

Identification of emotions on Twitter during the 2022 electoral process in Colombia

Juan Jose Iguaran Fernandez¹, Juan Manuel Perez², and Germán Rosati³

¹ Maestría en Data Mining, Universidad de Buenos Aires (UBA), Argentina
juanjose_if3@hotmail.com

² Instituto de Ciencias de la Computación, CONICET, Universidad de Buenos Aires,
Buenos Aires, Argentina
jmperez@dc.uba.ar

³ CONICET. Escuela IDAES, Universidad Nacional de San Martín, Argentina
grosati@unsam.edu.ar

Abstract. The study of Twitter as a means for analyzing social phenomena has gained interest in recent years due to the availability of large amounts of data in a relatively spontaneous environment. Within opinion-mining tasks, emotion detection is specially relevant, as it allows for the identification of people’s subjective responses to different social events in a more granular way than traditional sentiment analysis based on polarity. In the particular case of political events, the analysis of emotions in social networks can provide valuable information on the perception of candidates, proposals, and other important aspects of the public debate. In spite of this importance, there are few studies on emotion detection in Spanish and, to the best of our knowledge, few resources are public for opinion mining in Colombian Spanish, highlighting the need for generating resources addressing the specific cultural characteristics of this variety.

In this work, we present a small corpus of tweets in Spanish related to the 2022 Colombian presidential elections, manually labeled with emotions using a fine-grained taxonomy. We perform classification experiments using supervised state-of-the-art models (BERT models) and compare them with GPT-3.5 in few-shot learning settings. We make our dataset and code publicly available for research purposes.

Keywords: Emotion Detection · NLP · BERT · LLM

1 Introduction

Twitter is a microblogging platform that can be examined as a forum for opinion analysis on social phenomena, especially in politics, despite its representativeness biases. Previous studies have explored how analyzing discussions on Twitter using NLP tools can reflect subjective perceptions of social phenomena [1], [2], [3]. However, there are few studies in Spanish for political events outside Iberian Spanish and, to the best of our knowledge, no resources are available for Latin American varieties of this language.

The evolution towards the use of neural networks, particularly the Transformer architecture [5], has significantly improved the efficiency and accuracy of opinion mining tools by processing extensive texts and capturing the contextual complexity of language. This is particularly evident in BERT, a Transformer-based model pre-trained on large amounts of text, which is then fine-tuned for a specific task [4].

This study focuses on the analysis of emotions in tweets during the 2022 Colombian presidential elections using pre-trained language models. A total of 1,200 election-related tweets were collected, labeled, and used to fine-tune and evaluate pre-trained models. The dataset is offered as a resource for future research. Additionally, the same dataset was labeled using GPT-3.5, a large language model (LLM) created by OpenAI ⁴ from GPT-3 [7], with which one can interact with through text prompts. In this case, the task requested was the labeling of the tweets. The performance of these two labeling methods was compared.

This paper follows the following structure: Section 2 provides an overview of the most relevant works related to this study. Section 3 details how the tweets were collected and the framework and results of the labeling process. Section 4 describes the classification experiments carried out. Section 5 shows the results that were obtained for both the fine-tuned model and GPT 3.5. Finally, section 6 provides some conclusions regarding this study.

2 Previous works

In this section we describe the most relevant studies to date related opinion mining and machine learning models used for this area. We also describe some of their applications in the context of social networks and social related phenomena.

2.1 Supervised learning for emotions detection

The automatic detection of emotional responses in text has long been a subject of interest. The significance of systems capable of identifying negative or positive opinions using online movie reviews has been highlighted in the past [8]. Additionally, the exploration of the semantic orientation of words to determine the overall polarity of texts has also been studied [9].

To undertake the task of automatically detecting emotional states in text, it is usually necessary to have a labeled dataset from which automation can proceed. The creation of such resources has been undertaken in the past for individual words as well as more complex texts such as news articles or children's stories [10–12]. These resources later serve as training datasets for the task of supervised learning using traditional machine learning algorithms.

