

Article

The Influence of Fake News on Social Media: Analysis and Verification of Web Content during the COVID-19 Pandemic by Advanced Machine Learning Methods and Natural Language Processing

Andreea Nistor ^{1,*} and Eduard Zadobrischi ^{2,3,*} 

¹ Faculty of Economics and Public Administration, “Stefan cel Mare” University, No.13 Str. Universitatii, 720229 Suceava, Romania

² Department of Computers, Electronics and Automation, Faculty of Electrical Engineering and Computer Science, “Stefan cel Mare” University, No.13 Str. Universitatii, 720229 Suceava, Romania

³ Department of Computer Science, Technical University of Cluj-Napoca, Gh. Baritiu St. 26-28, 400027 Cluj-Napoca, Romania

* Correspondence: andreea.nistor@usm.ro (A.N.); eduard.zadobrischi@usm.ro (E.Z.)

Abstract: The purpose of this research was to analyze the prevalence of fake news on social networks, and implicitly, the economic crisis generated by the COVID-19 pandemic, as well as the identification of solutions for filtering and detecting fake news. In this context, we created a series of functions to identify fake content, using information collected from different articles, through advanced machine learning methods with which we could upload and analyze the obtained data. The methodology proposed in this research determined a higher accuracy of fake news collected from Facebook, one of the most powerful social networks for the dissemination of informative content. Thus, the use of advanced machine learning methods and natural language processing code led to an improvement in the detection of fake news compared to conventional methods.

Keywords: COVID-19; fake news content; economic impact; natural language processing; pandemic fake news; social networks; social relation



Citation: Nistor, A.; Zadobrischi, E. The Influence of Fake News on Social Media: Analysis and Verification of Web Content during the COVID-19 Pandemic by Advanced Machine Learning Methods and Natural Language Processing. *Sustainability* **2022**, *14*, 10466. <https://doi.org/10.3390/su141710466>

Academic Editors: Hyunchul Ahn and Maxim A. Dulebenets

Received: 29 May 2022

Accepted: 19 August 2022

Published: 23 August 2022

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Over time, social networks have become an indispensable tool in both the distribution of information and its provision by specialists [1] and also for ordinary citizens, all due to technological developments and smart devices. They have the ability to simplify the query process and also to broadcast them, characterized by texts, images, and videos, being used to collect data and information. In most cases, the speed of the transmission of information through social networks is much higher than that of traditional media, which specializes in the dissemination of breaking news [2]. At the same time, the number of users receiving online emergency information has risen significantly.

The distribution of content on social networks cannot be controlled, so false information reaches a large audience [3] without restrictions and without being able to determine the exact source. Fake news generates important changes in the economy, culture, and politics, and social networks are considered important channels for distributing content. Generally, both Twitter and Facebook are used to build the act of communicating, exchanging information, and delivering content to a large audience [4]. Over time, the two networks have been used as propaganda tools, especially in the political field, so nowadays, rumors are unfounded and unaccompanied by a statement from the person who is concerned, ignored, and trying to verify them.

Fake news is presented as verified news, but it is characterized by untrue ground, which is spread in an easy way with the help of social networks [5], their purpose being to

capitalize on and obtain economic benefits. Usually, the rumors have a negative influence from a social point of view, causing confusion. In this context, an attempt was made to reduce the phenomenon of fake news, identifying over time different solutions for stopping fake news and identifying them. Monther Aldwairi and Ali Alwahedi conducted a study to identify fake content through a tool that can be downloaded and attached to the browser [6]. Additionally, according to the study conducted by Md. Sayeed Al-Zaman, lately, more fake news has revolved around health due to the pandemic caused by COVID-19 [7]. At the same time, the phenomenon is an old, well-established one which has caused a lot of economic damage. In this context, fake news about the economic sector has spread in Romania. In this sense, it was said that Romania's development was intentionally kept in a stage of underdevelopment by the National Bank, this being later denied by the governor of the National Bank of Romania, Mugur Isărescu. Therefore, more and more countries have imposed legislation in order to avoid the spread of fake news [8], and numerous studies have been conducted in this regard, investigating in advance a number of features of fake news such as the pattern of fake content by checking posts of many Twitter users.

The study aimed to analyze the fake information exposed in the online environment by means and agglomerations of CNN and NLP processing with the ultimate goal of obtaining an encapsulated plug-in solution for browsers. Therefore, the ultimate goal was to outline the first version of a browser-like FakeBlocker solution in order to limit the sharing of fake information. The proposed method is based on several training sets with true and false information, which were assimilated from social media. The method takes into account the successive iteration of the source of information and the analysis of names, profile pictures, number of users, tags, the cadence of posts, organizing events, appreciation, or groups to which it belongs. These elements were part of a subsequent approach to increase the degree of accuracy. The temptation to resist fake news needs to be realized in the virtual environment through tools designed to verify the information. Thus, an amalgam of skills and abilities presided over by technology is needed to identify fake information and to distinguish it from true information, but in this case, the functional literacy of a population was applied, as well as a set of necessary skills to navigate the virtual environment.

Over time, a number of methods have been created, accompanied by algorithms to solve the problem of fake news, reaching tangible solutions through neural networks [9,10]. Fake information for entertainment purposes is extremely harmful to social behavior, so the COVID-19 pandemic has generated a lot of fake news, which has been circulated online and on social networks. More and more people have begun to spread fake news, exposing the information as coming from health professionals. Exposure of people to misinformation exacerbates insecurity, so the most vulnerable to fake news have been those with a tendency to depression [11], those with religious backgrounds, and those without clear principles and beliefs. The online environment has strengthened people's confidence in the information they provide, shaping their perspectives based on the topics they identify with. Additionally, from a quantitative point of view, fake news is spreading widely through social networks, such as Twitter, Facebook, YouTube, and Instagram [12], and an accentuated problem of them could be observed since 2016 since the US presidential election. In addition to the political sector, the economy is heavily affected by misinformation, producing colossal changes in the markets, and the owners are the most affected in this case. During the COVID-19 pandemic, there have been many controversies in the matter of fake news, being extended in many areas, which is why research groups have resumed research on the detection of fake news through automatic learning methods [13], logical regression techniques, and the use of key vectors in news classification. Therefore, the contributions presented in this manuscript refer directly to the analysis and identification of news through the accumulation of key anchors created with the help of automatic learning algorithms but incorporated at the browser level in the form of an extension. This aspect can limit the user's access to the news with inappropriate content. In Section 1, the article focuses on general elements and aspects, as well as the introduction of false claims into this theory. The Section 2 presents the review of the research carried out by other researchers and also

the methods used in data analysis and processing. In the Section 3 explains in detail the process of data detection and analysis based on a methodological standard. In the Section 4 exposes the practical component and highlights the results obtained after implementation. The Section 5 highlights the new elements brought by the proposed solution, analyzes the results, and draws a series of conclusions, as well as exposes some future directions.

1.1. Analysis of Fake News

Fake news is everywhere, and exploring and identifying it is a constant challenge, so many experts have tried successfully to solve this problem many times, either through machine learning or through experiments. To identify fake news, Ahmed et al. applied a model [13], proving that the linear models had a better resolution than the nonlinear ones. Shu et al. [14] also showed that social networks are an ideal environment for spreading fake information as there are many people with similar characteristics in terms of perspectives, so the theories of some may become the certainties of others [15]. Social networks have the ability to filter audience categories based on their own interests, and malicious accounts have the ability to influence users' opinions with the help of messages sent automatically. In this direction, Shu et al. used these characteristics to make a series of network representations in order to identify the fake content, demonstrating that, as a rule, the social content and the one with temporal characteristics form the analysis basis [14]. Thus, the social dimension was characterized by the identity of the user, whether human or robot; thus, researchers proposed a method in order to determine the type of user, focusing on the frame time at which the posts are made as well as the characteristics of the account from which the content is posted. Conroy et al., in order to detect fake news, analyzed linguistically the messages considered misleading, specifying that network data could be copied and misdirected [16]. Moreover, through the analysis of the language and the NLP technique, Hussein [17,18] wrote a series of articles that focused on the analysis of websites that contained fake information, and Bali et al. also approached the identification of fake news with the help of NLP and ML [19,20] and showed that the accuracy was quite good compared to other classifiers; nevertheless, it was below 90%.

Fake news is also perpetuated in the economic environment, integrated into all aspects of communication, so misinformation in society has become one of the greatest dangers of society. In this regard, there have been a number of attempts to overcome the effects of fake news, as well as to avoid them, through well-designed legislation and media literacy. One of these methods included the detection of fake news through crowdsourcing; however, its widespread dissemination caused its accentuation due to high traffic on social networks. 70 % of fake news also had the ability to be reproduced and a greater ability to reach a large audience compared to truthful news, as demonstrated by researchers Vosoughi, Roy, and Aral [21]. At the same time, Ahmad et al. researched certain language skills designed to differentiate between fake and truthful content using machine learning algorithms [22]. Abdullah et al. developed a multimodal learning model for identifying fake news through a neuronal network [23], which did not generate remarkable and lasting results; therefore, the study was oriented toward a series of data sets.

In order to detect fake news, data sets were developed that were meant to cover about seven areas of interest in the news area by Pérez-Rosas et. al, with one set intended to include six different news domains through crowdsourcing and another set involving the inclusion of a single news domain through the data collected on the web. The use of linguistic factors to identify the green news has been demonstrated in several studies, including the construction of a linear classifier, designed to validate the intertwining of the syntactic, signatory, and lexical through a crossover of linguistic features, but this generated an accuracy of 78%. Another proposed model is a multilayer perceptron, made by Davis and Proctor [23], which obtained a much better accuracy. In this case, a data set called the Fake News Challenge was used, described by a series of automated models for identifying fake news necessary in the classification of texts, and this model achieved an accuracy of over 90% [22,24]. The purpose of this introduction has as its main objective the

exposition of concepts and definitions of fake news in relation to social networks. Exposing useful information to readers and researchers in order to generate a better understanding, although there is no comprehensive definition of fake news [23], is related to perceptual issues. The Oxford Dictionary offers a definition of the field of fake news and that the most common topic is related to rumors that turn into news, which are often false. The whole concept is simplistic, especially when a news item starts from a rumor, and is not interpreted properly. It can be defined as an unverified statement or information created by users of social networks and that is extremely easy to propagate through the prism of private groups [24].

