https://doi.org/10.37815/rte.v36n2.1219
Original articles

Fake News Detection and Fact Checking in X posts from Ecuador Chequea and Ecuador Verifica using Spanish Language Models

Uso de modelos en idioma español para la detección de noticias falsas y verificación de hechos en tuits de Ecuador Chequea y Ecuador Verifica

Mariuxi Toapanta Bernabé[1 2] https://orcid.org/0000-0002-4839-7452,
Miguel Ángel García-Cumbreras[2] https://orcid.org/0000-0003-1867-9587, L. Alfonso Ureña-López[2] https://orcid.org/0000-0001-7540-4059

[1]*Universidad de Guayaquil*, Guayaquil, Ecuador
mctb0005@red.ujaen.es

[1]*Universidad de Jaén*, Jaén, España
magc@uja.es, laurena@ujaen.es

Sent: 2024/07/15
Accepted: 2024/12/20
Published: 2024/12/30

Abstract

Currently, verifying news content before its dissemination poses a significant challenge due to the rapidity with which it spreads and the ease of replication. These factors contribute to the proliferation of fake news. Collaborative initiatives like Duke Reporters' Lab and the International Fact-Checking Network (IFCN) have been established to enhance the accuracy of fact-checking to combat various forms of disinformation. The accredited fact-checking platforms in Ecuador are Ecuador Chequea and Ecuador Verifica.

This paper details the outcomes from five transformer-based models, namely BETO, MarIA, RoBERTuito, BERTuit, and BERTin, for classifying fake news in Spanish. The rating system of Ecuador Chequea and Ecuador Verifica validated the news gathered from these platforms' accounts on the social network X (Formally known as Twitter), including X posts generated between January 2020 and March 2024. The findings validate that in terms of accuracy, recall, precision, and F1 score, the MarIA language model outperforms other Spanish-based models such as BERTin, RoBERTuito, BETO, and BERTuit.

Summary: Introduction, Materials and Methods, Results, Discussion, Conclusions and Further Work..

How to cite: Toapanta , M., García-Cumbreras, M. & Ureña-López, L. (2024). Fake News Detection and Fact Checking in X posts from Ecuador Chequea and Ecuador Verifica using Spanish Language Models. *Revista Tecnológica - Espol, 36(2), 158-173.* https://rte.espol.edu.ec/index.php/tecnologica/article/view/1219

Keywords: Fact-checking grading system, fake news, fact-checker, Spanish Language Models, Text Classification.

Resumen

En el contexto actual, validar el contenido de las noticias previo a su publicación representa un desafío significativo debido a la inmediatez con que se difunden y la sencillez con la que pueden replicarse, condiciones que favorecen a la propagación de noticias falsas. Proyectos colaborativos como Duke Reporters'Lab y la International Fact-Checking Network IFCN han surgido para promover la veracidad en la verificación de hechos con el fin de combatir diversas manifestaciones de desinformación. En el Ecuador, los medios de verificación de hechos acreditados son Ecuador Chequea y Ecuador Verifica.

Este artículo presenta los resultados de cinco modelos basados en transformadores como BETO, MarIA, RoBERTuito, BERTuit y BERTin, para la clasificación de noticias falsas en idioma español. El sistema de calificación de Ecuador Chequea y Ecuador Verifica se utilizaron para verificar las noticias que se extrajeron de las cuentas de estos medios en X y contenían los tuits creados entre Enero-2020 y Marzo-2024. Los hallazgos muestran que en términos de exactitud, recuperación, precisión y puntuación F1, el modelo de lenguaje MarIA tiene un mejor desempeño que los modelos basados en el idioma español como BERTin, RoBERTuito, BETO y BERTuit.

Palabras clave: Sistema de Calificación para verificación de hechos, Noticias Falsas, Verificador de hechos, Modelos de lenguaje en español, Clasificación de Texto.

Introduction

Over the last decade, the phenomenon of "fake news"[1] has underscored the increasing global need for data verification, also known as Fact-Checking (Ireton & Posetti, 2020).

Fact-Checking Organizations

To support Fact-Checking initiatives, the Poynter Institute established the International Fact-Checking Network (IFCN)[2] in 2015, which adheres to the Code of Principles (IFCN Code of Principles, 2024a,b)[3].

On the other hand, the Duke Reporters' Lab[4] was founded in 2011 (Duke Reporters' Lab, 2024), and both organizations offer a verification map[5], that facilitates the identification of fact-checkers globally, thus providing a valuable resource for consultation.

[1] For the Collins Dictionary, it is "false information, often sensational, disseminated under the guise of news," while the Oxford Dictionary defines it as "news that conveys or incorporates false, fabricated, or deliberately misleading information, or that is characterized or accused of doing so".
[2] It brings together a growing community of data verifiers from around the world and advocates for factual information in the global fight against misinformation. (Poynter, 2024).
[3] Comprised of five guiding principles that signatories of the IFCN commit to uphold: 1) a commitment to independence and impartiality; 2) transparency of sources; 3) transparency in funding; 4) transparency with methodology; and 5) a commitment to open and honest correction. (IFCN Code of Principles, 2024a,b).
[4] It is a journalistic research institute of the Sanford School of Public Policy at Duke University, whose main projects focus on fact-checking. (Duke Reporters' Lab, 2024).
[5] It is a global geographical representation where sites accredited for verification by the IFCN or Duke Reporters' Lab are registered. (Universidad de Las Américas, 2021).

