ISSN 2007-9737

Exploring Political Polarization in México: Automatic Classification of Comments on You Tube

Alba Maribel Sánchez-Gálvez[1,*], Ricardo Álvarez-González[1],
Santiago Alejandro Molina-Iturbide[1], Francisco Javier Albores-Velasco[2]

[1] Benemérita Universidad Autónoma de Puebla, Puebla,
Mexico

[2] Universidad Autónoma de Tlaxcala, Tlaxcala,
Mexico

{alba.sanchez, ricardo.alvarez}@correo.buap.mx,
santiago.molina@alumno.buap.mx, fjalboresv@garzas.uatx.mx

Abstract. YouTube, the second-largest social network globally, hosts over two and a half billion monthly users, with content surpassing five hundred hours uploaded every minute [1]. Channels dedicated to news and political discourse facilitate interactive communication, enabling users to critique, express viewpoints, and protest anonymously. Partisan engagement on social media is highly controversial and can influence the attitudes and behaviors of individuals and organizations opposing views [2]. Amid growing concerns about political polarization in Mexico, the fourth country with the highest number of YouTube users, this study aims to understand digital communication patterns and their impact on user attitudes. Web Scraping and Natural Language Processing techniques were employed to gather and analyze comments from two antagonistic political channels on YouTube: "El Chapucero" and "Atypical TV". The objective was to identify key aspects of polarization in the comments of users of these YouTube channels to create a Machine Learning model capable of predicting a user's political stance. Distinct features in the dataset were highlighted to train four Machine Learning and Deep Learning classifiers: Naive Bayes, Logistic Regression, CNN, and Bidirectional LSTM. These classifiers were able to automatically infer the political leanings of users, the one that performed the best was CNN with a precision of 96%. The main contribution of this study lies in the word analysis that provide insights into the Mexican partisan dynamics on YouTube and in the precision of comment classification, which is achieved due to the polarization existing between these political opinion channels.

Keywords: Natural language processing, text classification, web scraping, youtube, political polarization.

1 Introduction

The growing dependence on social media for political communication is creating unprecedented opportunities to study the spread of political information and misinformation through communication networks [3]. Research on the use of social media for promoting and sharing news, rumors, and information has revealed that emotional responses facilitate diffusion, increasing the likelihood that emotional content will be shared and consumed by others on the network.

Politics has become increasingly emotional, and the resulting polarization has created echo chambers that favor narratives and stories that repeat a single point of view. Additionally, individuals with intense emotional responses to political content are more likely to share and consume news and political information on social media platforms.

The ability of social media to spread content to millions of users with just one click has positioned them as a fundamental ground for political marketing. Promoters seek to identify a small initial group of users on a social network to maximize the spread of persuasion [4].

Political influencers are users capable of widely disseminating media information through social networks. An influencer affiliated with a media organization could be a media company, an official media outlet, an established writer, reporter, or paid consultant. An influencer affiliated with a

doi: 10.13053/CyS-28-4-5295

political party could be a politician or a political campaign platform. YouTube is recognized as the second most popular website in the world by Alexa Internet [5]. The role of YouTube creators and their videos on political and social issues is becoming increasingly relevant.

Credibility is directly related to the evaluation of YouTubers and the content of their videos [6]. Alongside traditional news publishing practices, news agencies now share information via the internet, as the current audience preference is to read news online. Additionally, media outlets have established YouTube channels to disseminate visual stories, and readers actively engage by commenting to share their opinions below the corresponding news.

This interaction between news and comments has become a valuable source of information and research [7]. There are two main categories of content creators on YouTube: influencers and large corporate groups.

The content of each shared video and the corresponding user comments results in the YouTube channel acquiring a political bias. Political bias in the media occurs when they emphasize viewpoints, transmit information selected to promote their own political stance, and present only information that favors their political opinion [8].

There is a growing concern that social media sites contribute to political polarization by creating echo chambers that isolate people from opposing views on current events [9]. Political polarization also has negative consequences in terms of creating extreme political ideologies, which restrict discussions and exchanges with the opposing side [8].

Research on daily comments made by users on YouTube allows us to understand the contexts in which political polarization occurs in Mexico, which boasts 83.1 million active users on YouTube, ranking fourth in the number of users after India, the United States, Brazil, and Indonesia.

Leveraging the vast number of comments on YouTube channels with opposing perspectives in Mexico, the goal of the study is to utilize this information to predict the political stance of users, using Artificial Intelligence techniques such as Natural Language Processing and Machine Learning.

2 Related Works

In this section, we delve into the application of Natural Language Processing (NLP) and Artificial Intelligence (AI) for the processing and analysis of comments. Hate speech proliferates on Thai social media platforms, including YouTube. The study outlined in the article [10] extracted comments from the live chat on 11 Thai football news channels spanning from 2021 to 2023, utilizing the Chat Downloader script. For classification purposes, transformer-based language models like BERT, XLM-RoBERTa, DistilBERT, WangchanBERTa, and TwHIN-BERT were deployed, trained across multilingual and Thai languages.