A research group [25] managed to detach terms and concepts with similarities, such as rumors or satire, but they were based on other characteristics: authenticity, intentions, and type of information. When we analyze a rumor that contains information, the previous characteristics are not found in its main body. The intention is clear in fake news, especially since the information is presented in a twisted manner and the direction is extremely clear, focusing on a certain area [26,27]. According to all the definitions and identification characteristics, the relationship between the concepts of fake news and rumors is represented in Figure 1.



Figure 1. Correlation between rumors, news, and fake news.

In order to determine the fake news, it is necessary to outline certain elements of them in social media. The news circuit focuses on a three-stage process: making the news, publishing it, and last but not least, distributing it. Fake news is aimed at certain goals and can be achieved on or through social media, and this news usually focuses on a short and suggestive headline designed to attract the body of the text, which can sometimes be accompanied by images [25]. In addition, the news is published by social media publishers, such that the publisher is itself a user of social networks. Through the identity held in social media, the user benefits from a series of defining elements, such as friends, followers, posts, events, activities, and frequented locations, and by distributing fake news, friends and followers of the user are those who initially view the designed news [28]. Following this phase, the news is distributed, which is done through the actions of the recipients who can either ignore them or in turn pass on their claims, appreciate them, or add comments, so that the news can present much higher visibility. When fake news is distributed by recipients, their followers in turn distribute it more, and the process of spreading fake news is constantly moving. Fake news is detected either by consulting official sources or by specific detection methods, but the longer it takes to identify fake news, the more users are affected [29]; therefore, it is recommended that the detection be performed at the time of

news distribution or even at an earlier stage. When fake news is detected, the distribution process is complete. Figure 2 shows how fake news can be broadcast on social media.

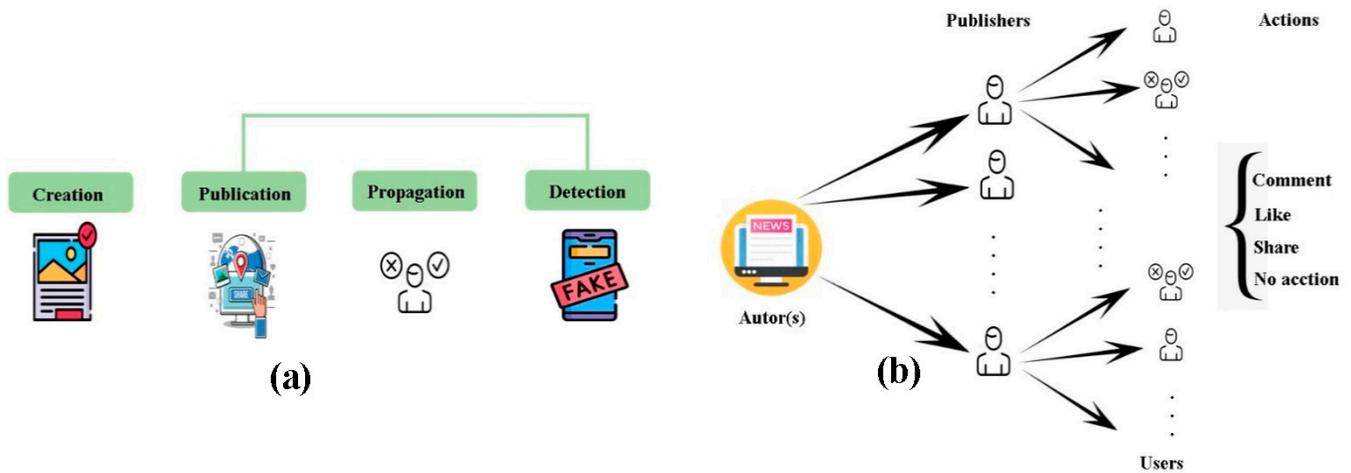


Figure 2. The stages of creation, publication, propagation, and detection of an article (a). Exposure regarding the spread of fake articles starts with a content creator and reaches the masses of people who consume content from social media (b).

An example of fake news that has set Europe on fire is the blackout, which has scared many European countries. Thus, the Austrian Ministry of Defense began in September 2021 a training campaign for the population to know how they should react in the event of a major power outage. This information spread such that from an informative material of the Austrian government officials became large fake news. The media interpreted the information, and people began to misrepresent the information, preparing for such an event and stocking up. In Romania, the information arrived four weeks later, with the media promoting a series of headlines such as: “Blackout in Europe! Huge power outage. The army has already intervened: Get ready!” or “Europe, one step away from remaining in total darkness for days on end. Austria is preparing for a blackout. The effects would be devastating.” See Figure 3 for the social impact.

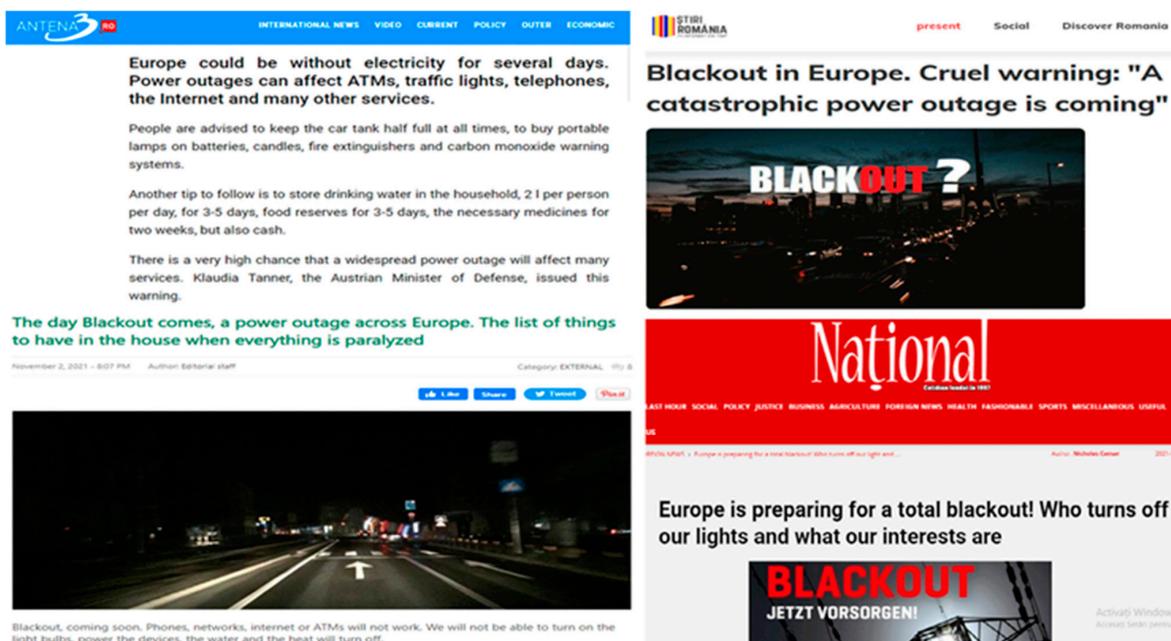


Figure 3. Examples of fake news. Image extracted from the media trusts antenna3, national, and stiririromania.

In terms of content, the news has a title, suggestive images, the content itself, the author, and the source. These are essential when determining the author's intentions so that the features of the news can be identified according to the complexity of the body text, including in this case the number of words used [30], features that express the emotions of the text, both positive and negative, as well as the existing text type. In the identification of fake news, the style of writing plays an important role, and as a rule, fake content focuses on important political events or financial interests, and often, the detection of non-truthful news is not possible to achieve only through existing methods. Figure 4 shows the main characteristics for identifying fake news on Facebook.



Figure 4. Features to identify fake news. Image extracted from the media trust dcnews.ro.

1.2. Phishing and Malware Can Generate Fake News

Following the COVID-19 pandemic, an increase in cybercrime cases has been reported, so pandemic crime is evolving and highly adaptable to change, facilitating public deception. These situations are highlighted by errors, such as a significant number of users who are not trained in the use of the virtual space, the elderly who encounter security problems when accessing social networks or websites, not updating devices on time, connecting accounts to devices without later logging out, employees who either work from home or remotely access company accounts via public Wi-Fi [31], and lack of training in accessing educational platforms.

In addition, since the beginning of the pandemic, the number of malware has increased, and the most frequent ones have been targeting health institutions [32], which have been forced to pay sums of money in order to avoid the downtime of online platforms and the loss of patient data. From 2019 to 2020, more than 490 such attacks occurred in the United States. Fake news is also distributed for malevolent purposes to manipulate certain audiences or to influence people's behavior concerning certain political, economic, or natural events [33]. In April 2020, more than 4000 fake news stories about the COVID-19 pandemic were spread, where out of 1 million tweets about SARS-CoV-2 that were analyzed on Twitter, about 80% of them were classified as false. In this context and a large number of false information, social networks have resorted to the implementation of tools designed to spread such misleading information. This fake content has been dubbed "Infodemia", such

that the World Health Organization has stated that the fight is not only with the pandemic but also with the fake news circulating about it [34]. Along with this type of content came articles with the wrong prescriptions, which contained tips on the proper medication for COVID-19 as well as on natural remedies that can be used to fight this virus.

A lot of fake news and conspiracy theories have circulated in Romania about recipes for treating COVID-19 as well as conspiracies, such as the connection between COVID-19 and 5G networks [35]. In this context, a series of videos appeared which benefited from a lot of views following the spread of false information, and as a result, information was spreading that 5G networks have the ability to accentuate the effects of COVID-19 in terms of symptoms. In April 2020, the European Commission detected over 2700 fake items on 110 subjects related to the pandemic. Most methods of identifying fake news focus on a number of features for identification [36]. Through a study (Hakak et al.) [37], the main features of the fake news were collected and classified by a model, which intertwines three machine learning methods, obtaining very good accuracy. The learning patterns used are usually based on features such as the body of the text or the title.

2. Materials and Methods

This section presents a number of methods used to detect fake news by exposing data sets, processing them, determining data, and last but not least, presenting algorithms to solve the spread of fake news. The data structure was made up of both fake and truthful news, with the set being outlined by a series of features, such as the title used in the articles, the body of the text, and the time of the news.

The data processing was done in the initial stage followed later by the testing and the data process, but before these stages, the news was juxtaposed, excluding the data that were not a necessity. Thus, pronouns were removed, and capitalized words were transformed into lowercase words through the lexicon, prepositions, psychological and linguistic features, and last but not least, semantic features. The lexicon was used to monitor the frequency with which keywords appeared in the news, the parts of the sentence used, and the punctuation marks in the text, the signatory was meant to follow the signatory aspect, and the sentence parts were based on the parts of speech. Additionally, in order to classify the news, a psycholinguistic component was included, which followed from a numerical point of view the words from the news texts through a dictionary-type software.