Fact-Checking in Ecuador

Ecuador Chequea

In October 2016, Ecuador Chequea was founded, becoming the first Ecuadorian media dedicated exclusively to verifying public discourse and deceptive content circulating on the Internet. This portal is part of a global data verification network that includes, among others, Chequeado from Argentina, a pioneering project in Latin America that has provided training to the Ecuador Chequea team. Additionally, it is a member of Latam Chequea and has obtained certification from the International Fact-Checking Network (IFCN) (Ecuador Chequea, 2024a,b).

Ecuador Verifica

It is an initiative coordinated by Ecuador Chequea that emerged to combat misinformation in the 2021 presidential elections.

The National Democratic Institute (NDI), an organization dedicated to combating misinformation, supports a coalition that unites media outlets, civil society organizations (CSOs), and universities to verify political discourse and promote transparency in public institutions (Ecuador Verifica, 2024).

Fact-Checking Rating System

The methodology applied by Ecuador Chequea for the verification process is internationally certified and endorsed by the International Fact-Checking Network (IFCN). As of July 19, 2021, Ecuador Verifica adopts the same methodology as Ecuador Chequea, and their websites[6] detail their rating system to verify news as "CIERTO", "Falso", "ALTERADO", "ENGAÑOSO", "IMPRECISO", "SÁTIRA" and "INVERIFICABLE".

Figure 1

Ecuador Chequea and Ecuador Verifica Rating System

[6] Methodology Used by Ecuador Chequea https://ecuadorchequea.com/metodologia/ and Ecuador Verifica https://ecuadorverifica.org/metodologia/

Research Objectives and Questions

- RO1. To conduct a rigorous and equitable evaluation of various Spanish language models using a dataset of verified news from Ecuador Chequea and Ecuador Verifica, to compare their quality and effectiveness in accurately detecting fake news. RQ1. How can these Spanish language models be evaluated to determine their performance in detecting fake news within the Ecuadorian context?
- RO2. To analyze the performance of these models through key metrics such as accuracy, recall, precision, and F1 score, identifying the most effective models in handling Spanish-language social media data. RQ2. When analyzed through these metrics, what are the strengths and limitations of the Spanish language models, and how do they detect fake news in X posts?
- RO3. To ensure that the dataset and code are accessible for further research, enabling reproducibility and future studies in this area. RQ3. How can the provision of an accessible dataset and code facilitate additional investigations and ensure the reproducibility of results in the detection of fake news in Spanish-language contexts?

These objectives and questions aim to rigorously evaluate, compare, and provide insights into the effectiveness and capabilities of Spanish language models in detecting fake news, using the fact-checking methodologies employed by "Ecuador Chequea" and "Ecuador Verifica". This research seeks to understand the performance of these models in accurately identifying misinformation within the Ecuadorian context and the broader Spanish-speaking communities, ensuring that the process aligns with established fact-checking standards.

The article is organized as follows: Section 2 provides a comprehensive review of the most significant research conducted with Spanish Language Models. Section 3 details the dataset used for training, the preprocessing techniques applied to the X posts and outlines the experimental setup. Section 4 presents the evaluation results of the language models. Section 5 offers an in-depth analysis of the model's performance and behavior based on the results. Finally, Section 6 concludes with the main findings and suggestions for future research directions.

Related Work of NLP

The field of Natural Language Processing (NLP) has undergone a paradigm shift in recent years, rendering techniques previously used in many tasks obsolete. Deep Learning (DL) techniques are currently being researched to detect and eliminate the proliferation of fake news (Martínez-Gallego, Álvarez-Ortiz, & Arias-Londoño, 2021).

English Language Models

Language models, which require computationally demanding hardware and unsupervised pre-training methods with large corpora, have become a basic component in the field of natural language processing. The transformer architecture has been utilized in a wide range of NLP tasks in recent years (Vaswani et al., 2017), outperforming previous models based on recurrent neural networks.

Recent NLP innovations have predominantly been driven by large companies, with a predominant focus on the English language due to the high costs associated with developing and training such models. Many pre-trained models are available for free in English, though there has been a growing effort to apply these techniques to other languages (Gutiérrez-Fandiño, et al., 2021), (Vaca Serrano, et al., 2022) and (Agerri & Agirre, 2022).

Spanish Language Models\

Despite being the second most spoken language in the world, Spanish lacks significant linguistic models. However, more models in this language have been released in recent years. Unfortunately, they are not as effective as the models in English due to the lack of Spanish corpora of the same quality and volume used by English models; they also have significant costs, ranging into the hundreds of thousands of dollars and are accessible only to large multinational corporations.

In addition to language, the task domain is another critical factor that degrades the performance of these types of models. The more different the target domain compared to the source domain, the more noticeable the degradation. This is particularly true for the X domain, where users typically communicate with each other in an informal manner and using social media slang.