Data labeling was conducted both manually and automatically using 11 distinct datasets with varied proportions of positive and negative classes. Remarkably, XLM-RoBERTa demonstrated superior performance, attaining an F1 score of 0.96 in detecting offensive language and hate speech. Sentiment analysis unveils latent emotional trends on social media, offering insights into people's opinions and sentiments.

The authors of [11] undertook the collection and analysis of comments on YouTube pertaining to the Thailand-China High-Speed Train and Laos-China Railway projects from October 2014 to May 2022. Each comment underwent manual classification as positive, neutral, or negative, followed by the application of the automatic learning classifiers: Logistic Regression, Naïve Bayes, Random Forest, Bidirectional long short-term memory (Bi-LSTM) and Bidirectional encoder representations from transformers (BERT), with the latter exhibiting the highest precision of 94.59%.

The authors in [12] introduces a dataset comprising of 762,678 public comments and responses to 16,016 video news releases spanning from 2017 to 2023 on a reputable Bengali news YouTube channel. The primary objective is to assist scholars in identifying patterns in public opinion and analyzing temporal shifts. The dataset is publicly accessible on Mendeley.

The authors in [13], for the measurement of political polarization, a comprehensive analysis was conducted on 11 million comments sourced from over 600,000 users across 37,000 videos from 77 YouTube channels. The model hinges on user activity, delineating feature functions such as coverage, duration, and enthusiasm. Experimental findings reveal the existence of 30 highly polarized YouTube channels (comprising 16 left-wing and 14 right-wing channels) exhibiting a measured bias rate surpassing 70%.

Lastly in [14], this paper scrutinizes a dataset for sentiment analysis concerning the ongoing conflict between Ukraine and Russia. The dataset was curated by gathering comments from videos featured on three prominent YouTube TV news channels in Bangladesh, covering reports on the ongoing conflict. A total of 10,861 comments were collected and labeled with three polarity sentiments: Neutral, Pro-Ukraine (Positive), and Pro-Russia (Negative).

A reference classifier was developed, leveraging various transformer-based language models pretrained on unlabeled Bangla corpora. These models underwent fine-tuning using the acquired dataset, with hyperparameter optimization conducted across five transformer language models, including: BanglaBERT, XLM-RoBERTa-base, XLM-RoBERTa-large, Distil-mBERT, and mBERT. The best-performing model achieved an impressive 86% accuracy with a F1 Score of 0.82.

Once the state of the art was reviewed, two Machine Learning methods were selected for the proposed model in this paper: Naive Bayes and Logistic Regression, from Deep Learning CNN, and Bi-directional LSTM were chosen.

3 Methodology

As shown in Figure 1, the classification process ranges from the collection of public opinion data on YouTube channels, the preprocessing of comments, statistical analysis, and feature extraction, to the application of Machine Learning and Deep Learning algorithms and evaluation of results.

3.1 Data Extraction

With the purpose of analyzing the opinions published daily on news and politics YouTube channels, two channels were selected, both with similar content and an acceptable number of subscribers, but with opposing political tendencies: "El Chapucero," a Mexican political channel known for its humor and subtle irony with 1.61 million subscribers, and "Atypical Te Ve," a channel focused on politics and entertainment with 890 thousand subscribers.

A YouTube comment-downloader extractor was utilized to collect 3000 comments from each channel in October 2022. The extracted data include user channel, and comment, were stored in a relational database, and exported to .csv format. The data characteristics included the date, comment, votes, replies, user image, channel, among others. However, Table 1 displays only three columns for processing: User, channel, and comment.

3.2 Data Cleaning and Preprocessing

Typically, users who express their opinions on these channels use bad words and insults, with spelling errors, in addition to referring to political actors by offensive nicknames, so the preprocessing of the comments allows us to eliminate unnecessary and irrelevant words and prepare the data for subsequent analysis and applying Machine Learning and Deep Learning algorithms. This previous step helps increase accuracy and reduce processing time.

The cleaning process involved removing various elements from the comments, such as URLs links, usernames, hashtags, emojis, punctuation marks, single letters, numbers, and Unicode characters. Additionally, multiple spaces were replaced with a single space.

After the elimination of repeated or empty comments, there were 2,950 comments from “El Chapucero” and 2,887 from “Atypcal Te Ve”. To obtain a balance, the analysis was carried out with 2,887 comments per channel.

