From Posts to Votes: Using RoBERTa to Analyze Social Media as an Indicator of Presidential Election Wins, through Comparisons between Presidential Elections of the Philippines and United States

A Thesis

Presented to the

Department of Information Systems

and Computer Science

Ateneo de Manila University

In Partial Fulfillment

of the Requirements for the Degree of

Bachelor of Science in Computer Science

by

Michael Gavin N. Del Castillo, Mary June Aubrey C. San Jose,

Shaira Jane L. Sy

2025

The thesis entitled:

From Posts to Votes: Using RoBERTa to Analyze Social Media as an Indicator of Presidential Election Wins, through Comparisons between Presidential Elections of the Philippines and United States

submitted by **Michael Gavin N. Del Castillo, Mary June Aubrey C. San Jose,**

Shaira Jane L. Sy has been examined and is recommended for Oral Defense.

———————————————
PATRICIA ANGELA R. ABU, Ph.D.
Chair

———————————————
KENNEDY E. ESPINA, M Sc.
Adviser

———————————————
RAPHAEL A. GUERRERO, Ph.D.
Dean
School of Science and Engineering

The Faculty of the Graduate School of Ateneo de Manila University accepts the thesis entitled:

From Posts to Votes: Using RoBERTa to Analyze Social Media as an Indicator of Presidential Election Wins, through Comparisons between Presidential Elections of the Philippines and United States

submitted by **Michael Gavin N. Del Castillo, Mary June Aubrey C. San Jose, Shaira Jane L. Sy** in partial fulfillment of the requirements for the Degree of Master of Science in Computer Science.

<table>
<tr><td>REENA ESTUAR, Ph.D.
Panelist</td><td>NOEL VICTORINO, Ph.D.
Panelist</td></tr>
</table>

KENNEDY E. ESPINA, M Sc.
Adviser

RAPHAEL A. GUERRERO, Ph.D.
Dean
School of Science and Engineering

Date: 2025

ABSTRACT

In recent decades, social media has had a major role as a platform through which political ideologies are spread and political discussions occur. One of the most prominent examples of the importance of social media in world politics is former United States (US) Vice President Kamala Harris' social media campaign for the 2024 US presidential elections. Mainly targeting younger audiences through viral "memes" and other social media trends, Harris was able to amass widespread support for her campaign, having just over 5 million followers supporting her endeavors on TikTok and X (formerly Twitter) combined.[19] Similarly, the 2022 Philippine elections saw the Angat Buhay campaign of former Vice President Leni Robredo. Similarly to Harris, Robredo was able to garner the attention of young audiences on social media. Rallies in support of Robredo alongside Robredo's track record as a politician made her a popular choice for millions as a capable presidential candidate.[18]

Despite massive online support, both Harris and Robredo had lost their respective elections, the former only garnering 226 electoral votes (against Donald Trump's 312 votes) and the latter garnering some 14.8 million votes (as opposed to Ferdinand 'Bongbong' Marcos, Jr., who gathered 31.1 million).[2, 8] Given a possible disparity between social media popularity and election votes, the aim of

this research is to provide a data-driven analysis on the effectiveness of social media as an indicator of election wins by observing social media trends at the time of both 2022 and 2024 elections, as well as comparing and contrasting these elections in terms of said trends.

Through Natural Language Processing (NLP) techniques, the researchers aim to analyze social media data on the aforementioned presidential candidates that had transpired in online spaces during pre-election seasons, namely Facebook, TikTok, and X (formerly Twitter). ML models for sentiment analysis, such as BERT models, will be used to analyze these different posts to determine whether or not social media support directly translates to election success.

TABLE OF CONTENTS

CHAPTER

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

In recent decades, social media has had a major role as a platform through which political ideologies are spread and political discussions occur. One of the most prominent examples of the importance of social media in world politics is former United States (US) Vice President Kamala Harris' social media campaign for the 2024 US presidential elections. Mainly targeting younger audiences through viral "memes" and other social media trends, Harris was able to amass widespread support for her campaign, having just over 5 million followers supporting her endeavors on TikTok and X (formerly Twitter) combined.[19] Similarly, the 2022 Philippine elections saw the Angat Buhay campaign of former Vice President Leni Robredo. Similarly to Harris, Robredo was able to garner the attention of young audiences on social media. Rallies in support of Robredo alongside Robredo's track record as a politician made her a popular choice for millions as a capable presidential candidate.[18]

Despite massive online support, both Harris and Robredo had lost their respective elections, the former only garnering 226 electoral votes (against Donald Trump's 312 votes) and the latter garnering some 14.8 million votes (as opposed to Ferdinand 'Bongbong' Marcos, Jr., who gathered 31.1 million).[2, 8] Given a possible disparity between social media popularity and election votes, the aim of

this research is to provide a data-driven analysis on the effectiveness of social media as an indicator of election wins by observing social media trends at the time of both 2022 and 2024 elections, as well as comparing and contrasting these elections in terms of said trends.