Many models on X have been trained in various languages for user-generated text. However, pre-trained models for user-generated text in Spanish are either not available or hard to find in popular model repositories, such as Hugging Face's model hub (Pérez et al., 2022).

Spanish pre-trained model

The use of pre-trained models for Spanish text has seen significant advancements in the last years.

BETO

It is a version of BERT trained on an extensive Spanish text corpus, marking the first pre-trained model for the Spanish language (Cañete, et al., 2023). Created in late 2019 by researchers from the University of Chile, the model was trained on a collection of corpora including Wikipedia and the OPUS Spanish corpus (Tiedemann & Thottingal, 2020), and was later evaluated on the GLUES dataset, comparing favorably to the multilingual BERT. With approximately 110 million parameters, the model is of similar size to a BERT Base.

TWil-BERT

This language model, named TWil-BERT, is a specialization of BERT for the Spanish language and the X domain. This specialization involves training a BERT model from scratch to obtain coherent and contextualized embeddings of X posts in Spanish (González, Hurtado, & Pla, 2021).

MarIA

It is a family of Spanish language models providing public resources for both the industry and the scientific community. MarIA currently includes Spanish language models RoBERTa-base, RoBERTa-large, GPT2, and GPT2-large, which can be considered the largest and best models for Spanish. The models were pre-trained on a massive 570GB corpus of clean, deduplicated texts, consisting of 135 billion words from the Spanish Web Archive compiled by the National Library of Spain between 2009 and 2019 (Gutiérrez-Fandiño, et al., 2021).

RoBERTuito

This language model is a specialization of BERT for both the Spanish language and the X domain, named TWil-BERT. This specialization consists of training a BERT model from scratch to obtain coherent and contextualized embeddings of X posts in Spanish (Pérez, Furman, Alonso Alemany, & Luque, 2021).

RigoBERTa

It is a state-of-the-art language model for Spanish developed by the Institute of Knowledge Engineering (IIC) capable of adapting to different language domains (legal, health, etc.) to enhance Natural Language Processing (NLP) applications in specific fields. RigoBERTa is trained on a well-curated corpus formed from different sub corpora with key characteristics. It follows the DeBERTa architecture, which has several advantages over other similar-sized architectures like BERT or RoBERTa. (Vaca Serrano, et al., 2022).

BERTuit

The largest transformer proposed to date for the Spanish language, BERTuit, has been pre-trained on a large dataset of 230 million Spanish X posts using the RoBERTa optimization.

The proposed resource enables the identification of fake news propagators on X and the detection of such users. It can be used in applications focused on this social network that aim to combat the spread of misinformation. (Huertas-Tato, Martín, & Camacho, 2022).

BERTin

It is a model produced during the Flax/Jax Community Week. The BERTIN model is a RoBERTa-large with 24 layers, 16 heads, 1024 hidden units, and 355M parameters. It was trained from scratch on the Spanish portion of the mC4 dataset. (De la Rosa, et al., 2022).

Pretrained Generalist Models vs. Domain-Specific Models

Domain-specific models like MarIA and BERTuit offer several key advantages over pretrained generalist models such as BERT or GPT-3. (Ding et al. 2023; Huertas-Tato, Martin, y Camacho 2022; Martinez-Rico, Araujo, and Martinez-Romo 2024; Sarker 2021, 2022; Sellami, Sadat, and Beluith 2018).

- Domain-Specific Context and Vocabulary. These models capture the nuanced and specialized vocabulary inherent to a specific domain, enabling them to learn the precise usage and meaning of terms and phrases. For instance, MarIA is tailored to the Spanish language, having been trained on a wide range of texts including Spanish Wikipedia, news articles, and diverse documents. At the same time, BERTuit is specialized in the informal, condensed language typical of Spanish X posts. In contrast, generalist models often lack this level of detail, resulting in less accuracy when dealing with domain-specific terminology.
- Enhanced Semantic Understanding. Domain-specific models demonstrate superior semantic comprehension, especially when handling ambiguous terms or specialized jargon. This reduces the likelihood of misinterpretation and increases accuracy when dealing with polysemous words or technical expressions, making them particularly effective in contexts such as legal, medical, or technical texts.
- Adaptation to Language Styles and Structures. Models like MarIA and BERTuit are better equipped to adapt to distinct communication styles and syntactic structures. For example, BERTuit is adept at understanding the informal, often fragmented language used on social media platforms, whereas MarIA is more proficient with the formal structure found in scientific articles or technical reports. Generalist models, on the other hand, often struggle with these stylistic variations, leading to less precise interpretations.
- Noise Reduction in Learning. By focusing exclusively on texts relevant to their domain, these models avoid learning unnecessary patterns, resulting in more efficient and targeted learning. This minimizes the noise often encountered in training generalist models, which are exposed to a wider range of irrelevant data,

thereby enhancing the overall performance and accuracy of domain-specific models.