Three dictionaries were created to facilitate text cleaning in the comments of the two YouTube channels. The first one contains slang words in Spanish, which helps identify and process colloquial terms with 602 words. The second dictionary focuses on abbreviations and short

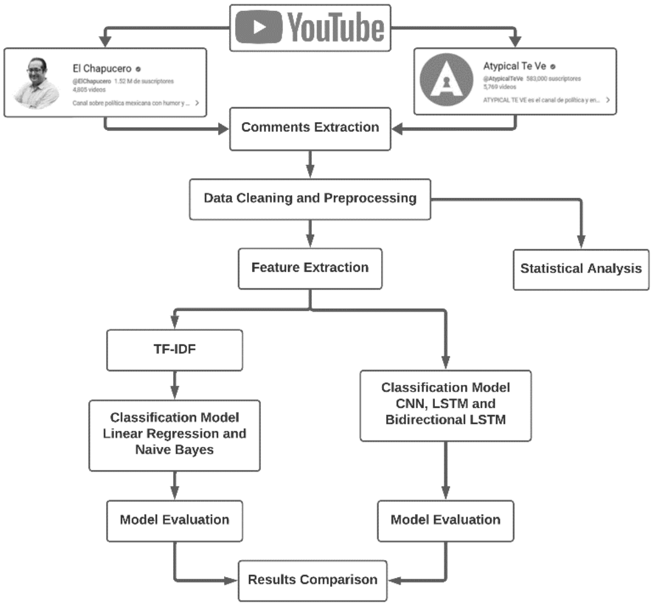

YouTube
El Chapucero
@ElChapucero 1.52 M de suscriptores
4,805 videos
Canal sobre política mexicana con humor y ...
Atypical Te Ve
@AtypicalTeVe 583,000 suscriptores
5,769 videos
ATYPICAL TE VE es el canal de política y en...
Comments Extraction
Data Cleaning and Preprocessing
Feature Extraction
Statistical Analysis
TF-IDF
Classification Model Linear Regression and Naive Bayes
Classification Model CNN, LSTM and Bidirectional LSTM
Model Evaluation
Model Evaluation
Results Comparison

Fig. 1. Methodology overview: Steps from data collection to classification evaluation

Table 1. Sample comments from the dataset

	Comment	**User**	**Channel**
0	Que pareja infernal, MALO y M-AL…	Carl-XM Mendez	atypical_te_ve
1	En realidad nos sorprende ?	Jaime Maldonado	atypical_te_ve
2	Mamadas de esta loca	Cristian raul Pech …	atypical_te_ve
3	Yo. Creo. Que. Uno. Es. Más. Pende…	Hilario Hernandez	atypical_te_ve
4	Yo. En. Mi. Opinion. Quien. Es. Ese…	Hilario Hernandez	atypical_te_ve
…	…	…	…
5996	El próximo president de México ser …	Leo Perez	el_chapucero
5997	JA JA JA JA JA, PUES SUS ARG …	Javier L	el_chapucero
5998	Yo no entiendo, xq salen de un part …	Facundo Alvarez	el_chapucero
5999	que desepcion la tía Tatis nos tenía e...	Elfego Ramos	el_chapucero

names to mention a word or a proper name with 12 words. For example, "Andres Manuel López Obrador" could be mentioned as "López Obrador", "Manuel López Obrador", "López", "Obrador", "presidente", or even informally as "el cacas", and "Cuarta Transformación" as "4t" or derogatorily as "cuatrote". The third dictionary was of misspelled words, with 1917 entries.

To normalize the text, lists of words were used, aiding in the identification and processing of colloquial terms, abbreviations, short names, and misspelled words. Misspelled words were rectified, and the text was standardized using a custom search dictionary for short words or abbreviations commonly found in informal writing but conveying the same meaning as shown in Tables 2 and 3. As part of preprocessing, comments were converted to lowercase, and stop words were removed.

Feature extraction involved segmenting the text into sentences and then into words, essentially

Table 2. Standardization of words - Abbreviations and short names dictionary

	Abbreviations and Short Names Dictionary
'amlo':	['andrés manuel lópez obrador', lópez obrador', 'presidente', 'cacas'],
'tatiana':	['tatiana clouthier', 'tía tatis', 'tatis'],
'lilly':	['lilly téllez'],
'xóchitl':	['xóchitl gálvez'],
germán':	['germán martínez'],
'cabeza':	['cabeza vaca'],
'4t':	['cuarta transformación', '4t', '4a', '4ta' '4tno', 'cuatrote'],
'alito':	['alito moreno'],

Table 3. Standardization of words – Misspelled words

	Misspelled Words
bravo	[['brabo'], ['bravoo'], ['bravooo'], ['bravooooioooooo'], ['bravoooooo'], …]
cabeza	[['acabaza'], ['acabesade'], ['acabesadebaca'], ['acabeza'], ['cabeca'], …]
clouthier	[['cvesa']] [['clohutier'], ['clothier'], ['clotier'], ['clouhtier'], …]
corrupción	[['corropcion'], ['corrucion'], ['corrupccion'], ['corrupcción'], …]
delincuentes	[['delicuentes'], ['delincuendades'], ['delincuentaso'], ['lelincuentes']]
jalife	[['galife'], ['galyfe'], ['halife'], ['jalif'], ['jalifes'], ['jaliffe'], ['jaliife']]
lilly	[['alili'], ['lile'], ['lili'], ['lilf'], ['lilia'], ['lili'], ['liliy'], ['lilli'], ['lilliy'], …]
lópez	[['lopes'], ['lopez'], ['lopez']]
méxico	[['amexico'], ['dméxico'], ['mejico'], ['mex'], ['mexi'], ['mexico'], …]
obrador	[['hobrador'], ['obradorrrr'], ['obradorsete'], ['obrafor']]

tokenizing it. Additionally, lemmatization and stemming techniques were applied to the comments.