⁴ <https://openai.com/>

2.2 Neural networks for text analysis

Traditional techniques for analyzing emotions in text have focused on the relationship between terms and emotional states. Recently, the field has shifted to algorithms capable of capturing contextual relationships, such as neural networks [13].

Initially, the most popular architecture for NLP tasks was Recurrent Neural Networks (RNNs) [22], due to their capacity to retain previous data outputs to predict new inputs. This architecture facilitated context analysis in text as it processes sequences of words. However, its sequential nature made it computationally expensive for handling large datasets. To address this issue, a new architecture called Transformers was introduced [5]. Transformers use a layer defined as self-attention for parallel processing, allowing for the consideration of each word's importance in context. This parallel processing capability enables the handling of large datasets, thus providing the model with sufficient training data to enhance language understanding.

Thanks to the parallelization and ability to retain distant word relationships in Transformers, highly predictive language models like BERT were developed [4]. These models are trained on vast amounts of data, such as Wikipedia, to be pre-trained and then use contextual language representations for various NLP tasks by fine-tuning on specific tasks. For Spanish, a dedicated BERT model trained with diverse texts has been proposed [6], which outperforms multilingual BERT in Spanish evaluations.

2.3 Sentiment and emotion on social networks

The analysis of sentiment in online text, particularly on blogging platforms like Twitter, is crucial across various sectors including advertising, finance, and academia [15]. Such platforms provide significant data for sentiment analysis [16], [14]. In relation to social phenomena, Twitter sentiment has been found to correlate with opinion polls [1] and significant events [17]. In politics, Twitter sentiment often reflects public perception [2].

S. M. Mohammad et al. extensively analyzed emotional content in tweets during the 2012 US presidential elections [3]. In their study, they developed a manual for labeling emotions which was provided to multiple labelers. These labelers were asked questions about the emotional content of a tweet and to select the emotion that best fit. Based on the resulting dataset, groups containing several emotions were created to provide a more robust dataset. From this dataset, a machine learning model was trained to identify emotions within tweets. This work has served as a primary influence on the present study.

Regarding Spanish resources for sentiment analysis and opinion mining in general, many resources have been created, particularly as part of workshops such as SemeEval and IberLEF. For instance, EmoEvent [24] is a dataset for emotion detection in Spanish tweets related to several, distinct events. TASS [25] is a sentiment analysis dataset created in the shared task of the same name. To the best of our knowledge, there are no resources for emotion detection in Colombian Spanish tweets.

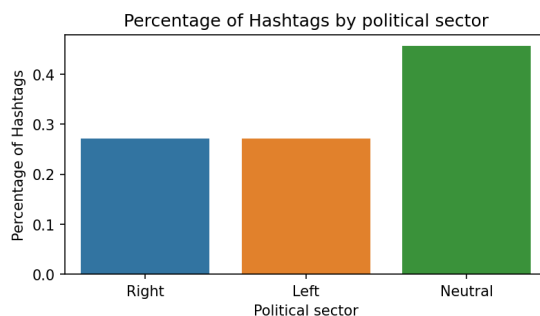


Fig. 1: Percentage of Hashtags According to Assigned Political Orientation

3 Data

This chapter discusses the data collection process carried out between the two electoral rounds, focusing on the utilization of political hashtags. It begins with an explanation of the filtering process aimed at retaining only relevant tweets, followed by an exploratory analysis of the data. Furthermore, it describes the labeling process, including the establishment of correlations between labels to form groups, and discusses the level of agreement among labelers.