- Effective Knowledge Transfer within the Domain. Domain-specific models effectively leverage and transfer knowledge across the primary domain and its subdomains. For example, a model trained on medical language can quickly adapt to specialized fields like cardiology or neurology, utilizing the foundational knowledge to enhance its performance in these subdomains.

Domain-specific models such as MarIA and BERTuit are highly advantageous in scenarios that require nuanced understanding and precision, particularly in tasks such as sentiment analysis, fact-checking, or the detection of fake news. They outperform generalist models by delivering more accurate and contextually relevant results in applications that demand a deep understanding of language within specific domains.

Size and Specificity of the Training Corpus in Domain-Specific Models

The size and specificity of the training corpus are crucial factors that impact the effectiveness of Spanish NLP models like MarIA and BERTuit in detecting fake news.

Larger training corpora. Allow models to capture a wider range of linguistic patterns, improving their ability to generalize and accurately detect fake news, especially on dynamic platforms like X.

Domain-specific corpora. Enhance precision by training models on the particular vocabulary, context, and patterns of a specific domain (e.g., politics or health), making them more effective in identifying misinformation in those areas.

Research supports that models trained on large, domain-specific data achieve superior performance in fake news detection (Peña et al. 2023). This approach maximizes accuracy and allows models to handle diverse misinformation more effectively. (Garrido-Muñoz, Martínez-Santiago, and Montejo-Ráez 2023)

Advantages and Disadvantages of Preprocessing and Fine-Tuning in NLP for Social Media Analysis

Advantages

- Better Handling of Informal Language: Fine-tuning models like BERTuit allows them to capture the nuances of social media language, improving accuracy in text analysis. (Peña et al. 2023)
- Adaptation to Specific Domains: Models like MarIA perform better when fine-tuned with domain-specific data, making them more effective in detecting fake news in targeted contexts. (Peña et al. 2023)
- Efficiency: Preprocessing reduces noise and saves computational resources, enhancing the training process for large datasets. (Garrido-Muñoz, Martínez-Santiago, and Montejo-Ráez 2023)

Disadvantages

- Loss of Important Information: Overly aggressive preprocessing can remove crucial context, such as emojis, affecting sentiment analysis. (Garrido-Muñoz, Martínez-Santiago, and Montejo-Ráez 2023)
- Risk of Bias: Fine-tuning can amplify biases present in the training data, leading to skewed results. (Garrido-Muñoz, Martínez-Santiago, and Montejo-Ráez 2023)

- Complexity: Fine-tuning requires high-quality data and expertise, making it more challenging to implement.

Preprocessing and fine-tuning techniques offer clear advantages in adapting NLP models to the unique and often complex language used on social media, making them particularly effective for tasks such as fake news detection in Spanish. Models like MarIA and BERTuit benefit from these techniques by achieving greater accuracy and domain-specific understanding. However, these methods carry the risk of amplifying biases, losing important contextual information, and requiring significant expertise and computational resources to implement effectively. Striking the right balance between thorough preprocessing and maintaining linguistic nuances is essential for ensuring that these models remain both accurate and unbiased, providing reliable results in dynamic environments like X.

Materials and Methods

Dataset

The dataset comprises posts on X that were verified by Ecuador Chequea @ECUADORCHEQUEA and Ecuador Verifica @ecuadorverifica between January 1, 2020, and March 25, 2024.

The Apify platform, which extracts data from websites, is used to collect data from posts on X. Tasks are created with the required filters for execution. The information from the X posts is downloaded in CSV format after completing the task.

Cleaning and Preprocessing

The cleaning and preprocessing process for the news from the @ECUADORCHEQUEA and @ecuadorverifica accounts begins once the dataset is loaded.

The preprocessing aims to make the data easier for the algorithm to process and is an important step involving data manipulation before execution to increase efficiency. Performance may decrease if the target domain is very different from the domain used in pre-training.

This is the case with X, where users communicate informally, using typical social media slang expressions, often with lexical-syntactic errors or adding special tokens like hashtags, user mentions, and emojis. Additionally, some of the most used NLP techniques such as Stop Words, Stemming, Tokenization, and Padding are included.

Data Cleaning

The data standardization process known as Text Normalization removed irrelevant elements, such as links, mentions, hashtags, emojis, and punctuation marks, in addition to using a series of regular expressions to preprocess the text.

For data cleaning, the X posts-preprocessor library is used instead of writing the X post-cleaning logic. The clean method performs X post cleaning by removing URLs, mentions, reserved words (RT, FAV), hashtags, emojis, smileys, and numbers by default unless you specify only some of the options.

Continuing with the cleaning process, the text is converted to lowercase, punctuation marks and extra spaces are removed.

Preprocessing

Next, the preprocessing process is carried out.

- Tokenization begins using TweetTokenizer, which is a specialized tokenizer designed to handle X posts and other social media texts.
- Stopwords uses the NLTK library to remove empty words using the Spanish corpus.
- Lemmatization uses the STANZA library. This text normalization technique aims to reduce words to their root (lemma).

