Design and development of an assisted communication system, using machine learning and natural language processing, for the detection of fake news on Twitter

Altamirano Perez, Brayan Omar y Valencia Valencia, Odalis Dayanna

Departamento de Ciencias de la Computación

Carrera de Ingeniería en Software

Artículo académico, previo a la obtención del título de Ingeniero en

Software

Dr. Carrillo Medina, José Luis

Dr. Marcillo Parra, Diego Miguel

17 de julio de 2023

Design and development of an assisted communication system, using machine learning and natural language processing, for the detection of fake news on Twitter.

Brayan Omar Altamirano Perez, Odalis Dayanna Valencia Valencia and José Luis
Carrillo Medina

[1] Universidad de las Fuerzas Armadas ESPE, Sangolquí, Ecuador
.
`(boaltamirano,odvalencia1,jlcarrillo)@espe.edu.ec`

Abstract. Misleading news has always existed in the world, it has become much easier to spread false information in new communication technologies, allowing users to be not only consumers, but also producers of discourses that circulate and that are often of dubious source of information, known as "Fake News" and these have proliferated around the world. Twitter, a social network that allows users to read and share information, is not free from this problem and, as it is used as an information tool in Latin American media, the spread of fake news is the order of the day, which is why the use of APIs provides a unique value by allowing the communication of various modules, improving the efficiency of use and reliability of the information, The objective of this research is the creation of an extension that allows the detection of false news at a Hispanic level through its own datasets and machine learning, in turn allowing communication with an order bot that reports the results obtained to the staff in charge, allowing to comply with an efficient flow of assistance.

Keywords: API, deception, diffusion, impact, machine learning, mitigate, order bot, social media, systems

1   Introduction

1.1   A Subsection Sample

A social network is a communication tool that allows the web to participate in an open environment for publishing and exchanging data that is constantly developing, information is generated through millions of web pages or blogs created by users and organizations, as well as through social networks and initiatives such as Open Data. In this way, information is shared in free environments that can be used in various areas ranging from everyday life to the development of new technologies. All platforms on the Internet, which allow users to communicate, share and generate information without references to sources, became popular in the early 1990s. [1]

With the increase in the use of online social networks, users' interaction is increasing on social networks as they share ideas, thoughts, feelings, knowledge, news and advertisements of all kinds. [2]

In today's world with the influx in the rate of smart phones and the increasing popularity of social networking platforms, people have started to consume more news from social networks such as Facebook, Twitter, than traditional forms of media [3]. A lot of factors have contributed to this growth, which includes with traditional one-way communication and presenting a platform where the audience can respond, almost instantly and directly [4], As such these networks allow the exchange of ideas and information through a network, of which are internet based and allow users to interact and share content such as personal information, documents, videos and images, users interact with social networks through personal computers, tablets or smartphones, through applications or web-based programs [5].

The use of social networks has been adopted by people all over the world and is now part of the daily routine, making these spaces a place where brands and companies also want to connect with their prospects and customers, providing many positive opportunities [6], for example: sharing the company's vision, personalizing information and interacting directly with customers, the ability to segment society and get to know each of their customers better when people share their likes, preferences and other information as they approach the target. for the group valuable to the company. [7]

The use of social media platforms such as Facebook and Twitter are ubiquitous in modern times, making them powerful platforms for the propagation and consumption of news. It spreads when someone propagates it through sponsorships such as replying, sharing or reposting, without validating its authenticity. [8]

The social network Twitter is a social platform, and its main service is two-way communication with which you can share information of various kinds in a quick, simple and freeway [9]. It was created in 2006, and is currently used by all kinds of people, from teenagers to senior citizens, and even companies, generally to share opinions or reflections. [10]

Therefore, this social network allows users to be producers and consumers of content at the same time, and has facilitated the dissemination of both real and misleading information. [11]. Fake news is a phenomenon that has taken over the media, especially social networks. [12]. This part is known as fake news, which is defined as information that has been published with the intention of deceiving and confusing the reader, the media that produce this news imitate the way of publishing reliable media, ignoring the verification processes or intentionally adulterating the content. in this way, pretending to pass off a lie, or at best an opinion, as an objective fact. [13] [14]. This term was chosen as word of the year 2016 by the Oxford Dictionary, and is defined as "referring to or denoting circumstances in which objective facts are less influential in the formation of public opinion than appeals to emotion and personal belief". Post-truth is a broad phenomenon that impacts on multiple political, economic and social aspects [15].