Through Natural Language Processing (NLP) techniques, the researchers aim to analyze social media data on the aforementioned presidential candidates that had transpired in online spaces during pre-election seasons, namely Facebook, TikTok, and X (formerly Twitter). ML models for sentiment analysis, such as BERT models, will be used to analyze these different posts to determine whether or not social media support directly translates to election success.

CHAPTER I

INTRODUCTION

1.1   Context of Study

In recent decades, social media has had a major role as a platform through which political ideologies are spread and political discussions occur. One of the most prominent examples of the importance of social media in world politics is former United States (US) Vice President KAMALA HARRIS' social media campaign for the 2024 US presidential elections. Mainly targeting younger audiences through viral "memes" and other social media trends, Harris was able to amass widespread support for her campaign, having just over 5 million followers supporting her endeavors on TikTok and X (formerly Twitter) combined.[19] Similarly, the 2022 Philippine elections saw the Angat Buhay campaign of former Vice President LENI ROBREDO. Similarly to Harris, Robredo was able to garner the attention of young audiences on social media. Rallies in support of Robredo alongside Robredo's track record as a politician made her a popular choice for millions as a capable presidential candidate.[18]

Despite massive online support, both Harris and Robredo had lost their respective elections, the former only garnering 226 electoral votes (against DONALD

Trump's 312 votes) and the latter garnering some 14.8 million votes (as opposed to Ferdinand 'Bongbong' Marcos, Jr, who gathered 31.1 million).[2, 8] Given a possible disparity between social media popularity and election votes, the aim of this research is to provide a data-driven analysis on the effectiveness of social media as an indicator of election wins by observing social media trends at the time of both 2022 and 2024 elections, as well as comparing and contrasting these elections in terms of said trends.

Through Natural Language Processing Algorithms, the researchers aim to analyze the conversations on the aforementioned presidential candidates that had transpired in online spaces during pre-election seasons, namely Facebook, TikTok, and X (formerly Twitter). Previous research endeavors have already shown the effectiveness of sentiment analysis in determining key themes behind social media posts, especially in the context of events such as elections. Thus, the researchers would like to push this idea further by not only contextualizing the data within a single setting. Rather, the researchers aim to compare and contrast the election periods of the Philippines and US, given that, as aforementioned, the two countries experienced a supposed upset in terms of electoral candidate votes relative to their presence on social media. The researchers aim to determine whether or not social media support directly translates to election success, or if other factors were present which had contributed to the losses of Harris and

Robredo. To provide a more thorough analysis, the researchers also intend to perform the same analysis on social media spaces during the 2016 elections for both the Philippines and US.

1.2 Research Questions

1. Can NLP techniques aid in predicting election outcomes through sentiment analysis on social media posts?

   (a) What were the different themes surrounding and keywords related to sentiments expressed by users on Facebook, X (formerly Twitter), and TikTok on certain presidential election candidates during election periods?

   (b) Do the frequency and sentiments of related keywords and phrases in Facebook posts and comments, posts from X (formerly Twitter), and Tiktok videos and comments indicate majority support for leading candidates?

   (c) Can a developed visualization effectively illustrate social media presences throughout the four elections (Marcos Jr. vs. Robredo, Duterte vs. Roxas in the 2022 and 2016 Philippine elections respectively; and Trump vs. Harris, Trump vs. Biden in 2024 and 2020 US elections

respectively), and whether or not these presences predicted their electoral results?

1.3 Research Objectives

1. Analyze different texts on social media platforms to draw insights on how different presidential candidates were perceived by social media users, and how these perceptions were shared online.