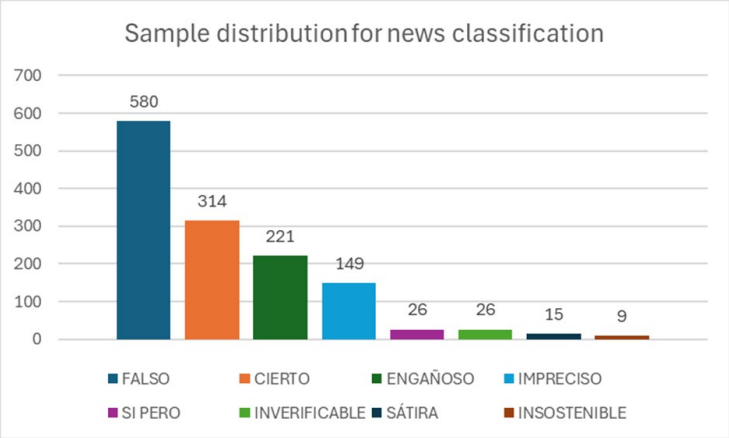
Experiments design

Dataset

The Spanish news corpus contains 1340 news items in Spanish, consisting of 580 "FALSO", 314 "CIERTO", 221 "ENGAÑOSO", 149 "IMPRECISO", 26 "SI PERO", 26 "INVERIFICABLE", 15 "SÁTIRA" y 9 "INSOSTENIBLE". Figure 2 shows through a chart the corresponding distribution for each news.

Figure 2

Sample distribution for Ecuador Chequea and Ecuador Verifica news classification



The split of the dataset was set to a ratio of train: 80% / test: 20%. Based on the outcomes of the initial analysis, among the techniques for balancing the dataset, oversampling of the minority classes was chosen, which led to the use of SMOTE (Synthetic Minority Over-sampling Technique). Table 1 shows the values after balancing the dataset.

Table 1

Training arguments for transformers-based models

Classification News	Number of News
FALSO	580
CIERTO	580
ENGAÑOSO	580
IMPRECISO	580
SI PERO	580
INVERIFICABLE	580
SÁTIRA	580
INSOSTENIBLE	580
Total	**4640**

Experiment setup

Fine-tuning a pre-trained model and adjusting TrainingArguments are crucial steps in a machine learning project, especially when you are working with language models like BERT, GPT, or RoBERTa.

TrainingArguments is a class provided by the Hugging Face's transformers library, offering a comprehensive set of parameters for configuring the training process of machine learning models, particularly those related to natural language processing (NLP). The most frequent values are explained below:

- output_dir: Specifies the directory where the model checkpoints and training outputs will be saved. This is essential for model persistence and later evaluation or use.
- per_device_train_batch_size: Determines the batch size for training on each device (GPU/CPU). Smaller batch sizes require less memory but can lead to longer training times and potentially more noise during the training process.
- per_device_eval_batch_size: Sets the batch size for evaluation on each device. This is similar to per_device_train_batch_size but used during the model evaluation phase.
- num_train_epochs: Defines the total number of training epochs. An epoch represents one complete pass through the entire training dataset. Adjusting the number of epochs can affect the model's performance and training time.
- warmup_steps: The number of steps to increase the learning rate from 0 to the initial learning rate set via learning_rate. This can help improve model performance and stability in the initial phase of training.
- weight_decay: A regularization parameter to prevent overfitting by applying penalties on large weights during optimization. Adjusting this can help improve the generalization of the model.
- learning_rate: The initial learning rate for the optimizer. The choice of learning rate can significantly impact model performance and convergence speed.
- evaluation strategy: Determines when the model evaluation should occur. Options include "no" (no evaluation), "steps" (after a set number of training steps), and "epoch" (after each training epoch).
- save_strategy: Similar to evaluation_strategy, but for saving model checkpoints. This can be set to "no", "steps", or "epoch".
- logging_dir: Directory where the training logs will be stored. Useful for monitoring the training process through tools like TensorBoard.
- logging_steps is a parameter within TrainingArguments that specifies the frequency at which training, and evaluation metrics are logged. This parameter is particularly useful for monitoring the model's performance and the training process in real-time or through logging frameworks like TensorBoard.
- load_best_model_at_end: Whether to load the model checkpoint with the highest evaluation metric or the final model at the end of training. Useful for automatically selecting the best model.
- metric_for_best_model: Specifies the metric to use when load_best_model_at_end is True. The model with the best value of this metric will be loaded at the end.

The Table 2 shows the parameterized values in TrainingArguments for Fine-tuning a pre-trained model.