Stemming aims to reduce words to their base or root form, simplifying them and thereby reducing the dimension of features. Unlike stemming, lemmatization considers the context and grammar of words.

3.3 Feature Extraction

The process of converting raw data into a machine-understandable format (numbers) is called feature engineering. The performance and accuracy of machine learning and deep learning algorithms fundamentally depend on this process.

TF-IDF, a renowned technique in text mining and information retrieval, quantifies the importance of specific terms within a corpus, contextualized by their frequency in individual documents and rarity across the entire dataset. This technique is bifurcated into two primary components: Term Frequency (TF) and Inverse Document Frequency (IDF).

Term Frequency (TF) computes the recurrence of a term within a single document, positing that the relevance of the term augments proportionally with its frequency of occurrence.

Inverse Document Frequency (IDF), complementing TF, ascertains the scarcity of a term across the corpus, assigning more weight to terms that provide higher discriminatory power due to their infrequency. The mathematical formulation of IDF mitigates the prominence of terms that are ubiquitous across documents, thus offering a balanced view of term significance [15]. Int this work TF-IDF was applied for feature extraction on the machine learning algorithm and word embedding for the Deep Learning algorithm.

3.4 Classification Model

Machine Learning, a subset of Artificial Intelligence, enables predictions, classifications, or clustering using large amounts of data. However, Machine Learning algorithms cannot directly process messages; they require conversion into numerical form using natural language processing techniques [16].

Deep Learning, another branch of Artificial Intelligence, is capable of learning from extensive datasets, leading to breakthroughs in areas such as Image Recognition and Natural Language Processing. The fundamental difference between Machine Learning and Deep Learning lies in their approaches to feature extraction. While Machine Learning models rely on manually extracting features, Deep Learning models utilize automatic feature extraction within deep layers, resulting in more efficient and accurate learning [16].

Various Machine Learning classifiers were utilized to analyze user comments on YouTube channels. Initially, Logistic Regression and Naive Bayes were employed. Subsequently, deep neural networks including Convolutional Neural Networks (CNNs) and Bidirectional Long Short-Term Memory (Bi-LSTM) were explored. Logistic Regression is a supervised learning algorithm used for predicting the probability of event occurrence. It outputs discrete outcomes based on selected independent variables [17].

The Naive Bayes was another algorithm we adopted for classification. It is a probabilistic classifier that uses Bayes theorem under the assumption of independence among predictors.

Despite its inherent assumption of feature independence, which simplifiHes linguistic relationships. In the specific context of comment classification, it evaluates the probability of a comment being categorized as from the “Atypical Te Ve” or “ElChapucero” channel, considering the presence of indicative terms or phrases [18].

LSTM (Long Short-Term Memory) networks address the challenges of learning long-term dependencies, crucial for tasks like text analysis, where understanding sequence and context is essential [19].

Convolutional Neural Networks (CNNs) are highly effective in recognizing patterns directly from image pixels, making them powerful for tasks such as image classification [20]. Bidirectional Long Short-Term Memory (Bi-LSTM) networks enhance understanding by processing data in both forward and backward directions, incorporating information from both past and future contexts [20].

3.5 Evaluation of the Model

To evaluate the results of the binary classification prediction, four metrics were used which are based on true positives (TP), false positives (FP), True Negatives (TN) and false negatives (FN).

The primary measure of a classifier's performance is accuracy, calculated as the ratio of correctly predicted examples to the total number of examples. While accuracy is commonly used, it may not be suitable in scenarios where the classifier's performance differs across different classes. In such cases, a more comprehensive analysis is required, often facilitated by the confusion matrix [21]:

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN}. \tag{1}$$

Precision represents the ratio of correctly predicted positive results to the total number of positive results predicted by the classifier:

$$Precision = \frac{TP}{TP+FP}. \tag{2}$$

Recall indicates the ratio of correctly predicted positive results to the actual number of positive results:

$$Recall = \frac{TP}{TP+FN.} \tag{3}$$

F1-score, the harmonic mean of precision and recall, provides a balanced measure of a classifier's performance across multiple classes:

$$F1-score = \frac{2}{\frac{1}{Precision}+\frac{1}{Recall}} = 2\frac{Precision * Recall}{Precision + Recall}. \tag{4}$$

4 Results

The statistical analysis is carried out prior to the application of the Machine Learning algorithms and is integrated by number of users per channel, number of comments per user, number of words

Table 4. Top 5 users by comment activity.