Fake news on social media is a problem; the spread of fake news and misinformation online is one of the top ten dangers to society Any news content that is false, fabricated, biased, slanted, skewed and factually incorrect [16].

Therefore, to mitigate the problems related to the spread of fake news an application programming interface is used allowing the detection of Fake News in the social network

twitter, through machine learning algorithms, which will contribute to reduce the disinformation of users with this type of networks. For this, the term API is mentioned, which is an abbreviation of Application Programming Interfaces. It is a set of definitions and protocols used to develop and integrate application software, enabling communication between two software applications through a set of rules [17].

An API is a kind of bridge that connects various types of software or applications and can be created in various programming languages. In addition to good development, an API must have clear and objective documentation in order to facilitate its implementation. In addition, a predefined data format is often used to share information between systems in order to achieve integration between them. The most commonly used are XML (Extensible Markup Language), YAML (originally Yet Another Markup Language, but officially YAML Ain't Markup Language) and JSON (JavaScript Object Notation) for web applications. There is also a pattern in web APIs called REST (Representational State Transfer), which is a set of rules and definitions that allows projects to be developed with well-defined interfaces [18].

Automatic detection of fake news on the social network, Twitter, has recently attracted attention, although, from a technical perspective, it can be considered as a straightforward binary classification problem [19], the biggest challenge is the collection of sufficiently large training corpora, as manual annotation of tweets as fake or non-fake news is a costly and tedious process [20].

Therefore, the Twitter API allows developers to interact with the platform and access its data, through this resource, developers can create applications that integrate with Twitter and execute actions such as posting tweets, collecting user data, reviewing trends and generally analyzing activity on the platform [21].

For this reason, the purpose of this research is the implementation of an extension that allows the generation of an information circuit using AI in the exposure and selection of news taken from the Twitter site.

2 Background

2.1 The impact of the internet on everyday life.

The development of the internet has set in motion a debate on how communication via the internet affects social relations. The internet frees the individual from geographical restrictions and unites people around new communities of interest that are not tied to a specific location. We live in a new networked and globalized society, united through new technologies. The Internet is our tool for relational interaction and also poses new challenges for privacy and security.

The Internet has evolved a great deal since its creation, which is, however, very recent from a historical perspective, and little has remained of that first static network designed to carry a few bytes or to send a small message between two terminals. Today, infinite amounts of information are uploaded and downloaded on this electronic giant. Until not so long ago, the Internet was a simple repository of information where only those people capable of understanding and manipulating code were in charge of publishing and maintaining content; now we are all fundamental participants, having the possibility of generating content and commenting on existing content.

2.2   Evaluation of information sources on the internet.

Understanding the Internet as a social process, in constant development and mutation and not as a defined and finished product, is fundamental for understanding the need to develop mechanisms that enable an optimized use of the available resources [22]. Since it is improbable to combine freedom of expression with prior selection, it is up to information professionals to try to guarantee their users a relative order in this chaos, by defining and developing instruments to control the quality of the information [22].

3   System Description

The design of the applied software outlines a two-section approach: the training of the neuron and the structure of the web extensions in communication with the order bot, presented in Figure 1, generating a technology stack focused on the optimal performance of an artificial intelligence application.
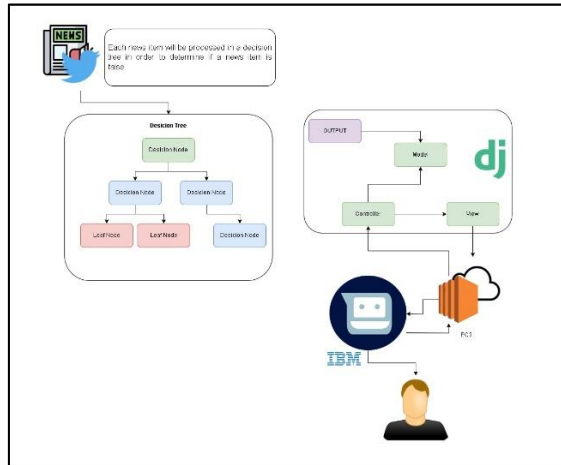

Fig. 1. Proposed Description Scheme.

