Technology Journal of Management , Accounting and Economics (TECH)
ISSN: 2311-3995

Vol. 13 No. 2 (2025)

Mistakes vs. Malice: Automatic Classification of Unintentional Fake News on Celebrity Deaths

Beter M. Roberto
Assist Prof.Dr. University of Phoenix
Roberto.der@gmail.com

Abstract

This study addresses the growing problem of information clutter, particularly in the context of celebrity death announcements. As the rapid spread of misinformation becomes a critical social issue, especially through mass and social media, it is essential to understand the mechanisms behind this phenomenon. The research employs an interdisciplinary approach, integrating computational, linguistic, sociological, and journalistic perspectives to analyze the characteristics of unintentional fake news. Using machine learning classification techniques, it seeks to differentiate between unintentional fake news, real news that debunks these false claims, and real news. The findings reveal significant linguistic features that contribute to the classification process. However, while there are models capable of classifying certain specific news types, none of them are able to correctly classify all news types considered, underscoring the complexities involved in distinguishing between correct, misinformation, and disinformation. This work not only sheds light on the nature of unintentional fake news but also emphasizes the need to improve fact-checking processes in journalism to combat the viral spread

of misinformation. Ultimately, the study calls for further research into the implications of these findings for media practices and the role of technology in addressing the challenges of information disorder in contemporary society.

Keywords: Fake news; mass communication; machine learning; natural language processing.

Introduction

Fake news is a message that has been deliberately created to misinform, including false information within a general context of truth, in order to manipulate public opinion among certain social groups. According to UNESCO [1], this term is in itself an oxymoron, since news should, by definition, be verifiable news. Fake news is an example of disinformation, a phenomenon that deliberately combines malicious information (malinformation) and misinformation (misinformation) [2]. This is best illustrated in Figure 1 . Malicious information is information created solely to defame or harm; for example, hate speech, harassment, and the leak of sensitive and private information. Misinformation refers to information that is misleading or false, but not deliberate, due to misinterpretations or translations, typing errors, or other problems on the part of its issuers. All of these types of information problems fit within the more general concept of information disorder [3]. While fake news has existed for centuries [4], the ubiquity and instantaneousness of current technology allows for its real-time spread on a global level, representing a critical problem for societies at different levels, such as the political [5], the health [6] or the societal [7].

Technology Journal of Management , Accounting and Economics (TECH)
ISSN: 2311-3995

Vol. 13 No. 2 (2025)

Source: Based on [3], [4] .

Figure 1 Taxonomy of information disorder.

In this context, the object of study of this work, much less studied than disinformation fake news, is what we will call unintentional fake news, a type of misinformation that occurs when a news story is adapted and shared based on a fake news story, believing that the original source is true. The frequent use of the Internet and social networks as the only source of information leads to a loss of quality of informative content, which makes it difficult for journalists to adequately verify the facts and circumstances that are published [8] . This type of phenomenon is increasingly occurring in mass media, where economic needs and scarce resources lead to fast-paced editorial environments, in which communicators do not have enough time to verify news facts. This can lead to

incorrect publications that fuel the spread of hoaxes or deceptions, erroneous claims, and unverified rumors [9] . Unintentional fake news stories often expand and complement the content of the original fake news story, transforming what was once a simple Twitter/X post or Instagram image into a seemingly full-blown press release. This problem contributes to the rapid viralization and consolidation of fake news stories, making them even more difficult to detect [10] .

A paradigmatic case of unintentional fake news is the announcement of celebrity deaths. This is the case with the news of Chomsky's fake death in June 2024, which contrasts with other real news stories of equivalent media impact, such as the death of Stephen Hawking in March 2018.

Since there is no clarity regarding the authorship of this type of fake news or its main motives, comparative literature raises some hypotheses. From political and behavioral psychology, studies refer to the return of the idea of the search for epistemological certainty [11] , that is, being certain that one knows something that "no one else knows," especially useful when thinking about news with a conspiratorial nature, and how readers engage with such information. Other works raise a possibility of content monetization, at a time when news consumption is directly related to reach and usage metrics [12] , [9] , [8] , [13] . It could even be suggested that they are part of social experiments, since the reasons are not as clear, unlike other fake news of a political nature or conspiracy theories.

It is in its nonspecific nature that we find the analytical possibility of considering this problem. In the case of news about celebrity deaths, there doesn't seem to be a purpose to establish or manipulate discourse, but the intention of the initial creator,

whether for "experimental" purposes or not, is equally conscious and generates negative effects. Therefore, the original news item does qualify as fake news and, therefore, as disinformation. In contrast, mass media reports that amplify and amplify this false content, usually originally posted on social media, qualify as misinformation.

The phenomenon can also be addressed from the urgency of the journalistic scoop [14] .

Specifically, sociological theory has presented the press as a field [15] under which competition prevails, causing a need to stay ahead of the curve [16] . While it is difficult to link it to the intentional generation of false content and disinformation, it can help understand its rapid and widespread transmission to other media and social media platforms.

Currently, five types of methods are distinguished for classifying fake news . [17] Two of these techniques are manual. Fact-checking is a journalistic practice in which experts on the topic of a given news item analyze its veracity, gathering information about the news. It is an effective verification practice, but inefficient in terms of time and human resources. Due to its slow verification speed, it does not prevent fake news from spreading and reaching a certain level of manipulation in the collective unconscious. The other manual technique is crowdsourcing, a task similar to the previous one, but carried out by a collaborative team, not necessarily experts in the subject. This technique can improve response times but decrease the quality of verification.

Two other automated techniques are more efficient by being able to deliver results almost in real time, but less effective. The first is the use of natural language processing (NLP) techniques, based on the application of linguistic features and other computational linguistic techniques to recognize the lexical, syntactical, grammatical, and semantic characteristics that distinguish real news from fake. The second refers to a set of machine learning (ML) and deep learning (DL) techniques, which involve the use of artificial intelligence and advanced statistics to teach an algorithm to identify patterns in said data using a large training dataset and thus be able to classify new news items with a certain degree of accuracy, precision, and confidence. There are many ML and DL methods currently available [9] , [18] . However, its main problems are, on the one hand, the dependence on contexts (topic, language, time period, historical circumstances, etc.) when training these news items, that is, they only learn to classify the specific types of news with which they were originally trained, and on the other hand, the low interpretability of its results, that is, the limited capacity to understand why a news item that was labeled as fake is actually so, and what distinguishes it from real news. Thus, the latter method is actually a set of hybrid methods; a combination of the previous ones. For example, the use of linguistic features specific to NLP can be used to help make ML models more interpretable [19] .