3.1 Data gathering

The initial dataset comprises 585,001 tweets collected between May 22nd and June 22nd, 2022, during the presidential elections in Colombia. These tweets were extracted using 173 political trends, i.e., hashtags per day, obtained from websites that store historical trends^{5 6 7}. After a filtering process, which removed tweets with fewer than 5 words, those with a proportion of mentions or hashtags exceeding 20

The hashtags were classified as Left, Right, or Neutral based on their content and perceived political trend. It is worth mentioning that even if a hashtag was associated with a particular political trend, it could still represent a different point of view. The distribution of hashtags was as follows: Neutral (40

Temporal analysis reveals that certain dates, such as May 24th (the date of a debate), May 29th (the first round of elections), June 9th (an event known as the "Petro videos" where a leaked video regarding political strategy was made public), and dates around June 19th (the second round of elections), experienced peaks of activity. These peaks were observed across all three political sectors, as illustrated in Figure 2.

⁵ <https://getdaytrends.com/>

⁶ <https://archive.twitter-trending.com/>

⁷ <https://www.exportdata.io/trends/worldwide>

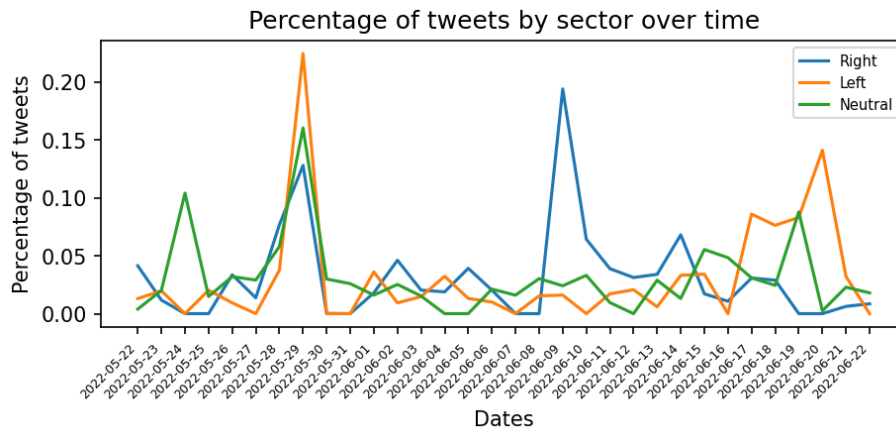


Fig. 2: Percentage of tweets according to political orientation over time

3.2 Labeling

A set of 1200 tweets was selected using stratified random sampling to preserve the original proportions of hashtags. This set was then subjected to manual classification by each of the authors, with labels corresponding to emotions identified in the tweets. Emotion, as defined by the APA [18], involves a complex reaction encompassing experiential, behavioral, and physiological elements. However, in this study, the focus was solely on the authors’ expressive responses to labeling emotions, as subjective and physiological components were not accessible.

To establish a satisfactory labeling framework, an iterative process was developed following the methodology of Mohammad et al. [3]. A manual was created to outline the task and describe the possible labels. Subsequently, the authors labeled some tweets, measured the agreement of those labels, and engaged in discussions regarding the task’s execution. This iterative process continued until the labeling and the output were deemed satisfactory. Details of this process are presented in Figure 3.

This process ultimately led to the development of a labeling interface using the web platform Label Studio ⁸. The interface enabled the assignment of one or multiple emotions to each tweet via a multiple-choice scheme featuring 14 emotions and an "Other" category. Prior to labeling, participants were prompted to determine if the tweet contained emotional content. The resulting interface is illustrated in Figure 4.

After labeling, a database containing the assigned labels for each tweet by each author was generated. This database was then utilized to calculate the correlation between labels assigned by the authors. It was observed that labelers assigned semantically similar labels to certain tweets, as depicted in Figure 5.

