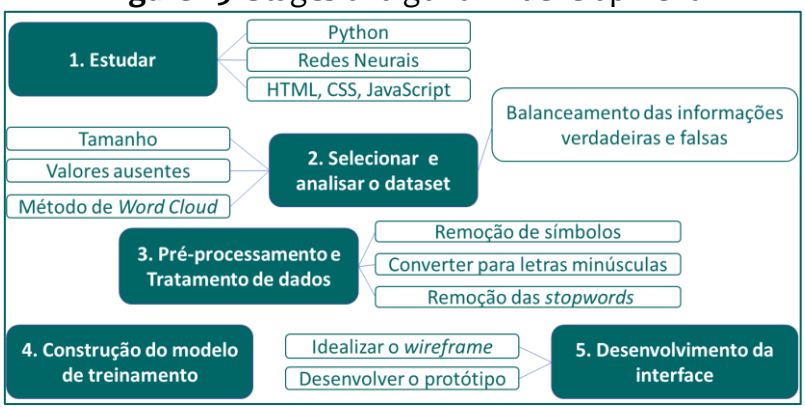
With regard to the frontend, all the screens of the application were built with the exception of the dashboard and the ReactJS library was used along with styled components for styling the page, the structure offers easy maintenance in the code. For the backend, an API was built in which routes were implemented to register infodemic information in both text and images, classify whether the information is false or true and consult the news already registered. For this, a Node.js environment with typeScript, a PostgreSQL database and the S3 cloud image storage system from Amazon Web Services - AWS were used (the free AWS plan was used).

- Development of the classification algorithm

For the artificial intelligence module, the libraries and dataset to be used were initially imported (available at: https://www.kaggle.com/datasets/saratchendra/fake-news), later, after the complete development of the model, the actual dataset is used. Figure 06 exposes some stages of development.

Through Figure 05 it is possible to observe some methods that were applied for the pre-processing and treatment of data. That is, the quality of the data that were imported for model training was checked, such as: dataset size, missing values, balancing true and false information. In addition, the Word Cloud method was applied, which captures the most frequently used words and creates a cloud. The missing values that were identified in the previous step were removed and the dataset index was reset. Data pre-processing was started by converting all data letters to lowercase and separating them into a string so that punctuation and symbols contained in the dataframe could be removed. Posteriorly, to reduce the size of the information to be read by the model to be developed, it was necessary to remove Stop Words, which are words that are not as relevant as prepositions and pronouns. Then, a function was created to remove words with less than three letters in case they were not eliminated along with the Stop Words. In view of this, for the aforementioned data analysis and pre-processing steps, several libraries and functions were used, such as: matplotlib.pyplot, nltk.corpus, nltk.tokenize, nltk.stem, wordcloud, STOPWORDS, and other simple developed functions.

Figure 05. Stages of algorithm development



Source: Authors (2022)

e-ISSN: 2595-4881

Another important method applied was the counting of words with the same origin, such as: programmer, programs, programming, programmers, etc. Furthermore, the idea was to sequence the words into numerical values and apply the Padding technique, which is responsible for leaving the input data sequences with the same length, either with padding or truncation. In addition, transforming the vectors into a matrix that would be the input of the model to be developed and, finally, the neural networks. Some of the libraries and functions used in this last step were: gensim, tensorflow.keras.models and tensorflow.kera.initializers.

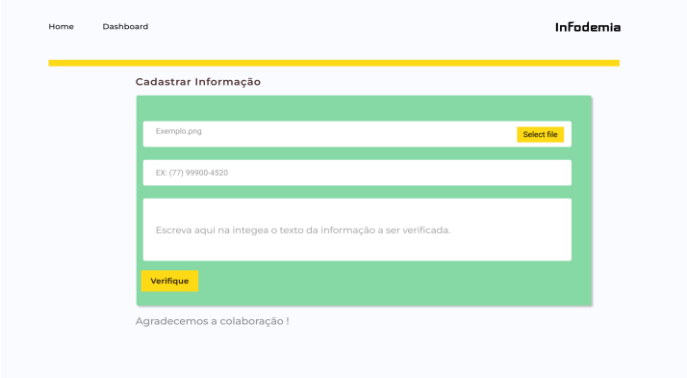
Results

The developed screens – presentation screen, registration screen, *dashboard*– are shown in Figures 06, 07, 08 and 09.