This work aims to automatically classify unintentional fake news. To do so, we address the problem of automated detection and characterization of unintentional fake news using a hybrid approach based on ML models and NLP linguistic features. The work also focuses on news stories originally written in Spanish, a

language widely used on the internet, being the third most used language, although considerably less explored than English. Its analysis presents complexities given its unique linguistic characteristics, such as a greater number of inflectional and derivational alternatives at the morphological level and greater flexibility at the syntactical level [20], which implies an in-depth investigation of its lexical and grammatical features in written texts. This avoids the loss of substantive information when translating only from English, a very common solution in this field of research. As a case study, the phenomenon of celebrity death news will be analyzed, contrasting unintentional fake news with real news that refutes them, as well as with news of actual deaths. The expected results are to identify lexical, grammatical, and discursive features that allow for the automatic differentiation of real news from unintentional fake news.

The framework is then presented, using an interdisciplinary approach that encompasses journalistic, sociological, computer science, and technological aspects of detecting fake news, disinformation, and misinformation. The methodology of the work is then detailed, presenting the experimental design, and then the experimental setup based on a real data set. The results are then presented and discussed, along with the main conclusions and future work.

Related Work

It is now widely accepted that the world is dominated by an "infocracy," which refers to an information regime controlled by algorithms that analyze human user data . [21] In this context, a veritable "information war" occurs, as the different voices interacting in the virtual public space seek to dominate one another. These

messages are emitted by various actors from their positions in different social fields [and] , therefore, reflect the power relations they exercise. In this context, the strategies used by the actors comprising these fields of dominance (political, economic, cultural) are diverse, using the media as a channel and, more recently, the digital platforms that have emerged in the last 30 years.

For years, the relationship between digital communication and fake news has been inseparable [17] . Various researchers have tried to understand this articulation, frequently referring to the present public sphere as "post-truth" [23] , [18] , [24] a phenomenon for which updates and new research programs have been proposed [25] . Thinking about the deaths of famous people as fake news is relevant in the sense of asking why these events are possible and in what contexts news of this type is shared. Sociology introduces the concept of the attention market [26] for this type of process , which seeks to help understand the generation and consumption of certain types of content over others. There is clear consensus that during the pandemic, the consumption of this type of content increased, fostered by exacerbated emotions, communication disorders due to echo chambers [27] , filter bubbles [28] , and even intentionality. The approach of this work aligns with that of others, which allows us to move from concern about the threat of the type of content to an understanding of why at certain times certain content is shared and not others [29] .

From a computational and technological perspective, increased computing and data collection capabilities, coupled with recent advances in artificial intelligence, have enabled progress toward partial solutions in specific contexts for the detection and classification of various types of information disorders. However, it is these same

technologies that have enabled greater sophistication and effectiveness in the creation of fake news and malicious information [30] . In the particular case of misinformation, machine learning methods have been applied in multiple contexts, such as social [31] or health crises [32] . One of the main current challenges is to move towards hybrid methods that use explainable artificial intelligence techniques [3] . In particular, the combination of machine learning and natural language processing seem an appropriate path to pursue greater interpretability of analysis results [19] .

Methodology

This work will employ a hybrid methodology, utilizing quantitative classification and analysis techniques from machine learning and computational linguistics, as well as qualitative discourse analysis techniques. This approach will allow for not only a descriptive but also an explanatory one, supported by greater interpretability of the analysis results. Specifically, rather than simply classifying news as real or fake—a complex problem highly dependent on context and data quality—the goal is also to understand the potential linguistic and discursive differences between real news and unintentional fake news.

This paper aims to classify news stories about celebrity deaths, which are rife with unintentional fake news. The general procedure is described in Figure 2 and exemplified in Figure 3 , both of which are detailed below.

Technology Journal of Management , Accounting and Economics (TECH)
ISSN: 2311-3995

Vol. 13 No. 2 (2025)

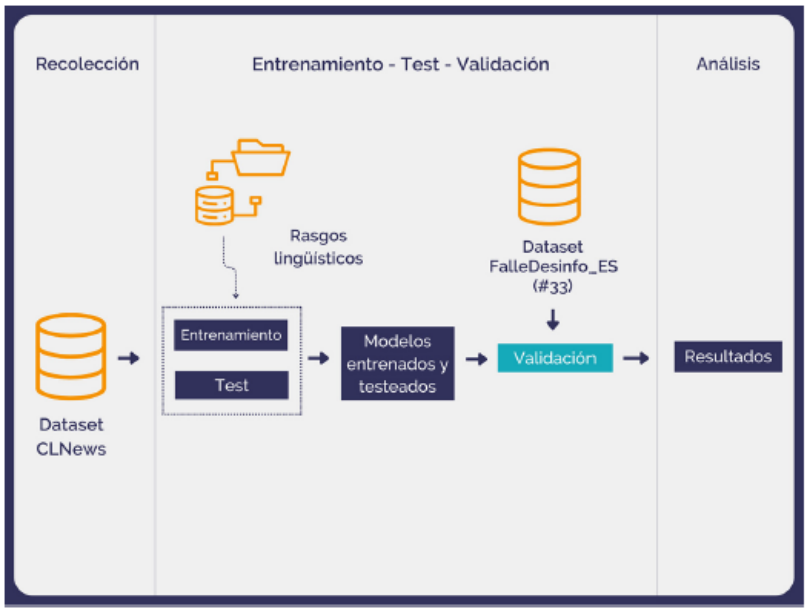


Figure 2 Data manipulation process.