⁸ <https://labelstud.io/>

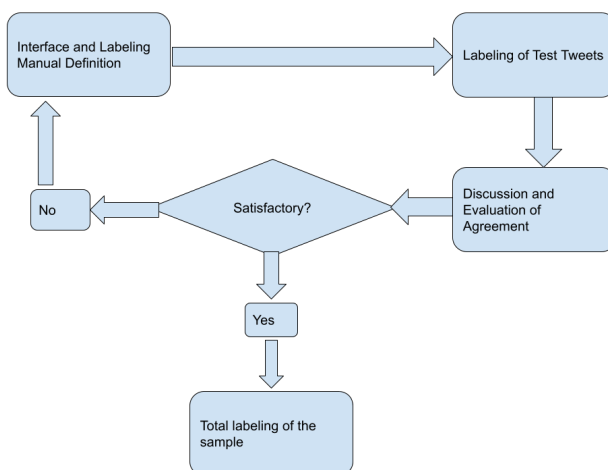


Fig. 3: Annotation workflow

Based on these results, labels were grouped into four categories based on their correlation: Joy, Fear, Sadness, and Disgust. Final labels were assigned if at least two labelers assigned some of the granular emotions that make up a specific group. The Fleiss Kappa index was then used to measure agreement among labelers, with results shown in Table 1.

Joy and Disgust exhibited higher scores compared to Sadness and Fear. It is noteworthy that the emotions with better performance also had a higher number of tweets. Additionally, during the labeling process, overlap was observed between fear and disgust, as well as between sadness and disgust, leading to instances where only one labeler identified one of these emotions, resulting in disagreement. The final number of tweets labeled for each emotion, as well as the overlap between them, is illustrated in Figure 6. The final dataset containing the labels assigned for each tweet is available for the community⁹.

⁹ https://huggingface.co/datasets/jjiguaran/tweets_emotions_elections_colombia

Table 1: Fleiss Kappa index for each emotion

	alegria	miedo	tristeza	asco
Number of Tweets	464	98	103	580
Fleiss k index	0.69	0.47	0.4	0.62

En mi casa esta prohibido hablar de política, por múltiples diferencias, pero por aquí, QUE VIVA PETRO!!!
#PetroPresidenteEnPrimera #PetroEsElCambio

¿Es relevante el presente tweet para el analisis de emociones durante la campaña presidencial?

Sí^[1] No^[2]

¿Existe contenido emocional en este tweet?

Sí^[3] No^[4]

¿Esta el texto escrito de manera ironica?

Sí^[5] No^[6]

¿Que emocion describe mejor el texto?

Admiración^[7] Asombro^[8] Ira^[9] Decepción^[10] Asco^[11] Desagrado^[12] Miedo^[13] Odio^[14] Incertidumbre^[15] Alegria^[16]

Agrado^[17] Tristeza^[18] Sorpresa^[19] Confianza^[20] Otra^[21]

¿Si en la pregunta anterior eligio Otra, con que emocion describiría mejor el tweet?

Fig. 4: Labeling interface

4 Classification experiments

In this section, two approaches for classifying the dataset are explained: pre-trained models and the Large Language Model (LLM), detailing their training process and evaluation metrics.

4.1 Pre-trained models fine-tuning

Pretrained language models are accessible on the Hugging Face platform ¹⁰, utilizing the Transformers library ¹¹. For fine-tuning, three pre-trained language models were chosen: RoBERTuito [19], BETO [20], and RoBERTa [21]. These models were selected because they were specifically trained for Spanish, and in the case of RoBERTuito, for the Twitter context.

These models underwent training and evaluation using K-fold cross-validation, where the dataset was partitioned into train and test sets k times. The model was trained on the train partition and evaluated on the test set. In this case, k was set to 5.

The performance of the model was assessed using the F1 score metric, defined as follows:

