Attention-based CNN-Bi-LSTM for the Sentiment Analysis of 2022 Philippine Post-Election Tweets

Abigail C. Nadua and Marie Betel B. de Robles

***Abstract*— Twitter is a popular social media platform for users to freely express their thoughts and opinions, making it a big source of data for companies and researchers to gather user sentiments. In the recent 2022 Philippine Elections, many users also turned to Twitter to express their support towards their preferred candidates. This election was mostly focused on the top two presidential candidates, the current president Bongbong Marcos, and the former vice president Leni Robredo. This study gathered and labeled post-election tweets into positive, negative, and neutral for both Bongbong Marcos and Leni Robredo. These tweets were processed and trained using an Attention-based CNN-Bi-LSTM model with pre-trained GloVe word embeddings. The model had the best performance out of all the other baseline models on average, with an 80.86% accuracy, 81.47% precision, 80.03% recall, and 80.74% f-score. Contrary to the results of the elections, only 13.7% of the tweets expressed positive sentiments towards Bongbong Marcos, 20.4% had negative sentiments, and the rest were neutral (65.9%). Meanwhile, 44% of the tweets were positive sentiments towards Leni Robredo, only 8.0% were negative, and the remaining 48% were neutral.**

***Index Terms*— sentiment analysis, tweets, Twitter, elections, attention, cnn, bi-lstm, glove, deep learning**

I. Introduction

A. Background of the Study

The 2022 Philippine Elections that took place on May 9 had been talked about by many people, not only by Filipinos but also by people all around the world, sharing and expressing their thoughts through social media. Since the beginning of the campaign period, voters and non-voters alike had been mostly talking about the top two presidential candidates, the current President, Pres. Ferdinand Marcos Jr. (Bongbong Marcos), former senator and the son of the late President Ferdinand Marcos Sr., and Atty. Maria Leonor Gerona Robredo (Leni Robredo), the Vice President of the Philippines from the previous administration. This election had been considered by analysts as one of the most significant elections in South East Asian history [1]. Some of the reasons stated were that Bongbong Marcos (BBM), the son of the late dictator who was ousted during the 1986 People Power revolution, had filed for candidacy as a president and was a consistent frontrunner in election surveys. Another reason was Leni Robredo (Leni), a constant second-placer in surveys and the head of the opposition at the time. She had openly expressed her concerns regarding another Marcos presidency, and was able to group people with the same concerns together.

Presented to the Faculty of the Institute of Computer Science, University of the Philippines Los Bañosin partial fulfillment of the requirements for the Degree of Bachelor of Science in Computer Science

Prior to the elections, based on the final Pulse Asia Survey conducted on the 16th to 21st of April 2022 with 2,400 respondents, BBM remained in an overwhelming lead with 56% of likely voters with valid responses, while Leni had 23% [2]. True to the surveys, the election resulted in a landslide victory by BBM, who secured over 31 million votes, compared to the total of the candidate with the second highest votes, Leni, with only over 15 million total votes. Following these results, many people had turned to social media to share their sentiments.

At the beginning of 2022, the Philippines has been reported to have 82.4% of its total population use social media sites like Facebook, Twitter, and Instagram [3]. Social media has been used by users to express their sentiments or opinions like sharing news, discussing about major events and political campaigns, and promoting social and development works [4]. These are done through making reviews, comments, and posts about different topics, resulting in a large amount of data that can be further analyzed for research [5]. These opinions from many kinds of users are analyzed using sentiment analysis [6]. Sentiment Analysis (SA) is a field of study under natural language processing (NLP) which focuses on analyzing people's opinions and sentiments, among others, about a specific entity or topic [7].

The two traditional and most common approaches for sentiment analysis are the lexicon-based and machine learning approaches [8]. Lexicon-based approaches use pre-compiled sentiment lexicons containing different words and their corresponding polarity in order to classify a word in a set of data to either positive or negative. This approach doesn't require a dataset for training because word classification is based on the lexicon used. Machine learning approaches are based on machine learning algorithms to label data. As opposed to lexicon-based approaches, machine learning approaches require a training dataset to automate classification and labeling of data. There have also been studies that perform sentiment analysis by using a hybrid of both approaches, or different methods of the same approach.

Another approach is deep learning, a sub-field of machine learning consisting of neural networks [9]. A neural network is based on the neurons of a human brain, and it is used in a lot of applications such as text generation, vector representation, among others [10]. It includes networks like the Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Deep Neural Networks (DNN), and many others. Through various studies, deep learning approaches have been proved to perform better than the traditional ones [8], [11].

In this study, sentiment analysis was performed to post-election tweets using an Attention-based CNN with Bidirectional Long Short-Term Memory (Bi-LSTM) model. Bi-LSTM is an extension of LSTM, which is a special type of RNN. The tweets from after the elections would also be compared to the results of the actual elections.

B. Statement of the Problem

Before the elections, Pulse Asia Research Inc. (Pulse Asia), the public opinion polling body in the Philippines, had been conducting surveys about the Filipinos' presidential bets monthly. However, the surveys conducted by Pulse Asia were criticized to be flawed in terms of representation of the general population [12]. Dr. Peter Cayton from the University of the Philippines (UP) also believed that there were certain sectors that were over- and underrepresented [13]. Although Pulse Asia defended that they were baseless claims [14], there was still a need to provide a new and unbiased dataset telling the public's opinions and sentiments before the elections, and gain a new perspective on public sentiments regarding the event.

On the other hand, there had been many issues that were shown in the news during and after the elections. For instance, thousands of vote counting machines (VCMs) had malfunctioned from different areas during the election which caused massive delays and voters had to stay until the VCMs are fixed or replaced to protect their votes [15]. Despite these issues, this election had the fastest result transmission compared to previous elections, which people, including the political analyst Ramon Casiple, found surprising [16], [17]. Some people found it suspicious and claimed that the election was rigged, which the Commission on Elections (COMELEC) shrugged off because it would be difficult to prove [18]. Moreover, Séverine De Laveleye, a Belgian parliamentarian, also commented that the election was "marred by a higher level of failure of the electronic voting system than ever before", and that there were rampant vote buying, red tagging, and incidents of violence [19], [20].

Comparing sentiments from another platform to the results of the actual elections would provide a new point of view on what the public really thinks, regardless of who they voted for. Social media may not be able to represent all classes of voters but it could still provide a sample space of what the public thinks [4]. Chaudhry et al [4] also said that social media "provides the nearest approximation of public sentiment". In this study, posts from Twitter, or tweets, were collected and analyzed to represent public sentiments.

C. Objectives

This study generally aims to analyze the sentiments of Twitter users after the 2022 Philippine elections using an attention-based CNN-Bi-LSTM model. Specifically, it aims to:

- label a unique dataset consisting of tweets about the results of the 2022 Philippine Elections
- compare the results of the labeled dataset with the actual elections
- evaluate the performance of the attention-based CNN-Bi-LSTM model in classifying 2022 Philippine Post-Election tweets by calculating its accuracy, precision, recall, and f-score

D. Significance of the Study

This study would help in providing a new labeled dataset consisting of public sentiments taken from the Twitter social media site. The labeled dataset would give a new and unbiased perspective regarding the public's sentiments about the recent 2022 Philippine Elections. It would also help in the investigation regarding the legitimacy of the recent elections.

This would also contribute to the study of attention-based sentiment analysis, which adds weight to words that can greatly influence the classification of a given statement and identifies its most important semantic information [21], by classifying the dataset using CNN and Bi-LSTM.

E. Scope and Limitations

The data that would be gathered in this study came from users' tweets, starting from 11th May to 25th May 2022, a total of 15 days. The tweets from 10th May 2022 would not be included to take voters who weren't able to vote on the actual day of elections into consideration. The tweets were collected on 10th October 2022, so any deleted tweets or tweets from private accounts prior to the time of collection could not be included. News outlets and retweets (RTs) would be removed in the dataset to be analyzed. Tweets from users without a profile picture and accounts created one month before the elections would also be removed to minimize tweets from possible trolls. However, there was no guarantee that all the gathered tweets came from non-troll users. The tweets would be labeled into two categories which would be based on their stance concerning the top two presidential candidates, BBM and Leni. For each category, the tweets were labeled as positive, negative, or neutral. All the tweets were labeled as agreed upon by three annotators for both categories, BBM and Leni. The labels for both categories are mutually-exclusive. A positive label in one category does not mean a negative label for the other, and vice versa.