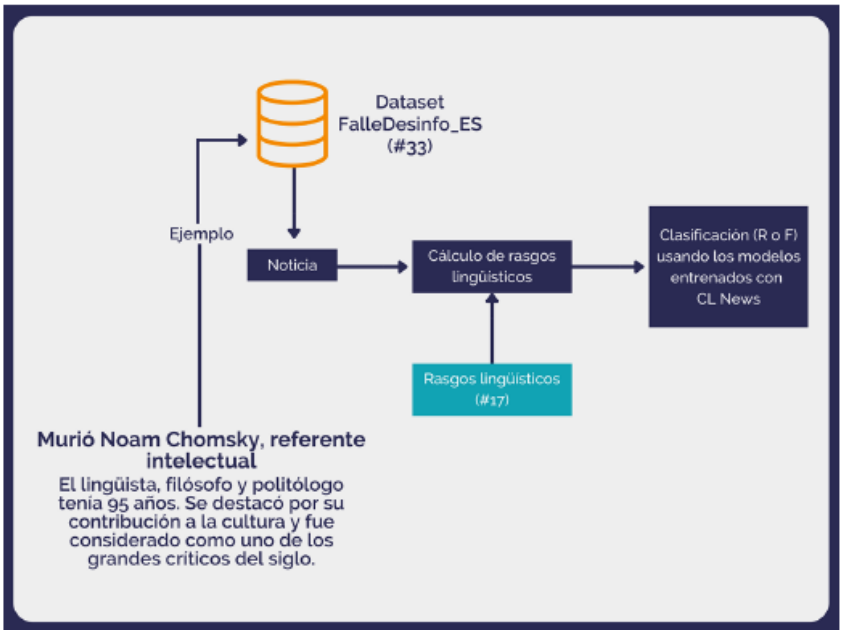


Figure 3 Example of data manipulation process.

Technology Journal of Management , Accounting and Economics (TECH)
ISSN: 2311-3995

Vol. 13 No. 2 (2025)

Data collection

Supervised machine learning techniques require training, testing, and validation stages. For the training and testing stages, the existing CLNews database, used for the classification of false rumors in Spanish [3], was used . CLNews contains 300 content propagation trees on Twitter (currently X), related to trending topics of social, political, and sporting events in Chile. These trees,

The data collected between June 2019 and January 2020 were manually labeled, resulting in 87 true rumors, 53 false rumors, 49 unverifiable rumors, and 111 non-rumors. Note that after these training and testing stages, a model was obtained that was ready to be validated with new news.

For the validation stage, the dataset was manually constructed from a series of celebrity death news stories written exclusively in Spanish, without translations, and published in Spanish-language media. Three types of news stories from online media outlets were considered, including the headline, subheading, and body of the news:

- T1: Unintentional fake news.
- T2: Real news, which refutes the misinformation in T1.
- T3: Real news.

Group T2 is proposed as a methodological innovation, as it deals with real news, but with different characteristics than those of a real news story about a person's death. The newsworthy event, in these cases, is not the death of a celebrity, but rather the fact that their death has been denied.

Each group contained the same number of news stories. For the same celebrity, the news stories in T1 and T2 must have the same publication date, to prevent news from later days from expanding the content by adding additional context and generating additional noise to the comparisons. It should be noted that the validation dataset does not need to be large, as this does not represent a theoretical or practical limitation for the described purpose. The validation dataset used, called FalleDesinfo_ES, is detailed later in the experimental setup section.

Data analysis

For news classification, five classic supervised machine learning models were used, based on linguistic features [19] for the Spanish language: support vector machine (SVM), linear kernel and radial basis function (RBF); random forest (RF), XGBoost, and logistic regression (LR) [33] . As mentioned above, these models were trained and tested with the CLNews database, and validated with the new FalleDesinfo_ES dataset.

Linguistic features are algorithmically implemented metrics that allow the quantification of specific characteristics of a text, such as: statistics related to its words (either all of them or specific types, such as adjectives, adverbs, nouns, etc.), sentences or paragraphs; counts of terms belonging to specific dictionaries, among others. This work considers various existing metrics, based on previous work [19] and implemented specifically for the Spanish language, identified through the PACTE [45] platform for quantitative analysis of Spanish texts .

Technology Journal of Management , Accounting and Economics (TECH)
ISSN: 2311-3995

Vol. 13 No. 2 (2025)

To compare the news items, the classifiers and linguistic features were applied to all news items of each type (T1, T2, and T3). The classification results were then compared within each type, between types, and between classifiers.

The classifiers allow each news item in the dataset to be labeled as "real" or "fake." The performance of these classifiers is evaluated using the traditional metrics of accuracy, precision, recall, and F1-score, indicated in equations (1)–(4):

$$accuracy = (TP + TN) / (TP + FP + TN + FN) \quad (1)$$

$()^2$

$()^3$

$$F1 = 2\,precision / (precision + recall) \quad (4)$$

where TP corresponds to true positives, TN to true negatives, FP to false positives, and FN to false negatives.

These metrics provide insight into the degree of confidence with which the classifier performed its labeling. It should be noted that the F1 score is conceived as the harmonic mean between precision, the value that considers correctly identified examples (TP) and those identified as corresponding to another class (FP); and recall, the value that considers correctly identified examples (TP) and those that were not identified for the respective class. This allows both measures to be combined into a single value, which is useful for comparing the combined performance of precision and recall across multiple classification results.

For linguistic features, these yield numerical values. As part of classifier analysis, identifying the most relevant linguistic features helps us identify the most

significant features for the model's predictions. For linear SVM and logistic regression, these values are based on the model's coefficients, so higher values indicate greater relevance. In a Random Forest, the amount of information each feature contributes by reducing uncertainty in the trees is measured. XGBoost evaluates importance based on how frequently each feature is used to split the data in the model's trees. For SVM with an RBF kernel, the importance permutation is used, which measures how much the model's accuracy drops when the values of a feature are altered. These metrics allow us to identify key features that influence the model's results.

Due to the nature of the news, significant differences are expected between the news stories of groups T1 and T2. However, the differences between groups T1 and T3 may be less evident.

Experimental Setup

Next, the experimental design described in the previous section will be presented. The dataset, the performance of the previously tested automatic classifiers, and a description of the linguistic features that were significant for this work and that have implications for the following Results section are detailed.