¹⁰ <https://huggingface.co/>

¹¹ <https://huggingface.co/docs/transformers/index>

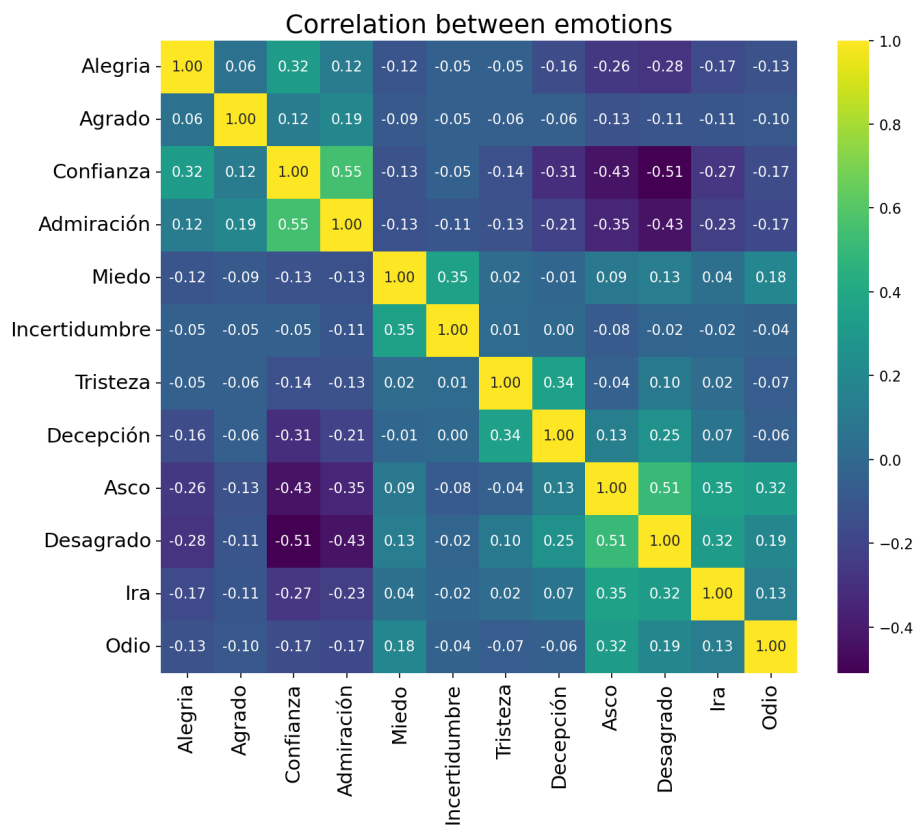


Fig. 5: Correlation index between emotion labels assigned to the tweets

$$F1 = \frac{TP}{TP + \frac{1}{2}(FP + FN)} \quad (1)$$

Where TP represents the number of true positives, FP stands for the number of false positives, and FN indicates the number of false negatives across all classes. Similarly, the Micro F1 score is calculated, but it considers all present classes. This metric was chosen because the model allows for multiple classifications, enabling the simultaneous evaluation of the performance of different classes.

The main hyperparameters employed include the implementation of AdamW, an optimization algorithm proposed by [23]. A learning rate of 5e-05 was utilized. Additionally, 3 training epochs were conducted on the training set, with a batch size of 8.

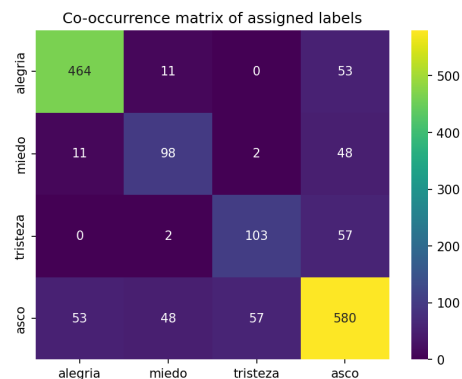


Fig. 6: Co-occurrence matrix between assigned labels

4.2 Few-shot learning

The few-shot approach was conducted using OpenAI’s API ¹² to establish a connection to their GPT-3.5 model. Once connected, a prompt is submitted to the model, defining a request. In this case, the prompt requested the model to label the provided tweet with one of the four emotions and to label it as "other" if it did not match any of them. A brief description of the emotion was included within the prompt, along with a few examples of tweets already classified with their respective labels. The model was also asked to provide an explanation of how it arrived at the given label.