II. Review of Related Literature

A. Election Analysis

There were various studies performing sentiment analysis regarding the elections from different countries. Most of the purpose of these studies had been classified by Chauhan et al [22] into two categories: election prediction through social media and political stance detection. The first category focuses on predicting the possible results of an upcoming election based on the gathered data, most of it coming from social media. The second one focuses on identifying the users' political stance or voting preference based on their posts in social media platforms or political debates. It can include but are not limited to the users' voting intentions, and whether they are in favor or not about a certain topic, party, or electoral.

One of the studies in the first category is [23], which aimed to predict the results of the 2016 US presidential election based on tweets. The study gathered data regarding the three candidates, Donald Trump, Hillary Clinton, and Bernie Sanders, and classified the tweets into positive, negative, and neutral for each candidate. The study also measured the average polarity and average subjectivity for each candidate. The authors used a lexicon-based approach to classify the data. The sentiment of each tweet was determined by the average polarity of each meaningful word present in the tweet (1 for positive, -1 for negative, 0 for neutral). The results showed that Hillary had the most number of positive tweets which gave her the best average polarity score. On the other hand, Bernie had the best average subjectivity score.

Budiharto & Meiliana from the study [24] also collected data from Twitter, focusing on tweets about the two presidential candidates of the 2019 Indonesian elections, Jokowi and Prabowo, in order to predict the results using sentiment analysis. The authors got the public opinion based on the collected hashtags related to the country's presidential elections. The authors gathered data from Twitter and accessed the Twitter API using the R programming language. They also used a lexicon-based approach to classify the tweets. The results of the analysis match with the four survey institutes in Indonesia.

The study of Chaudhry et al [4] on the other hand, aimed to capture the opinions of social media users from every state. The authors analyzed the sentiments from before, during, and after the 2020 US election for each state, labeled them into positive and negative for both Joe Biden and Donald Trump, then compared them with the actual election results. The study used the machine-learning algorithm, Naive Bayes (NB), to classify the tweets. The results matched with the results of the actual 2020 US Elections, with the exception of four outlier states which were Arizona, Wisconsin, Georgia, and Pennsylvania. This study can be considered to be under the second category as defined in the study [22], which is political stance detection.

A study about sentiment analysis regarding Philippine elections, specifically the 2016 Elections is [25]. The study focused on providing a sentiment analysis classifier specializing in mixed-language tweets while also improving the analysis for mainstream media. The authors identified the subjective and objective tweets, comparative tweets, rational and emotional tweets, and its polarity. They used a Multinomial NB and Support Vector Machine (SVM) classifier, which are popular machine learning algorithms for sentiment analysis, to classify the tweets.

The study by [26] is another study that focused on the recent 2022 Philippine elections. The authors annotated 114,851 English and Filipino tweets and performed sentiment analysis on them. The tweets were classified into three classes: positive, negative, and neutral, using a self-trained Multinomial NB classifier. Using 30% of the total data for testing, the classifier yielded an accuracy of 84.33%.

For this study, Twitter sentiments from the period after the 2022 Philippine Elections would be analyzed using sentiment analysis. The tweets would be labeled into positive, negative, and neutral for both Bongbong Marcos and Leni Robredo, similar to the studies [4], [23]. The labeled data would also be compared to the actual results of the elections like the study [4], based on the results of the annotation. This study would be considered under the political stance detection category, aiming to determine whether the public was in favor, not in favor, or neutral, of the two presidential candidates by the end of the elections.

B. Data Gathering

Most studies get their data to be analyzed from social media sites, specifically Twitter. Posts made in Twitter, also known as tweets, are short since it can only contain 280 characters, but it can also be substantial. It is also one of the most widely used social media platform and its users can just post whatever comes to mind [27]. Twitter users can easily express themselves which makes the social media platform suitable for studies that require public sentiments [28]. Data collection is also easier because Twitter's application programming interface (API) and database is available to the public [28].

In the studies [23], [29], the tweets were retrieved using Python's Twython library, a Python wrapper for the Twitter API. On the other hand, the studies [30]–[32] retrieved the tweets using Python's Tweepy library, which is also used to easily access the Twitter API. There are also studies who used the R programming language like the study [33], which used the twitteR package to extract tweets.

Python was the programming language used for this study because of its excellent capabilities for processing texts, simple syntax, and multiple available libraries. For data gathering, Python's Tweepy library was used because of its rich documentation and active community.

C. Preprocessing Techniques

Preprocessing techniques are equally important in sentiment analysis. Preprocessing is defined in the study [34] as the "procedure of cleansing and preparing texts that are going to be classified". Twitter data are considered to be unstructured and informal, which mostly contain unnecessary characters or texts. This is also called noise, which doesn't contribute any useful information to the data.

The study [34] compared popular preprocessing techniques and identified which ones would give better accuracy. The preprocessing baseline was removing unicode strings and noise. The other techniques would be considered as a necessary step if its own accuracy exceeds that of the baseline's. The study compared each technique and also a combination of the high-accuracy techniques in various algorithms and dataset. In the used Neural Networks algorithm, the combination of the following preprocessing techniques led to the best accuracy: replacing URLs and user mentions, replacing contractions, replacing repetition of punctuation, and lemmatization.

Emojis are also widely used in tweets to express the user's sentiments. An emoji is defined in the study [35] as "a small digital image or icon used to express an idea or emotion". For instance, the emoji 😂 can express happiness while the emoji 😡 can express anger. However, emojis can also represent the opposite meaning of the statement. An example is the statement *Wow, that's great* 😒 . Since the given emoji was used, the positive statement became negative instead.

Different studies have different approaches with regards to handling emojis. Many studies simply removed the emojis [33], [36]. However, prior studies had proven that including emojis into the database and classifiers led to a more accurate performance of up to 70%, compared to excluding emojis which was only about 60% [37].

There were studies that use emoji embeddings like [38], [39]. However, some of the challenges presented in the study [39] with this method was that emojis could be used differently depending on the culture, new emojis would get introduced from time to time, and it was difficult to identify complex emotions like sarcasm, dark humor, etc., to name a few.

In the studies [35], [40], emojis were replaced by their textual descriptions instead. According to the authors of [35], "while emojis are common, the words in their descriptions are more common". There is more data to pre-train for word embeddings than for emoji embeddings. The textual descriptions of various emojis also have a lot of words in common, which makes it easier to represent them for embeddings.

When it comes to handling hashtags (#), many studies also just remove them during the preprocessing [40], [41]. However, there are also studies that include hashtags as a feature. The study [25] included and compiled hashtags with positive connotations for the classification of Philippine 2016 Elections Twitter data. If one of these hashtags was found in the text, then it would be classified as positive.

The study [35] acknowledged the presence of hashtags in tweets and chose to remove the hash symbol expand the words in the hashtag. For instance, the hashtag *#VoteWisely* would become *Vote Wisely* after this process. In the study [42], the hashtag and other special characters like mention tags (@) and asterisks (*) were replaced by their text values instead of removing them.

The preprocessing techniques that would be used in this study were the same as the best techniques described in the study [34]: replacing contractions, replacing repetition of punctuation, and lemmatization, except that URLs and user mentions would be totally removed because it's not relevant for the study to measure the influence of a user through mentions and URLs. Additionally, emojis would be handled the same as the studies [35], [40], where the emojis would be replaced by their textual descriptions. Hashtags would be handled similarly to the studies [35], [42], where the hashtag symbol would be replaced by the word hashtag, and the words would be expanded.

D. Word Embeddings

According to the study [43], Word embedding is an NLP technique used for language modeling and feature learning. It maps each word into a vector where words with similar meanings have similar representation. The same study used word embeddings and TF-IDF and compared them with different deep learning networks (RNN, CNN, and DNN). TF-IDF, or Term Frequency-Inverse Document Frequency, is a statistical measure of how important a word is depending on its frequency in the whole document, and its frequency in the other documents in the collection. The word is considered as high-importance if it has high frequency in its own document, and low frequency in others. In the case of RNN, the use of TF-IDF resulted in a poor performance in all evaluation metrics (accuracy, precision, recall, F-score) compared to the other networks, but using word embeddings showed a much better performance than the others in all evaluation metrics. CNN performed slightly better when using word embeddings as well, while DNN showed an almost similar performance for both.