Dataset

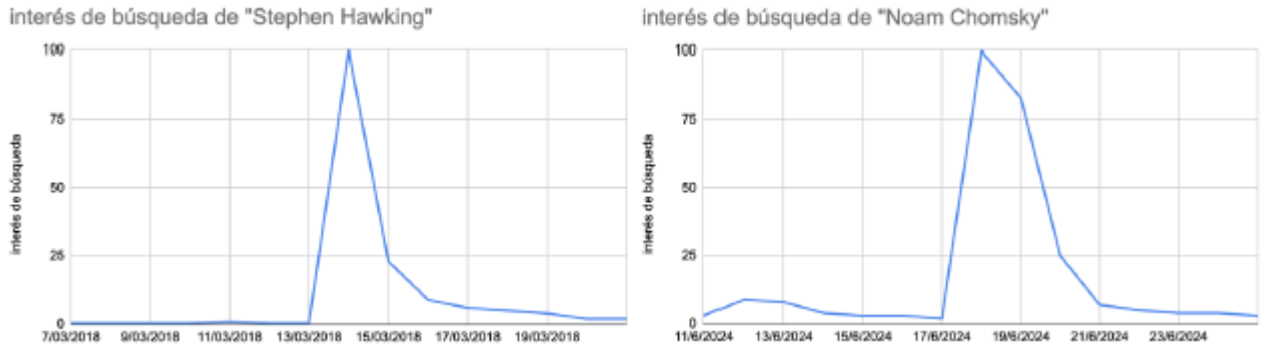
For the classifier validation dataset, 11 news items of each type were manually collected, totaling 33 news items in total. This dataset, called FalleDesinfo_ES, has been published for free use [34]. Group T1 consists of news items related to the alleged death of linguist and intellectual Noam Chomsky, disseminated on June 18, 2024, which amplified a fake news item initially spread by social media platforms

Technology Journal of Management , Accounting and Economics (TECH)
ISSN: 2311-3995

Vol. 13 No. 2 (2025)

(Twitter/X and Facebook). Group T2 contains news items published on the same day that refute the previous report. Moreover, some media outlets in T1 and T2 are repeated. Finally, group T3 contains real news items about the death of physicist Stephen Hawking on March 14, 2018.

Figure 4 shows the normalized global internet search interest generated by Stephen Hawking and Noam Chomsky, according to Google Trends, between one week before and one week after the news of their deaths. The less abrupt decrease in Chomsky's case is due to news reports in the following days denying his death.



Fuente: Google Trends.

Figure 4 Search interest for Stephen Hawking (left) and Noam Chomsky (right) on the Internet, including the date of the news of their deaths.

Figure 5 shows the geolocation of the same searches over the same period. Hawking's normalized searches are led by Honduras, Paraguay, the United Kingdom, Australia, and Panama.

Technology Journal of Management , Accounting and Economics (TECH)
ISSN: 2311-3995

Vol. 13 No. 2 (2025)

Fuente: Google Trends.

Figure 5 Geolocation of searches for Stephen Hawking (left) and Noam Chomsky (right) on the Internet during the same time period as above.

For Chomsky, the top ten countries are all Latin American, with Honduras, Nicaragua, El Salvador, Paraguay, and Uruguay leading the list.

It is worth mentioning that all the countries on the list have faced or been exposed to disinformation campaigns, especially in the context of elections and social crises, and in some cases, they have even been carried out (and financed) by government agencies [35] , [36] . This has been done with the aim of influencing public opinion, gaining greater citizen support, and/or damaging the reputation of political adversaries. It is in this digital ecosystem that the cases of Honduras [37] and Nicaragua [38] stand out , where disinformation strategies are systematized and institutionalized, and the techniques of cybertroop [36] and farm troll [39] are mainly used . Thus, in this situation, it could be conjectured that the constant exposure of Hondurans to fake news has generated a certain skepticism that leads them to question the credibility of certain sources (mainly social media) and to verify

information in other media, using the Google search engine. More focused studies would be necessary to better understand this ranking.

The news sources collected in the FalleDesinfo_ES dataset for each group were as follows:

- T1: Continental, Debate, e-consultation, El Cronista, Página12, La Tercera (2), Línea Directa, LT10 (original source: Cadena 3), Radio UChile, RDN (original source: ABC).
- T2: API, Cadena 3, Debate, El Comercio de Perú, The Republic Newspaper, The Journalist, La Tercera, MSN, Portfolio, T13, Wired.
- T3: As, Clarín, CNN Español, DW, El Mundo, Euro News, El Heraldo, Infobae, La Tercera, Público, T13.

Some T1 news had to be extracted from the history of websites stored on the Internet Archive (archive.org), while it is noteworthy that others remain available online at the time of writing.

Automatic classifiers

As mentioned above, the five classifiers used were pre-trained using the CLNews database of false rumors in Spanish [40] . 90% of the data was used for training and 10% for testing, with the latter achieving the performances illustrated in Table 1 . Among all of them, the linear SVM model stands out, with RF and XGBoost showing considerably lower performances than the others. This information is relevant when interpreting the results obtained in the following section for the new validation data.

Table 1 Performance of previously tested classifiers.

Modelo	Accuracy	Precisión	Recall	F1-Score
SVM (lineal)	0,857	0,883	0,857	0,846
SVM (RBF)	0,714	0,802	0,714	0,645
RF	0,714	0,802	0,714	0,645
XGBoost	0,643	0,629	0,643	0,632
LR	0,643	0,629	0,643	0,632

Linguistic features

The linguistic features considered can be grouped into five categories: surface variables, grammatical category variables, lexical variables with a discursive function, polarity lexicon, and emotions and feelings lexicon.

Surface variables consider the linguistic units that can be recognized on the surface of the text, such as sentences, paragraphs, words, and characters. Each of these units can be measured in relation to the total, maximum, minimum, and median number. For the first category, five features are considered: total number of sentences, minimum and maximum number of words, maximum number of characters, and maximum number of non-space characters.

Grammatical categories refer to the classification of words in a text. These include verbs, nouns, adjectives, adverbs, and others. Like the previous variable, they can be measured in relation to the total, maximum, minimum, mean, and median number. Added to these are the standard deviation, the idf (inverse document frequency), and the tf-idf (term frequency-inverse document frequency). For the second category, eight features are considered: maximum, minimum, mean, and median number of prepositions, maximum number of auxiliaries and tf-idf, standard deviation of numerals, and idf of proper nouns.