The prompt we used was as follows:

Asigna una o varias de siguientes las emociones segun correspondan al tweet dentro de la siguiente lista:

- Alegria: Positividad, entusiasmo, apoyo, confianza, celebracion o gratitud.
- Tristeza: El tweet expresa emociones de dolor emocional, desanimo, decepcion o pesar
- Asco: Expresiones intensas de aversion, desprecio o ataques negativos.
- Miedo: Sensacion de amenaza, inquietud, incertidumbre o ansiedad

Los tweets a analizar estan en el marco de las elecciones colombianas del año 2022. Si un tweet no encaja con ninguna emociion descrita o el contenido emocional no es suficientemente evidente, etiquetalo como "Otra".

Describe paso a paso el razonamiento que tuviste para llegar a esas emociones.

¹² <https://openai.com/blog/openai-api>

Table 2: Performance metrics of models

Model	Joy	Disgust	Fear	Sadness	Micro F1
RoBERTa	81.2	82.5	28.1	35.6	73.4
BETO	78.1	77.8	31.3	32.4	68.3
RoBERTuito	82.0	80.5	35.8	40.6	72.9
GPT-3.5	84.1	78.8	52.0	45.5	75.2

5 Results

Table 2 displays the results obtained by each model evaluated in the dataset. It shows that, among the supervised algorithms, the best-performing models are RoBERTa and RoBERTuito, both of which exhibit higher performance for joy and disgust compared to fear and sadness. This aligns with the agreement scores observed during the labeling process, indicating that the overlap between disgust and fear/sadness may pose a challenge for the model.

We can observe that in the case of GPT-3.5 the best-performing labels were joy and disgust, while fear and sadness performed the worst. However, it is noticeable that GPT’s performance is considerably better than that of the trained models for these emotions. The performance for joy is slightly better and for disgust slightly worse compared to RoBERTa and RoBERTuito. This demonstrates that the Large Language Model (LLM) seems to perform better when distinguishing between negative emotions, and is quite better for underrepresented emotions such as fear and sadness.

To have a better understanding of the weak points of GPT-3.5, we conducted an error analysis. Table 3 shows some examples of the errors made by the model. While this task is highly subjective and—in some cases—disagreement might not necessarily mean an error, in some cases the predictions show a clear misunderstanding of the tweet. For this specific task, we can observe that, in some cases, the model is struggling with some aspects of ironic/sarcastic tweets or not identifying the use of some slang words or expressions. In other selected examples, GPT-3.5 fails to detect mixed emotions in tweets (e.g. both disgust and joy).

6 Conclusions

This study aimed to compare the performance of fine-tuned pre-trained language models, namely RoBERTuito, RoBERTa, and BETO, against the Large Language Model (LLM) GPT-3.5 through few-shot learning in identifying emotions present in tweets related to the 2022 presidential elections in Colombia.

This comparison was facilitated by the creation of a dataset of 1200 tweets manually labeled by the authors. The labeling task utilized a web interface and followed an internal manual, allowing the assignment of one or several of the 14 available emotions to each tweet. To assign an emotion to a tweet, at least two annotators had to agree on it. Finally, each tweet was classified with one of the

Text	Labels	CoT Prediction
Irony not detected		
@lvelez @lafm an alternative to exercise the right to be informed as voters, don't hide the old man, don't feel sorry for them. @petrogustavo is ready.	Disgust	The author expresses his support towards the idea of a mandatory debate to exercise the right to be informed and criticizes the attitude of hiding the older candidate. The final response is: joy.
From very early in the morning supporting Marbelle in her dream of leaving the country.	Disgust	The author expresses his support and enthusiasm for Marbelle in her decision to leave the country, which reflects positivity and celebration. The final response is: joy.
As it is already known that the Retrograde Rodolfo Hernandez is going to win by order of the great mafia elites of Colombia, so let the great mafia elites of Colombia tremble, because he, Rodolfo, is going to put an end to CORRUPTION	Disgust	The author expresses his confidence in the victory of Rodolfo Hernandez and his contempt for the mafia elites, and also shows support for the fight against corruption. The final answer is: joy, disgust.
Missing mixed emotion		
It is not objective. It seems to me that the moderators are biased. I expected a little more from @NoticiasCaracol but well...What else is expected from a private Colombian media.	Sadness, Disgust	The author expresses his dissatisfaction with what he considers a biased debate, showing his disappointment and distrust towards the moderators and the Colombian private media. The final answer is: sadness.
EiQueLosDELATA Tomorrow we will go to the polls to defeat this mafia!	Joy, disgust	The author expresses enthusiasm and support towards the next day's election day, with the goal of defeating those he considers a gang of mafiosi. The final response is: joy.