To define the issue of fake news, the content of news articles presents a number of features which can distinguish between true and fake news by analyzing the title, the body of the text, the images used, or the author, thus rendering a function (f):

$$f(a) = \begin{cases} 0 & \text{if } a \text{ is false} \\ 1 & \text{if } a \text{ is true} \end{cases} \quad (1)$$

Manuscripts reporting large data sets that are deposited in a publicly available database should specify where the data have been deposited and provide the relevant accession numbers. If the accession numbers have not yet been obtained at the time of submission, please state that they will be provided during review. They must be provided prior to publication. Interventional studies involving animals or humans and other studies that require ethical approval must list the authority that provided approval and the corresponding ethical approval code; see Figure 5.

Mentioning again the aspect of the COVID-19 pandemic, most users of social networks, blogs, and consumers of video content have become vulnerable and much more easily influenced. There are studies and analyses that attribute all this fake news to mistakes or accidental problems. Therefore, some consumers are directly affected by social or educational shock, as was the case with the Nepal earthquakes or the Asian crisis [38]. All this fake news in the period 2020–2022 exposed the health of the inhabitants globally and endangered the medical field. The field of real data analysis and the separation of fake ones is a difficult one and requires an extremely well-developed analysis and documentation.

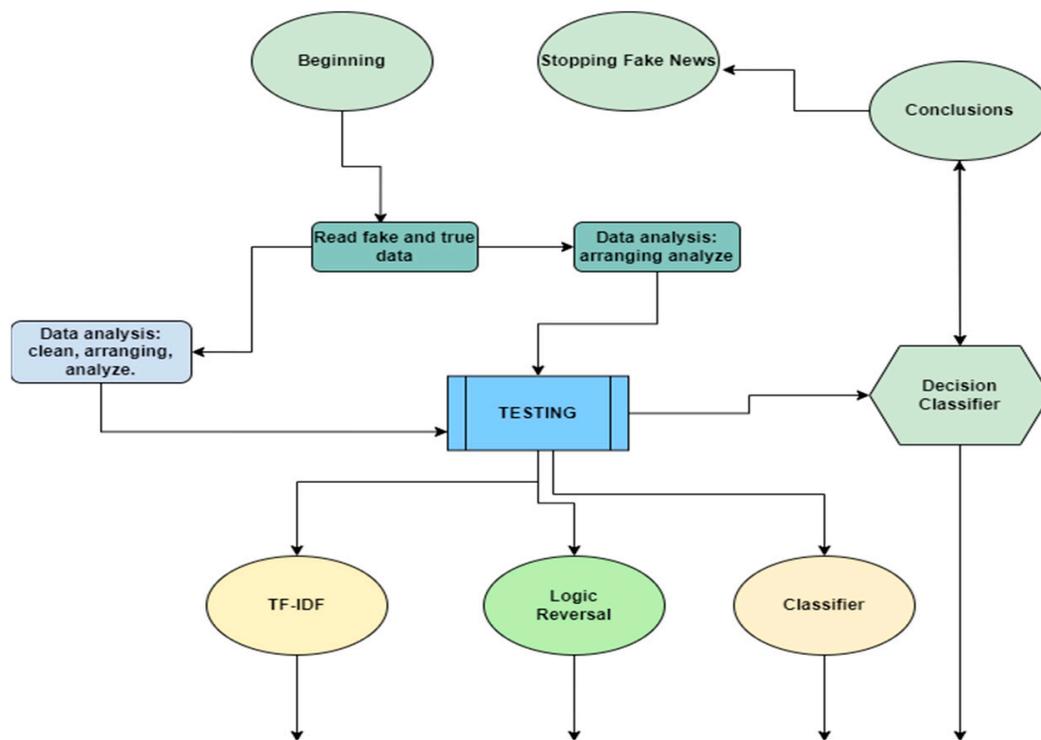


Figure 5. Data processing in the fight against fake news.

We must emphasize the use of natural language processing, which has an extremely high weight in the formulation of fake news, especially by applying it in as many areas as possible. Thus, natural language model and processing (NLP) offers the ability of an algorithmic system to form word combinations that generate the process of understanding and expose the speech process. This method can perform and detect actions and deviations of language, extraction of tones, and pronunciations of other nationalities. It is also possible to highlight the ways of recognizing and detecting speech and emotions (NER) or labeling speech fragments (POS).

2.1. Utility of Decision Tree

One of the most important tools that we can use in a decision-making structure and is based on a diagram-type architecture, which in the end can be considered a classifier, is the decision tree. Each internal node that specifies and indicates a “valid” condition or not, subsequently forming an attribute that branches the decision based on the test conditions formulating a result, structures it. Therefore, a class-type label that results from calculations performed on all attributes identifies the leaf node. Thus, we have a distance from the root to leaf and identify it as a classification, which can function variably as dependencies even if they belong to different categories. The usefulness in identifying the most important variables and in describing the relationship between them becomes a strong point, especially in terms of creating new variables and characteristics.

Tree-based learning algorithms are found everywhere in the predictive methods used for supervised learning, recognized for their extremely high degree of accuracy. It is recommended in the analysis and processing of nonlinear relationships in mapping, as they can solve regression problems; see Figure 6.

```

// Decision Tree Pseudocode
GenerateDecisionTree(Sample s, fetures F)
if stop_conditions(S,F) = true them
a. leaf = crate_Node()
b. Leaf.lable = classify(s)
c. Return leaf
root = create_Node()
root.testcondition = find_bestSplit(s,f)
v= {v | v a possible outcommme of root.testconditions}
for each value v? value:
sv: = { s ?root.testcondition(s) = v and s? S}
child = Tree_Growth(Sv , F);
Grow child as a descent of roof and label the edge (root -> child) as value
return root
}

```

Figure 6. Decision tree pseudo-code structure.

2.2. Utility of Naïve Bayes

In the case of this technique, the algorithm works on the basis of Bayesian theories, which state that it is assumed that there are no predictive elements and is most often used in multiple machine learning processes and problems [39]. An extremely simple example is that when analyzing a fruit or a vegetable, the case of the tomato is identified by color and a diameter between 3 and 5 inches. The dependence of different functions is undertaken by other functions, and the algorithmic notion of the naïve Bayes method can assume that all these functions prove separately that the analysis is performed on a tomato. The reference equation of this method reads as follows:

$$P(c | x) = \frac{P(x|c)P(c)}{P(x)} \quad (2)$$

$$P(c | X) = P(x_1 | c) \times P(x_2 | c) \times \dots \times P(x_n | c) \times P(c)$$

where $P(c | X)$ is the subsequent probability, $P(x | c)$ represents the likelihood, $P(c)$ is the class priority probability, and $P(x)$ represents the prediction of prior probability.

In the case of the naïve Bayes pseudo-code syntax, the training data set T includes the following, $F = (f_1, f_2, f_3, \dots, f_n)$, these being the values of the predictor variables in the test data set.

The analysis process is carried out as follows: read the training data set T then calculate the average and the norm of the predictor variables for each class, repeating the iteration. The next step is to calculate the probability of using the Gaussian density analysis equation for each class, repeating this cycle until estimation of the probability of all prediction variables $(f_1, f_2, f_3, \dots, f_n)$. Finally, the probability for that class is calculated and the results with the highest weight are taken into account. The naïve Bayes model can be used and is ideal for dynamic data, having the quality of being easily remodeled when there is a load of new data and requiring a reconstruction of the nodes.

2.3. Utility of k -NN (k -Nearest Neighbors)

The k -nearest neighbors algorithm is used in classifying new positions using reciprocal class elements and overlapping nearest neighbors K to measure distance [40].

It can fall into the category of supervised learning algorithms, and the main recognized feature is that of intrusion detection and recognition of objects and shapes. k -NN is nonparametric—see Figure 7—it cannot be assigned a specific distribution that specifies data or makes predictions about it. When we want to get the most viable results in terms of taxonomy, it is necessary to plan algorithms and systems for detecting the paradigm we are in. For this reason, there may be different planners who perform the classification for certain models and actions, but they can make progress, allow analysis, and have training sets. When a model can perform an exact execution, the classified data sets and styles are correct because they have variable classifiers and they do not allow overlapping. Therefore,

classification planners supplement the information on the analysis and make the process easier. Classifiers such as naïve Bayes have a wider coverage by detecting fake news, especially on social media, with results leading to an accuracy of about 80%. The results can be improved by introducing bordered decision trees and improving the gradient that supports vectorization.

```
// KNN Pseudocode
Classify (X,Y, x) // X-training data, Y-classlabels of X, x-unidentified
for i = 1 to m do
{
  calculate distance d(Xi, x)
end for
Calculate set I containing indices for k smallest distance d (Xi, x).
return majority label for {Yi where i∈I}
} while (true);
```

Figure 7. k-NN pseudo-code structure.

3. Results

The direct correlation between the methodology used and the classification is based on different models, and the implementation of algorithms dedicated to the detection of fake articles becomes a challenge and a necessity because the influence of fake news on the population sometimes results in loss of life. Whether we are talking about using machine learning or supervised learning methods to classify data sets, the first step is to collect the data set, preprocess it, implement the selected features, then create the training data sets and the final test, data, and execution of classifiers. In Figure 8, the methodology by which the testing and analysis were performed is transposed. This methodology was based on exposing information by conducting experiments on social media data sets using algorithms such as those presented above, only adjusting to the level of analysis and classifiers.