In order to obtain a trained data set focused on detecting fake news, we proceeded to: Collect and prepare Twitter data, including both real and fake news, where you can use the Twitter API or third-party libraries to collect data, with the data collected we proceeded to make their respective preparation where the irrelevant ones were removed, normalized and filtered according to the base specifications [23].

With the data prepared, the data is divided into two sets, one to train the model and one to test its accuracy. Typically, a 70/30 ratio is used, meaning that 70% of the data is used to train the model and 30% is used to test the accuracy of the model.

4 System Description

4.1 A Subsection Sample

Please note that the first paragraph of a section or subsection is not indented. The first paragraphs that follows a table, figure, equation etc. does not have an indent, either.
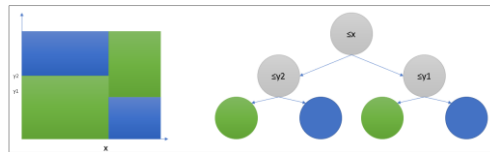
Subsequent paragraphs, however, are indented.

Fig. 2. Exemplification of the training of a decision tree.

Once the trained model is obtained, it is possible to detect whether or not it is a fake news, which is configured using an api created in DJANGO that will be in charge of managing tweets. In this section, the connection is made with the order bot designed in Watson, which will manage the information sent by the API and will determine whether the news is approved for transmission.

5 System Development

5.1 Development Process

The development of Detectweet assisted extension, developed for the detection of fake news in Spanish, started with the division of its processes, starting with:

Design and development of the dataset: Being a dataset, the objective of the dataset is to present a complete set focused on news stories in Spanish language that affected the country both locally and internationally.

To do so, we proceeded to investigate and take news from reliable sources within the country and from dubious channels, comparing the characteristics that make a real news from a false one.

Finally, an ordered list was obtained through a .csv file, compiling around 310 sources divided in 186 fake and 124 real news, as shown in figure 3.
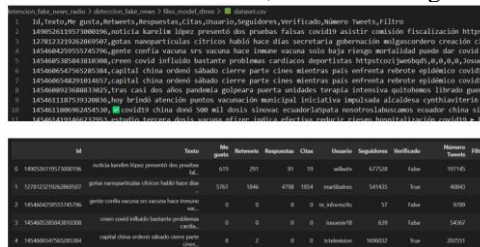

Fig. 3. Dataset presentation..

Validation criteria and algorithms: By a correct validation and implementation, several training algorithms were taken into account, obtaining better results and processing of two, decision trees, random forest and k-Nearest Neighbors(k-NN).

In order to validate the correct functioning of the algorithm, we sought to validate the following criteria:

- *Validity:* The reliability and trustworthiness of information.
- *Accuracy:* Closely linked to validity, it refers to the correctness of the information.
- *Authority and reputation of the source:* Specialty and status of the producer.
- *Uniqueness:* Amount of primary information not available from other sources.
- *Completeness:* Degree of completion of the information available.
- *Coverage:* Depth and breadth of the source.

Selection of decision algorithms: With the selected algorithms, we proceeded to carry out the specific tests using a prepared dataset, detailing below the results obtained for both the k-NN algorithm and for random forest and decision trees.

k-Nearest Neighbors (k-NN): supervised classification algorithm used to classify objects based on their similarity and the training objects.

The learning process of the algorithm proved not to be the best option when training the dataset, obtaining the following results (see Figure 4).
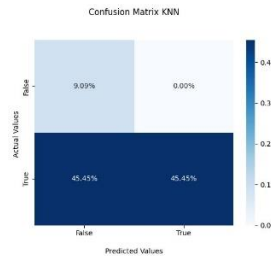

Fig. 4. Percentage of prediction applying KNN.

As can be seen in the image, the detection percentages do not exceed 50% at the time of applying KNN, and for this reason the aforementioned algorithm will be discarded.

Random Forest: a joint learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time.
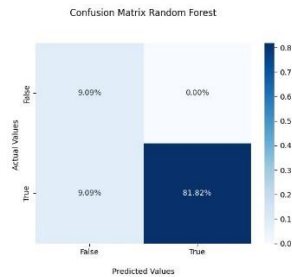

Fig. 5. Percentage of prediction applying Random Forest.