Lexical variables with a discourse function focus on words that serve a discursive function within the text, such as markers or modalizers. They are measured through total quantification. For this third category, two features are analyzed: counterargumentative textual discourse markers and generalizing opinion connectors.

The polarity lexicon refers to words associated with an emotional polarity. They will be negative polarity when associated with a negative emotion, and positive polarity when the emotion is positive. Polarity classification is carried out using dictionaries that offer varying degrees. For the trait in this category, the dictionary created by Finn Ârup Nielsen (Affin) is used. The fourth category highlights the very positive trait Affin.

Finally, the emotions and feelings lexicon consists of words associated with emotions and feelings in a text. Like the previous category, words are classified using dictionaries. For the trait analyzed regarding the emotion of anger, the NRC dictionary is used. Developed by Saif Mohammad and Peter Turney, it establishes eight basic emotions: anger, fear, anticipation, trust, surprise, sadness, joy, and disgust, to classify the words in the text.

Results and Discussion

Linguistic characteristics of news of deaths

The linguistic features relevant to each classification model are illustrated in Table 2. In the model columns, a 1 means that the metric is among the ten most relevant for that model, and a 0 means that it is not. The "Total" column represents the number of models for which the linguistic feature was relevant, considering the

Technology Journal of Management , Accounting and Economics (TECH)
ISSN: 2311-3995

Vol. 13 No. 2 (2025)

entire dataset. The features are listed in order of importance according to this value. The last three columns then show the number of models for which the linguistic feature is relevant, segmented by each news type: T1, T2, and T3.

Table 2 Relevant linguistic features for each classification model.

ID	Rasgo lingüístico	Modelo					Total	T1	T2	T3
		SVM (lineal)	SVM (RBF)	RF	XG Boost	LR				
R1	**promedio de adposiciones por oración**	**1**	**1**	**1**	**1**	**1**	**5**	**3**	**4**	**1**
R2	máximo núm. de caracteres por oración	1	1	1	1	1	5	1	3	1
R3	máximo núm. de caracteres sin espacios por oración	1	1	1	1	1	5	3	2	2
R4	idf de sustantivos propios en toda la noticia	1	1	1	1	1	5	2	4	4
R5	mediana de adposiciones por oración	1	1	1	0	1	4	3	2	4
R6	tf-idf de auxiliares en toda la noticia	1	0	1	1	1	4	4	1	5
R7	máximo núm. de palabras por oración	1	0	1	1	1	4	1	3	1
R8	frecuencia de conectores de opinión generalizadores	1	1	0	0	1	3	3	2	5
R9	mínimo núm. de palabras por oración	0	1	1	1	0	3	3	5	4
R10	máximo núm. de adposiciones por oración	0	1	0	1	0	2	3	2	0
R11	mínimo núm. de adposiciones por oración	0	1	0	0	1	2	2	3	0
R12	medida de sentimiento enojo según léxico NRC	0	0	1	1	0	2	3	1	2
R13	frecuencia de marcadores contraargumentativos	1	0	0	0	1	2	4	2	4
R14	núm. total de oraciones en toda la noticia	0	0	1	1	0	2	4	4	4
R15	máximo num. de auxiliares por oración	0	1	0	0	0	1	4	5	3
R16	desviación estándar de núm. numerales por oración	1	0	0	0	0	1	4	4	5
R17	medida de sentimiento muy positivo según léxico Affin	0	0	0	0	0	0	3	3	5

As observed, the four most relevant linguistic features for news classification, considering the complete validation dataset, are related to sentence length in terms of characters (R2-R3), the use of adpositions (R1), and proper nouns (R4). Further features are related to sentence length (R7, R9) and text length (R14), as well as the use of adpositions (R5, R10-R11). Other relevant features are added regarding the frequency of auxiliaries (R6), generalizing opinion connectors (R8), and

counterargumentative markers (R13). For two of the models, vocabulary that reflects anger (R12) is also relevant.

If the news types are analyzed separately, feature R17, which lexicons of very positive sentiments, emerges as relevant for each of them, and the relevance overlaps also increase for features R15 and R16, which consider auxiliaries and numerals, respectively. For the real news in T3, a lower relevance of adpositions stands out, and a greater relevance in the frequency of auxiliaries, generalizing opinion connectors, numerals, and very positive sentiments.

This analysis allows us to identify linguistic differences between news items in T1 and T2, but especially between them and T3. Therefore, as will be seen below, it allows us to find differences between T1 and T3 that are unclear when considering only automatic classification models.

Automatic classification of death news

The classifier results are presented in Table 3. Green indicates a news item that is correctly classified according to its type, and red indicates an incorrect one. The "Unification" columns refer to the joint classification results of all classifiers. The "Correctness" column indicates "Yes" if the majority of classifiers correctly labeled the news item according to its type. This result is often useful for decision-making, since greater coincidence in the model results translates into greater confidence for analysts.

Table 3 News classification results. A red color means the label is incorrect, and a green color means the label is correct.

Technology Journal of Management , Accounting and Economics (TECH)
ISSN: 2311-3995

Vol. 13 No. 2 (2025)

