Towards the Comprehensive Detection of Fake News in Socio-digital Media in Mexico with Machine Learning

Carlos Augusto Jiménez Zarate and Leticia Amalia Neira Tovar

Abstract Contemporary society has made information and communication technologies (ICT) a fundamental axis for socio-digital interaction, we live increasingly hyperconnected through the Internet, and although there are various platforms for socio-digital interaction, only some have achieved globalization and a massive reach of users and have established themselves among the users of social networks on the Internet, to disseminate news content. But the massive use of socio-digital platforms has brought with it harmful practices such as the dissemination of fake news. This research is a review of the literature, of investigations that have used machine learning and other techniques or methods such as natural language processing, text classification and neural networks, for the detection of fake news, to develop an algorithm for the automatic detection of fake news, in the socio-digital content broadcast in Mexican Spanish.

Keywords Socio-digital networks · Fake news detection · Machine learning

1 Introduction

Between 2004 and 2006, Facebook and Twitter were founded; both platforms represented a turning point in social communication. Since then, new platforms for socio-digital interaction have been generated; these new social networks are also known as social media. It is clear that each year more users of social networks are added, and with it the increase in the flow of content and information between people. According to the "15th Study on the Habits of Internet Users in Mexico" [1] presented by the Internet MX Association, it revealed that Internet users in Mexico were 82.7 million and that of these 82% (67.8 million) are users who use the Internet to access social networks and 76% (62.8 million) of users use the Internet to search for information.

C. A. Jiménez Zarate (✉) · L. A. Neira Tovar
Universidad Autónoma de Nuevo León, San Nicolas de Los Garza, Nuevo Leon, Mexico

F. Torres-Guerrero et al. (eds.), *2nd EAI International Conference on Smart Technology*, EAI/Springer Innovations in Communication and Computing,
https://doi.org/10.1007/978-3-031-07670-1_5

Socio-digital interaction has brought with it an increase in negative practices to social communication, such as misinformation and false news or fake news; an investigation carried out by Knight Foundation (2018) revealed that more than one million false tweets are published per day on Twitter. Various civil and government organizations have called this misinformative phenomenon as infodemic; within the context of the COVID-19 pandemic, international health organizations such as the World Health Organization (WHO) and the Pan American Health Organization (PAHO) issued on May 1, 2020, a document to explain the infodemic [2].

The fight against fake news has taken on greater relevance; due to the COVID-19 pandemic, on May 21, 2020, the UN issued a tweet from its official account showing the launch of a tool to verify news through the portal "Shareverified.com" under the motto "There has never been a greater need for accurate and verified information."

On Twitter, content dissemination campaigns have been generated or developed with negative aspects, such as the spread of disinformation, fake news, the use of bots or automated digital positioning systems, or the dissemination of negative or hateful content, as indicated by Stella et al. [3].

Mønsted et al. [4] determined that socio-digital networks generate structures that propagate content or information, which can be analyzed with greater efficiency through complex contagion models, and also assume that the probability of content adoption depends on the number of unique sources of information. In this context, the complex contagion of content propagation bots can be targeted as news disseminators with erroneous information, disinformation, or fake news.

In the "Report on disinformation campaigns, fake news and their impact on the right to freedom of expression," which was prepared by the National Human Rights Commission of Mexico [5], warns about the use of content with information that is not attached to objective or inaccurate facts, and that exploits the emotions or beliefs of the audiences, to attract more "likes" or "retweets" on the Facebook and Twitter platforms; it also mentions that most of the citizens do not have the time, resources, or instruments to verify the content or information they receive in an increasingly connected society.

In the bulletin UNAM-DGCS-318, the study entitled "Radiography on the Dissemination of Fake News in Mexico" is mentioned [6], prepared by UNAM, which ensures that at least 89% of users on Twitter in Mexico have been exposed to this type of content. Fake news affects various areas, such as business marketing, as determined by Visentin et al. [7], who investigated the negative effect on brands that advertise in media or sites that spread fake news; their results indicate that marketing managers should be encouraged to monitor, since the proliferation of fake news, constitutes there is an increasing risk for the business sector. The increase in the use of socio-digital platforms has brought with it an increase in harmful practices to social communication, such as the spread of fake news, the use of bots, trolling, and artificial positioning strategies. Vosoughi et al. [8] determined that fake news spreads further, faster, and deeper than real news on social networks like Twitter, as well as fake news on political issues; they have a higher level of spread than other social topics.