2. Determine if these perceptions directly translated into electoral wins, or if other factors contributed to the outcome in the case that social media support did not indicate an electoral win

3. Develop a dashboard providing a comprehensive overview of social media sentiments in relation to election results by comparing results drawn from sentiment analysis algorithms and actual elections results

1.4 Scope and Limitations of the Study

This paper consists of two case studies to provide a rich analysis of social media data extracted from presidential election periods. First, a case study analysis comparing the recent 2022 Philippine presidential election, with leading candidates Ferdinand "Bongbong" Marcos Jr. and Maria Leonor "Leni" Robredo,

against the recent 2024 US presidential Election, with leading candidates Donald John Trump and Kamala Devi Harris. Second, a comparison between the 2016 Philippine presidential election, with leading candidates Rodrigo Roa Duterte and Manuel "Mar" Araneta Roxas II, and the 2020 US presidential elections, with leading candidates Donald John Trump and Joseph Robinette "Joe" Biden Jr. To add to the discussion, social media data about each president's respective vice presidential running mate candidate will be considered as well. The paper aims to analyze how the respective campaign periods of each candidate (and their running mates) were reflected on different social media spaces. This is so that the researchers can determine the effectiveness of social media platforms as indicators of electoral wins, and determine whether electoral results can be foreseen based on online traction and popularity.

This paper will focus only on social media platforms such as Facebook, Twitter, TikTok, in data collection due to their popularity within the United States and the Philippines, which significantly increases the chances of attaining a statistically large dataset in comparison to other social media platforms. Only posts made prior to the actual election days will be collected, as the goal is to compare pre-election social media data to post-election results. Thus, posts made during the election days and beyond will be excluded from the study. It is also worth mentioning that, for all elections under the scope of this study, only

the winner and first-runner up will be considered as collecting enough data on all candidates might not be feasible.[21] Finally, "posts" include any and all tweets made by the general public on the selected candidates alongside any posts made by the candidates themselves.

Finally, attached to the names of some candidates are certain criminal cases. The Marcos Family was responsible for a series of atrocities and human rights violations in the 1970s, and Donald Trump is currently facing multiple criminal cases.[17, 16] The paper will not explore such topics in-depth as they are outside of the scope and focus of the study; however, these may be touched upon briefly if it is a popular discussion point among users in the data collected.

1.5   Significance of the Study

This paper aims to contribute to the field of social computing by analyzing the behavior of users in online spaces during election periods. By way of sentiment analysis and other NLP techniques, the researchers aim to gather statistically large amounts of social media data that represent the general public, and after which make use of models that process such data and draw insights from the sentiments behind social media posts concerned with presidential elections.

The analytics and findings of this study will benefit the general public's knowledge of how social media runs in both Philippine and American contexts,

especially during the election seasons. These findings will also provide them with more context between the spheres of social media in terms of various political ideologies the said candidates have. The researchers hope to impart realizations on how social media might not be the definitive factor behind electoral wins.

CHAPTER II

REVIEW OF RELATED LITERATURE

2.1 Analytics for Elections

2.1.1 Philippine Elections (2016, 2022)

In his study on the 2016 Philippine presidential elections, after analyzing pre-election surveys, the candidates' campaign strategies and their advocacies, and their supporters' age demographic and news tracking, Holmes (2016) observed that the elections in the Philippines are political clan-dominated, personality-oriented, and media driven.[15]

Rodrigo Duterte's victory in the election was believable to the public and was attributed to: the clarity of his campaign slogan, his significant support from a geographic area, and how he criticized and questioned the character and competence of his fellow candidates. However, one of the most significant observations in the study is the importance of the media, which is updated in real-time and where voter preference was shaped and reformed by Duterte's critiques and 'bashing'.[15]

Following Duterte's term in office, Ferdinand 'Bongbong' Marcos Jr.'s electoral win has caused much uproar among the nation's scholars. The general

resurgence of the Marcos Clan in politics can be attributed to 3 factors: (1) the people's nostalgia of the Marcos era, (2) Duterte's political influence, and (3) the Marcos' years-long digital disinformation campaign on social media.[23] Duterte's consequent influence on Marcos' resurgence cannot be dismissed, as signs have pointed out that Duterte's indirect endorsement led to his win.[13]

2.1.2 US Elections (2020, 2024)

At the cusp of the COVID-19 pandemic, the 2020 US presidential election had taken a major hit– particularly for one of its leading candidates, Donald Trump, whose vote share is largely affected by COVID-19-related cases.[7] It is likely that Trump was viewed negatively for how he handled the pandemic as the most affected counties and states are ones without stay-at-home orders, in swing states, or states that Trump won in 2016. This mismanagement is what likely led to changes in voter preferences and Joe Biden's eventual electoral win.