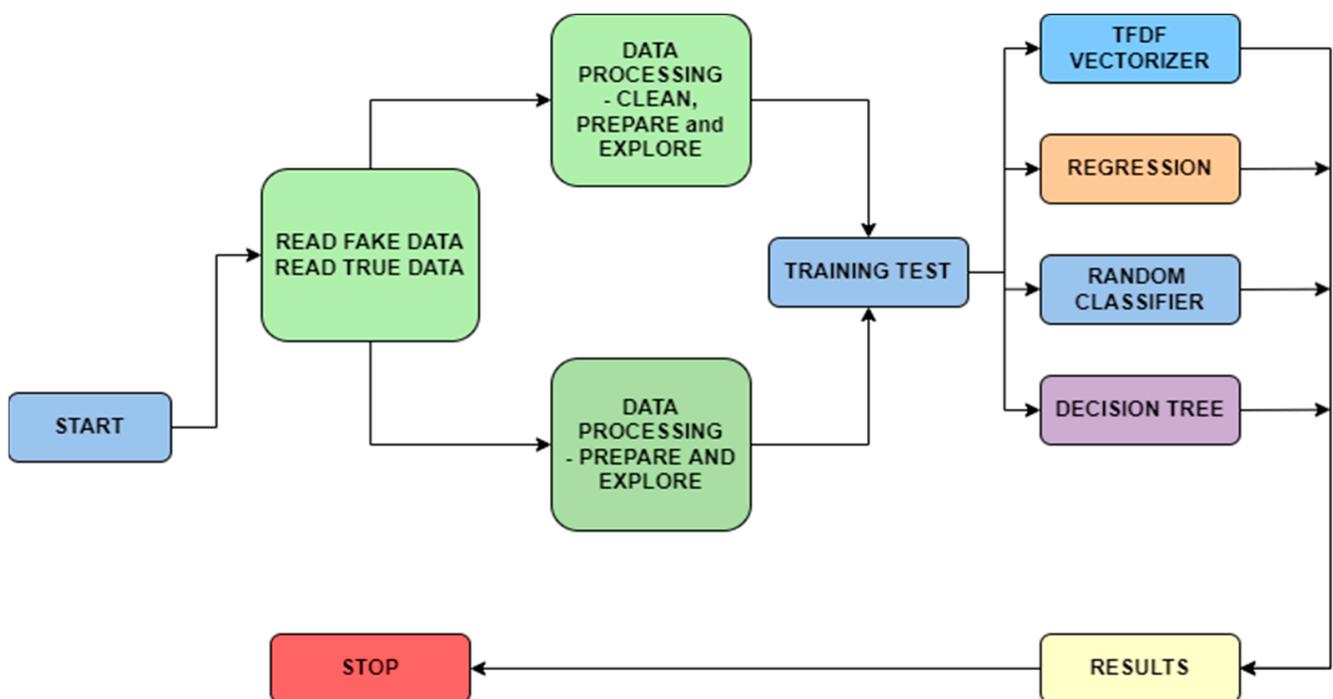


Figure 8. Descriptive diagram of the methodology regarding the proposal of the analysis algorithm.

Reducing the number of fake news becomes a major goal, and applying a set of algorithms for ranking is the first step that is urgently needed, just by integrating them into search engines or introducing filters that automatically scan fake news by analyzing details

and keywords. Incorporating Python language models into existing algorithms can perform the analysis and discovery of false information with a high degree of accuracy compared to other approaches. Prior to analyzing and proposing a detection method, Python code was refactored and improved to achieve the desired optimization. The basics in shaping the algorithm were structured on the k-nearest neighbors (k-NN), linear regression, naïve Bayes, decision tree, and vector analysis models. The use of these models in a single architecture could anticipate and increase the degree of detection with an accuracy up to 90%.

The analyzed data sets contained both fake and real news, and the files consisted of over fifteen thousand examples of news, divided equally. Therefore, the data set took into account characteristics such as title, text, subject, and also the date on which the materials were posted to increase the accuracy of the analysis process. Thus, Figure 8 shows the proposed architecture through which the first stage of content analysis was performed. In the data analysis process, we used preprocessing scales, and the files tested the content in several data cascades. This aspect can facilitate the application of several classification models, and as in the present case, aspects that increase the degree of machine learning. In view of shaping the model, a new approach was outlined by collecting and filtering data or more precisely eliminating background noise. Instantiation and iteration of data sets were divided into feature matrices, classified as components of natural language.

Once the data set was shared, a first iteration was applied to the information brushing that was not part of the training sets, creating a branched classifier model, which in this case was necessary. As shown in Figure 9, it was observed that the preprocessing was outlined with a message confirming the application of the algorithm on the drive portion. By responding to the system with the help of naïve Bayes and branched trees, an iterative sequence of cascading responses was outlined. Testing using data sets prepares the vector for monitoring and decision acceptance. Data sets are usually created with equal weights on the data, in which case half are real items and the other half are fake items, making the accuracy 50%. This aspect is as normal as possible because in the analysis, everything is done randomly and the final weighting coefficient exceeds about 85% of the data, with the rest being left as a test set for the proposed model. Analyzing text data takes time and preprocessing because the ambiguity of the information or expression can make an interpretation become imperceptible and that noise can take time in the process of elimination. Therefore, the use of processing using natural language processing exposes tokenization of words and encodes them into categories, such as integer results, as well as values with a floating scale in order to be acceptable as input data for ML-type algorithms. This whole process results in the extraction and vectorization of the entire feature set, outlining a library called Python scikit-learning-fail that exposes a calling token that contains tools such as a vectorized counter and a vectorized data map. In the light of the presented situation, Figure 10 was made, which shows the data in the form of a confusion matrix. The use of accuracy in the degree of obtaining a result makes direct reference to the embedding layer of the characteristics, these vectors being fed layer by layer in a dimension. We could not form a convolutional neural network (CNN), which requires the application of characteristics and at least 60 filters with 6–7 different sizes. Being a script dedicated to generating a plug-in that can be installed in the browser, the preprocessing or processing factors must also be taken into account. The calculation of semantic similarities within sentences remains an extremely used variant, and the deep neural network has the ability to convert text into a vector sequence with a fixed length. Therefore, the two sequences were analyzed in terms of textual items for each title–body pair. Labeling and validation focused on linear transformation in order to reduce the feature set.

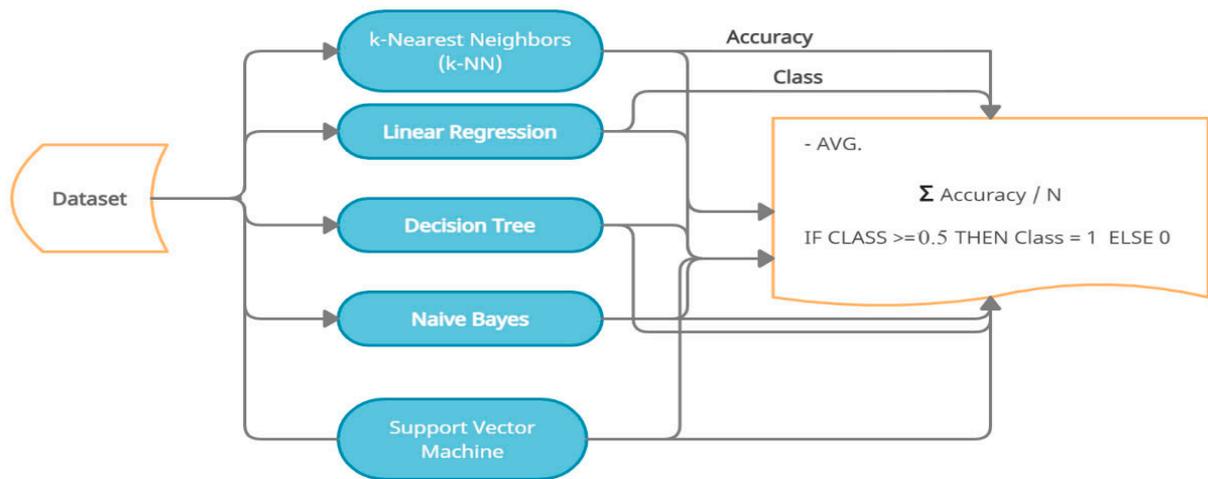


Figure 9. Descriptive diagram of classification algorithms in architecture.

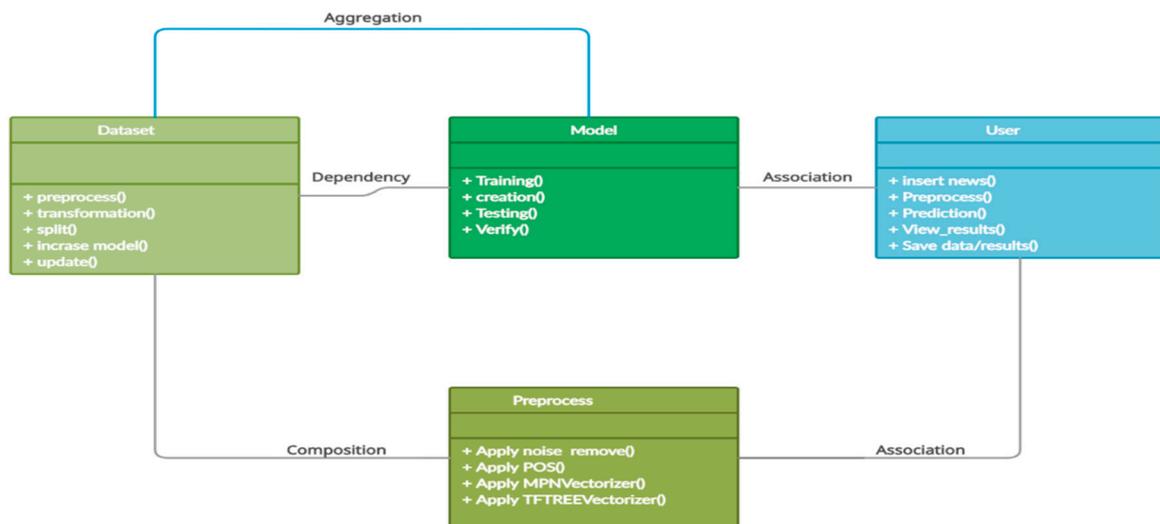


Figure 10. Diagram of detection model for fake information.

The ID, MstFake-Plus, identifies the preparation of the training data set, and this set automatically extracts training sentences from the analyzed article based on the submitted exposure. In order to obtain an answer or verdict on the text, the data were concatenated and truthful features were outlined. The basis of these features was consolidated on approximately four defining features (nouns, verbs, sentences, and prepositions), these being identified by a unique ID with a class tag (0, 1, 2, or 3) correlating with the model involved. In order to more accurately set out the procedure, the following steps were taken to assess the quality of the news.