The popular word embeddings used by many studies are the Word2Vec and GloVe models. Word2Vec is a word embedding tool which includes two methods for learning word representations which are the Continuous Bag-of-Words (CBOW) model and the Skip-gram model [44]. In CBOW, the context words are used to predict the representation of the target word while in Skip-gram, the word representations of the surrounding words in a sentence or document are predicted [45]. Meanwhile, GloVe, or Global Vectors, is an unsupervised method of learning word representations using the statistics of word co-occurences in the whole collection of data (corpus) [46].

The study [45] is a comparative study for different word embedding models. After comparing each word embedding model (CBOW, Skip-gram, GloVe, and Hellinger-PCA) with each other, the authors concluded that GloVe is the best model because of its scalability, and also works well with smaller data. It requires less training time but it does have quadratic computational complexity. For these reasons, the GloVe model would be used for this study and the pre-trained word embeddings would be utilized for efficiency.

E. Attention-based CNN-Bi-LSTM Model

Bidirectional Long Short-Term Memory, or simply Bi-LSTM, is a variation of the RNN. It is also an extension of the standard LSTM, which was first proposed to solve the gradient vanishing problem of RNN. The LSTM model [47] is used to capture long range dependencies in sequences, composed of multiple LSTM cells modeling the memory in a neural network. It also consists of gates for storing and accessing information over time. However, LSTM processes in a temporal sequence, meaning inputs are processed in succession, and ignores the future context. Bi-LSTM networks were created as an extension of the standard unidirectional LSTM where a second layer is added to flow on the opposite direction to utilize both the past and future information [21]. Bi-LSTM is used to learn the representation of sentences, which will be used to classify these sentences [38].

The study [48] proposed a stacked Bi-LSTM model for the sentiment analysis of Chinese microblogs. The model simply stacks two Bi-LSTM layers so there would be two forward layers and two backward layers. With the CBOW word embedding, it resulted to have the best performance compared to the other models used in the study with 89% average accuracy.

The disadvantage of the Bi-LSTM model is that it has weak capabilities on capturing the most important features [49]. This weakness can be solved by CNN, which is good at capturing relevant features based on context features that can improve the classification accuracy. CNN consists of a convolutional layer and a pooling layer [49]. The convolution layer is the "core" of CNN, which consists of convolution kernels to extract features from an input. The pooling layer is for reducing the dimensions of a vector and avoiding overfitting. This results to less number of parameters to learn and less amount of computation performed, which increases efficiency. By maximum pooling (max-pooling), the model can retain key features in the text.

The study [50] compared different deep learning models to perform sentiment analysis on drug reviews with two types of classifications. One was with three classification (positive, negative, neutral), and one with ten classifications. The results for both types showed that the combination of Bi-LSTM and CNN had the best performance. For 3-class, CNN+Bi-LSTM showed the best performance in terms of F1-score. For 10-class, Bi-LSTM+CNN was the best model.

The attention mechanism came from the idea of how humans mostly focus on a certain region of an image or word in one sentence [51]. The visual attention of humans focuses on a part with "high attention", while the surrounding parts are viewed as "low attention". Simply put, the purpose of an attention mechanism is to give more focus to the most relevant information or features, and it is widely applied in Computer Vision, Speech Recognition, Machine Translation, and Image Caption Generation [52].

The study [53] used a bidirectional CNN and RNN model with an attention mechanism for polarity-detection of a document-level sentiment analysis (ABCDM). The model resulted in an average accuracy of around 90% across the eight datasets tested. On the other hand, the study [54] also used an attention-based CNN and Bi-LSTM model for sentiment analysis. This time, the study is based on TF-IDF and GloVe word embedding. It achieved an average accuracy of around 92% across four standard datasets. The study [49] also used an attention-based CNN and Bi-LSTM model with the addition of a gating mechanism to assign weights for the important features captured by the CNN layer and the context features captured by the Bi-LSTM layer, to further improve the performance of the model. The model's best performance achieved an F1-score of 98% on two standard datasets.

This study would also be using a deep learning approach, specifically CNN and Bi-LSTM, to classify the dataset. It would also have an attention mechanism to give more importance to words that could be the deciding factor for the classification.

F. Performance Evaluation

Most studies use accuracy, precision, recall, and F1-score (can also be F-score) as their evaluation metrics [5]. These metrics are computed as:

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

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

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

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

where TP stands for *True Positive*, FP is for *False Positive*, TN is for *True Negative*, and FN is for *False Negative*. True positives are positive data that is correctly classified as positive while false positives are data that are incorrectly classified as positive. The same thing applies for the true and false negatives, but in terms of negative data.

Grandini et al [55] defined accuracy as the overall measure of how much the model correctly predicts the classification of the given data. Meanwhile, they stated that precision measures how much of the predicted data are actually positive (TP) among all those that were predicted to be positive (TP+FP). They also defined recall as the measure of how much of the data were predicted to be positive (TP) compared to the number of data that should be positive (TP+FN). Lastly, they interpreted F1-score as a weighted average between precision and recall, which would be "useful to find the best trade-off between the two".

This study would also be computing for the model's accuracy, precision, recall, and f-score to evaluate its performance. For comparison, the dataset would also be trained using the following models: CNN-Bi-LSTM, Attention-based CNN, Attention-based Bi-LSTM, CNN, and Bi-LSTM. For each model, accuracy, precision, recall, and f-score would be computed and compared to the Attention-based CNN-Bi-LSTM model.

III. Materials and Methods

A summary of processes done in this study is shown in Fig. 1. A device with AMD Ryzen 5 4600H processor and NVIDIA GeForce GTX 1650 GPU was used. The specific details were explained in the following subsections.

A. Data Gathering

A total of 56,000 tweets were collected on 10^{th} October 2022 through the Twitter API using Python's Tweepy library. The study collected tweets from 11^{th} May to 25^{th} May 2022. Search queries and hashtags related to the 2022 Philippine Elections were used such as #Eleksyon2022, #Halalan2022, #KulayRosasAngBukas, and #Uniteam. Gathered tweets from news outlets, retweets, users without a profile picture, and user accounts created within one month prior to the elections were removed. Based on the criteria for the tweets, 47,873 tweets were filtered out and 8,127 tweets remained. The remaining tweets were saved into a CSV file which were then labeled.

B. Data labeling

Three annotators manually labeled the gathered tweets according to two categories: BBM and Leni. In each category, the tweets were labeled as one of the three classifications (positive, negative, and neutral). During labeling, tweets that were out of context and those that do not focus on BBM, Leni, or the

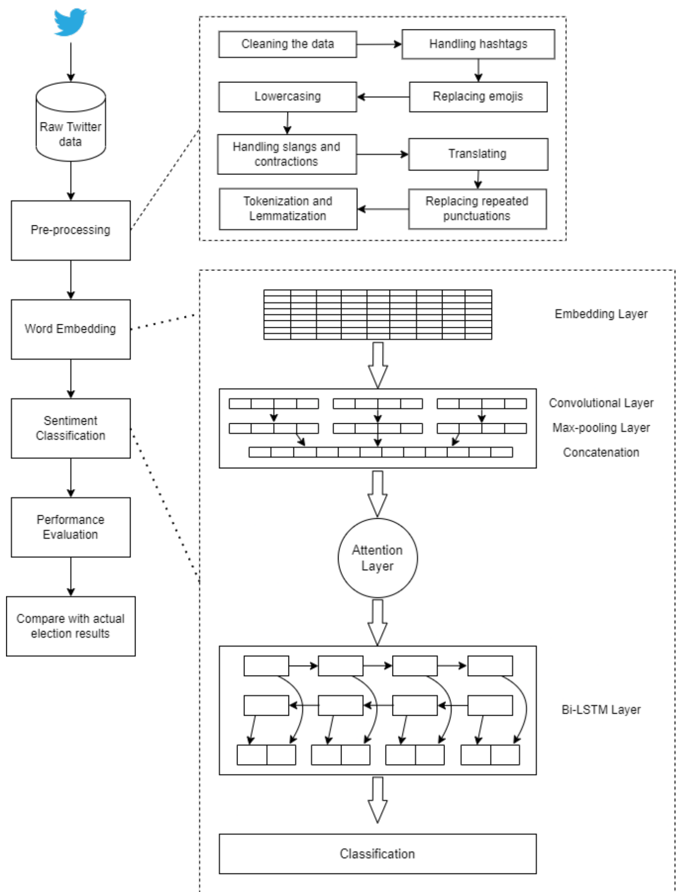


Fig. 1: System flow

elections in general were removed, as agreed upon by the three annotators. With these criteria, 3,311 tweets out of the 8,127 tweets were agreed to be removed by the annotators, which left a total of 4,816 tweets. 80% of which were used for training, while the rest were for testing. The final labels of each tweet for the two categories were also agreed upon by the three annotators.