Figure 06. Landing page for viewing infodemic data

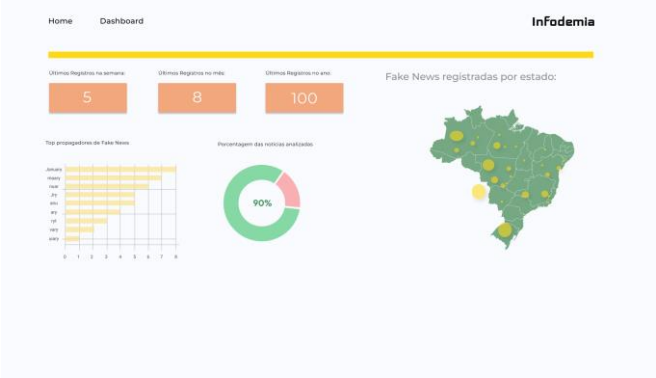
Source: Authors (2022)

Figure 07. Registration page



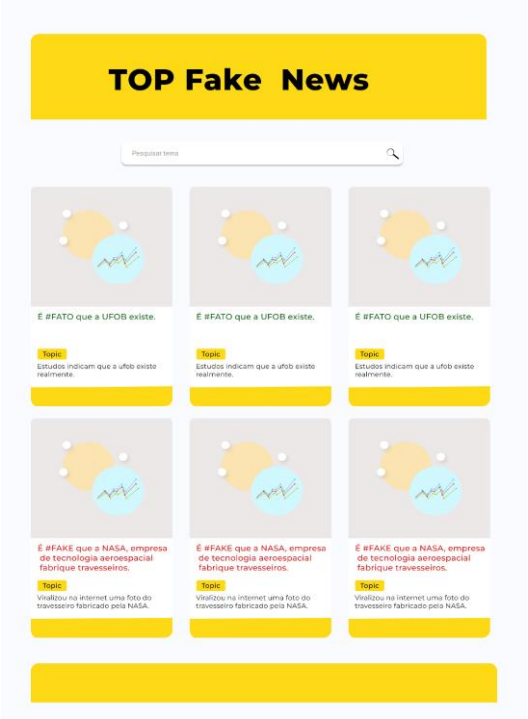
Fonte: Autores (2022)

Figure 08: Dashboard page of the software for recording infodemics



Source: Authors (2022)

Figure 09. Wireframe for viewing infodemic records



Source: Authors (2022)

e-ISSN: 2595-4881

Figure 10 shows the WordCloud method used in the data processing step for the development of the computational model. It is possible to observe that the most frequent words in the texts classified as true are New York, Donald Trump, President and others.

Figure 10. Word cloud

Source: Authors (2022)

Figure 11 exposes the application of the Gensim library to check the similarity of words, such as the word "person". As for the results regarding the development of neural networks, the expected result was not obtained, since it was not possible to complete this step due to errors detected during the execution of the code. Solutions were sought to resolve the errors found, but were unsuccessful.

Figure 11. Count of similar words

```
#Checando relação de similaridade
modelVetor.wv.most_similar("person")
```

```
[('someone', 0.6731433272361755),
 ('man', 0.5903780460357666),
 ('individual', 0.587113082408905),
 ('circumstance', 0.5861910581588745),
 ('people', 0.5819699168205261),
 ('thing', 0.5683997273445129),
 ('manner', 0.549304187297821),
 ('somebody', 0.5465718507766724),
 ('anyone', 0.5427058339118958),
 ('otherwise', 0.5427022576332092)]
```

Source: Authors (2022)

Conclusion

The present work presents, in its methodological route, the experience of the classroom of engineering students with the proposal to carry out the interdisciplinarity of the disciplines of algorithms and numerical calculation, using computational modeling and mathematical modeling, to carry out research for the area of education and health.

In this way, an algorithm capable of offering the communication interface with users was implemented, using the connection with the data stored in the file so that the user can use them to carry out academic learning experiments. The student will be able to carry out the classification manually, using a selector to analyze the veracity of the data, and then display the data in a panel, after which all the data are displayed in a Dashboard format.

It is worth mentioning that in the methodological path, there was an effort on the part of the team to carry out the Modeling of an AI, capable of automatically classifying the data, however, this effort exceeded the project schedule. However, it is emphasized that the effort will be carried out as part of future works.

Thanks

The authors would like to thank UFOB which supported the research through: Institutional Program of Initiation Scholarships in Technological Development and Innovation; Institutional Scientific Initiation Scholarship Program sponsored by the State of Bahia Research Support Foundation; to the guiding professors Dr. Leandro Brito Santos and Drª Fernanda Vasques Ferreira for their support, encouragement and guidance throughout the period; in addition to the support of colleagues and friends.