In the context of data analysis and extraction, the transformation of incomplete raw data that did not have an ideal consistency into fragmented FDC data sets was addressed. Thus, the direct task of the NPL technique was applied by converting the data into texts and formatting the characters in small letters, eliminating the link words and roots and also calling the Keras libraries. Some of the common words found in the texts had minor importance in relation to the analyzed characteristics. When link words are removed, the processing time and space occupied by meaningless words decrease. Texts with similarities can form inconclusive sentences with a common basic form and the process can be defined as stemming, which in order to parse a stemmer of the NLTK algorithm was implemented. Thus, the execution of the preprocessing steps considerably reduced the number of itera-

tions per library and the tokenization with the help of the Keras library that divides each title into a word factor. After the preprocessing process, the data were vectorized in a word map and even a list of textual vectors. The presented aspects are the basis of the characteristics introduced in the architecture of the CNN with co-evolutionary layers. One of the most important contributions of the work refers to the analysis of language characteristics by modeling methods and obtaining new hybrid deep learning models. These are based on two layers of neural understanding, i.e., what we already know about CNN and keywords from the common language. The proposed approach produced much better predictive analyses, but the central aspect confers the feasibility of implementation in a navigator, as we need voluminous training sets. The development relationships between the models presented in the manuscript were established, but the classification was optimized in the process preceding the preprocessing. In mixed use, the set of unreduced features was incorporated into a new dimensionality reduction procedure. Therefore, the characteristics were selected after obtaining the training sets and after feeding the proposed architectures. The first layer of the model analyzes the input data, titles, and bodies of the article, the degree of conversion for each word in a vector form of size 100. Thus, the characteristics were found in an approximate number of over 5000, obtaining a matrix layer reported to previous dates. The output matrix contained raw data through the matrix multiplication procedure, producing a vector for each word. All these vectors were passed through the CNN layer in order to obtain contextual feature extractions.

Although studies showed that chi-square statistics can be the most efficient feature of the selection algorithm [41], it tests the relationships between all variables with a certain category. It tests and estimates the degree of independence between two words, comparing the quadratic distribution with a degree of freedom and judging the extremity [42]. The creation of independent features converges to the correlation between the keywords and the created tags, and the rest of the remaining features are eliminated. Afterward, the independence of the classes is calculated against the predefined threshold.

The data used was imported from the Kaggle platform so it can be used by anyone without limitation, there are two dedicated files, one with fake news and one with true news, and imports can be made in an international language. To begin, we tried to solve the problems using the methods described above through the Python solution, where in the first stage, we read and concatenated the data. Thus, we obtained a first import, after which we established labels and flags to be able to follow closely the false or true characteristics. The following process was concatenating the data and data frames to expose a viable iteration set in the mentioned issue. For ethical reasons and to obtain truthful information, all data were mixed up to avoid any problems of authenticity and to exclude bias.

Through this brushing process, a final form of the text was obtained, which then went through the exploitation of the data according to the subject and the priority that prevails. For this stage, the information was sorted according to the category it belongs to and grouping, and topics were plotted and displayed according to a diagram.

As we can see in Figure 11, the diagram provides information about the categories identified in the training files that contain the items to be analyzed. We had about 40,000 articles in different fields, and most of them covered the COVID-19 pandemic and events related to government tensions across the country. The period analyzed in this study was the first part of 2021 until the end of the year with including only material that generated impact and interactions at the level of the platforms where they were listed and analyzed.

The influence of this information on the pandemic or government tensions put enormous pressure on both the population and the neighboring countries in the areas affected by these data. The economic factor suffered and was directly influenced by the materials spread in the online environment and containing unfounded information, such as the period when no information was known about COVID-19 and most media trusts announced a food crisis, leading to the population depleting food stocks and strictly necessary items from the big chain stores. Lately, fake news in addition to misinformation of the population

has created a reluctance of people and stopped normal and human reactions, with a lack of empathy being a consequence of fake news.

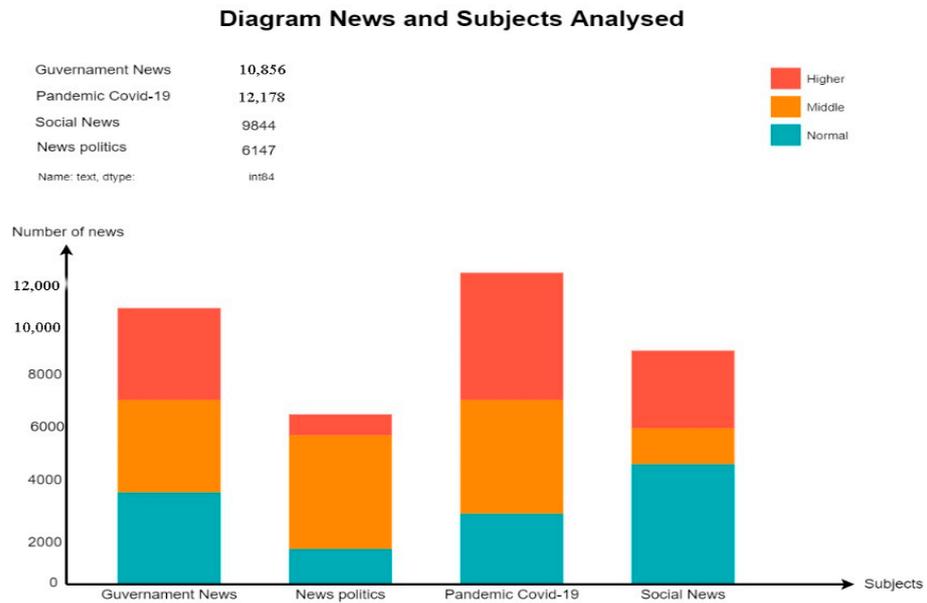


Figure 11. Diagram news and subject analyzed.

Thus, after sorting the information according to the desired categories, the decision-making process was prepared, which returned a true or false variable that aimed to identify fake news. A grouped matrix was established, and the text was sorted according to the established order; the plot result is found in Figure 12. Therefore, in order to highlight the keywords that were the basis of the real articles and the fake articles, we made based on the obtained data two taxonomy-type diagrams in which we identified the words that were most often used. In this process, we needed to import from the cloud where the data were processed and iterate it into two categories to target them properly. First, we imported the data from the fake data file and created a join for all the words, and the final process was performing a bilinear interpolation that ordered the repetitive words in the analyzed content and executed a number of four plots, with the result shown in Figure 13.

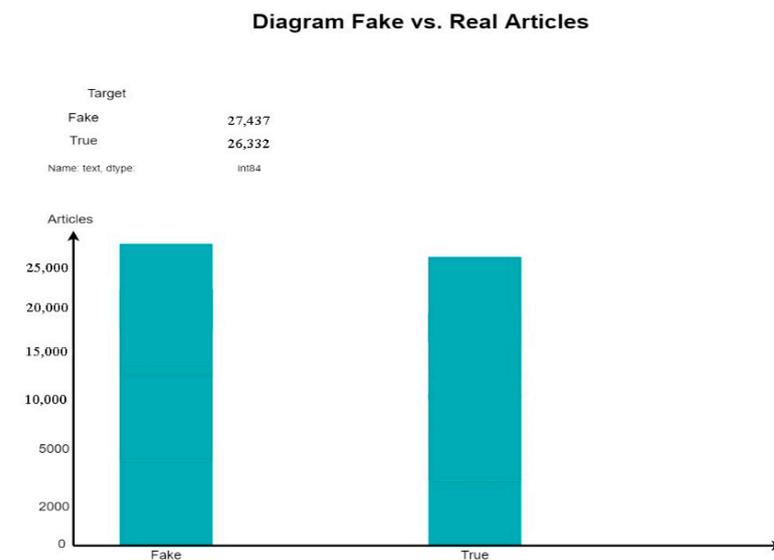


Figure 12. Diagram Fake vs. Real Articles.

were counted. Each word received a primary key and a unique ID to set different values, and the columns with the highest frequencies and the highest amounts were distinctly outlined by the others. Executing a vectorization of the columns with the most common words and their barring according to density can highlight the identified word and by a rotation at the level of the matrix label.

The density of identifying words in the data sets analyzed was an alarming one with a significant imprint on the political and social directions and health, with these being the most important for the economy of the countries. If the false information reaches as many people as possible, it can lead people to act out of self-defense, and the defensive and conservative phenomenon occurs because a person tries to save themselves. The impact on the economies of the countries where fake news prevails is found in most areas of the planet, and at the top are countries such as Romania, Bulgaria, Moldova, Hungary, Poland, Austria, and Greece. Therefore, if a piece of news leaves a piece of real information from a country such as France or Germany, this information becomes distorted in other countries, trying to distort either the situation or alarm the population in order to create panic and financial and social instability. The most common words in the fake news could be found by counting the total data, iterating each class to the number of targets set with a false value and referring to the number of words or their density; the same was done in the case of keywords for articles with real content. Therefore, Figure 15 is the chart showing the most common words in the fake news, and Figure 16 shows the most common words in the real news.

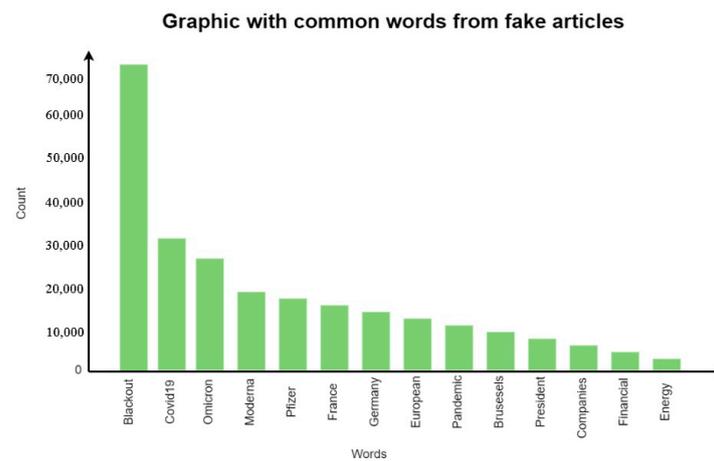


Figure 15. Graphic with common words from fake articles.

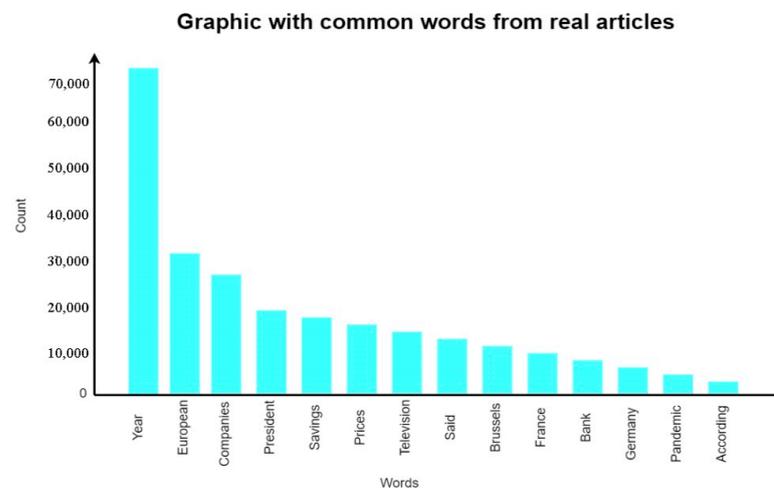


Figure 16. Graphic with common words from real articles.