The 2024 US presidential election was predicted to be one of the most competitive in modern history with a tight competition between candidates Donald Trump and Kamala Harris, the new face of the Democratic Party.[24] In the end, Trump had managed to win the race, but not without escalating the political polarization of the nation– Democrats and Republicans.[24]

2.2   Social Media Use and the Elections

2.2.1   Candidate Activity

The 2016 Philippine presidential contest is widely considered the first "social media election" in the Philippines.[25] The two Philippine presidential elections, 2016 and 2022, are undoubtedly linked as they involve something other than the rise of the Marcos-Duterte alliance: the prominence of social media as a means to bolster their presence in people's lives and boost their popularity.

Despite having the most engagement, Duterte's online presence during his presidential campaign is nothing short of lackluster and underwhelming.[25]

On the other hand, Marcos Jr's campaign has been well-established and maintained in the years leading up to his campaign. His pitch throughout the campaign calls for national unity, featuring the glorification of his father's legacy. In Rappler's three-part study on "networked propaganda" back in 2019, there was a rise in many pro-Marcos pages and channels on different social media platforms, notably on TikTok. There was less activity from Marcos Jr. himself, however, those channels were particularly full of pro-Marcos content.[22]

2.2.2 Public Opinion

Duterte's successful campaign can be attributed to his aggressive supporters– most of whom are vocal online and active offline. As observed by Sinpeng et al (2020), despite Duterte's unprofessional online presence, his supporters are committed and constantly rallied to his defense against the criticism of other candidates.[25] There are also prospects of the heavy involvement of informal actors like paid trolls and influencers as having major roles in mobilizing (and agitating) digital communities, which helped spread his popularity.[25]

"The recent election has been the most social media-active and engaging campaign in the country's democratic history." In the 2022 Philippine presidential race, both leading candidates (Marcos Jr. and Robredo) have taken great leads on social platforms like Facebook and X, respectively.[4] Electoral campaigns are aimed at spreading awareness about the candidate's identity and, over the years, Marcos Jr. has amassed a large number of supporters on TikTok based on the top 4 trending hashtags related to him: #bongbongmarcos (3.4 billion views), #bbmsarah2022 (2.3 billion views), #uniteam (2.5 billion views) and #bbm2022 (2 billion views).[22]

2.3 Social Media Usage

2.3.1 Scraping Methodology

In the Philippine context, as Filipino is a low-resource language, a lack of Filipino datasets of tweets has been a pressing issue from previous studies. In addition, in the American context, because of the recency of the 2024 US Presidential Elections, established public datasets are sparse. Thus, the researchers had to collect data through APIs and open-sourced scrappers. One of the APIs used in previous studies to extract tweets is from X (former Twitter) Developer Application Programming Interface v2, then run a series of Python codes to get relevant tweets from different public accounts.[6, 5]

Ways of extracting posts from Facebook and TikTok are different from ways of extracting tweets from X (former Twitter) because of the unique nature of their postings, which, for example, in the case of TikTok, consist of only images or videos. In posts from Facebook, in the study from Grujic, et al. (2014), they employ the Facebook Graph API, an HTML-based API to access information from Facebook.[14] It utilized PHP, HTML, and jQuery to query data, post new stories, upload photos, and more. In the study by Alashri et al (2016), they used Python codes to extract Facebook posts and comments from the pages of presidential candidates for the 2016 US Presidential Elections.[3]

Meanwhile, in extracting data from TikTok, Vassey, et al (2022) gathered data related to little cigar and cigarillo products by scraping public posts with hashtags containing high engagements.[26] Researchers then developed a codebook to analyze themes within the videos and captions, establishing interrater reliability through subsample coding. Similar methods are also used by Abbas et al (2022) wherein they used the top three hashtags (#SaveSheikhJarrah, #SavePalestine, #FreePalestine) to search TikTok video content related to youth activism in the Israel-Palestine conflict, however, it also used to find videos via non-random sampling method to code frame by frame to see what is the message of the video.[1] Lastly, Cheng and Li (2023) converted the audio of news videos in TikTok to text via a Google API to train the text into a sentiment classifier.[10] Then, they took images of them every second to calculate the second-person view ratio, essential for studying the prevalence of that person of view in every news video. Then, they took every frame-per-second of a video to calculate the second-person view ratio, essential for studying the prevalence of that viewpoint.