For the analysis of this growing and continuous amount of digital information, it is necessary to implement machine learning algorithms, as well as other artificial intelligence tools, because these tools provide a great capacity for data processing and have become an indispensable support for the development of highly competitive predictive models.

2 Literature Review

To guide the investigation on the construction of a system for the detection of fake news on Twitter in Mexico, this investigation has condensed the most relevant investigations that can help its development. Shao et al. [9] analyzed 14 million tweets, where 400 thousand articles were shared during 10 months between 2016 and 2017; through their analysis, they managed to find evidence that much of the misinformation is due to super propagators that are social bots that publish automatically links to articles; the analysis tools they used were the Hoaxy and Botometer verification systems, which were developed by researchers at Indiana University. Davis et al. [10], developers of the BotOrNot system, claim that the classifier generates more than 1000 characteristics through the use of metadata and information extracted from patterns and the content of the interaction.

The Hoaxy system collects public tweets that contain links to news; the platform is freely accessible and allows systematic studies on a large scale, on topics or hashtags that are part of a fake news dissemination strategy. Shao et al. [11] used the Hoaxy platform to carry out an investigation of the dissemination of erroneous information before and after the US presidential election in 2016, the study was based on the analysis of the core of the propagation networks, determined that the network of users is polarized between true or false information. The dissemination and propagation of fake news cover different areas or themes of society, but they can also be categorized by dimensions as described by Shu et al. [12], who determined three types of dimensions (content, social, and temporal); their research made it clear that fake news is not an insignificant matter, since they are built to lie to readers and propose from an analysis point of view of social networks, a method of inoculation before the spread of fake news, which consists of identifying the nodes, routes, or main propagation links, and with this information, strategies for inoculation, blocking, or containment can be created. Ahmed et al. [13] focused on the detection of spam and fake news through text classification, for which they developed a new n-gram model.

The detection of fake news is not easy; it requires models and systems that can summarize and compare the news with reliable sources to be able to categorize them; that is why alternatives are sought such as identifying the position through the automatic detection of the relationship between two pieces of a text. Thota et al. [14] developed a model where they used the deep learning architecture of neural networks, with vectorization through a bag of words with a dense neural network, to be able to categorize the positions; the model showed good results to categorize the headings and new articles or news. Altunbey et al. [15] compared more than 20 supervised artificial intelligence algorithms for the classification of fake news and determined that the decision tree algorithm obtained a better result.

Oehmichen et al. [16] determined the characteristics of the accounts that spread fake news; in their research, they started from the creation of a dataset, for which they collected for 4 months the tweets related to the hashtags of the presidential election in the USA of 2016; they took into account tweets greater than 1000 retweets and managed to create a dataset of 9001 tweets. They determined that the fake news spreading accounts are recently created; the vast majority are unverified, have fewer updates, use strange characters in the name and description, have few followers and follow many more, and are generally dedicated to interact with retweets.

Fake news can be used to stifle social protest movements. Zervopoulos et al. [17] used various machine learning techniques, such as naive Bayes, support vector machine, C4.5, and random forest, to be able to classify the characteristics linguistics of the fake news; for this, they took the tweets in English and Chinese from a Twitter database to be able to classify the fake news. Zhou et al. [18] proposed a multimodal analysis system, which integrates the textual and visual analysis of the news; for this, they resorted to the construction of a dataset with information from news verification sites in the USA.

An important space for analysis in the Spanish language has been the development of the Semantic Analysis Workshop (TASS) which is part of the actions of the Spanish Society for Natural Language Processing (SPNL), which aims to encourage semantic analysis in Spanish language. In this effort, the IBERLEF (Iberian Languages Evaluation Forum) has been integrated, where a competition is developed to encourage research for word processing for Iberian languages such as Spanish, Portuguese, Catalan, Basque, and Galician. Salas et al. [19] implemented an analysis scheme, using a system of machine learning algorithms, a model to determine Spanish and Mexican satire on Twitter. The results of their research showed a high accuracy for detecting satire and that there is no significant difference in satire from both countries.

Posadas et al. [20] conducted an investigation to detect fake news in the Spanish language, for which they created a new dataset of the content broadcast on Twitter by formal media and media that regularly publish false content; they used four algorithms for classification machine learning, which were support vector machine, logistic regression, random forest, and boosting. Within this review, very few investigations focused on the detection of fake news for the Mexican Spanish language were found.