C. Preprocessing

This was a necessary step to make the text easier to classify by removing the unnecessary and repetitive parts. It would also help to avoid compromising the results of the analysis. Moreover, social media language doesn't follow the standard and formal way, especially in Twitter where each tweet could only have a maximum of 280 characters. To make the tweets more consistent and standardized, the following techniques were done to each tweet:

1) Data Cleaning: Some components were removed from the gathered tweets including user mentions (eg. @johndoe) and URLs, while elongated words (eg. helloooo) were shortened and some symbols were replaced. This process resulted to a cleaner data which would increase the model's efficiency.

2) Handling Hashtags: Different hashtags were prominently used during and after the election period, which would certainly affect the polarity of the tweet (eg. #LeniKiko2022, #BBMSara2022). The hashtag symbol were separated and replaced by the word *hashtag* instead, similar to the study [42]. For example, the hashtag *#VoteWisely* became *hashtag Vote Wisely* after this process.

3) Replacing Emojis: Emojis were replaced by their textual descriptions, which were transformed using Python's `emoji` library. Colons and underscores that resulted from this process were replaced by a space.

4) Lowercasing: All letters were transformed to lower case for consistency.

5) Replacing slangs and contractions: Slangs and contractions were replaced by its original word. For example, the word *I'm* will be replaced by *I am*. Contractions were replaced using Python's `contractions` library. A slang dictionary containing the most commonly used slangs in both English and Filipino was created, with their equivalent replacements. The slang dictionary also contained terms that were largely used during the elections. For example, the term "kakampink" was replaced by "leni supporter", "uniteam" was replaced by "marcos supporter", and other similar terms, focusing on the support for the top two presidential candidates. This was so that these words would be included in the word embeddings, and lessen the out-of-vocabulary (OOV) words.

6) Translation: Since the context of the data was about the Philippine Elections, the tweets would not purely be written in English. These tweets must also be captured and included during the word embedding process. The study [25] resolved this issue by translating non-English words to English. The sentences were translated by integrating the Google Translate API through Python's `googletrans` library.

7) Replacing Repetition of Punctuation: Punctuation marks that were repeated multiple times, like in the statement *That's awesome!!!*, were replaced. The resulting statement was *That's awesome multiple exclamation point*, similar to the study [34].

8) Lemmatization: Lemmatization is the process of converting a word to its base form. It is similar to stemming which only removes the suffix of the word [56]. Lemmatization takes the word's part of speech into consideration which adds meaning to the text. For example, if the word *changing* is lemmatized, it would return *change* but if it's stemmed, it would only return *chang* [57]. After tokenizing each tweet, only non-digit tokens and tokens with more than two characters were lemmatized and included as a feature.

D. Word Embedding Layer

The GloVe word embedding model was used for this study, which is an unsupervised learning algorithm for obtaining vector representation for words [46]. It generates word embeddings by using word co-occurrence matrices from a given corpus. A co-occurrence matrix tells how a word pair occurs together. The study utilized a pre-trained GloVe word embedding model with 2 billion tweets, 27 billion tokens, 1.2 billion vocabulary, and 200 dimensions, similar to the study [54], to generate a 200-dimensional word vector for each word in the given corpus. An embedding index was created using the embedding model. The embedding index consisted of the pre-trained words and their corresponding 200-dimensional word vector.

From the lemmatized tokens, only those that exist in the embedding index were considered as useful tokens. Otherwise, the spelling of the token was checked using Python's `textblob` library. If the token still didn't exist in the embedding index, it would not be considered as a useful token. Each token was assigned their own index and word vector to the embedding matrix. The embedding matrix was entered into the embedding layer as weights. The Python library `Keras`, a deep learning API, was used for producing the embedding layer.

E. Sentiment Analysis

The `Keras` Python library was also used for this whole section. The output of the embedding layer was then used as an input for the attention-based CNN-Bi-LSTM model.

1) CNN Layer: The input was alternately passed through three convolutional and max-pooling layers. The convolutional layers had filter values of 200, 150, and 100, with a kernel size of 1, 2, and 3 respectively and a ReLU activation. Meanwhile, the three max-pooling layers had a pool size of 2. Gaussian Dropouts with a rate of 0.2 were applied as regularization to avoid overfitting after each pair of convolutional and max-pooling layers, similar to the study [54].

2) Attention Layer: The CNN layer would capture the important features in the corpus. These features would be given more attention to emphasize their significance in the attention layer. The attention mechanism used in this study was based on the study [58]. After the Attention layer, a Gaussian Dropout with a rate of 0.2 was also used.

3) Bi-LSTM Layer: A Bi-LSTM layer with 128 units would receive the generated attention scores from the attention layer. A Gaussian Dropout with a rate of 0.2 was also used.

4) Dense Layers: An additional Dense layer with 32 units and a ReLU activation would be added, followed by a Gaussian Dropout. Finally, a dense layer with a *softmax* activation would serve as the output layer. The *softmax* classifier is used for multi-class classifications, where a single instance has more than two possible mutually-exclusive classes. The number of units in the dense layer was set to 3 because a single category would be classified into 3 classes (positive, negative, neutral). A chronological order of layers and its corresponding output shapes is shown in Fig. 2.

5) Training: Categorical cross-entropy was used as a loss function to compute the discrepancy between the predicted and actual classification of the text, which was best used for multi-class classifications. An adam optimizer with a learning rate of 0.0001 was also used, similar to the study [54]. The dataset was trained for 100 epochs and because of the imbalanced dataset, additional weights for each class were computed and included into the training. The weights were dependent on the class distribution in the training data.

F. Performance Evaluation

Accuracy, Precision, Recall, and F-score were computed for each category (BBM and Leni) and were averaged in order to evaluate the performance of the chosen classifier.

Layer (type)	Output Shape	Param #
embedding (Embedding)	(None, 76, 200)	1385200
conv1d (Conv1D)	(None, 76, 200)	40200
max_pooling1d (MaxPooling1D)	(None, 38, 200)	0
gaussian_dropout (GaussianD ropout)	(None, 38, 200)	0
conv1d_1 (Conv1D)	(None, 37, 150)	60150
max_pooling1d_1 (MaxPooling 1D)	(None, 18, 150)	0
gaussian_dropout_1 (Gaussia nDropout)	(None, 18, 150)	0
conv1d_2 (Conv1D)	(None, 16, 100)	45100
max_pooling1d_2 (MaxPooling 1D)	(None, 8, 100)	0
gaussian_dropout_2 (Gaussia nDropout)	(None, 8, 100)	0
custom_attention (CustomAtt ention)	(None, 8, 100)	108
gaussian_dropout_3 (Gaussia nDropout)	(None, 8, 100)	0
bidirectional (Bidirectiona l)	(None, 256)	234496
gaussian_dropout_4 (Gaussia nDropout)	(None, 256)	0
dense (Dense)	(None, 32)	8224
dense_1 (Dense)	(None, 3)	99

Fig. 2: Sequential Model Layers

These metrics were calculated as shown in Formulas 1–4 respectively.

The study conducted sentiment analysis using Attention-based CNN-Bi-LSTM, CNN-Bi-LSTM, Attention-based Bi-LSTM, Attention-based CNN, Bi-LSTM, and CNN classifiers for cross-validation and compared their metric scores.

G. Comparison with the Actual Election Results

To compare the results of data annotation with the actual election results, the distribution of classes for both BBM and Leni categories would be compared and analyzed.