References

AVAAZ 2020. **O Brasil está sofrendo uma infodemia de Covid-19 Os brasileiros acreditam mais em notícias falsas que os italianos e os estadunidenses**. Disponível em: https://avaazimages.avaaz.org/brasil_infodemia_coronavirus.pdf Acesso 25 abr. 2021.

CANAVILHAS, J.; COLUSSI, J.; MOURA, Z. Desinformación en las elecciones presidenciales 2018 en Brasil: un análisis de los grupos familiares en WhatsApp. **El profesional de la información**, v.28, n.5, 2019.

CHOLLET, F. **Deep Learning with Python.** Manning Publications: New York, 2018.

BUENO, W. C. Empurroterapia na imprensa: esta doença tem remédio? Disponível em: https://docplayer.com.br/39606910-Empurroterapia-na-imprensa-esta-doenca-tem-remedio.html Acesso em: 11 mai 2021.

CAMBRA, U. C.; ITURRIZAGA, A. U. & HEVIA, T. M. (Coord). **Comunicación y Salud. Nuevos escenarios y tendências**. Madri: Editorial Complutens, 2011.

EXPRESS, **Framework web rápido, flexível e minimalista para Node.js**, versão 17.0.2 Disponível em: <https://expressjs.com/pt-br/org Acesso em: 15 ago 2022>

FLANAGAN, D. **JavaScript**: O Guia Definitivo. 6ª ed. São Paulo: Novatec, 2012.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**, Massachusetts: MIT Press, 2016.

HENRIQUES, C. M. P. A dupla epidemia: febre amarela e desinformação. **Revista Eletrônica de Comunicação, Informação & Inovação em Saúde**, v.12, n.1, 2018. Disponível em: https://www.reciis.icict.fiocruz.br/index.php/reciis/article/view/1513. Acesso em: 15 ago 2022.

KAO, A.; POTEET, S. R. **Natural Language Processing and Text Mining**. Springer-Verlag: London, 2007.

VARÃO, R. **Há alguma novidade na ideia de fake news?** Brasília: Blog SOS Imprensa, 2017.

VOLKOFF, V. **Pequena história da desinformação: do cavalo de Tróia à Internet**. Lisboa: Editorial Notícias, 2000.

MONTEIRO, R. A. et al. **Contributions to the Study of Fake News inPortuguese**: New Corpus and Automatic Detection Results, Computational Pe Portuguese Language, Cham: Springer International Publishing, 2018.

GOMES, W. **O que são Fake News?** [vídeo]. Publicado pelo canal INCT em Democracia Digital, 2020. 1 vídeo (38 min). Disponível em: https://www.youtube.com/watch?v=8tvJ4cMtYXY. Acesso em: 22 abr. 2021.

ORGANIZAÇÃO PAN-AMERICANA DA SAÚDE. **Entenda a infodemia e a desinformação na luta contra a COVID-19**, Disponível em: iris.paho.org/handle/10665.2/52054?locale-attribute=pt Acesso em: 27 março. 2022.

OLIVEIRA, C. L. V.; ZANETTI, H. A. P. **Node.js:** Programe de forma rápida e prática. Editora Saraiva, 2021.

REACT. **Uma biblioteca JavaScript para criar interfaces de usuário,** versão 17.0.2 Disponível em: <https://pt-br.reactjs.org Acesso em: 15 ago 2022>

SHCHUR, A. **Fake news detector with deep learning approach (Part-II) Modeling,** Disponivel em: <https://towardsdatascience.com/fake-news-detector-with-deep-learning approach-part-ii-modeling-42b9f901b12b Acesso em: 15 ago 2022.

VARÃO, R. **Há alguma novidade na ideia de fake News?** Observatório da Imprensa, Brasil 2019. Disponível em: <//bit.ly/2WlmrUo> Acesso em 01 abr 2022.

WE ARE SOCIAL AND HOOTSUITE. DIGITAL 2021. DISPONÍVEL EM: HTTPS://WWW.AMPER.AG/POST/WE-ARE-SOCIAL-E-HOOTSUITE-DIGITAL-2021-RESUMO-E-RELATORIO-COMPLETO. ACESSO EM: 30 ABR. 2021.

SANTOS, N. **Uso e efeitos de fake News na pandemia de covid-19.** 2020. 1 vídeo (29 min). Publicado pelo canal INCT em Democracia Digital. Disponível em: https://www.youtube.com/watch?v=ajJyjpFRE04. Acesso em: 25 abr. 2021.