In order to finalize the analysis process, the modeling was performed, which consisted of a way of vectorizing the entire corpus that was stored in the text-type instance, and the machine learning and classification algorithm was applied. The data set was preprocessed in a variation of 14.8 K, the texts that were out of context were labeled differently, and using the NLP and NLTK libraries, the whole context was separated from the basic terms. Filtering and analyzing using POS (speech parts) transformed information and data sets into statistical values and data sets. We extracted lexical features and the number of words or their length through the vectorization process using the unigram and bigram iterations through the TF-IDF Vectorization function in Python with sklearn standard.

Therefore, if the data sets were separated in equal weights in order to concatenate the files, we obtained a common portion that produced a data set and then a confusion matrix influence between fake and real news. In order to highlight the appearance of the confusion matrix and how it behaves in the case of analysis, we imported the data in the form of a matrix and normalized both the titles and the content, taking into account the length of words and interpolations performed. Through normalization by the asymmetry between the sum of the word density on the decision axes, we obtained a confusion matrix; otherwise, the matrix was not normalized, true or false labels were set (see Figure 17) where content was excluded and the coarse versus classified content was plotted, and in Figure 18, the confusion matrix is highlighted.

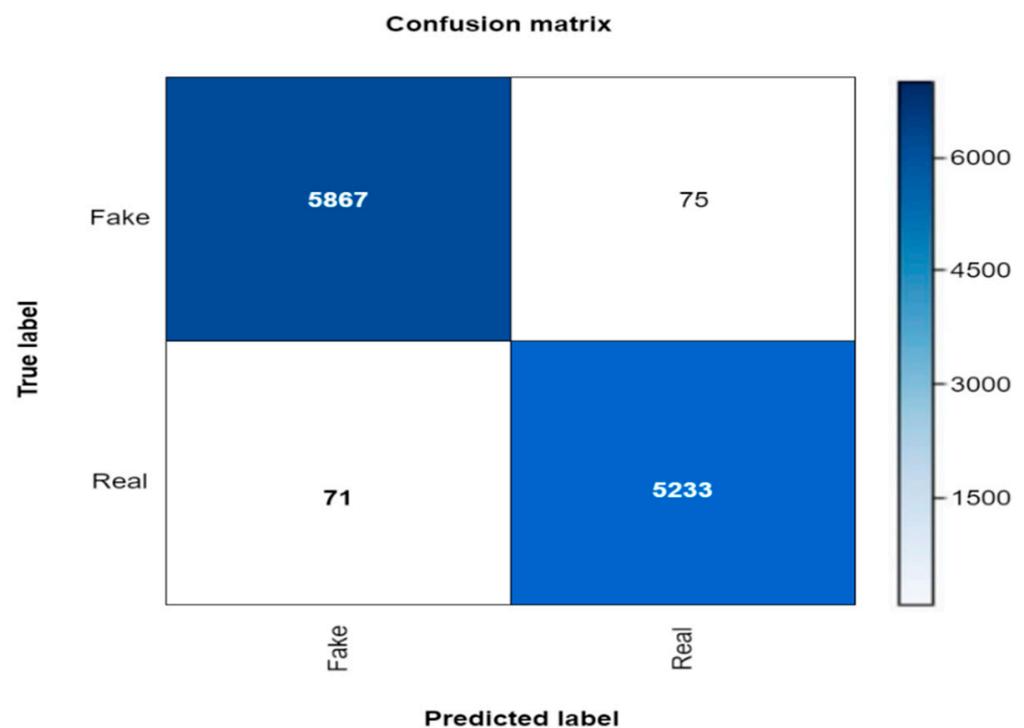


Figure 17. Confusion matrix without normalization.

Therefore, the data were divided into two test categories with randomized labels, logical regression was performed by vectorizing the data and applying RF-IDF, the data were vectorized, the models were transformed according to the degree of fit, and the logical regression model was applied which in its final form becomes a predictive model. In the last process, the data were formatted and a degree of accuracy regarding the analysis and shaping of the valid information within the training set was exposed. In the case of Figure 17, the accuracy is shown in the analysis performed, followed by the confusion matrix by the same predictive plotting procedure from the confusion matrix, outperforming the true or false type classes with an accuracy of 98.88%, and in Figure 18 is the confusion matrix for the decision tree with an accuracy of 99.55%.

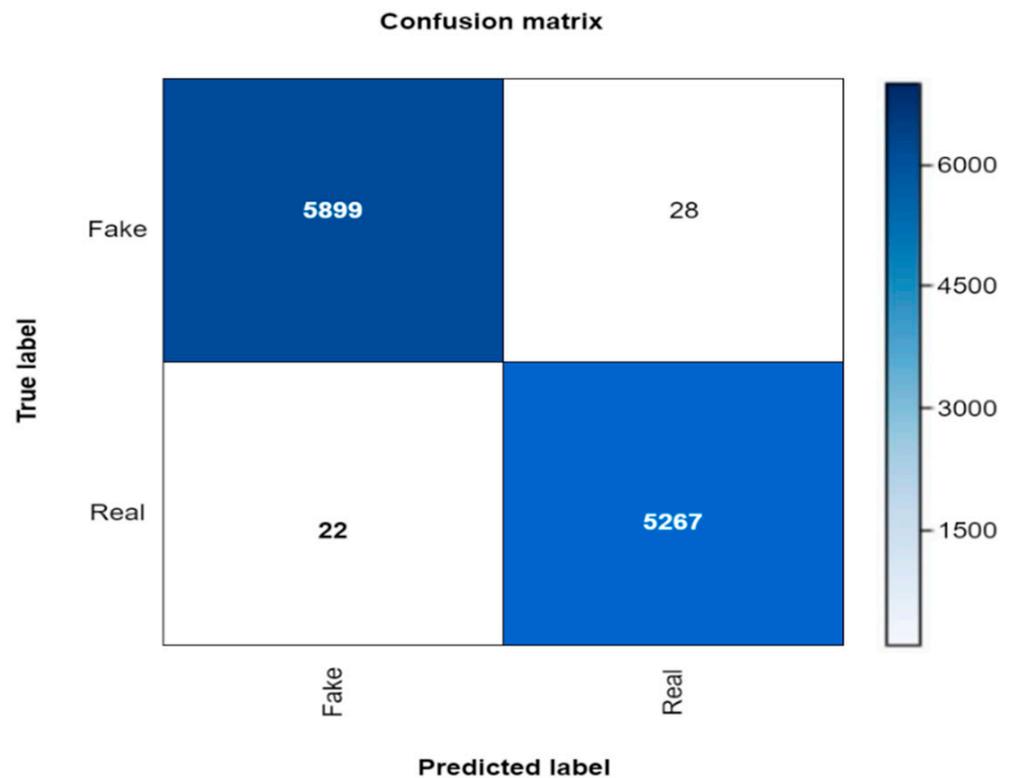


Figure 18. Confusion matrix for tree decision.

We evaluated the accuracy, precision, and recall of information with a relatively high impact score between fake and actual news. The proposed solution could identify and analyze the text by standard NLP methods in order to identify fake news directly in the browser, with the only limitation that the process was less automated and the impact created was minimal. This approach has the ability to turn into a plug-in that can be installed for most non-users in order to filter the information before it reaches the reader.

The construction of the machine learning model using data samples and training sets with false information helped to shape it, but there are a few issues when it comes to large-scale implementation and the use of such modules by users. Even if the implementation of Python is not so simple, it may become necessary to filter out false information that harms the image of public figures, states of the world, or their economies by the mere fact that it is false information that repeatedly reached the people who are easily influenced.

As we know, most browsers have a developer mode that allows those who use them to interact with some of the source code and the information behind the interface. In this case, a code reset was converted to JavaScript to extract false information and to outline the keyword density of an article. This method is one used only at the research level to present distinctly the mode of action of the interactions presented previously in another form. Figure 19 shows the extraction of word density from a trusted article that has not been highlighted by false information or news that puts pressure on the reader, although most articles have elements that hint at the use of subliminal messages for the reader and could not so drastically influence human thinking. The most common words and those that highlight the current situation in a contrasting note were extracted.