IV. Results and Discussion

A. Labeled Data

In the BBM category, 65.9% of the data were classified as Neutral, 20.4% were classified as Negative, and 13.7% were classified as Positive, as shown in Fig. 3. Meanwhile, Fig. 4 shows that tweets were also mostly classified as Neutral at 48%, then 44% were labeled as Positive, and 8% were labeled as Negative.

The distribution of classes in both the BBM and Leni categories were heavily imbalanced, which means that there was an uneven distribution among the three classes (Positive, Negative, Neutral). The BBM category was heavily biased towards the Neutral class, while the Negative class in the Leni category was heavily under-represented.

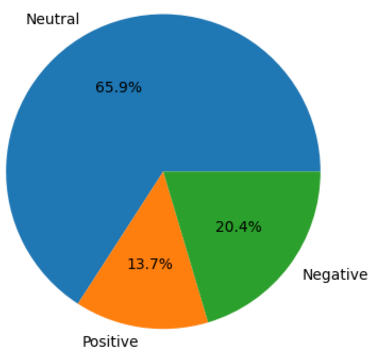


Fig. 3: Distribution of classes in BBM category

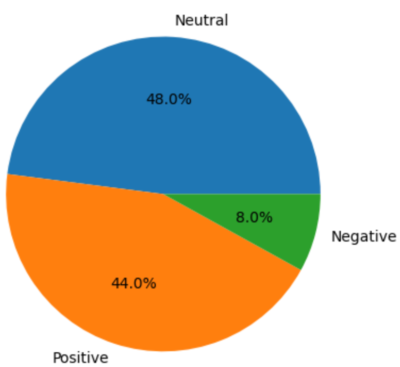


Fig. 4: Distribution of classes in Leni category

We can see in the word cloud for the BBM category shown in Fig. 5 the most used words of tweets that expressed positivity, neutrality, and negativity. The larger a word or phrase was in the word cloud, the more frequently it showed up on the tweets. For both positive and negative classes in the BBM category, words related to BBM were mostly used such as "bbm", "uniteam", "marcos", to name a few. The negative class also used many of the word "leni" based on its size in the word cloud. This could denote that the two candidates were compared to each other frequently, where Bongbong Marcos was the one receiving negative sentiments. Under the neutral class of the same category, there were also many mentions of words related to Leni Robredo such as "leni", "leni robredo", "kakampink", etc. This could denote that many tweets that are neutral towards Bongbong Marcos were talking about Leni Robredo instead.

Fig. 6 shows the most used words of the tweets classified into each of the three classes. Similar to the BBM category, the neutral class for the Leni category also mostly mentioned about the other candidate, but the word "leni" was also included in the tweets frequently. The negative class also mentioned the word "bbm" frequently, signifying the comparison between the two. The positive class in the Leni category did not have as much words related to the other candidate, compared to the positive class in the BBM category. Based on the word cloud, most tweets were talking about Leni Robredo herself, the *Angat Buhay* program or NGO, and her supporters (*kakampink*).

Although the overall word cloud had many mentions of both candidates, as shown in Fig. 7, it was mostly composed of words related to Leni Robredo such as "leni", "leni robredo", "madam leni", "kakampink", "angat buhay", and others. For Bongbong Marcos, most tweets used the word "bbm", but there were also mentions of the words "marcos" and "uniteam". This result was similar to the word cloud of tweets from the study [26], where some of the most mentioned words were "leni robredo" and "kiko pangilinan" in all three classes (positive, negative, neutral).

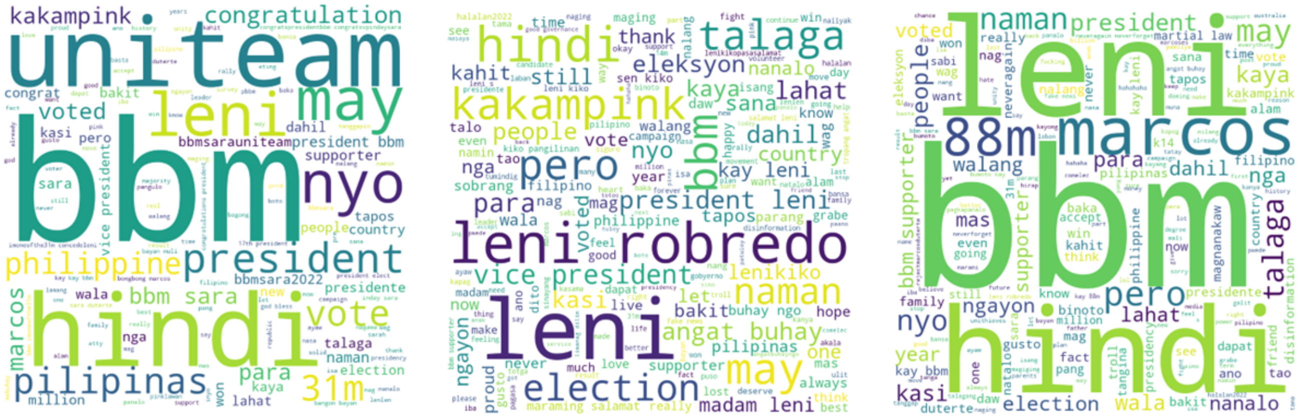
B. Comparison of Labeled Data with the Actual Results

Contrary to the results of the elections, only 13.7% of the 4,816 tweets had positive sentiments towards Bongbong Marcos the voting period. This percentage is much less than the surveys (56%) and the results of the elections (58.77%). Meanwhile, Leni Robredo got more positive sentiments at 44%, which is greater than the surveys (23%) and the results of the elections (27.94%). However, it is important to take note that it is also possible that by the time of collection, some tweets from after the election period had already been deleted, or the accounts had become private or had been erased.

C. Analysis of the Different Models

The best results of the analysis were shown in Tables I–III after training the dataset in each model five times. Fig. 8 also shows a visual representation of the comparison of each performance metric for all models. "ACB" and "CB" correspond to Attention-based CNN-Bi-LSTM and CNN-Bi-LSTM respectively.

For the BBM category, shown in Table I, the Attention-based CNN-Bi-LSTM model produced the best accuracy of 80.81%, then closely followed by the Attention-based CNN model with 80.71% accuracy. However, the Attention-based CNN model performed best in terms of precision (81.85%), recall (79.98%), and f-score (80.9%), which was then closely followed by the Attention-based CNN-Bi-LSTM model with 81.27% precision, 79.67% recall, and 80.46% f-score. The CNN model got the next best results with 78.53% accuracy, 79.32% precision, 78% recall, and 78.66% f-score. It is followed by the CNN-Bi-LSTM model with 78% accuracy, 78.25% precision, 78% recall, and 78.12% f-score. Meanwhile, the Bi-LSTM model got 76.76% accuracy, 76.92% precision, 76.76% recall, and 76.84% f-score. Lastly,



Positive Neutral Negative

Fig. 5: BBM Category Word Cloud

Positive Neutral Negative

Fig. 6: Leni Category Word Cloud

the Attention-based Bi-LSTM model got 76.66% accuracy, 78.25% precision, 76.66% recall, and 76.7% f-score.

The Leni category in Table II shows that the Attention-based CNN-Bi-LSTM performed best only in terms of precision with 81.66%. In terms of accuracy, recall, and f-score, the CNN-Bi-LSTM model performed best with the Attention-based CNN-Bi-LSTM model at a close second. The Attention-based CNN-Bi-LSTM model got an 80.81% accuracy, 79.67% recall, and 80.46% f-score. Meanwhile, the CNN-Bi-LSTM model got 81.33% accuracy, 81.44% precision, 81.02% recall, and 81.23% f-score. It is followed by the CNN model with 80.81% accuracy, 81.29% precision, 80.19% recall, and 80.73% f-score. Next is the Attention-based CNN model which got 78.84% accuracy, 79.21% precision, 78.63% recall, and 78.92% f-score. Then, followed by the Bi-LSTM model with 78.22% accuracy, 78.36% precision, 78.11% recall, and 78.23% f-score. Finally, the Attention-based Bi-LSTM got an accuracy of 76.66%, a precision of 76.74%, a recall of 76.66%, and an f-score of 76.7%.