	User	Number of comments	Channel name
1	eduardo Montiel Barrientos	40	Atypical Te Ve
2	Yuca Teco	35	Atypical Te Ve
3	Angel Chavez	28	Atypical Te Ve
4	Raiana Sg	16	Atypical Te Ve
5	MEDUSA	16	El Chapucero

Table 5. Distribution of comment length by channel and entire corpus

Comment's length	Atypical Te Ve	El Chapucero	Both Channels
1	205	106	311
2	303	243	546
3	320	283	603
4	300	244	544
5	240	245	485
6	183	226	409

Table 6. Total words per channel

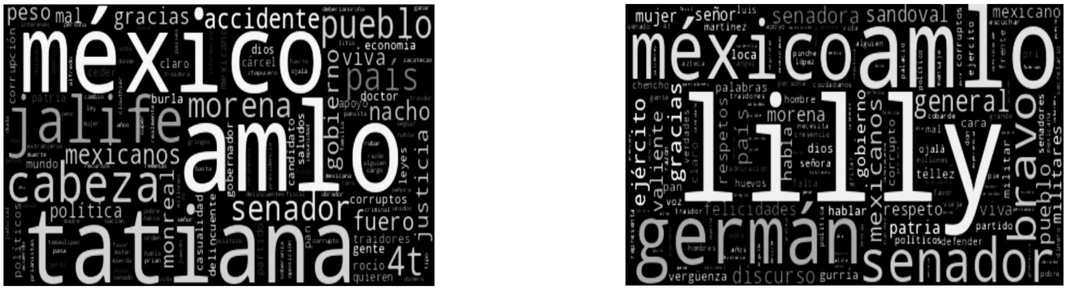
	Comments	Words
El Chapucero Channel	2887	28457
Atypical Te Ve Channel	2887	24596
Total	5774	53053

Table 7. Top 10 most frequent words in both channels

	Atypical Te Ve	El Chapucero	Both Channels
1	lilly	amlo	amlo
2	amlo	mexico	mexico
3	german	tatiana	senador
4	mexico	jalife	tatiana
5	senador	cabeza	lilly
6	bravo	4t	german
7	general	pueblo	mexicanos
8	mexicanos	pais	pais
9	senadora	senador	pueblo
10	gracias	morena	jalife

per comment, number of words per channel, frequent words in the channels and its frequency.

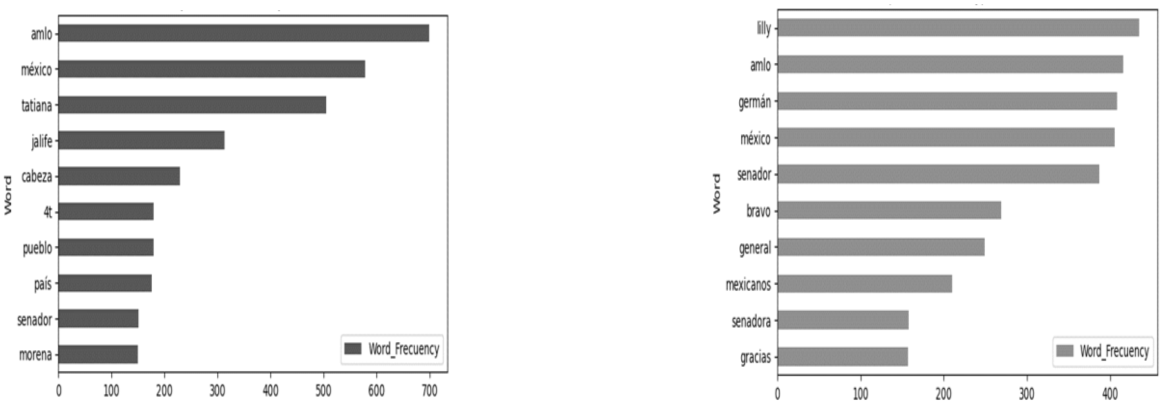
The analysis shows that the 6,000 were issued by 4,469 users, 2,301 affiliated with the “El Chapucero” channel and 2,168 users subscribed to the “Atypical Te Ve” channel, which means that several users published multiple comments, according to the Table 4, also showing the top 5 users with the highest number of comments on the two channels, highlighting the greater activity of followers of the “Atypical Te Ve” channel, suggesting a higher engagement with this channel.