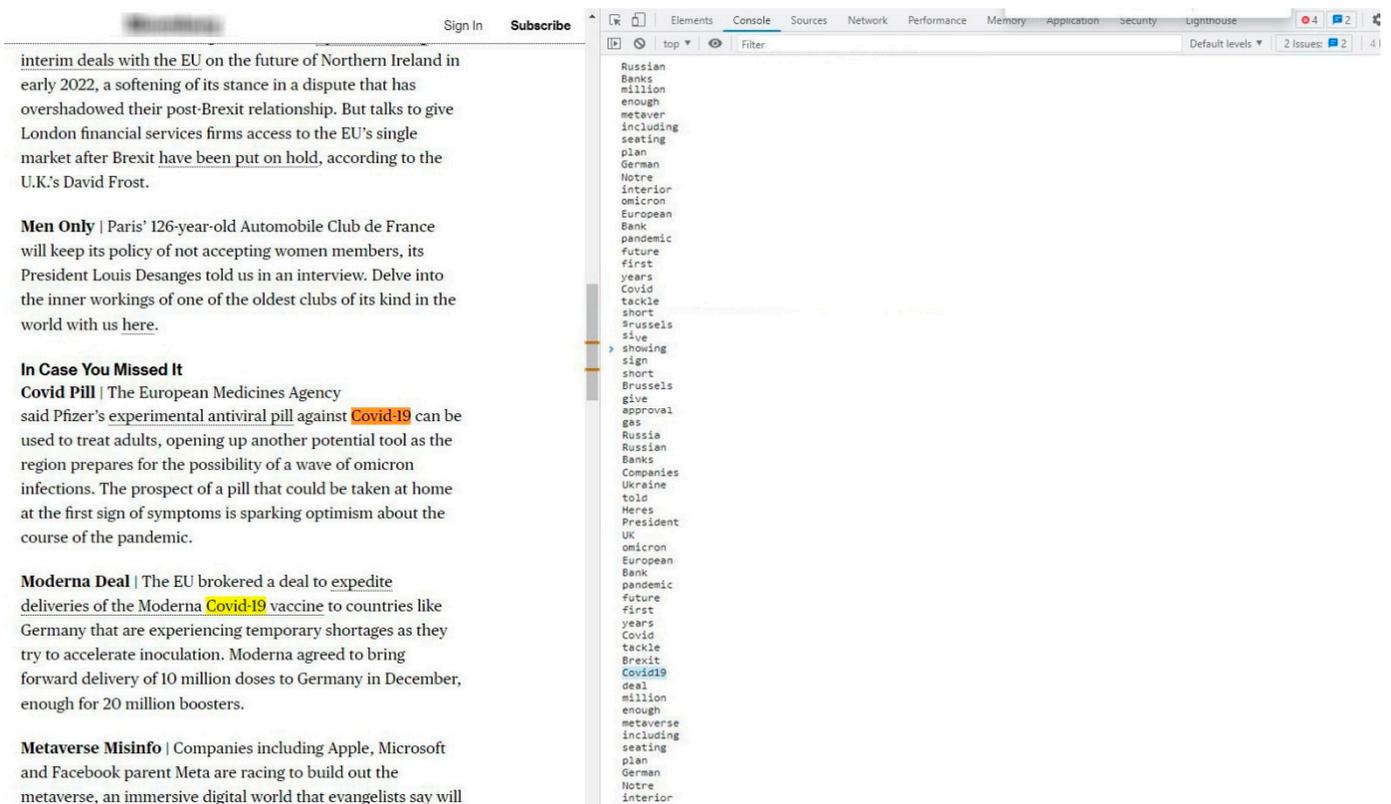


Figure 19. Extracted keyword density from a normal article using the NLP method in the browser using JavaScript. The exposure of the keyword identified and around which the entire article revolves COVID-19 highlighted following the query.

Regarding the accuracy of the data selection and the identification of false and true news, Table 1 shows the results obtained from the data cleaning process.

Table 1. The accuracy obtained from the data cleaning process.

No.	Machine Learning Models	Real News Identified	Fake News Identified	Accuracy
1	NLP	25,891	26,368	97.55%
2	k-NN	26,805	24,632	98.65%
3	TF-IDF	26,332	27,437	99.63%

A pop-up script was created, installable as an extension, based on a root folder containing all the files needed to extract and identify the fake content. To be able to access the content of web pages, we need a script dedicated to the content, which is equivalent to a JavaScript file. The script iterates each address, title, and content in the vector created based on keywords, and later alerts you when the content is less credible. In order to optimize the script, scraping extensions were created in order to call the JavaScript libraries and manipulate the DOM much more efficiently; see Figure 20a. An event page manages certain tasks much faster without downloading the active pages. The extension imports the existing data in the learning matrices and creates a JavaScript object to instantiate the pop-up window that notifies the user. The events are temporarily saved in a local JavaScript object after they are filtered. Within the script, you can find the filtering algorithm that uses match-type parameters. This aspect leads to the display of a percentage that gradually exposes the state of the content on that screen or web page; see Figure 20b.



Figure 20. General presentation of the mode of action of the scripts (a). Exposure regarding the degree of detection of the extension installed in the navigator (b).

4. Discussion

According to the study, the classifiers provided good results in detecting fake news, with an accuracy of over 90%. Thus, in terms of results of over 80% accuracy, they are considered successful studies, and the recent literature has shown that fake news continues to be a major problem in social and economic life. The present research showed that similar articles addressing this issue need to focus on the completed data set. Fake news can be controlled in the future, and future studies should focus on creating an interface, necessary for building a possible application, designed to attract through complex and accessible graphics. Such an application could facilitate the detection of fake news, and the user is able to differentiate between true and false information through machine learning methods. These methods, therefore, include TF-IDF, NLP, using JavaScript injection, NLTX, and Python, and by choosing this type of design, the methods used aim to generate results with very high accuracy [42].

From our study, we can see that the classifiers used ensured good results, obtaining an accuracy of over 90%. Previous research is considered to be successful, showing purity of over 80%, and our study provided very good results in the quick identification of fake news. The reduction in fake news has become an extremely important goal, and the application of a set of algorithms for ranking is the first step in their integration within search engines or the introduction of filters intended to automatically scan fake news. In order to identify a direct comparison of previous works with our study, a series of additional analyses were performed. In this study, the basic elements used in the outline of the algorithm were structured on k-nearest neighbors (k-NN), linear regression, naïve Bayes, decision tree, and vector analysis models. Thus, through these models incorporated in a single architecture, the degree of precision increased up to 90%. In order to obtain better accuracy, the cleaning process of the data set must be taken into account. The accuracy obtained in our work was compared with the values obtained in previous works, as shown in Table 2. Oshikawa et al. [43] used machine learning and NLP technology, and in terms of binary classification, they obtained results of over 90%. Additionally, Zervopoulos et al. [44] used a smaller data set, which included both fake and real tweets regarding the same event, using only linguistic features. They had an accuracy of 93.6% with random forest on average. Another comparison was that with the study carried out by Volkova and Jang [45], which focused on the creation of predictive models with the aim of reducing false negatives by introducing psycholinguistic signals, with very good accuracy. Their study focused on the importance of linguistic predictive models. Table 2 shows a comparison of recently applied methods. Thus, in this table, we are able to observe a higher accuracy of our study

compared to other research, and a limitation of this study is represented by a series of fragile data sets.

Table 2. The accuracy of our study compared with other studies.

No.	The Most Used Machine Learning Models	Accuracy	Research
1.	TF-IDF vector using Python and JavaScript	99.63%	Our research
2.	BI-LSTM	91.51%	[43]
3.	NLP	97.55%	[44]

The COVID-19 pandemic has generated strong changes in the economic sector by imposing quarantine since the beginning of the pandemic and social distancing in order to limit the virus. Many companies stopped working during that period, so this led to a decrease in productivity and supply, uncertainty among people, salary cuts, and layoffs. All these factors have led to the closure of a significant part of the economy, and the internal consequences of a company led to disruption of trade, with a strong impact on business partners as well as the global value chain. The fake news phenomenon spread with the COVID-19 pandemic, so that economically, the fake news led to distrust and uncertainty among the people, and the COVID-19 pandemic compared to the 2008 crisis had a much stronger impact on the economy and GDP in various countries. Table 3 shows the impact of COVID-19 on GDP in various countries.

Table 3. Comparison between the economic impact generated by the 2008 crisis and the crisis generated by the COVID-19 pandemic.

GDP Growth Rate	2008	2009	2020	2021
World	3.1	−0.1	−3.1	6.1
China	9.6	9.4	2.2	8.1
Japan	−1.2	−5.7	−4.5	1.6
United States	0.1	−2.6	−3.4	5.7
European Union	1.6	−4.8	−5.6	5.4
Africa	4.4	3.1	−1.6	6.9

Source: Own processing according to the atlanticcouncil.org (accessed on 6 April 2022) [46].

As can be seen from Table 3, due to the global economic crisis of 2008, global GDP was reduced by 0.1%, and in 2020, due to the pandemic crisis, GDP was reduced by 3.1%. Additionally in China, economic growth has been positive, so in 2008, GDP was 9.6%, in 2020, it was reduced by 2.2%, and in 2021, it increased to 8.1%. Moreover, in the case of Japan, the impact of the crisis was negative, where in 2008, the GDP was reduced to 1.2%, and in 2020 to 4.5%. In the case of the United States, in 2008, the GDP was 0.1%, and in 2020, it decreased by 3.4%. In terms of the European Union, GDP fell by 4.8% in 2009 due to the crisis and by 5.6% in 2020. In Africa, GDP stood at 4.4% in 2008 and fell by 1.6% in 2020. Therefore, the COVID-19 pandemic, in addition to spreading fake news, also generated an economic crisis in which the global GDP was reduced by 3% and some of the most developed countries and regions underwent significant changes.

5. Conclusions

Social networks generate more fake news and at the same time real dangers for society unlike traditional media. The amplification of fake news with the help of social platforms has grown considerably in the last decade. Therefore, the result of the conducted study provides a clear insight by adjusting the analysis and machine learning classifiers. The study offered a success rate of over 90, differentiated and exposed fake news, and at the same time limited the user's interaction with fake material. The applied techniques and the adjustment of automatic learning algorithms and natural language processing contributed substantially to the results obtained. One of the most important aspects is closely related to

the implementation of an extension that can be installed on the browser, and this may limit the interactions of users with the material that distorts the truth. Future developments must focus on the development of a perfectly functioning GUI-type application and also the completion of the extension and its synchronization with most APIs from social networks. The final application can also be made in the form of a web platform in which to insert text that will later be processed, generating decision matrices and word clouds with the density and frequency of words. These two variants of further development may be able to provide an increased degree of safety and exposure to a material that distorts the truth. The involvement of new types of classifiers, decision trees, parameterized logical regression, and also new ways of classifying decision trees are part of the future research perspectives. The generation of new, more efficient models is necessary because some news involving negative aspects have enough ways to bypass the filtering and analysis of the algorithms. Therefore, the conducted study exposed the degree of danger to which users of social platforms and consumers of electronic content are exposed, as well as the way in which they are affected by false information.