In the case of Random Forest, the prediction for positives is superior, so by making modifications it could be taken as an alternative as shown in figure 5.

Decision Tree: a method used to perform an analysis that consists of graphically outlining the possible outcomes, costs and consequences of a complex decision.
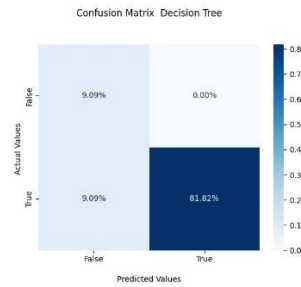

Fig. 6. Percentage of prediction applying Decision Tree.

The application of decision trees (see Figure 6) shows a similar performance to Random Forest, although its difference is in the level of performance, being faster than the previous one mentioned.

With the application of each one of the algorithms, it is possible to determine in a better way the best means for the training of the neuron, discarding KNN for an accuracy equal to 0.54 over 1 to determine if a news item is false or not and keeping as options the decision trees as random forest for its performance higher than 90% of accuracy, as it is seen in the image 7.
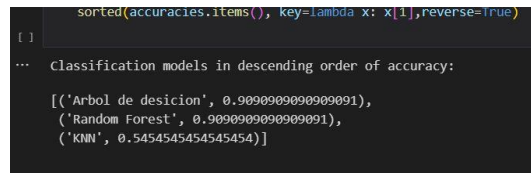
```
      sorted(accuracies.items(), key=lambda x: x[1],reverse=True)
[ ]

···   Classification models in descending order of accuracy:

      [('Arbol de desicion', 0.9090909090909091),
       ('Random Forest', 0.9090909090909091),
       ('KNN', 0.545454545454545454)]
```

Fig. 7. Percentage of success applying each of the selected algorithms.

The final step for the selection of the appropriate algorithm is the measurement of the response times, obtaining as the best result 68ms, the response time obtained when applying decision trees and remaining as the optimal option over the other options, as shown in figure 8.

```
'Arbol de desicion': 0.9090909090909091 --- 68ms
'Random Forest', 0.9090909090909091 --- 377 ms
'KNN', 0.5454545454545454----- 2.6 segundos
```

Fig. 8. Response times in each algorithm.

Integration of the Trained Model and API: In order to link the neuron with the API, the connection is made by means of a view in Django, which will process the requests from the API and call the neural network to perform the desired task. Once the view receives the data, the neural network will proceed to analyze it and send the requested result.

Creation of the chrome extension: With the API running correctly we proceed to package and debug the extension, making sure it works perfectly. Once done we create an EC2 instance in AWS making sure we have the ports open, once done we connect via SSH and configure the web server, finally we secure the permissions and check the URL of the EC2 instance in Chrome.

Communication with the Order bot: Once the communication with the API and the extension is stable, we proceed to make the communication between the results obtained and the requests defined in the order bot, for this we designed and implemented a midelware to make the communication between both modules, we also proceeded to establish the necessary rules and secrets in IMB Watson.

6   Result and Discussion

As Detectweet is a browser extension, its operation is simple and the only additional requirement apart from having the extension in the browser is to be logged in to Twitter.

To start using it, the first thing to do is to search for a news item on the platform. Once found, the extension will notify if you want to analyze the information, and if you accept, it will start the process of analyzing the news item as shown in figure 9.
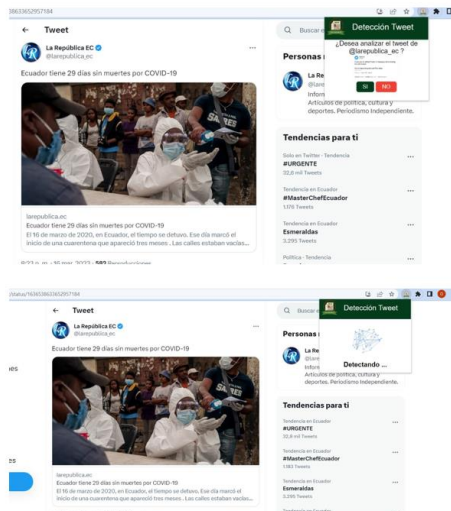

Fig. 9. Twepy execution process.