		Clasificadores					Unificación		
Grupo	Noticia	SVM (lineal)	SVM (RBF)	RF	XGBoost	LR	Real	Falsa	Correctitud
T1	T1-1	Falsa	Real	Real	Real	Falsa	3	2	No
	T1-2	Falsa	Real	Real	Real	Falsa	3	2	No
	T1-3	Falsa	Real	Real	Real	Falsa	3	2	No
	T1-4	Falsa	Real	Real	Real	Falsa	3	2	No
	T1-5	Falsa	Real	Falsa	Real	Falsa	2	3	Sí
	T1-6	Falsa	Real	Real	Real	Falsa	3	2	No
	T1-7	Falsa	Falsa	Real	Real	Falsa	2	3	Sí
	T1-8	Falsa	Real	Real	Real	Falsa	3	2	No
	T1-9	Falsa	Real	Falsa	Falsa	Falsa	1	4	Sí
	T1-10	Real	Real	Real	Real	Real	5	0	No
	T1-11	Falsa	Real	Real	Real	Falsa	3	2	No
"Falsas" T1		10	1	2	1	10			27%
"Reales" T1		1	10	9	10	1			
Correctas T1		91%	9%	18%	9%	91%			
T2	T2-1	Falsa	Falsa	Real	Real	Falsa	2	3	No
	T2-2	Falsa	Real	Real	Real	Falsa	3	2	Sí
	T2-3	Falsa	Falsa	Real	Real	Real	3	2	Sí
	T2-4	Real	Real	Real	Real	Real	5	0	Sí
	T2-5	Falsa	Real	Real	Real	Falsa	3	2	Sí
	T2-6	Real	Real	Falsa	Falsa	Real	3	2	Sí
	T2-7	Real	Real	Real	Real	Real	5	0	Sí
	T2-8	Real	Real	Falsa	Real	Real	4	1	Sí
	T2-9	Falsa	Real	Falsa	Real	Falsa	2	3	No
	T2-10	Falsa	Real	Real	Real	Falsa	3	2	Sí
	T2-11	Real	Real	Real	Real	Real	5	0	Sí
"Falsas" T2		6	2	3	1	5			82%
"Reales" T2		5	9	8	10	6			
Correctas T2		45%	82%	73%	91%	55%			
T3	T3-1	Real	Real	Falsa	Real	Real	4	1	Sí
	T3-2	Falsa	Real	Real	Real	Falsa	3	2	Sí
	T3-3	Real	Real	Falsa	Falsa	Real	3	2	Sí
	T3-4	Falsa	Real	Falsa	Real	Falsa	2	3	No
	T3-5	Falsa	Real	Falsa	Real	Falsa	2	3	No
	T3-6	Real	Real	Falsa	Falsa	Real	3	2	Sí
	T3-7	Falsa	Real	Falsa	Real	Falsa	2	3	No
	T3-8	Real	Real	Falsa	Real	Real	4	1	Sí
	T3-9	Falsa	Real	Falsa	Real	Falsa	2	3	No
	T3-10	Falsa	Real	Falsa	Real	Falsa	2	3	No
	T3-11	Falsa	Real	Falsa	Real	Falsa	2	3	No
"Falsas" T3		7	0	10	2	7			45%
"Reales" T3		4	11	1	9	4			
Correctas T3		36%	100%	9%	82%	36%			
"Falsas" Total		23	3	15	4	22			
"Reales" Total		10	30	18	29	11			

Technology Journal of Management , Accounting and Economics (TECH)
ISSN: 2311-3995

Vol. 13 No. 2 (2025)

As expected, the results vary considerably depending on the news group. For group T1 (unintentional fake news about Chomsky), the linear SVM and LR models are the best (91% classification success), while RF only got three labels right (18% success) and SVM (RBF) and XGBoost only got one label right (9% success). Since the latter three are the majority, the combined results only achieve a 27% joint success rate. News item T1-10 is noteworthy, incorrectly classified by all models as real. In terms of linguistic features, it is a short news item, devoid of emotional vocabulary, generally made up of statements and referencing other sources; otherwise, no counterargumentative textual discursive markers are observed, all of which justifies the difficulty in classifying it as fake.

For group T2 (real rectification news), the results are reversed. In this case, the XGBoost, SVM (RBF), and RF models, in that order, are the most successful (91%, 82%, and 73% success, respectively), with LR and linear SVM relegated to a low 55% and 45%, respectively. For the same reason as above, in this case the successful models are the majority, so a combined success rate of 82% is obtained.

Finally, for group T3 (real news about Hawking), a phenomenon distinct from the previous two occurs. Here, SVM (RBF) and XGBoost perform very well (82% and 100% success, respectively), just as in the previous case, but RF is very poorly positioned, with only 9% success. Therefore, with only two successful classifiers, the shared success rate is low, i.e., only 45%.

In Abstract, we can conclude that none of the models is successful for all news types. Of the five, the SVM and XGBoost models appear to be the best, although they are useful for classifying different types of news. Among the three, it should

also be noted that XGBoost does not have as good performance metrics as the SVM models (Table 1).

On the other hand, a marked difference is observed in the correctness of the models for type T1 and type T3 news stories, which indicates that, in the opinion of the automatic classifiers, unintentionally fake news stories are perceived differently than real ones, even though they all talk about the death of a celebrity and are worded in a similar way. This difficulty in distinguishing between reality and falsehood is (only partially) mitigated by the analysis of linguistic features, such as the one performed in the previous section.

Interpretation of the Results

The results highlight considerable variability in the accuracy of automatic classifiers depending on the type of news, reflecting the complexity of distinguishing between true and false information in a context of rapid distribution and consumption. This highlights the vulnerability of the digital public sphere to disinformation, underscoring a crisis of trust in the media and traditional institutions that has national [41] and global [42] scope . The difficulty in correctly classifying unintentional fake news shows how the rapid spread of unverified information, driven by competition in the journalistic field and the attention economy, exacerbates the problem of disinformation. Furthermore, in the context of widespread use and dissemination through social media platforms, it is difficult to trace the initial source of information, as well as to identify whether it is denied or corroborated. Sociologically, it prompts us to question its impact on public

discussion and the ways in which we communicatively relate. In this sense, it also urges us to problematize the power of the media over the constitution of truth.

Likewise, the dependence of machine learning and natural language processing models on the context in which they were trained limits their ability to generalize to new contexts or topics, indicating that technology alone cannot solve the problem of misinformation. Note how the lack of a "how" question limits the possibilities of understanding the problem. This suggests that a deep understanding of the social and cultural context in which information is produced and consumed is required. That is, going from the phenomenon to the field in which it is constituted.

Regarding journalistic practice, during the search for news, it became evident that many media outlets gather news from multiple sources and do not remove fake news once it has been debunked. Furthermore, some outlets maintain fake news they themselves debunk, even when it was written by the same journalist. This exacerbates the difficulties in distinguishing between reality and falsehood, and highlights the "post-truth" phenomenon discussed at the beginning of this article. In fact, the question remains as to what happens when a fake news story has been massively debunked, but a news outlet, for one reason or another, still decides to maintain the erroneous information that helped amplify the falsehood. Given this lack of correction, this fake news, initially unintentional, transforms into intentional, going from misinformation to disinformation. This is another example that highlights the enormous difficulties in automatically distinguishing, even using advanced artificial intelligence techniques, between reality and falsehood.