After getting the average results for both the BBM and Leni categories, the Attention-based CNN-Bi-LSTM model was shown to have the best results in all evaluation metrics with an average accuracy of 80.86%, an average precision of 81.47%, an average recall of 80.03%, and an average f-score of 80.74%. It is followed by the Attention-based CNN model, which got the second best results in all metrics, except recall, with 79.77% accuracy, 80.53% precision, 79.31% recall, and 79.91% f-score. Next is the CNN model with the next best f-score, with an average accuracy of 79.67%, 80.3% precision, 79.1% recall, and 79.7% f-score. Then, followed by the CNN-Bi-LSTM model with 79.67% accuracy, 79.85% precision, 79.51% recall, and 79.68% f-score. Then, the Bi-LSTM model got an average accuracy of 77.49%, average precision of 77.64%, average recall of 77.44%, and average f-score of 77.54%. Lastly, the Attention-based Bi-LSTM model got an average accuracy of 76.66%, 76.74% precision, 76.66% recall, and 76.7% f-score. These results are also shown in Table III.

Based on the performances of all the models in each

BBM Category Performance Evaluation				
Model	Accuracy	Precision	Recall	F-score
Att-CNN-BiLSTM	**80.8091%**	81.2698%	79.6681%	80.4610%
CNN-BiLSTM	78.0083%	78.2518%	78.0083%	78.1210%
Att-BiLSTM	76.6598%	76.7394%	76.6598%	76.6995%
Att-CNN	80.7054%	**81.8471%**	**79.9793%**	**80.9024%**
BiLSTM	76.7635%	76.9231%	76.7635%	76.8432%
CNN	78.5270%	79.3249%	78.0083%	78.6611%

TABLE I: Performance evaluation of all models in the BBM category

Leni Category Performance Evaluation				
Model	Accuracy	Precision	Recall	F-score
Att-CNN-BiLSTM	80.9129%	**81.6649%**	80.3942%	81.0246%
CNN-BiLSTM	**81.3278%**	81.4390%	**81.0166%**	**81.2273%**
Att-BiLSTM	76.6598%	76.7394%	76.6598%	76.6995%
Att-CNN	78.8382%	79.2059%	78.6307%	78.9172%
BiLSTM	78.2158%	78.3559%	78.1120%	78.2338%
CNN	80.8091%	81.2829%	80.1867%	80.7311%

TABLE II: Performance evaluation of all models in the Leni category

Average Performance Evaluation				
Model	Accuracy	Precision	Recall	F-score
Att-CNN-BiLSTM	**80.8610%**	**81.4674%**	**80.0311%**	**80.7428%**
CNN-BiLSTM	79.6681%	79.8454%	79.5125%	79.6786%
Att-BiLSTM	76.6598%	76.7394%	76.6598%	76.6995%
Att-CNN	79.7718%	80.5265%	79.3050%	79.9098%
BiLSTM	77.4896%	77.6395%	77.4378%	77.5385%
CNN	79.6681%	80.3039%	79.0975%	79.6961%

TABLE III: Average performance evaluation of all models

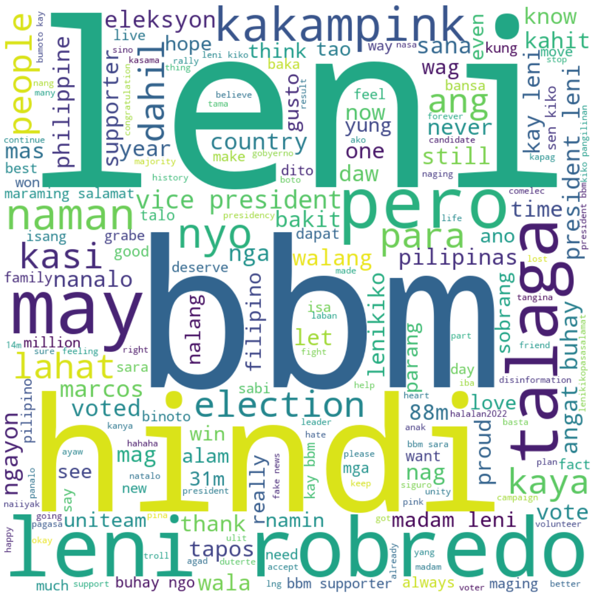


Fig. 7: Word cloud of all tweets

category for the given dataset, the Attention-based CNN model performed best on data that was heavily biased towards one class (BBM). On the other hand, the CNN-Bi-LSTM model had the best performance on data where one class was under-represented (Leni). However, the Attention-based CNN-Bi-LSTM model had the most consistent evaluation for both categories and its performance on both categories also closely followed the two previously mentioned models. Due to the model's consistency regardless of the data's class distribution, it achieved the best average performance among all the tested models in all evaluation metrics (accuracy, precision, recall, f-score).

V. Conclusion and Future Work

This study collected data regarding the sentiments of Twitter users after the recent 2022 Philipine National Elections. The collected data were filtered and labeled on two categories focusing on the top two presidential candidates at the time: the winning candidate, Pres. Bongbong Marcos, and Atty. Leni Robredo. Majority of the sentiments expressed neutrality on both categories (65.9% for BBM, 48% for Leni), but there were more positive sentiments about Leni Robredo (44%) and more negative sentiments about Bongbong Marcos (20.4%).

The labeled data were trained into an Attention-based CNN-Bi-LSTM model with pre-trained GloVe word embeddings. On average, the model achieved the best performance in all evaluation metrics: accuracy (80.86%), precision (81.47%), recall (80.03%), and f-score (80.74%). However, the Attention-based CNN model performed best on data with heavy bias on one class (BBM category) in terms of precision (81.85%), recall (79.98%), and f-score (80.9%). Meanwhile, the CNN-Bi-LSTM model performed best on data with an under-represented class (Leni category) in terms of accuracy (81.33%), recall (81.02%), and f-score (81.23%).

In future elections, it would be preferable to gather a larger dataset and having more annotators to get a higher-quality

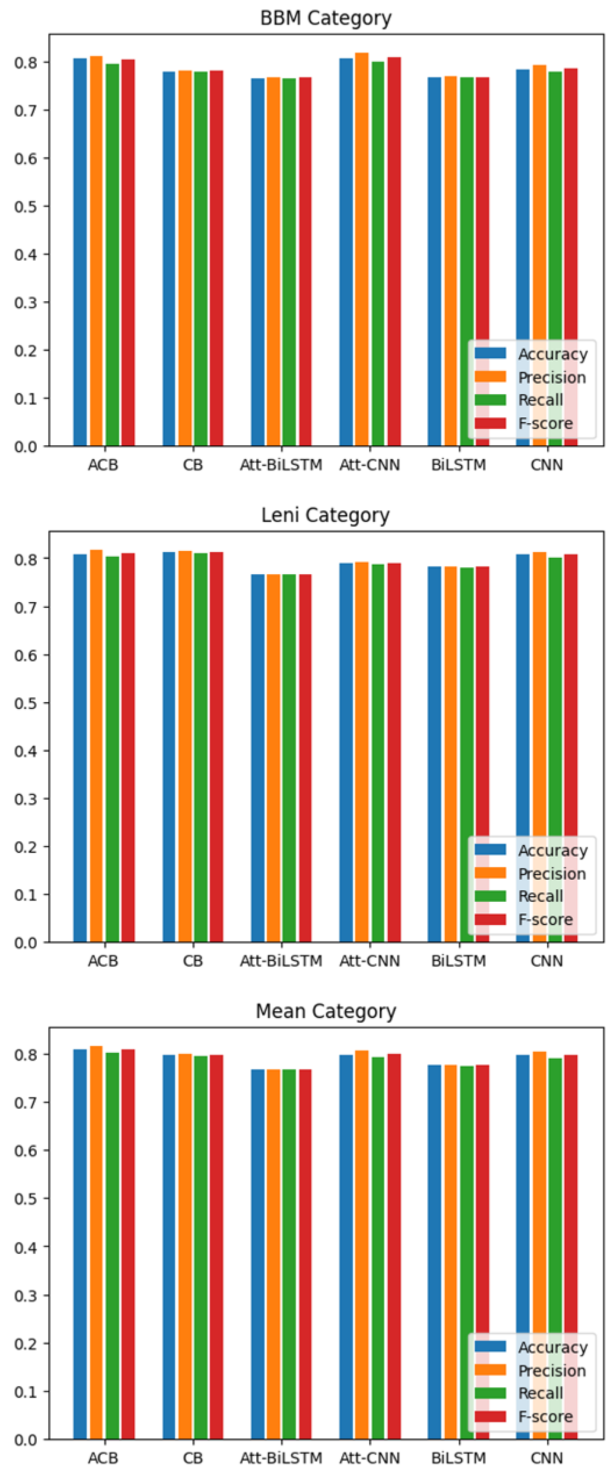


Fig. 8: Comparative analysis

data. In addition to having positive, negative, and neutral classes for the candidates, classifying the tweets' political stance in terms of their preferred candidate in a certain position would also be interesting. However, it's important to keep in mind that a tweet does not necessarily support one candidate if it's against another because there could always be more than two candidates in one position. Dividing the tweets per region is also recommended to get a better comparison between the results of the annotations and the actual elections. If possible, using a balanced dataset for training can also improve the model's performance.