Author Contributions: Conceptualization, E.Z. and A.N.; methodology, E.Z.; software, E.Z.; validation, E.Z. and A.N.; formal analysis, A.N.; investigation, A.N.; resources, E.Z.; data curation, E.Z.; writing—original draft preparation, A.N.; writing—review and editing, A.N.; visualization, E.Z.; supervision, E.Z.; project administration, A.N. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kapoor, K.K.; Tamilmani, K.; Rana, N.P.; Patil, P.; Dwivedi, Y.K.; Nerur, S. Advances in Social Media Research: Past, Present and Future. *Inf. Syst. Front.* **2018**, *20*, 531–558. [[CrossRef](#)]
2. Pérez-Escoda, A.; Pedrero-Esteban, L.; Rubio-Romero, J.; Jiménez-Narros, C. Fake News Reaching Young People on Social Networks: Distrust Challenging Media Literacy. *Publications* **2021**, *9*, 24. [[CrossRef](#)]
3. Baptista, J.P.; Gradim, A. Understanding Fake News Consumption: A Review. *Soc. Sci.* **2020**, *9*, 185. [[CrossRef](#)]
4. Casero-Ripollés, A. Influencers in the Political Conversation on Twitter: Identifying Digital Authority with Big Data. *Sustainability* **2021**, *13*, 2851. [[CrossRef](#)]
5. Stitini, O.; Kaloun, S.; Bencharef, O. Towards the Detection of Fake News on Social Networks Contributing to the Improvement of Trust and Transparency in Recommendation Systems: Trends and Challenges. *Information* **2022**, *13*, 128. [[CrossRef](#)]
6. Aldwairi, M.; Alwahedi, A. Detecting Fake News in Social Media Networks. *Procedia Comput. Sci.* **2018**, *141*, 215–222. [[CrossRef](#)]
7. Zaman, A. COVID-19-Related Social Media Fake News in India. *J. Media* **2021**, *2*, 100–114. [[CrossRef](#)]
8. Rusu, M.-L.; Herman, R.-E. Legislative Measures Adopted at the International Level Against Fake News. *Int. Conf. Knowl. Based Organ.* **2019**, *25*, 324–330. [[CrossRef](#)]
9. Nasir, J.A.; Khan, O.S.; Varlamis, I. Fake news detection: A hybrid CNN-RNN based deep learning approach. *Int. J. Inf. Manag. Data Insights* **2021**, *1*, 100007. [[CrossRef](#)]
10. Canavilhas, J.; Jorge, T.D.M. Fake News Explosion in Portugal and Brazil the Pandemic and Journalists' Testimonies on Disinformation. *J. Media* **2022**, *3*, 52–65. [[CrossRef](#)]
11. Romea, A.C.; Valero, D.; Elboj, C.; Melgar, P. Mental Health and Access to Information in Times of COVID-19: The Role of Social Work. *Int. J. Environ. Res. Public Health* **2022**, *19*, 4483. [[CrossRef](#)] [[PubMed](#)]
12. Alnazzawi, N.; Alsaedi, N.; Alharbi, F.; Alaswad, N. Using Social Media to Detect Fake News Information Related to Product Marketing: The FakeAds Corpus. *Data* **2022**, *7*, 44. [[CrossRef](#)]
13. Sa, A.; Hinkelmann, K.; Corradini, F. Combining Machine Learning with Knowledge Engineering to detect Fake News in Social Networks—a survey. *arXiv* **2019**, arXiv:2201.08032.
14. Shu, K.; Sliva, A.; Wang, S.; Tang, J.; Liu, H. Fake News Detection on Social Media. *ACM SIGKDD Explor. Newsl.* **2017**, *19*, 22–36. [[CrossRef](#)]
15. Ahmad, I.; Yousaf, M.; Yousaf, S.; Ahmad, M.O. Fake News Detection Using Machine Learning Ensemble Methods. *Complexity* **2020**, *2020*, 8885861. [[CrossRef](#)]

16. Ahmed, M.; Riaz, M.Q.; Qamar, M.; Asghar, R. Fake News Shared on WhatsApp During COVID-19: An Analysis of Groups and Statuses in Pakistan. *Media Educ. Mediaobrazovanie* **2021**, *17*, 4–17.
17. Hussein, D.M.E.-D.M. A survey on sentiment analysis challenges. *J. King Saud Univ. Eng. Sci.* **2018**, *30*, 330–338. [[CrossRef](#)]
18. Conroy, N.K.; Rubin, V.L.; Chen, Y. Automatic deception detection: Methods for finding fake news. *Proc. Assoc. Inf. Sci. Technol.* **2015**, *52*, 1–4. [[CrossRef](#)]
19. Bali, A.P.S.; Fernandes, M.; Choubey, S.; Goel, M. Comparative performance of machine learning algorithms for fake news detection. In Proceedings of the International Conference on Advances in Computing and Data Sciences, Ghazibad, India, 12–13 April 2019.
20. Vosoughi, S.; Roy, D.; Aral, S. The spread of true and false news online. *Science* **2018**, *359*, 1146–1151. [[CrossRef](#)]
21. Abdullah, A.; Awan, M.; Shehzad, M.; Ashraf, M. Fake news classification bimodal using convolutional neural network and long short-term memory. *Int. J. Emerg. Technol.* **2020**, *11*, 209–212.
22. Pérez-Rosas, V.; Kleinberg, B.; Lefevre, A.; Mihalcea, R. Automatic detection of fake news. In Proceedings of the 27th International Conference on Computational Linguistics, Santa Fe, NM, USA, 20–26 August 2018.
23. Davis, R.; Proctor, C. Fake News, Real Consequences: Recruiting Neural Networks for the Fight against Fake News. 2017. Available online: <https://web.stanford.edu/class/archive/cs/cs224n/cs224n.1174/reports/2761239.pdf> (accessed on 2 April 2020).
24. Endo, P.T.; Santos, G.L.; Xavier, M.E.D.L.; Campos, G.R.N.; de Lima, L.C.; Silva, I.; Egli, A.; Lynn, T. Illusion of Truth: Analysing and Classifying COVID-19 Fake News in Brazilian Portuguese Language. *Big Data Cogn. Comput.* **2022**, *6*, 36. [[CrossRef](#)]
25. Buchanan, T. Why do people spread false information online? The effects of message and viewer characteristics on self-reported likelihood of sharing social media disinformation. *PLoS ONE* **2020**, *15*, e0239666. [[CrossRef](#)] [[PubMed](#)]
26. Zhou, X.; Zafarani, R. Fake news: A survey of research, detection methods, and opportunities. *arXiv* **2018**, arXiv:1812.00315.
27. Bondielli, A.; Marcelloni, F. A survey on fake news and rumour detection techniques. *Inf. Sci.* **2019**, *497*, 38–55. [[CrossRef](#)]
28. Buchanan, T.; Benson, V. Spreading Disinformation on Facebook: Do Trust in Message Source, Risk Propensity, or Personality Affect the Organic Reach of “Fake News”? *Soc. Media Soc.* **2019**, *5*, 2056305119888654. [[CrossRef](#)]
29. Preston, S.; Anderson, A.; Robertson, D.J.; Shephard, M.P.; Huhe, N. Correction: Detecting fake news on Facebook: The role of emotional intelligence. *PLoS ONE* **2021**, *16*, e0258719. [[CrossRef](#)]
30. Molina, M.D.; Sundar, S.S.; Le, T.; Lee, D. “Fake News” Is Not Simply False Information: A Concept Explication and Taxonomy of Online Content. *Am. Behav. Sci.* **2021**, *65*, 180–212. [[CrossRef](#)]
31. Egelhofer, J.L.; Lecheler, S. Fake news as a two-dimensional phenomenon: A framework and research agenda. *Ann. Int. Commun. Assoc.* **2019**, *43*, 97–116. [[CrossRef](#)]
32. Jang-Jaccard, J.; Nepal, S. A survey of emerging threats in cybersecurity. *J. Comput. Syst. Sci.* **2014**, *80*, 973–993. [[CrossRef](#)]
33. International Political Science Abstracts. *Int. Politi Sci. Abstr.* **2021**, *71*, 1–162. [[CrossRef](#)]
34. López-García, X.; Costa-Sánchez, C.; Vizoso, Á. Journalistic Fact-Checking of Information in Pandemic: Stakeholders, Hoaxes, and Strategies to Fight Disinformation during the COVID-19 Crisis in Spain. *Int. J. Environ. Res. Public Health* **2021**, *18*, 1227. [[CrossRef](#)] [[PubMed](#)]
35. Buturoiu, R.; Udrea, G.; Oprea, D.-A.; Corbu, N. Who Believes in Conspiracy Theories about the COVID-19 Pandemic in Romania? An Analysis of Conspiracy Theories Believers’ Profiles. *Societies* **2021**, *11*, 138. [[CrossRef](#)]
36. de Oliveira, N.; Pisa, P.; Lopez, M.; de Medeiros, D.; Mattos, D. Identifying Fake News on Social Networks Based on Natural Language Processing: Trends and Challenges. *Information* **2021**, *12*, 38. [[CrossRef](#)]
37. Hakak, S.; Khan, W.Z.; Bhattacharya, S.; Reddy, G.T.; Choo, K.-K.R. Propagation of Fake News on Social Media: Challenges and Opportunities. In *International Conference on Computational Data and Social Networks*; Springer: Cham, Switzerland, 2020; pp. 345–353. [[CrossRef](#)]
38. Sarker, I.H. Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Comput. Sci.* **2021**, *2*, 160. [[CrossRef](#)]
39. Alenezi, M.N.; Alqenaie, Z.M. Machine Learning in Detecting COVID-19 Misinformation on Twitter. *Futur. Internet* **2021**, *13*, 244. [[CrossRef](#)]
40. Zhai, Y.; Song, W.; Liu, X.; Liu, L.; Zhao, X. A chi-square statistics based feature selection method in text classification. In Proceedings of the 2018 IEEE 9th International Conference on Software Engineering and Service Science (ICSESS), Beijing, China, 23–25 November 2018; pp. 160–163.
41. Xia, X.; Lo, D.; Qiu, W.; Wang, X.; Zhou, B. Automated configuration bug report prediction using text mining. In Proceedings of the 2014 IEEE 38th Annual Computer Software and Applications Conference, Vasteras, Sweden, 21–25 July 2014; pp. 107–116.
42. Ghinadya; Suyanto, S. Synonyms-Based Augmentation to Improve Fake News Detection using Bidirectional LSTM. In Proceedings of the 2020 8th International Conference on Information and Communication Technology (ICoICT), Yogyakarta, Indonesia, 24–26 June 2020; pp. 1–5. [[CrossRef](#)]
43. Oshikawa, R.; Qian, J.; Wang, Y.W. A Survey on Natural Language Processing for Fake News Detection. In Proceedings of the 12th Language Resources and Evaluation Conference, Marseille, France, 11–16 May 2020; pp. 6086–6093.
44. Zervopoulos, A.; Alvanou, A.G.; Bezas, K.; Papamichail, A.; Maragoudakis, M.; Kermanidis, K. Hong Kong Protests: Using Natural Language Processing for Fake News Detection on Twitter. In *IFIP International Conference on Artificial Intelligence Applications and Innovations*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 408–419.

45. Volkova, S.; Jang, J.Y. Misleading or Falsification: Inferring Deceptive Strategies and Types in Online News and Social Media. In Proceedings of the The Web Conference 2018 (WWW '18). International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, Lyon, France, 23–27 April 2018; pp. 575–583.
46. Strauss-Kahn, M.O. Can We Compare the COVID-19 and 2008 Crises? The Table Is Based on the Study of the Article Published in Atlantic Council. Available online: <https://www.atlanticcouncil.org/blogs/new-atlanticist/can-we-compare-the-covid-19-and-2008-crises/> (accessed on 6 April 2022).