El Chapucero

Atypical Te Ve

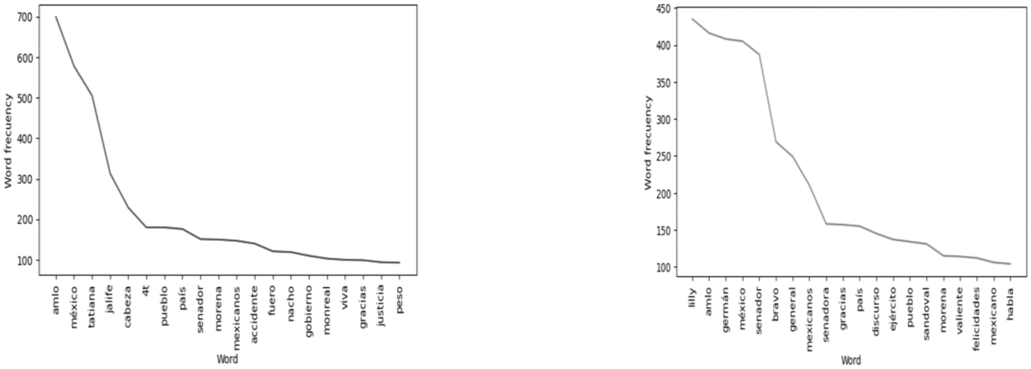
Fig. 2. Word clouds of most frequent words



El Chapucero

Atypical Te Ve

Fig. 3. Bar charts of most frequent words



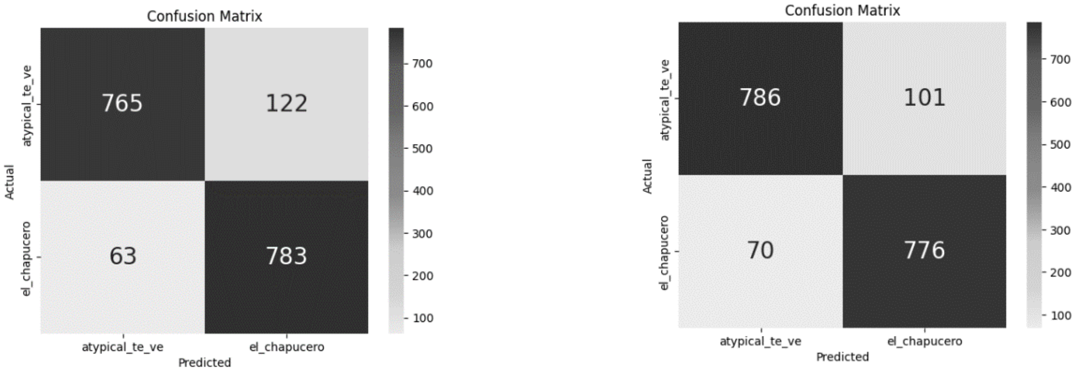
El Chapucero

Atypical Te Ve

Fig. 4. Graphics of most frequent words

Table 8. Performance comparison of classification algorithms

	Channel	Precision	Recall	F1 Score	Support	Accuracy
CNN	El Chapucero	0.96	0.98	0.81	886	0.83
	Atypical Te Ve	0.70	0.56	0.71	847	
Bidirectional LSTM	El Chapucero	0.92	0.87	0.86	886	0.87
	Atypical Te Ve	0.82	0.85	0.87	847	
Logistic Regression	El Chapucero	0.92	0.86	0.89	887	0.89
	Atypical Te Ve	0.87	0.93	0.89	846	
Naive Bayes	El Chapucero	0.92	0.89	0.90	887	0.90
	Atypical Te Ve	0.88	0.92	0.90	846	



Logistic Regression

Naive Bayes

Fig. 5. Confusion matrices for comment classification by logistic regression and Naive Bayes algorithms

The maximum number of comments per user is forty, while the mode is one and the media is 1.3.

According to Table 5, comments with a length equal to three are the most common in both channels and in the entire corpus. Likewise, comments in the third place are also predominant in both channels and in the entire corpus. However, in the “Atypical Te Ve” channel and in the entire corpus, comments of length two occupy the second place, while in “El Chapucero” channel, comments of length five occupy that position.

Although the number of comments is the same on both channels, Table 6 highlights that there is a greater number of words in the comments from “El Chapucero” channel. This is likely attributed to the average words per comment in “El Chapucero” channel being 9.8, with the longest comment reaching 200 words, while in the “Atypical Te Ve” channel, the average is 8.5 words with the longest comment being 111 words. Therefore, we can conclude that on “El Chapucero” channel, comments tend to be longer.

The top 10 most frequent words on both channels are shown in Table 7, and these words are related to political and social topics, reflecting the audience's discussion and interest in certain issues and prominent figures.

While words like 'AMLO' and 'Mexico' are prevalent in both channels, each channel also possesses its distinctive vocabulary. For example, terms such as Lilly Tellez and German Martinez, notable figures on the right in Mexico, and Tatiana Clouthier and Alfredo Jalife, prominent figures on the Mexican left, are specific to each channel.

These unique terms likely mirror the preferred topics and themes of their respective audiences.

The most frequent word in the corpus is "AMLO," occurring 1,115 times.