The elections proved how divisive the Filipinos could be, but also how strongly they could stand to what they believe in. After the elections, the researcher hoped that all Filipinos would be able to do what they could to contribute to the development of the country and the betterment of the lives of their fellow countrymen.

Acknowledgment

I would like to express my deepest gratitude to my adviser, **Prof. Marie Betel B. de Robles**, for her support, valuable comments, suggestions, advice, and guidance, that have been very helpful and made the completion of this study possible. For going out of her way to consult other professors regarding the study, for taking the time to reach out amidst the hard times, and for being patient even when I was having difficulty with my many research topics, I am very thankful.

I am also thankful to all the panelists during the presentations, **Dr. Maria Art Antonette D. Clariño**, **Prof. Miyah D. Queliste**, **Prof. Rizza D. Mercado**, and **Prof. Joseph Anthony C. Hermocilla** for their wonderful comments and suggestions, that made a great contribution for the completion of the study.

Special thanks to the annotators as well for taking the time and giving their very best to provide an unbiased and honest judgement to the data given. This study wouldn't have been possible without your help.

I would like to acknowledge my partner and college friends, **John Christopher Angeles**, **Rene Jotham Culaway**, **John Joseph Macalalad**, **Ryan Andrei Cruz**, **John Michael Tugay**, **Ma'am Bernadette Pelaez**, **Ma'am Erika Cunanan**, **Sir Perico Dan Dionisio**, **Sir Prince Karlo Aragones**, and **Ma'am Katherine Loren Tan**, for making my university life full of fun memories, for being within reach whenever I want to rant, for being my friendly academic rivals, and for being my family far away from home.

For the best and my last roommate in UPLB, **Julia Sabroso**, thank you so much for the philosophical talks, for being an emotional pillar and a pseudo-financial advisor, and for making my last days in UPLB more fun and meaningful.

Most of all, I would like to thank my parents, **Jinkee Nadua** and **Almin Nadua**, for your love, care, and unconditional support. Thank you so much for all the things you've done for me, for providing for the family, for allowing me to receive a good education, and for being patient about everything. This is all for you.

References

[1] A. J. Staff, "Why the 2022 philippines election is so significant," May 2022. [Online]. Available: https://www.aljazeera.com/news/2022/5/8/why-the-2022-philippines-election-is-so-significant

[2] "April 2022 nationwide survey on the may 2022 elections," May 2022. [Online]. Available: https://www.pulseasia.ph/april-2022-nationwide-survey-on-the-may-2022-elections/

[3] S. Kemp, "Digital 2022: The philippines," Feb 2022. [Online]. Available: https://datareportal.com/reports/digital-2022-philippines

[4] H. N. Chaudhry, Y. Javed, F. Kulsoom, Z. Mehmood, Z. I. Khan, U. Shoaib, and S. H. Janjua, "Sentiment analysis of before and after elections: Twitter data of us election 2020," *Electronics*, vol. 10, no. 17, p. 2082, 2021.

[5] A. Yadav and D. K. Vishwakarma, "Sentiment analysis using deep learning architectures: a review," *Artificial Intelligence Review*, vol. 53, no. 6, pp. 4335–4385, 2020.

[6] P. Tyagi and R. Tripathi, "A review towards the sentiment analysis techniques for the analysis of twitter data," in *Proceedings of 2nd international conference on advanced computing and software engineering (ICACSE)*, 2019.

[7] B. Liu, "Sentiment analysis and opinion mining," *Synthesis lectures on human language technologies*, vol. 5, no. 1, pp. 1–167, 2012.

[8] O. Habimana, Y. Li, R. Li, X. Gu, and G. Yu, "Sentiment analysis using deep learning approaches: an overview," *Science China Information Sciences*, vol. 63, no. 1, pp. 1–36, 2020.

[9] K. Jain and S. Kaushal, "A comparative study of machine learning and deep learning techniques for sentiment analysis," in *2018 7th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO)*. IEEE, 2018, pp. 483–487.

[10] Q. T. Ain, M. Ali, A. Riaz, A. Noureen, M. Kamran, B. Hayat, and A. Rehman, "Sentiment analysis using deep learning techniques: a review," *International Journal of Advanced Computer Science and Applications*, vol. 8, no. 6, 2017.

[11] Y. Chandra and A. Jana, "Sentiment analysis using machine learning and deep learning," in *2020 7th International Conference on Computing for Sustainable Global Development (INDIACom)*. IEEE, 2020, pp. 1–4.

[12] Z. Sarao, "Pulse asia defends sampling methods amid criticisms on latest survey," May 2022. [Online]. Available: https://newsinfo.inquirer.net/1591737/pulse-asia-defends-sampling-methods-amid-criticisms-on-latest-survey

[13] K. Subingsubing, "Surveys under scrutiny; pulse asia stands by method," May 2022. [Online]. Available: https://newsinfo.inquirer.net/1592530/surveys-under-scrutiny-pulse-asia-stands-by-method

[14] G. P. Lalu, "Pulse asia: False accusations, baseless claims vs survey erode ph democracy," May 2022. [Online]. Available: https://newsinfo.inquirer.net/1592208/pulse-asia-false-accusations-baseless-claims-vs-survey-erode-ph-democracy

[15] C. Chavez, "43 malfunctioning vcms reported on election day-año," May 2022. [Online]. Available: https://mb.com.ph/2022/05/09/43-malfunctioning-vcms-reported-on-election-day-ano/

[16] T. G. Santos, "Fastest results, highest turnout, says comelec of 2022 polls," May 2022. [Online]. Available: https://newsinfo.inquirer.net/1595907/fastest-results-highest-turnout-says-comelec-of-2022-polls

[17] A. Parrocha, "Fast transmission of poll results 'surprising': analyst," May 2022. [Online]. Available: https://www.pna.gov.ph/articles/1174141

[18] D. de Leon, "'mahirap i-prove': Comelec shrugs off claims of automated election rigging," May 2022. [Online]. Available: https://www.rappler.com/nation/elections/comelec-response-claims-automated-rigging-2022/

[19] K. Dela Peña, "Int'l observers: May 9 ph vote falls short of free, fair election standards," May 2022. [Online]. Available: https://newsinfo.inquirer.net/1600309/intl-observers-may-9-ph-vote-falls-short-of-free-fair-election-standards

[20] "International observers say philippine polls did not meet 'free and fair' standards," May 2022. [Online]. Available: https://www.licas.news/2022/05/20/international-observers-say-philippine-polls-did-not-meet-free-and-fair-standards/

[21] P. Zhou, W. Shi, J. Tian, Z. Qi, B. Li, H. Hao, and B. Xu, "Attention-based bidirectional long short-term memory networks for relation classification," in *Proceedings of the 54th annual meeting of the association for computational linguistics (volume 2: Short papers)*, 2016, pp. 207–212.

[22] P. Chauhan, N. Sharma, and G. Sikka, "The emergence of social media data and sentiment analysis in election prediction," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 2, pp. 2601–2627, 2021.

[23] F. Nausheen and S. H. Begum, "Sentiment analysis to predict election results using python," in *2018 2nd international conference on inventive systems and control (ICISC)*. IEEE, 2018, pp. 1259–1262.