Table 1 summarizes the investigations that are considered relevant due to their methods for the construction of an efficient system for the detection of fake news in Twitter Mexico networks.

Table 1 Literature review. Most relevant papers for detecting fake news with machine learning

Author(s)	Research title	Method
Shao et al.	Hoaxy: A Platform for Tracking Online Misinformation (2016)	Extraction of tweets containing the URLs of the websites
Shao et al.	The Spread of Low-Credibility Content by Social Bots (2017)	Extraction of articles shared on Twitter, verification of users who shared the articles or content of low credibility
Ahmed et al.	Detecting Opinion Spams and Fake News Using Text Classification (2017)	They used 3 existing datasets and created a new one made up of 12,600 fake news and 12,600 legitimate news. In addition, they used two extraction techniques and six learning classification techniques (stochastic gradient descent, linear support vector machines, K-nearest neighbor, logistic regression, and decision trees)
Salas et al.	Automatic Detection of Satire in Twitter: A Psycholinguistic-Based Approach (2017)	They created a dataset with satirical and non-satirical news and obtained data from Mexican and Spanish sites. And for the analysis they used machine learning algorithms for their classification
Vosoughi et al.	The Spread of True and False News Online (2018)	They created a dataset with 500K reply tweets that included the links from the verification sites to other tweets (original tweet). Later, from the original tweet, I determine its cascades of propagation. This in order to quantify the spread of rumors or false news
Thota et al.	Fake News Detection: A Deep Learning Approach (2018)	They used three different neural network architectures, with the TF-ID vector representation of combined words being the best performing one
Altunbey et al.	Fake News Detection Within Online Social Media Using Supervised Artificial Intelligence Algorithms (2019)	They developed a two-phase method. The first is to convert unstructured data into structured data, using weighting vectors. The second phase is an experimental evaluation of 23 supervised algorithms
Oehmichen et al.	Not All Lies Are Equal. A Study into the Engineering of Political Misinformation in the 2016 US Presidential Election (2019)	They created a dataset of 9001 tweets with more than 100 retweets, referring to the 2016 presidential election in the USA. They used syntax and sentiment analysis
Posadas et al.	Detection of Fake News in a New Corpus for the Spanish Language (2019)	They created a news dataset of formal media and sites that regularly post fake news. They used machine learning algorithms such as support vector machine, logistic regression, random forest, and boosting
Zaizar et al.	ITCG's Participation at MEX-A3T 2020: Aggressive Identification and Fake News Detection Based on Textual Features for Mexican Spanish (2020)	They used the dataset from the MEX-A3T Natural Language Processing Assessment Forum, and for classification they used the Natural Language Toolkit (NLTK) for tokenization and stopword removal. For the training of their model, they made use of cross validation with learning algorithms
Zhou et al.	SAFE: Similarity-Aware Multi-modal Fake News Detection (2020)	They used a multimodal approach that integrates the visual textual analysis of the news, to be able to categorize them as true or false; the dataset was obtained from news verification sites in the USA

3 Proposed Method

The monitoring and analysis of socio-digital interaction have become essential for the analysis and planning of communication strategies and socio-digital interaction, either to know social opinion, or in the development of strategies for digital marketing campaigns in business sectors, social or political, as Antoniadis et al. [21].

This research proposes the implementation of machine learning algorithms for data processing and analysis, since they are capable of systematically analyzing large dataset and being able to categorize them without the interference of human bias.

After reviewing the literature and the state of the art regarding the detection of fake news on Twitter Mexico, this research proposes to use text categorization techniques for the body or text of the news (URL in tweet) indexed in the tweet, for the classification of users using the random forest algorithm. For the analysis of the spread of content, it will be done using the concepts of social network analysis and the method to determine diffusion cascades of Goel et al. [22].

The independent variables for the development of this research will be:

(a) Tweet text
(b) Text of the news (URL in tweet)
(c) Broadcast users
(d) Propagation of the tweet

The dependent variables will be the following for the text of the tweet and the text of the news that is contained in the URL of the tweet:

(a) Fake news
(b) Satire
(c) Propaganda
(d) Real news

The dependent variables for the independent variable of @user will be:

(a) Bot
(b) Troll
(c) Human

The dependent variables for tweet propagation will be:

(a) Viral
(b) Non-viral

The independent and dependent variables proposed will be integrated into a system that may be capable of detecting false news and other variants of the news content broadcast on Twitter Mexico (Fig. 1).