After the analysis process the extension will present two states, the first one notifying if the entered news is true and the second one if the news is false, sending these results to the order bot that will determine the value of the news and inform the radio announcer about it.

With the extension running, we proceeded to analyze the performance of the extension by testing it with fake and real news, each one checked and recently raised, obtaining the following results.

Table 1. Test comparison test between a specialized web site and Detectweet.

Ord	Tweet Id	Result on newtral.es	Detectweet prediction	Description
1	https://twitter.com/Africamar/status/1634288274998022191	Fake	True	Filipinas: emite una orden de arresto contra Bill Gates por "asesinato premeditado" en relación con el lanzamiento de la vacuna COVID-19.
2	https://twitter.com/dejanirasilveir/status/1632479592438022145	Fake	True	Putin ordena destrucción de todas las vacunas contra el C0VID-19 en Rusia.
3	https://twitter.com/MakerlyYenny/status/1582335544755093504	Fake	Fake	Con las #vacunas que ha comprado la UE, cada europeo tocaría a 178 dosis. ¿Qué responsabilidad tiene Úrsula Von der Leyen, cuyo marido trabaja en una farmacéutica?
4	https://twitter.com/DIWWOOO/status/1603774311034015746	Fake	Fake	Fallo de la Corte Suprema de EE. UU.: las vacunas contra el covid no son vacunas.
..	…….	….	…..	…….
13	https://twitter.com/Newtral/status/1595073374304481281	True	True	No hay evidencias de que la subvariante XBB de la covid provoque una enfermedad "más grave y mortal"
...	……………………..	….	…..	……..
19	https://twitter.com/Newtral/status/1621184368675880961	True	True	La consejería de sanidad de tu comunidad decide cómo incorporar esta vacuna.
20	https://twitter.com/Newtral/status/1624390092625149954	True	True	El papel de las mujeres en el combate contra la COVID-19.

Test one resulted in 5 hits out of 10 for false news and 10 out of 10 hits for true news (see table 1).

With a better calibration, further tests were carried out, obtaining as final results the values presented in table 2.

Table 2. Second test comparison test between a specialized web site and Detectweet

Ord	Tweet Id	Result on newtral.es	Detectweet prediction

1	https://twitter.com/Africamar/status/1634288274998022191	**Fake**	**Fake**
2	https://twitter.com/dejanirasilveir/status/1632479592438022145	**Fake**	**Fake**
3	https://twitter.com/MakerlyYenny/status/1582335544755093504	**Fake**	**Fake**
4	https://twitter.com/DIWWOOO/status/1603774311034015746	**Fake**	**Fake**
..	…….	….	…..
13	https://twitter.com/Newtral/status/1595073374304481281	**True**	**True**
…	…………………..	….	…..
19	https://twitter.com/Newtral/status/1621184368675880961	**True**	**True**
20	https://twitter.com/Newtral/status/1624390092625149954	**True**	**True**

Demonstrating the correct functioning of the extension and thereby adding another tool to the veracity of the information.

Conclusions and Future Work

The application of artificial intelligence tools are a step towards freedom of expression and the veracity of information, they have now expanded, so that combating the lack of information and malicious publications are a means of defense for the innocent, also to implement AI today is increasingly accessible to the general population, allowing a steady growth in its use.

As future work, it is proposed to improve the order bot in order to allow direct communication with the radio audience, allowing the automation of programming and generating a means of emergency in case there is no active staff, or an emergency arises.

References

1. I. Dongo, Y. Cardinale y A. Aguilera, «Análisis de credibilidad de las fuentes de información disponibles en la web: Una revisión y una contribución,» 20 noviembre 2019.
2. M. Ramezani, M. Rafiei, S. Omranpour y H. Rabiee, «Etiquetado de noticias lo antes posible: ¿real o falso?,» 27 agosto 2019.
3. S. A. S. Moreira, N. S. Brandão, M. P. P. d. S. Belcavello y G. d. O. Andrade, The presence of TIC in the teaching-learning process of youth and adults with Down syndrome: The game as didactic-pedagogical resource, IEEE, 2017.
4. A. Bedi, N. Pandey y S. Kumar, «Un marco para identificar y asegurar los problemas de noticias falsas y rumores en las redes sociales,» 18 octubre 2019.