From Figures 3 and 4 we can see that the words that have a similar frequency in the case of “El Chapucero” are: 4t, pueblo and país, and in the case of “Atypical Te Ve” are: amlo, germán an méxico.

The results of applying Machine Learning and Deep Learning are shown in Table 8, where the performance of four algorithms was compared using accuracy, precision, recall, and F1 score to find the most suitable model architecture for comment classification.

The results show that Bidirectional LSTM, Logistic Regression and Naive Bayes have almost the same precision and that CNN outperforms the deep learning algorithms. Also it is shown that “El Chapucero” has the best Precision in all the classifiers. However, the best accuracy corresponds to Naïve Bayes.

The confusion matrices of the Logistic Regression and Naive Bayes algorithms are shown in Fig. 5. It is observed that the sum of false positives and false negatives is lower when the Naive Bayes algorithm is applied. Then it is concluded that in this case Naive Bayes is better than Logistic Regression.

5 Conclusions

The automatic classification algorithms used in this work are commonly used according to what is reported in the literature, as seen in the article [11] where the best of them reported a precision of 94.57% and in the case of our work the best algorithm obtained a precision of 96%.

It can be concluded that the most frequent words in both channels are related to political topics, public figures, and discussions about the government and Mexican society, but with different focuses. Although there are some common words, such as "AMLO" and "Mexico", each channel has its own set of distinctive words, reflecting the preferences and topics favored by each audience, which has different interests.

The diversity of words suggests that the audience of each channel is interested in a variety of political and social issues and is actively engaged in discussions about them. Lastly, the words used in each channel may reflect political polarization in Mexico and the ideological preferences of each channel's audience, according to the following analysis:

Words and figures such as "AMLO", "Tatiana", "Jalife", and "Morena" are closely related to left-wing political figures and movements in Mexico. The frequent presence of these terms in the comments on the "El Chapucero" channel indicates an interest and support for the Mexican left-wing politics. The absence of terms associated with right-wing politics in Mexico in the list of the most frequent words on the "El Chapucero" channel suggests an ideological bias towards the left in the audience of that channel.

The prominence of specific left-wing terms and the absence of right-wing terms in the comments may indicate a division and polarization in online political and social discourse. In summary, the terms and figures mentioned suggest an ideological inclination towards the left in the audience of the "El Chapucero" channel, which could reflect political polarization in Mexico and preference for certain topics and political leaders within that ideology.

On the other hand, Lilly Téllez and Germán Martínez are prominent political figures in Mexico, primarily affiliated with the right. Both have been known for their critical positions towards the current government and for expressing opinions that align more with the ideology of right-wing politics in the country. In the context of comments on the "Atypical Te Ve" channel, where these political figures are highlighted, users are likely discussing topics related to the opinions and actions of Lilly Téllez and Germán Martínez in relation to the current government, public policies, and other relevant topics for the political situation in Mexico.

Words like "bravo" and "gracias" could be related to the overall tone of comments on the "Atypical Te Ve" channel, which appears to be critical towards the president and left-wing politics. Here are some possible interpretations: The word "bravo" is commonly used to express approval, support, or admiration. In the context of critical comments towards the president or left-wing politics, "bravo" could be used sarcastically or ironically to highlight actions or speeches considered negative or controversial. And the word "gracias" could be used sarcastically or ironically to express gratitude for something negative, such as criticism or a problematic situation.

In the context of critical comments, "gracias" could be used to emphasize dissatisfaction or discontent with certain events or policies. In summary, both "bravo" and "gracias" could be used in a sarcastic or ironic context within critical comments towards the president or left-wing politics. These words could reflect the provocative and controversial nature of online political discourse, especially in environments where controversial topics are discussed, or strong opinions are expressed.

The prominent presence of these figures in the comments suggests that users of the "Atypical Te Ve" channel are interested in political topics and in expressing opinions related to the stances and actions of right-wing politicians like Téllez and Martínez.

The polarization effect generates "echo chambers" in which some individuals are exposed to a single point of view, thus favoring narratives and stories that reinforce the channel's perspective. This environment allows for the creation of a model to train Artificial Intelligence algorithms capable of predicting a user's political tendencies through their comments with remarkable accuracy. This study lays the groundwork for future research on how artificial intelligence, through Natural Language Processing and Machine Learning, reveals the implications of polarization in the Mexican digital environment.