[24] W. Budiharto and M. Meiliana, "Prediction and analysis of indonesia presidential election from twitter using sentiment analysis," *Journal of Big data*, vol. 5, no. 1, pp. 1–10, 2018.

[25] Z. C. Turla and J. Caro, "Sentiment analysis of philippine national elections 2016 twitter data," in *Theory and Practice of Computation: Proceedings of Workshop on Computation: Theory and Practice WCTP2016*. World Scientific, 2018, pp. 194–207.
[26] J. J. E. Macrohon, C. N. Villavicencio, X. A. Inbaraj, and J.-H. Jeng, "A semi-supervised approach to sentiment analysis of tweets during the 2022 philippine presidential election," *Information*, vol. 13, no. 10, p. 484, Oct 2022. [Online]. Available: http://dx.doi.org/10.3390/info13100484
[27] Y. Ndasauka, J. Hou, Y. Wang, L. Yang, Z. Yang, Z. Ye, Y. Hao, A. J. Fallgatter, Y. Kong, and X. Zhang, "Excessive use of twitter among college students in the uk: Validation of the microblog excessive use scale and relationship to social interaction and loneliness," *Computers in Human Behavior*, vol. 55, pp. 963–971, 2016.
[28] X. Guo and J. Li, "A novel twitter sentiment analysis model with baseline correlation for financial market prediction with improved efficiency," in *2019 Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS)*. IEEE, 2019, pp. 472–477.
[29] N. Bhaskar, P. Sandeep, N. Amarnath, and J. Thankachan, "A study: Sentimental analysis for election results by using twitter data," *Turkish Journal of Physiotherapy and Rehabilitation*, vol. 32, p. 3.
[30] A. Shelar and C.-Y. Huang, "Sentiment analysis of twitter data," in *2018 International Conference on Computational Science and Computational Intelligence (CSCI)*. IEEE, 2018, pp. 1301–1302.
[31] K. H. Manguri, R. N. Ramadhan, and P. R. M. Amin, "Twitter sentiment analysis on worldwide covid-19 outbreaks," *Kurdistan Journal of Applied Research*, pp. 54–65, 2020.
[32] B. P. Pokharel, "Twitter sentiment analysis during covid-19 outbreak in nepal," *Available at SSRN 3624719*, 2020.
[33] S. A. El Rahman, F. A. AlOtaibi, and W. A. AlShehri, "Sentiment analysis of twitter data," in *2019 international conference on computer and information sciences (ICCIS)*. IEEE, 2019, pp. 1–4.
[34] S. Symeonidis, D. Effrosynidis, and A. Arampatzis, "A comparative evaluation of pre-processing techniques and their interactions for twitter sentiment analysis," *Expert Systems with Applications*, vol. 110, pp. 298–310, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417418303683
[35] A. Singh, E. Blanco, and W. Jin, "Incorporating emoji descriptions improves tweet classification," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 2019, pp. 2096–2101.
[36] S. Fransiska, R. Rianto, and A. I. Gufroni, "Sentiment analysis provider by. u on google play store reviews with tf-idf and support vector machine (svm) method," *Scientific Journal of Informatics*, vol. 7, no. 2, pp. 203–212, 2020.
[37] M. Buytaert, "A machine learning approach to sentiment analysis and stance detection for political tweets," Ph.D. dissertation, Ghent University, 2018.
[38] Y. Lou, Y. Zhang, F. Li, T. Qian, and D. Ji, "Emoji-based sentiment analysis using attention networks," *ACM Transactions on asian and low-resource language information processing (TALLIP)*, vol. 19, no. 5, pp. 1–13, 2020.
[39] V. Grover, "Exploiting emojis in sentiment analysis: A survey," *Journal of The Institution of Engineers (India): Series B*, pp. 1–14, 2021.
[40] H. T. Phan, V. C. Tran, N. T. Nguyen, and D. Hwang, "Improving the performance of sentiment analysis of tweets containing fuzzy sentiment using the feature ensemble model," *IEEE Access*, vol. 8, pp. 14 630–14 641, 2020.
[41] J. P. D. Delizo, M. B. Abisado, and M. I. P. De Los Trinos, "Philippine twitter sentiments during covid-19 pandemic using multinomial naïve-bayes," *International Journal*, vol. 9, no. 1.3, 2020.
[42] C. Villavicencio, J. J. Macrohon, X. A. Inbaraj, J.-H. Jeng, and J.-G. Hsieh, "Twitter sentiment analysis towards covid-19 vaccines in the philippines using naïve bayes," *Information*, vol. 12, no. 5, p. 204, 2021.
[43] N. C. Dang, M. N. Moreno-García, and F. De la Prieta, "Sentiment analysis based on deep learning: A comparative study," *Electronics*, vol. 9, no. 3, 2020. [Online]. Available: https://www.mdpi.com/2079-9292/9/3/483
[44] W. Ling, C. Dyer, A. W. Black, and I. Trancoso, "Two/too simple adaptations of word2vec for syntax problems," in *Proceedings of the 2015 conference of the North American chapter of the association for computational linguistics: human language technologies*, 2015, pp. 1299–1304.
[45] S. Bhoir, T. Ghorpade, and V. Mane, "Comparative analysis of different word embedding models," in *2017 International conference on advances in computing, communication and Control (ICAC3)*. IEEE, 2017, pp. 1–4.
[46] J. Pennington, R. Socher, and C. D. Manning, "Glove: Global vectors for word representation," in *Empirical Methods in Natural Language Processing (EMNLP)*, 2014, pp. 1532–1543. [Online]. Available: http://www.aclweb.org/anthology/D14-1162
[47] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
[48] J. Zhou, Y. Lu, H.-N. Dai, H. Wang, and H. Xiao, "Sentiment analysis of chinese microblog based on stacked bidirectional lstm," *IEEE Access*, vol. 7, pp. 38 856–38 866, 2019.
[49] J. Deng, L. Cheng, and Z. Wang, "Attention-based bilstm fused cnn with gating mechanism model for chinese long text classification," *Computer Speech & Language*, vol. 68, p. 101182, 2021.
[50] C. Colón-Ruiz and I. Segura-Bedmar, "Comparing deep learning architectures for sentiment analysis on drug reviews," *Journal of Biomedical Informatics*, vol. 110, p. 103539, 2020.
[51] S. Sharmin and D. Chakma, "Attention-based convolutional neural network for bangla sentiment analysis," *Ai & Society*, vol. 36, no. 1, pp. 381–396, 2021.
[52] M. Zhang, V. Palade, Y. Wang, and Z. Ji, "Attention-based word embeddings using artificial bee colony algorithm for aspect-level sentiment classification," *Information Sciences*, vol. 545, pp. 713–738, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S002002552030949X
[53] M. E. Basiri, S. Nemati, M. Abdar, E. Cambria, and U. R. Acharya, "Abcdm: An attention-based bidirectional cnn-rnn deep model for sentiment analysis," *Future Generation Computer Systems*, vol. 115, pp. 279–294, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167739X20309195
[54] M. Kamyab, G. Liu, and M. Adjeisah, "Attention-based cnn and bi-lstm model based on tf-idf and glove word embedding for sentiment analysis," *Applied Sciences*, vol. 11, no. 23, p. 11255, 2021.
[55] M. Grandini, E. Bagli, and G. Visani, "Metrics for multi-class classification: an overview," *arXiv preprint arXiv:2008.05756*, 2020.
[56] M. Anandarajan, C. Hill, and T. Nolan, "Text preprocessing," in *Practical Text Analytics*. Springer, 2019, pp. 45–59.
[57] D. Khyani, B. Siddhartha, N. Niveditha, and B. Divya, "An interpretation of lemmatization and stemming in natural language processing," *Journal of University of Shanghai for Science and Technology*, 2021.
[58] T. Rocktäschel, E. Grefenstette, K. M. Hermann, T. Kočiskỳ, and P. Blunsom, "Reasoning about entailment with neural attention," *arXiv preprint arXiv:1509.06664*, 2015.

Abigail C. Nadua Abigail C. Nadua is a Computer Science undergraduate student from the University of the Philippines - Los Baños (UPLB). She is interested in the fields of natural language processing, data science, artificial intelligence, machine learning, and deep learning. She is also a member of the Systems Research Group (SRG) from the Institute of Computer Science - UPLB.