5. A. Martín, «Las Redes Sociales más utilizadas: cifras y estadísticas,» 26 octubre 2022. [En línea]. Available: https://www.iebschool.com/blog/medios-sociales-mas-utilizadas-redes-sociales/.

6. D. R. Manjunath y B. S. Hadimani, Hierarchical Clustering and Regression Classification based Review analysis on Movie based Applications, IEEE, 2019.

7. Rd Station, «¿Qué son las redes sociales?,» 2023. [En línea]. Available: https://www.rdstation.com/es/redes-sociales/#:~:text=El%20principal%20prop%C3%B3sito%20de%20las,una%20categor%C3%ADa%20de%20social%20media..

8. B. Rath, W. Gao y J. Srivastava, «Evaluación de la vulnerabilidad a las noticias falsas en las redes sociales: un modelo de evaluación de la salud comunitaria,» 27 agosto 2019.

9. Webempresa,, «¿Qué es twitter? ¿Cómo funciona? ¿Cómo puedo usarlo para mi organización?,» 1 marzo 2018. [En línea]. Available: https://www.webempresa.com/blog/que-es-twitter-como-funciona-2.html.

10. L. Berlanga, «Qué es Twitter, cómo funciona y qué te puede aportar esta red social,» 30 julio 2022. [En línea]. Available: https://www.ciudadano2cero.com/que-es-twitter/.

11. FIP, «¿Qué son las Fake News?,» 2023. [En línea]. Available: https://www.ifj.org/fileadmin/user_upload/Fake_News_-_FIP_AmLat.pdf.

12. B. Borja, «Así afectan las noticias falsas a las redes sociales y medios de comunicación,» 2023. [En línea]. Available: https://www.utadeo.edu.co/es/articulo/crossmedia-lab/277626/asi-afectan-las-noticias-falsas-las-redes-sociales-y-medios-de-comunicacion.

13. S. Nikam y R. Dalvi, «Modelo basado en algoritmo de aprendizaje automático para la clasificación de noticias falsas en Twitter,» 07 octubre 2020.

14. G. Usha, P. Madhavan, M. V. Cruz, N. A. S. Vinoth, Veena y M. Nancy, Enhanced Ransomware Detection Techniques using Machine Learning Algorithms, IEEE, 2022.

15. P. Haya, «¿Cómo se viralizan las noticias falsas? Algoritmos y bots sociales,» 2020. [En línea]. Available: https://www.iic.uam.es/innovacion/como-se-viralizan-noticias-falsas-algoritmos-bots-sociales/.

16. M. Hazim and. S. Ahsan, «Uso de noticias falsas y redes sociales por parte de los principales canales de noticias de la India,» 28 febrero 2020.

17. Y. Fernández, «API: qué es y para qué sirve,» 23 agosto 2019. [En línea]. Available: https://www.xataka.com/basics/api-que-sirve.

18. SYDLE, «¿Qué es API? Ejemplos, ventajas y tipos,» 24 febrero 2022. [En línea]. Available: https://www.sydle.com/es/blog/api-6214f68876950e47761c40e7.

19. H. A. Jannah y D. Hermawan, Analysis of Indonesian Society's Perceptions of the COVID-19 Vaccine in Youtube Comments Using Machine Learning Algorithms, IEEE, 2022.

20. S. Helmstetter y H. Paulheim, «Aprendizaje débilmente supervisado para la detección de noticias falsas en Twitter,» 28 agosto 2018.

21. N. Verdejo, «La Api de twitter dejará de ser gratuita, a excepción de una prueba limitada,» 2 febrero 2023. [En línea]. Available: https://wwwhatsnew.com/2023/02/02/la-api-de-twitter-dejara-de-ser-gratuita-a-excepcion-de-una-prueba-limitada/#:~:text=La%20API%20de%20Twitter%20es,y%20acceder%20a%20sus%20datos..

22. M. Alzyout, E. A. Bashabsheh, H. Najadat y A. Alaiad, Sentiment Analysis of Arabic Tweets about Violence Against Women using Machine Learning, 2021.

23. A. Unnikrishnan, U. Narayanan y S. Joseph, Performance analysis of various supervised algorithms on big data, IEEE, 2018.