In this regard, it is worth highlighting the controversy surrounding the fact that such false information, whether intentional or unintentional, is not removed from digital media platforms. Publishers argue their right to freedom of expression over the so-called "right to be forgotten" [43] associated with news that could damage people's image, as is the case of the false death of a public figure, which was reviewed in this investigation. However, the issue of the right to be forgotten digitally remains controversial, since although it is being rigorously addressed in Europe, in countries such as the United States, "search engines do not bear any type of responsibility for the information that appears in their search results" [43] . For its part, in the case of Chile, where the legal system is still not emphatic or categorical on the subject, a bill is being processed that could allow the elimination of information "that is considered obsolete due to the passage of time, or that in some way affects the development of any of the fundamental rights" [44] .

Conclusions

This paper addressed the problem of detecting unintentional fake news, a type of news that has been little analyzed in the literature. To this end, at least three methodological innovations were proposed. First, the use of a hybrid classification method, which combines supervised machine learning with natural language processing, using linguistic features specific to the Spanish language. Second, the use of a rigorous validation dataset, which does not consider real and fake (intentional) news, as is usual, but rather real news, unintentional fake news, and real news that refute the latter; something closer to the reality of traditional online

media. Finally, an interdisciplinary approach from computer science, linguistics, sociology, and journalism is used for the analysis and discussion of the results.

The results of the analyzed case study, on news about celebrity deaths, demonstrate the enormous difficulty in classifying unintentionally fake news. In linguistic terms, it is observed that some linguistic features, such as those related to the use of adpositions (prepositions), sentence and text length, the use of proper nouns, counterargumentative markers, positive and negative emotions, among others, are relevant for this purpose. Regarding the classifiers, while some were able to correctly classify certain types of news, none were able to correctly classify all three types of news considered.

A major limitation in the fake news classification problem, which also applies to this work, is the problem of multicontextuality. That is, the difficulty in ensuring that models are robust and maintain their performance, demonstrated for both training and test data, in new validation data. Indeed, contextual variation can negatively impact the accuracy of automatic classification models.

Despite the identified difficulties, this work opens a new perspective in the analysis of information disorders, distinguishing between what constitutes and represents intentional and unintentional fake news, the former being more prevalent on social media platforms, and the latter in mass media and traditional online news outlets. To continue experiments along these lines, the generation of more data for validation is left as future work.

References

[1] Unesco, Journalism, fake news & disinfor mation: handbook for journalism education and training, C. Ireton, and J. Posetti, Eds. París, Francia: Unesco, 2018.

[2] K. Shu, S. Wang, D. Lee, and H. Liu, "Mining Disinformation and Fake News: Concepts, Methods, and Recent Advancements Disinformation," in Misinformation, and Fake News in Social Media. Emerging Research Challenges and Opportunities, K. Shu, S. Wang, D. Lee, and H. Liu, Eds. Cham, Switzerland: Springer International Publishing, 2020, pp. 1-19, doi: 10.1007/978-3-030-42699-6.

[3] E. Puraivan, E. Godoy, F. Riquelme, and R. Salas, "Fake news detection on Twitter using a data mining framework based on explainable machine learning techniques" in 11th International Conference of Pattern Recognition Systems (ICPRS 2021), Talca, Chile, Mar. 17-19, 2021, pp. 157-162, doi: 10.1049/icp.2021.1450.

[4] I. Stavre and M. Puntí, "Fake news, something new?," Sociology and Anthropology, vol. 7, no. 5, pp. 212-219, 2019, doi: 10.13189/sa.2019.070504.

[5] S. van der Linden, C. Panagopoulos, and J. Roozenbeek, "You are fake news: political bias in perceptions of fake news," Media, Culture & Society, vol. 42, no. 3, pp. 460-470, Apr. 2020, doi: 10.1177/0163443720906992.

[6] N. Poulose, "Fake news in health and medicine," Data Science for Fake News, vol. 42, 2021, pp. 193-204, doi: 10.1007/978-3-030-62696-9_9.

[7] F. Olan, U. Jayawickrama, E.O. Arakpogun, J. Suklan, and S. Liu, "Fake news on social media: the impact on society," Information Systems Frontiers, vol. 26, no. 2, pp. 443-458, Jan. 2022, doi: 10.1007/s10796-022-10242-z.

[8] E. Papadogiannakis, P. Papadopoulos, E.P. Markatos, and N. Kourtellis, "Who funds Misinformation? A systematic analysis of the ad-related profit routines of fake news sites," Proceedings of the ACM Web Conference, 2023, pp. 2765-2776, doi: 10.1145/3543507.3583443.

[9] J.A. Braun and J.L. Eklund, "Fake news, real money: Ad tech platforms, profit-driven hoaxes, and the business of journalism," Digital Journal, vol. 7, no. 1, pp. 1-21, Jan. 2019, doi: 10.1080/21670811.2018.1556314.

[10] J. Alghamdi, S. Luo, and Y. Lin, "A comprehensive survey on machine learning approaches for fake news detection," Multimedia Tools and Applications, vol. 83, no. 17, pp. 51009-51067, Jan. 2024, doi: 10.1007/s11042-023-17470-8.

[11] J. Rose, "To believe or not to believe: an epistemic exploration of fake news, truth, and the limits of knowing," Postdigital Science and Education, vol. 2, no. 1, pp. 202-216, Jan. 2020, doi: 10.1007/s42438-019-00068-5.

[12] T. Lelo and R. Fígaro, A materialist approach to fake news, in Politics of Disinformation: The Influence of Fake News on the Public Sphere, G. López-García, D. Palau, M.B. Torres, E. Campos, and P. Masip, Eds. New York, USA: Wiley-Blackwell, 2021, pp. 23-34, doi: 10.1002/9781119743347.ch2.