Table 2

Training arguments for transformers-based models

Model	Training arguments
BETO[7]	output_dir='./results', per_device_train_batch_size=16, per_device_eval_batch_size=64, num_train_epochs=4, warmup_steps=500, weight_decay=0.01, learning_rate=1e-5, evaluation_strategy="epoch", save_strategy='epoch', logging_dir='./logs', logging_steps=100, load_best_model_at_end=True, metric_for_best_model="accuracy"
MarIA[8]	output_dir='./results', per_device_train_batch_size=16, per_device_eval_batch_size=16, num_train_epochs=3, warmup_steps=500, weight_decay=0.01, learning_rate=1e-5, evaluation_strategy="epoch", save_strategy='epoch', logging_dir='./logs', logging_steps=100, load_best_model_at_end=True, metric_for_best_model="accuracy"
RoBERTuito[9]	output_dir='./results', per_device_train_batch_size=8, per_device_eval_batch_size=8, num_train_epochs=3, warmup_steps=500, weight_decay=0.01, learning_rate=2e-5, evaluation_strategy="epoch", save_strategy='epoch', logging_dir='./logs', logging_steps=100, load_best_model_at_end=True, metric_for_best_model="accuracy"
BERTuit[10]	output_dir='./results', per_device_train_batch_size=8, per_device_eval_batch_size=8, warmup_steps=500, weight_decay=0.01, learning_rate=1e-5, evaluation_strategy="epoch", save_strategy='epoch', logging_dir='./logs', logging_steps=100, load_best_model_at_end=True, metric_for_best_model="accuracy"
BERTin[11]	output_dir='./results', per_device_train_batch_size=8, per_device_eval_batch_size=8, num_train_epochs=3, warmup_steps=500, weight_decay=0.01, learning_rate=1e-5, evaluation_strategy="epoch", save_strategy='epoch', logging_dir='./logs', logging_steps=100, load_best_model_at_end=True, metric_for_best_model="accuracy"

Results

Model fine-tuned

According to the results, the MarIA model outperformed the other models, achieving an Accuracy of 0.9601, indicating its high efficiency in correctly classifying fake news, as shown in Table 7. Additionally, the metrics obtained during the training of the remaining models are detailed in Table 3, 4, 5, 6, 7 and 8, providing a comprehensive comparison of their performance.

Table 3

Train Output BETO Model

Epoch	Training Loss	Validation Loss	Accuracy	Precision	Recall	F1 Score
1	1.9099	1.460454	0.575431	0.519867	0.575431	0.517765
2	0.9221	0.627142	0.789871	0.792705	0.789871	0.780918
3	0.5125	0.306508	0.911638	0.914206	0.911638	0.910657
4	0.2155	0.22554	0.935345	0.934864	0.935345	0.934516

[7] bert-base-spanish-wwm-uncased https://huggingface.co/dccuchile/bert-base-spanish-wwm-uncased
[8] PlanTL-GOB-ES/roberta-base-bne https://huggingface.co/PlanTL-GOB-ES/roberta-base-bne
[9] pysentimiento/robertuito-base-cased https://huggingface.co/pysentimiento/robertuito-base-cased
[10] bertin-project/bertin-roberta-base-spanish https://huggingface.co/bertin-project/bertin-roberta-base-spanish
[11] bertin-project/bertin-roberta-base-spanish https://huggingface.co/bertin-project/bertin-roberta-base-spanish

Table 4

Train Output MarIA Model

Epoch	Training Loss	Validation Loss	Accuracy	Precision	Recall	F1 Score
1	0.0398	0.185121	0.954741	0.954894	0.954741	0.953861
2	0.0156	0.35076	0.929957	0.943936	0.929957	0.926845
3	0.028	0.214479	0.960129	0.95985	0.960129	0.959735

Table 5

Train Output RoBERTuito Model

Epoch	Training Loss	Validation Loss	Accuracy	Precision	Recall	F1 Score
1	0.8486	0.59158	0.782328	0.827286	0.782328	0.772328
2	0.2734	0.257701	0.90625	0.908591	0.90625	0.901669
3	0.1322	0.211828	0.935345	0.935905	0.935345	0.933532

Table 6

Train Output BERTuit Model

Epoch	Training Loss	Validation Loss	Accuracy	Precision	Recall	F1 Score
1	0.3602	0.348519	0.881466	0.887611	0.881466	0.872059
2	0.1818	0.259604	0.923491	0.924281	0.923491	0.92049
3	0.1032	0.254212	0.934267	0.933133	0.934267	0.932568

Table 7

Train Output BERTin Model

Epoch	Training Loss	Validation Loss	Accuracy	Precision	Recall	F1 Score
1	1.1841	0.801565	0.693966	0.750559	0.693966	0.664146
2	0.2706	0.205788	0.9375	0.938318	0.9375	0.936755
3	0.064	0.150867	0.954741	0.954394	0.954741	0.954364

Table 8

Train Output Models

Model	Accuracy	Precision	Recall	F1 Score
BETO	0.9353	0.9349	0.9353	0.9345
MarIA	0.9601	0.9599	0.9601	0.9597
RoBERTuito	0.9353	0.9359	0.9353	0.9335
BERTuit	0.9343	0.9331	0.9343	0.9326
BERTin	0.9547	0.9544	0.9547	0.9544

Discussion

Research involving Spanish language models for text classification, specifically in the detection of fake news, has achieved notable results using datasets drawn from verified X posts by Ecuador Chequea and Ecuador Verifica. The experiments were designed to assess how effectively these models can classify news based solely on text data from X, without incorporating multimedia elements, following the fact-checking criteria established by these Ecuadorian verifiers.