Table 3: Error analysis for GPT-3.5 Chain-of-Thought (CoT) predictions. The table shows some examples where the model fails to detect irony or mixed emotions.

four labels resulting from grouping the original labels based on their correlation. It is important to note the inherent difficulty of the labeling task, as it aimed to achieve an objective classification of a subjective activity, necessitating an iterative process to develop a manual and a labeling interface that approached this purpose. Additionally, although all annotators are native Spanish speakers, only the author is originally from Colombia, which led to certain language usages or specific contextual situations being clearer to him than to the other annotators.

The results of the models revealed a much greater presence of joy and disgust than fear and sadness, consistent with observations during the labeling process, where these two emotions were less frequent and often accompanied by the emotion of disgust. These factors also explain the relatively lower predictive capacity for less prevalent emotions compared to joy and disgust.

It is noteworthy that GPT-3.5 significantly outperformed the fine-tuned models in the emotions of fear and sadness. This demonstrates the LLM's capability to perform as well as fine-tuned models for the most prevalent emotions, as well as its ability to excel in distinguishing negative emotions. Also, GPT-3.5 seems to be struggling with some aspects of irony and sarcastic tweets.

7 Limitations

One of the main limitations of this study is the small size of the dataset and the short period of time in which the tweets were collected. This has led to a lack of diversity in the data, which is somehow reflected in the distribution of perceived emotions. Additionally, the labeling process was performed by native Spanish speakers, but only one of them was Colombian, which could have led to some misunderstandings in the interpretation of the tweets.

Finally, the comparison between the fine-tuned models and GPT-3.5 was not exhaustive, as the latter was only used in few-shot learning without trying several prompts. This might have led to suboptimal performance for the LLMs.

References

1. Brendan O'Connor, Ramnath Balasubramanian, Bryan R. Routledge, Noah A. Smith, "From tweets to polls: Linking text sentiment to public opinion time series," in *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*, 2010.
2. Andranik Tumasjan, Timm Sprenger, Philipp Sandner, Isabell Welp, "Predicting elections with twitter: What 140 characters reveal about political sentiment," in *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 4, no. 1, pp. 178–185, 2010.
3. Saif M. Mohammad, Xiaodan Zhu, Svetlana Kiritchenko, Joel Martin, "Sentiment, emotion, purpose, and style in electoral tweets," *Information Processing & Management*, vol. 51, no. 4, pp. 480–499, 2015, Elsevier.
4. Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," *arXiv preprint arXiv:1810.04805*, 2018.

5. Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, Illia Polosukhin, “Attention is All You Need,” *Advances in Neural Information Processing Systems*, vol. 30, 2017.
6. José Cañete, Gabriel Chaperon, Rodrigo Fuentes, Jou-Hui Ho, Hojin Kang, Jorge Pérez, “Spanish pre-trained BERT model and evaluation data,” *PML4DC at ICLR*, vol. 2020, pp. 1–10, 2020.
7. Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D. Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, y otros, “Language models are few-shot learners,” *Advances in Neural Information Processing Systems*, vol. 33, pp. 1877–1901, 2020.
8. Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. *Thumbs up? Sentiment classification using machine learning techniques*. arXiv preprint cs/0205070, 2002.
9. Peter D. Turney. *Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews*. arXiv preprint cs/0212032, 2002.
10. Carlo Strapparava, Alessandro Valitutti, and others. *Wordnet affect: an affective extension of wordnet*. In *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC)*, volume 4, pages 1083–1086, 2004. Lisbon, Portugal.
11. Janyce Wiebe, Theresa Wilson, and Claire Cardie. *Annotating expressions of opinions and emotions in language*. *Language Resources and Evaluation*, 39(2-3):165–210, 2005. Publisher: Springer.
12. Cecilia Ovesdotter Alm and Dan Roth. *Emotions from text: machine learning for text-based emotion prediction*. In *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pages 579–586, 2005. Publisher: Association for Computational Linguistics.
13. Emmanuel Acheampong, Rezwanaul Haque, and Rifat Khondoker. *A review of transformer architectures for natural language processing tasks*. arXiv preprint arXiv:2101.01069, 2021.
14. Alexander Pak and Patrick Paroubek. “Twitter as a corpus for sentiment analysis and opinion mining,” In *Proceedings of the 7th conference on International Language Resources and Evaluation (LREC 2010)*, pages 1320–1326, 2010.
15. Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. “Opinion mining and sentiment analysis,” *Foundations and Trends in Information Retrieval*, vol. 2, no. 1-2, pp. 1–135, 2008.
16. Saghar Aman and Stan Szpakowicz. “Identifying expressions of emotion in text,” In *Proceedings of the AAAI Fall Symposium on AI and Consciousness: Theoretical Foundations and Current Approaches*, pp. 1–4, 2007.
17. Johan Bollen, Huina Mao, and Xiao-Jun Zeng. “Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena,” *arXiv preprint arXiv:1103.0028*, 2011.
18. G. R. VandenBos, *APA Dictionary of Psychology*, American Psychological Association, 2007.
19. Juan Manuel Pérez, Damián Ariel Furman, Laura Alonso Alemany, and Franco M. Luque. *RoBERTuito: a pre-trained language model for social media text in Spanish*. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 7235–7243, Marseille, France, June 2022. European Language Resources Association. <https://aclanthology.org/2022.lrec-1.785>
20. José Cañete, Gabriel Chaperon, Rodrigo Fuentes, Jou-Hui Ho, Hojin Kang, and Jorge Pérez. *Spanish Pre-Trained BERT Model and Evaluation Data*. In *PML4DC at ICLR 2020*, 2020.

21. Asier Gutiérrez Fandiño, Jordi Armengol Estapé, Marc Pàmies, Joan Llop Palao, Joaquin Silveira Ocampo, Casimiro Pio Carrino, Carme Armentano Oller, Carlos Rodriguez Penagos, Aitor Gonzalez Agirre, and Marta Villegas. *MarIA: Spanish Language Models*. Procesamiento del Lenguaje Natural, 68, Sociedad Española para el Procesamiento del Lenguaje Natural, 2022. DOI: 10.26342/2022-68-3. <https://upcommons.upc.edu/handle/2117/367156/#.YyMTB4X9A-0.mendeley>
22. Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, Yoshua Bengio. “Learning phrase representations using RNN encoder-decoder for statistical machine translation.” *arXiv preprint arXiv:1406.1078* (2014).
23. Ilya Loshchilov and Frank Hutter. “Decoupled weight decay regularization.” *arXiv preprint arXiv:1711.05101* (2017).
24. Plaza del Arco, Flor Miriam and Strapparava, Carlo and Urena Lopez, L. Alfonso and Martin, Maite “EmoEvent: A Multilingual Emotion Corpus based on different Events” *Proceedings of the 12th Language Resources and Evaluation Conference*, 2020.
25. Garcia-Vega, Manuel and Garcia-Cumbreras, Miguel Angel and Plaza del Arco, Flor Miriam and Urena Lopez, L. Alfonso and Montejo-Raez, Arturo “Overview of TASS 2020: Introducing Emotion Detection” *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2020)*, 2020.