References

1. **YouTube for press (2024).** https://blog.you tube/press/.
2. **Picanço-Rodrigues, V., Leonel-Caetano, M. A. (2023).** The impacts of political activity on fires and deforestation in the Brazilian Amazon rainforest: An analysis of social media and satellite data. Heliyon, Vol. 9, No. 12.
3. **Flamino, J., Galeazzi, A., Feldman, S., Macy, M. W., Cross, B., Zhou, Z., Szymanski, B. K. (2023).** Political polarization of news media and influencers on Twitter in the 2016 and 2020 US presidential elections. Nature Human Behaviour, Vol. 7, No. 6, pp. 904–916. DOI: 10.1038/s41562-023-01550-8.
4. **Magdaci, O., Matalon, Y., Yamin, D. (2022).** Modeling the debate dynamics of political communication in social media networks. Expert Systems with Applications, Vol. 206, DOI: 10.1016/j.eswa.2022.117782.
5. **Statista Homepage (2023).** http://www. statista.com.
6. **Zimmermann, D., Noll, C., Gräßer, L., Hugger, K. U., Braun, L. M., Nowak, T., Kaspar, K. (2020).** Influencers on YouTube: a quantitative study on young people’s use and perception of videos about political and societal topics. Current Psychology, Vol. 41, No. 4, pp. 6808–6824. DOI: 10.1007/s12144-020-01164.7.
7. **Chowdhury, L. H., Islam, S., Shatabda, S. (2024).** A bengali news and public opinion dataset from YouTube. Data in Brief, Vol. 52. DOI: 10.1016/j.dib.2023.109938.
8. **Tran, G. T., Nguyen, L. V., Jung, J. J., Han, J. (2022).** Understanding political polarization based on user activity: a case study in korean political YouTube channels. Sage Open, Vol. 12, No. 2, DOI: 10.1177/21582440221094587.
9. **Bail, C. A., Argyle, L. P., Brown, T. W., Bumpus, J. P., Chen, H., Hunzaker, M. F., Volfovsky, A. (2018).** Exposure to opposing views on social media can increase political polarization. Proceedings of the National Academy of Sciences, Vol. 115, No. 37, pp. 9216–9221.DOI: 10.1073/pnas.1804840115.
10. **Pookpanich, P., Siriborvornratanakul, T. (2024).** Offensive language and hate speech detection using deep learning in football news live streaming chat on YouTube in Thailand. Social Network Analysis and Mining, Vol. 14. No. 1, p. 18. DOI: 10.1007/s13278-023-01183- 9.
11. **Nokkaew, M., Nongpong, K., Yeophantong, T., Ploykitikoon, P., Arjharn, W., Siritaratiwat, A., Surawanitkun, C. (2023).**

Analyzing online public opinion on Thailand-China high-speed train and Laos-China railway mega-projects using advanced machine learning for sentiment analysis. Social Network Analysis and Mining, Vol. 14, No. 1, p. 15. DOI: 10.1007/s13278-023-01168- 8.

12. **Hasan, M., Islam, L., Jahan, I., Meem, S. M., Rahman, R. M. (2023).** Natural language processing and sentiment analysis on bangla social media comments on Russia–Ukraine war using transformers. Vietnam Journal of Computer Science, Vol. 10, No. 03, pp. 329-356. DOI: 0.1142/S2196888823500021.
13. **Hou, R., Han, S., Wang, K., Zhang, C. (2021).** To WeChat or to more chat during learning? The relationship between WeChat and learning from the perspective of university students. Education and Information Technologies, Vol. 26, pp. 1813–1832. DOI:10.1007/s10639-020-10338-6.
14. **Behzadidoost, R., Mahan, F., Izadkhah, H. (2024).** Granular computing-based deep learning for text classification. Information Sciences, Vol. 652. DOI: 10.1016/j.ins.2023. 119746.
15. **Bengesi, S., Oladunni, T., Olusegun, R., Audu, H. (2023).** A machine learning-sentiment analysis on monkeypox outbreak: An extensive dataset to show the polarity of public opinion from Twitter tweets. IEEE Vol. 11, pp. 11811–11826. DOI: 10.1109/ACCESS. 2023.3242290.
16. **Kazbekova, G., Ismagulova, Z., Kemelbekova, Z., Tileubay, S., Baimurzayev, B., Bazarbayeva, A. (2023).** Offensive language detection on online social networks using hybrid deep learning architecture. International Journal of Advanced Computer Science & Applications, Vol. 14, No. 11. DOI: 10.14569/IJACSA.2023.0141180.
17. **Karwa, R. R., Gupta, S. R. (2022).** Automated hybrid deep neural network model for fake news identification and classification in social networks. Journal of Integrated Science and Technology, Vol. 10, No. 2, pp. 110–119.
18. **Kumar, A., Sachdeva, N. (2022).** Multi-input integrative learning using deep neural networks and transfer learning for cyberbullying detection in real-time code-mix data. Multimedia systems, Vol. 28, No. 6, pp. 2027–2041. DOI: 10.1007/s00530-020-00672- 7.
19. **Igual, L., Seguí, S. (2017).** Introduction to data science. Berlin, Germany: Springer.
20. **Chatterjee, S., Krystyanczuk, M. (2017).** Python social media analytics. Packt Publishing Ltd.

Article received on 20/05/2024; accepted on 04/08/2024.
**Corresponding author is Alba Maribel Sánchez-Gálvez.*