The study focused on rigorously evaluating the models' ability to classify news as “CIERTO”, “Falso”, “ALTERADO”, “ENGAÑOSO”, “IMPRECISO”, “SÁTIRA” and “INVERIFICABLE”, adhering strictly to the verification methodologies employed by “Ecuador Chequea” and “Ecuador Verifica”. This process is crucial for understanding the nuances of misinformation in the Spanish language, especially within the Ecuadorian context, where fake news can impact public discourse.

The performance metrics—such as accuracy, precision, recall, and F1 score—offer a comprehensive view of how effectively the Spanish language models can handle text-based analysis on X. These metrics not only provide valuable insights into the strengths and potential improvements of the models when applied to real-world fact-checking and verification tasks but also highlight their potential as powerful tools in the fight against disinformation. The ability of these models to distinguish between factual and fabricated content with such high accuracy and precision is crucial for automating the detection of fake news, thereby enhancing the reliability of information disseminated across digital platforms.

Accuracy

The model achieved an accuracy rate of 0.9601, demonstrating a high degree of reliability in correctly identifying the veracity of news across the entire test dataset. This metric reflects the model's overall effectiveness in accurately classifying both genuine (true) and misleading (fake) news, making it a strong candidate for real-world applications.

Precision

With a precision score of 0.9599, the model effectively returns relevant results, indicating that when it classifies news as fake, it is accurate 95.99% of the time. This high precision is essential for minimizing false positives, ensuring that legitimate news is not incorrectly flagged as fake, which is critical for maintaining credibility in content moderation.

Recall

The model’s recall score of 0.9601 indicates its strong ability to identify nearly all instances of fake news within the dataset. In practical terms, this means the model successfully captures 96.01% of fake news cases, showcasing its efficiency in detecting a broad spectrum of misleading content with minimal oversight.

F1 Score

The model achieved an F1 score of 0.9597, which represents the harmonic mean of precision and recall. This balanced metric confirms the robustness and overall effectiveness of the model, indicating that it maintains both high precision and recall. An F1 score close to 1 underscores the model’s reliability in distinguishing between truthful and false content.

Conclusions and Further Work

The research demonstrates that Spanish language models, particularly MarIA and BERTuit, offer significant advantages in accurately detecting fake news on social media

platforms like X.These models, trained on large, domain-specific datasets—such as X posts verified by Ecuador Chequea and Ecuador Verifica—show exceptional performance compared to generalist language models, particularly in capturing the nuances of the Spanish language and the cultural context of misinformation.

The study's key findings highlight the importance of domain-specific training and fine-tuning techniques, which enable the models to adapt effectively to the informal and diverse language used on social media. The achieved accuracy of 0.9601 by the MarIA model, demonstrates its effectiveness in identifying false information within the Spanish-speaking context.

However, the integration of fine-tuning techniques allowed the model to adapt to the informal language, abbreviations, and slang typical of social media posts, particularly X posts, further enhancing its ability to detect misleading content.

Overall, this study confirms the potential of domain-specific models as robust solutions for automating fake news detection, contributing to enhancing the reliability of information disseminated across digital platforms in Spanish-speaking contexts.

For further work the following is suggest:

- Expand the dataset by collecting a broader range of verified X posts from @ECUADORCHEQUEA and @ecuadorverifica to enhance the robustness of the analysis.
- Incorporate additional X post elements that were not covered in this study, such as metadata or user interaction patterns, to provide a more comprehensive understanding of the context.
- Develop a comprehensive labeling guide to standardize the annotation process for the dataset used in this research, ensuring consistency and reliability in future analyses.
- Integrate a labeling feature within the Support System for Fact-Checkers, allowing fact-checkers to directly label X posts. This feature should align with the rating criteria used by “Ecuador Chequea” and “Ecuador Verifica” to maintain uniformity
- Utilize the trained language models within the proposed Support System, providing “Ecuador Chequea” and “Ecuador Verifica” with advanced tools for fact-checking and enhancing the accuracy of X post verification.

Acknowledgements

This research is part of the proposal presented at the Call for Research Project Proposals of the Internal Competitive Fund (FCI) 2023, which was approved on February 22, 2023, by the members of the Scientific Council who attended the Ordinary Session No. 1-2023 held under a hybrid modality according to the Opinion of the Research Council of the Faculty of Industrial Engineering and is currently in process due to the extension of the call according to RESOLUTION NO. CCIPI-SE-003-08-2022. Also, CONSENSO (PID2021-122263OBC21), MODERATES (TED2021-130145B-I00), and SocialTOX (PDC2022-133146-C21) funded by Plan Nacional I+D+i from the Spanish Government.

The authors declare equal contribution and sharing of authorship roles for this publication.