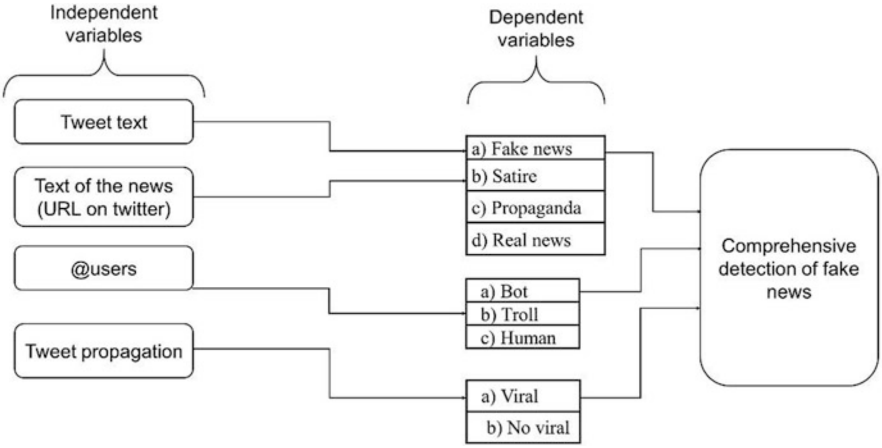

Fig. 1 Block diagram of the independent and dependent variables for the model proposed for the detection of fake news in the socio-digital media in Mexico

4 Development Phases

The first phase of the proposed system for detecting fake news on Twitter Mexico with machine learning will be the review of the literature on the various automatic techniques for detecting fake news.

The second phase consists of creating a dataset with tweets in Mexican Spanish for training and testing. For this, you will need to partition the dataset into two datasets:

- The first training partition will be by manual classification of the tweets and is composed of 70% for the dataset.
- The second partition will contain 30% of the dataset for the test dataset.

The training dataset will be processed and analyzed with various machine learning word processing algorithms. With the analysis of the text, the tweet can be classified to determine if it is fake news, satire, propaganda, or real news.

The third phase will consist of designing and building a comprehensive algorithm for the detection of fake news in Mexican Spanish. That you will have to analyze and categorize fake news automatically and massively.

The fourth phase consists of testing the algorithm and analyzing the results and verifying if the proposed algorithm has an acceptable degree of efficiency in execution time and of efficiency in the detection of fake news.

The fifth phase will consist of communicating the results in a completed research paper.

In Table 2, you can see the phases to carry out the project of the algorithm for the detection of fake news in the socio-digital networks in Mexico. Currently, the project is in the phase of building the dataset for further training and testing. The later stage is the design and construction of the algorithm for the detection of fake news.

Table 2 Project calendar for the construction of the algorithm for the detection of fake news in the socio-digital networks in Mexico. The boxes in green are the actions that have already been carried out

Phases	2020		2021		2022	
	Time Interval I: January to June	Time Interval II: August to December	Time Interval III: January to June	Time Interval IV: August to December	Time Interval V: January to June	Time Interval VI: August to December
Phase 1: Requirements Analysis and Existing Investigations	🟩					
	🟩	🟩				
Phase 2: Build the dataset for training and testing			🟩			
Phase 3: Design and construction of the algorithm for the detection of fake news				▨		
Phase 4: Testing and analysis of results					▨	
Phase 5: Communication of results					▨	▨

5 Conclusion

The literature review found that studies and research on the detection of fake news with machine learning or artificial intelligence techniques have been carried out mainly in the English language or with translators. The most used techniques for the classification and detection of fake news have been those of SVM, logistic regression, decision tree, and naive Bayes.

During the literature review, it was observed that throughout various investigations, various datasets have been created that have ranged from official news media, alternative media, journalists' accounts, as well as fake news verification websites and also fake news websites and satirical content. The topics that have covered the papers read have been mainly on politics and society.

For the Mexican Spanish language on Twitter, there is very little research that has been carried out in recent years, and they are limited to analyzing the text of the news that is attached through a link or link in a tweet; this limits the scope for detection, since it leaves out add-ons that can be integrated for a better detection of fake news.