[13] J. Gray, L. Bounegru, and T. Venturini, "'Fake news' as infrastructural uncanny," New Media & Society, vol. 22, no. 2, pp. 317-341, Jan. 2020, doi: 10.1177/1461444819856912.

[14] P. Bourdieu, On Television, Barcelona, Spain: Anagrama, 2006.

[15] P. Bourdieu, Course in General Sociology 1. Fundamental Concepts, 1st ed. Buenos Aires, Argentina: Siglo XXI Editores, 2020.

[16] D.L. Swartz, "Spotlight," in Journalism and Truth in an Age of Social Media, 1st. ed., J. Katz and K. Mays, Eds. New York, USA: Oxford University Press, 2019, pp. 36-38, doi: 10.1093/oso/9780190900250.003.0003.

[17] E.C. Tandoc, "The facts of fake news: A research review," Sociology Compass, vol. 13, no. 9, Art. no. e12724, pp. 1-19, Jul. 2019, doi: 10.1111/soc4.12724.

[18] J. Compton, S. van der Linden, J. Cook, and M. Basol, "Inoculation theory in the post-truth era: Extant findings and new frontiers for contested science, misinformation, and conspiracy theories," Social and Personality Psychology Compass, vol. 15, no. 6, Art. no. e12602, pp. 1-16, May 2021, doi: 10.1111/spc3.12602.

[19] E. Puraivan, R. Venegas, and F. Riquelme, "An empiric validation of linguistic features in machine learning models for fake news detection," Data & Knowledge Engineering, vol. 147, Art. no. 102207, Sep. 2023, doi: 10.1016/j.datak.2023.102207.

[20] J. Valenzuela, "Contrastive Linguistics English-Spanish: A General View," Carabela, no. 51. pp. 27-45, 2002.

[21] B. Ch. Han, Infocracy. Digitalization and the Crisis of Democracy, Bogotá, Colombia: Taurus, 2021.

[22] P. Bourdieu, "Social Space and Symbolic power," Sociology Theory, vol. 7, no. 1, pp. 14-25, 1989.

[23] S. Sismondo, "Post-truth?," Social Studies Science, vol. 47, no. 1, pp. 3-6, Feb. 2017, doi: 10.1177/0306312717692076.

[24] B. Çanakpinar, M. Kalelioglu, and V. D. Günay, "Semiotics and political discourse in the post-truth era," Language and Semiotic Studies, vol. 10, no. 1, pp. 65-82, Jan. 2024, doi: 10.1515/lass-2023-0040.

[25] J. Angermuller, "Truth after post-truth: for a Strong Programme in Discourse Studies," Palgrave Communications, vol. 4, no. 1, Art. no. 30, Mar. 2018, doi: 10.1057/s41599-018-0080-1.

[26] H. Rao and H.R. Greve, "The Plot Thickens: A Sociology of Conspiracy Theories," Annual Review of Sociology, vol. 50, no. 1, Feb. 2024, doi: 10.1146/annurev-soc-030222-031142.

[27] D. Rivera López, "Truth and Catastrophe in the Algorithmic Present," Latin American Journal of Humanities and Educational Divergences, vol. 3, no. 1, pp. 100-114, June 2024, doi: 10.5281/zenodo.12629874.

[28] J. Buschman, "Fake news as systematically distorted communication: an LIS intervention," Journal of Documentation, vol. 80, no. 1, pp. 203-217, Jul. 2024, doi: 10.1108/JD-03-2023-0043.

[29] O.L.T. Mai et. al, "Factors Affecting Students' Fake News Identification during Covid-19 in Vietnam: Access from sociological study and application of PLS-SEM model," Wseas Transactions on Business and Economics, vol. 20, Art. no. 126, pp. 1422-1438, Jun. 2023, doi: 10.37394/23207.2023.20.126

[30] E. Aímeur, S. Amri, and G. Brassard, "Fake news, disinformation and misinformation in social media: a review," Social Network Analisys and Mining, vol. 13, no. 1, pp. 13-30, Feb. 2023, doi: 10.1007/s13278-023-01028-5.

[31] K. Hunt, P. Agarwal, and J. Zhuang, 'Monitoring Misinformation on Twitter During Crisis Events: A Machine Learning Approach," Risk Analysis, vol. 42, no. 8, pp. 1728-1748, Nov. 2020, doi: 10.1111/risa.13634.

[32] Y. Zhao, J. Da, and J. Yan, "Detecting health misinformation in online health communities: Incorporating behavioral features into machine learning based approaches," Information Processing & Management, vol. 58, no. 1, pp. 1-24, Jan. 2021, doi: 10.1016/j.ipm.2020.102390.

[33] G. James, D. Witten, T. Hastie, and R. Tibshirani, An introduction to statistical learning, New York: Springer, 2021, doi: 10.1007/978-1-0716-1418-1.

[34] F. Riquelme et al., "FalleDesinfo_ES: Dataset of real and fake Spanish-language news about celebrity deaths," Zenodo, 2024, doi: 10.5281/ZENODO.13117808 .

[35] S. Bradshaw et al., "Country Case Studies Industrialized Disinformation: 2020 Global Inventory of Organized Social Media Manipulation," demtech.oii.ox.ac.uk. Accessed: Sep. 2024.

[36] S. Bradshaw et al., "Industrialized Disinformation 2020 Global Inventory of Organized Social Media Manipulation. Computational Propaganda Research Project," demtech.oii.ox.ac.uk.

[37] E. Cryst and I. García Camargo, "#VivaJOH o #FueraJOH. An analysis of Twitter's takedown of Honduran accounts," Standford Internet Observatory, Apr. 2020, doi: 10.25740/vz964sv5963.

[38] E. Culliford, "Facebook says it removed troll farm run by Nicaraguan government," reuters.com.

[39] V. Bergengruen, "Honduras Shows How Fake News Is Changing Latin American Elections," time.com.

[40] E. Providel, D. Toro, F. Riquelme, M. Mendoza, and E. Puraivan, "CLNews: The First Dataset of the Chilean Social Outbreak for Disinformation Analysis," Proceedings of the 31st ACM International Conference on Information & Knowledge Management, Atlanta, GA, USA, Oct. 2022, doi: 10.1145/3511808.3557560.