References

Agerri, R., & Agirre, E. (2022). Lessons learned from the evaluation of Spanish Language Models. *arXiv preprint arXiv:2212.08390.* https://doi.org/10.48550/arXiv.2212.08390

Cañete, J., Chaperon, G., Fuentes, R., Ho, J.-H., Kang, H., & Pérez, J. (2023). Spanish pre-trained bert model and evaluation data. *arXiv preprint arXiv:2308.02976*. https://doi.org/10.48550/arXiv.2308.02976

De la Rosa, J., Ponferrada, E. G., Villegas, P., González de Prado Salas, P., Romero, M., & Grandury,

M. (2022). Bertin: Efficient pre-training of a spanish language model using perplexity sampling. *arXiv preprint arXiv:2207.06814*. https://doi.org/10.48550/arXiv.2207.06814

Ding, J., Li, B., Xu, C., Qiao, Y., & Zhang, L. (2023). Diagnosing crop diseases based on domain- adaptive pre-training BERT of electronic medical records. *Applied Intelligence*, *53*(12), 15979-15992. https://doi.org/10.1007/s10489-022-04346-x

Duke Reporters' Lab. (2024). *About the Lab*. https://reporterslab.org/about-the-lab/

Duke Reporters' Lab. (2024). *Global Fact-Checking sites*. https://reporterslab.org/fact- checking/

Ecuador Chequea. (2024a, September 3). *Metodologia - Ecuador Chequea*. https://ecuadorchequea.com/metodologia/

Ecuador Chequea. (2024b, September 13). Nuestra historia - Ecuador Chequea. https://ecuadorchequea.com/historia/

Ecuador Verifica. (2024, February 17). *QUIÉNES SOMOS - Ecuador verifica*. https://ecuadorverifica.org/quienes-somos/

Garrido-Muñoz, I., Martínez-Santiago, F., & Montejo-Ráez, A. (2023). MarIA and BETO are sexist:

Evaluating gender bias in large language models for Spanish. *Language Resources and Evaluation 58,* 1387–1417. https://doi.org/10.1007/s10579-023-09670-3

González, J., Hurtado, L.-F., & Pla, F. (2021). TWilBert: Pre-trained deep bidirectional transformers for Spanish. *Neurocomputing, 426*, 58-69. https://doi.org/10.1016/j.neucom.2020.09.078

Gutiérrez-Fandiño, A., Armengol-Estapé, J., Pàmies, M., Llop-Palao, J., Silveira-Ocampo, J., Pio Carrino, C., . . . Villega, M. (2021). Maria: Spanish language models. *{arXiv preprint arXiv:2107.07253. https://doi.org/10.26342/2022-68-3*

Huertas-Tato, J., Martín, A., & Camacho, D. (2022). BERTuit: Understanding Spanish language in X through a native transformer. *arXiv preprint arXiv:2204.03465*. https://doi.org/10.48550/arXiv.2204.03465

IFCN Code of Principles. (2024a). *The commitments of the code of principles*. https://www.ifcncodeofprinciples.poynter.org/

IFCN Code of Principles. (2024b). *Verified signatories of the IFCN code of principles*. https://www.ifcncodeofprinciples.poynter.org/signatories

Ireton, C., & Posetti, J. Eds. (2020). *Periodismo,"noticias falsas" & desinformación: Manual de Educación y Capacitación en periodismo.* París; Santo Domingo, Francia: Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura UNESCO y el Ministerio de la Presidencia de la República Dominicana. https://unesdoc.unesco.org/ark:/48223/pf0000373349

Martínez-Gallego, K., Álvarez-Ortiz , & Arias-Londoño, J. (2021). Fake News Detection in Spanish Using Deep Learning Techniques. *arXiv preprint* arXiv:2110.06461v1 https://doi.org/10.48550/arXiv.2110.06461

Martinez-Rico, J. R., Araujo, L., & Martinez-Romo, J. (2024). Building a framework for fake news detection in the health domain. *PLOS ONE, 19*(7), e0305362. https://doi.org/10.1371/journal.pone.0305362

Peña, A., Morales, A., Fierrez, J., Serna, I., Ortega-Garcia, J., Puente, Í., Córdova, J., & Córdova, G. (2023). Leveraging Large Language Models for Topic Classification in the Domain of Public Affairs. En M. Coustaty & A. Fornés (Eds.), *Document Analysis and Recognition – ICDAR 2023 Workshops* (pp. 20-33). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-41498-5_2

Pérez, J. M., Furman, D. A., Alemany, L. A., & Luque, F. (2022). *RoBERTuito: A pre-trained language model for social media text in Spanish* (No. arXiv:2111.09453). arXiv. https://doi.org/10.48550/arXiv.2111.09453

Poynter. (2024). *Red Internacional de Verificación de Datos (IFCN)*. https://www.poynter.org/ifcn/

Sarker, I. H. (2021). Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions. *SN Computer Science*, *2*(6), 420. https://doi.org/10.1007/s42979-021-00815-1

Sarker, I. H. (2022). AI-Based Modeling: Techniques, Applications and Research Issues Towards Automation, Intelligent and Smart Systems. *SN Computer Science*, *3*(2), 158. https://doi.org/10.1007/s42979-022-01043-x

Sellami, R., Sadat, F., & Beluith, L. H. (2018). Building and Exploiting Domain-Specific Comparable Corpora for Statistical Machine Translation. En K. Shaalan, A. E. Hassanien, & F. Tolba (Eds.), *Intelligent Natural Language Processing: Trends and Applications* (pp. 659-676). Springer International Publishing. https://doi.org/10.1007/978-3-319-67056- 0_31