The proposed research will be useful in the first place for the detection of false news that is disseminated on social networks, and it will also be a tool that helps to report content detected as false news in the different socio-digital networks. The system can also be of great help in the development and monitoring of social communication strategies for any type of organization, be it business, social, governmental, or political. The use of this algorithm will be useful to improve the veracity of the contents in the socio-digital networks.

References

1. Movilidad en el Usuario de Internet Mexicano. 1–25.
2. E.S. Decisivo, G. Scholar, Entender la infodemia y la desinformación en la lucha contra la COVID-19. **395** (2020)
3. M. Stella, E. Ferrara, M.D. Domenico, Los bots aumentan la exposición a contenido negativo e inflamatorio en los sistemas sociales en línea. **1–21** (2020)
4. B. Mønsted, P. Sapieżyński, E. Ferrara, S. Lehmann, Evidence of complex contagion of information in social media: an experiment using Twitter bots. PLoS One **12**, 1–14 (2017)
5. C. Itzel, P. Farfán, Reporte sobre las campañas de desinformación, "no ticias falsas (fake news)" y su impacto en el derecho a la libertad CNDH. México (2019)
6. UNAM-DGCS-318. Radiografia de la propagacion de fake news en México. 1–2 (2020)
7. M. Visentin, G. Pizzi, M. Pichierri, ScienceDirect fake news, real problems for brands: the impact of content truthfulness and source credibility on consumers' Behavioral intentions toward the advertised brands. J. Interact. Mark. **45**, 99–112 (2019)
8. S. Vosoughi, D. Roy, S. Aral, News On-line. Science (80-.). **1151**, 1146–1151 (2018)
9. C. Shao, G.L. Ciampaglia, A. Flammini, F. Menczer, The spread of low-credibility content by social bots. Nat. Commun. (2017). https://doi.org/10.1038/s41467-018-06930-7
10. C.A. Davis, O. Varol, E. Ferrara, A. Flammini, F. Menczer, BotOrNot: A System to Evaluate Social Bots. 4–5 (2016). https://doi.org/10.1145/2872518.2889302
11. C. Shao, G.L. Ciampaglia, A. Flammini, F. Menczer, Hoaxy: A Platform for Tracking Online Misinformation (2016). https://doi.org/10.1145/2872518.2890098
12. K. Shu, H.R. Bernard, H. Liu, Studying Fake News via Network Analysis: Detection and Mitigation. 43–65 (2019). https://doi.org/10.1007/978-3-319-94105-9_3
13. S. Ahmed, K. Hinkelmann, F. Corradini, Combining machine learning with knowledge engineering to detect fake news in social networks – a survey. CEUR Workshop Proc. **2350** (2019)
14. A. Thota et al., Fake news detection: a deep learning approach. SMU Data Sci. Rev. **1**, 10 (2018)
15. B. Altunbey, Fake news detection within online social media using supervised artificial intelligence algorithms. Phys. A Stat. Mech. Appl. **540**, 123174 (2020)
16. A. Oehmichen et al., Not all lies are equal. A study into the engineering of political misinformation in the 2016 US presidential election. IEEE Access **7**, 126305–126314 (2019)
17. A. Zervopoulos et al., Hong Kong protests: using natural language processing for fake news detection on twitter. IFIP Adv. Inf. Commun. Technol. **584 IFIP**, 408–419 (2020)
18. Zhou. Syracuse: Detección multimodal de noticias falsas con reconocimiento de similitudes Abstracto. 1–21 (2020)
19. M.d.P. Salas-Zárate, M.A. Paredes-Valverde, M.Á. Rodriguez-García, R. Valencia-García, G. Alor-Hernández, Automatic detection of satire in Twitter: A psycholinguistic-based approach. Knowledge-Based Syst. **128**, 20–33 (2017)
20. J.P. Posadas-Durán, H. Gomez-Adorno, G. Sidorov, J.J.M. Escobar, Detection of fake news in a new corpus for the Spanish language. J. Intell. Fuzzy Syst. **36**, 4868–4876 (2019)
21. I. Antoniadis, P. Serdaris, A. Charmantzi, The application of social networking analysis in marketing: a case study of a product ' s page in Facebook. *2nd Int. Conf. Contemp. Mark. Issues* (2014)
22. S. Goel, A. Anderson, J. Hofman, D.J. Watts, The structural virality of online diffusion. Manage. Sci. **62**, 150722112809007 (2015)