2.4   Sentiment Analysis for Social Media and Elections

According to Liu (2012), the study of people's views, sentiments, assessments, appraisals, attitudes, and emotions about goods, services, organizations, people, problems, events, subjects, and their characteristics is commonly referred

to as sentiment analysis or opinion mining.[20] With the explosive growth of social media, it has become a hotspot of opinions, shaping our decisions, especially in an important political event like elections. In the field of social computing, election seasons are one of the widely researched topics, especially on how the interaction in social media affects society in terms of making decisions on who to vote for.

There are studies using sentiment analysis to analyze the social media activity on some platforms during the Philippine election season. In a study by Macrohon, et al. (2022) and Demillo, et al. (2023), they used the Naïve Bayes classifier, a probabilistic learning method, to determine the probability of a tweet belonging to the best class—applicable in determining the polarity of a post.[21, 12] Then, previous studies showed the usage of bidirectional encoder representation from transformers (BERT) models, modified to handle emojis and Tagalog language tweets. Aquino, et al (2025) introduced the emotion-infused BERT-GCN model for sentiment analysis, which includes emoji semantics into the models, treating them as sentiment representation;[6] meanwhile, Cruz, et al. (2022) used the RoBERTa-tagalog-cased model to get the vectorized version of Tagalog embeddings, essential to map echo chambers on Twitter via K-Means modeling.[11] Lastly, the Support Vector Machines (SVM) Classifier model was used by Demillo, et al. (2023) to handle binary classification of data, classifying

them as either a negative or positive sentiment.[12]

In addition to the established methodology in the previous studies, Belcastro et al (2020) proposed a new method to analyze the polarization of tweets—IOM-NN. IOM-NN or Iterative Opinion Mining - Neural Network uses a feed-forward neural network to classify posts iteratively, refining the classification rules each time.[9]. By focusing on individual user polarization based on the frequency of posts where they are positive for a certain faction, it avoids biases common in some social media analyses and provides a more accurate alternative compared to traditional polling methods.

BIBLIOGRAPHY

[1] ABBAS, L., FAHMY, S. S., AYAD, S., IBRAHIM, M., AND ALI, A. H. Tiktok intifada: Analyzing social media activism among youth. *Online Media and Global Communication 1*, 2 (2022), 287–314.

[2] ABS-CBN. Halalan results 2022. `https://halalanresults.abs-cbn.com/`, 2022.

[3] ALASHRI, S., KANDALA, S. S., BAJAJ, V., RAVI, R., SMITH, K. L., AND DESOUZA, K. C. An analysis of sentiments on facebook during the 2016 us presidential election. In *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)* (2016), IEEE, pp. 795–802.

[4] AMPON, R. K. S., AND SALATHONG, J. 'battleground of philippine elections': Political message strategies of bongbong marcos and leni robredo on social media. *International Journal of Social Science and Human Research 6*, 07 (2023).

[5] ANCHETA, J. R., GORRO, K. D., AND UY, M. A. D. Walangpasok on twitter: Natural language processing as a method for analyzing tweets on class suspen-

sions in the philippines. In *2020 12th International Conference on Knowledge and Smart Technology (KST)* (2020), pp. 103–108.

[6] AQUINO, J. A., LIEW, D. J., AND CHANG, Y.-C. Graph-aware pre-trained language model for political sentiment analysis in filipino social media. *Engineering Applications of Artificial Intelligence 146* (2025), 110317.

[7] BACCINI, L., BRODEUR, A., AND WEYMOUTH, S. The covid-19 pandemic and the 2020 us presidential election. *Journal of population economics 34* (2021), 739–767.

[8] BBC. Us presdiential election results 2024. https://www.bbc.com/news/election/2024/us/results, 2024.

[9] BELCASTRO, L., CANTINI, R., MAROZZO, F., TALIA, D., AND TRUNFIO, P. Learning political polarization on social media using neural networks. *IEEE Access PP* (03 2020).

[10] CHENG, Z., AND LI, Y. Like, comment, and share on tiktok: Exploring the effect of sentiment and second-person view on the user engagement with tiktok news videos. *Social Science Computer Review 42*, 1 (2024), 201–223.

[11] CRUZ, L. C., DELA CRUZ, J. N., MAGLANGIT, S. F., MAGTIRA, M., IMPERIAL, J. M., AND RODRIGUEZ, R. Is twitter an echo chamber? connecting online public

sentiments to actual results from the 2019 philippine midterm elections. In *2022 International Conference on Asian Language Processing (IALP)* (2022), IEEE, pp. 57–62.

[12] DEMILLO, R. E., SOLANO, G., AND OCO, N. Philippine national elections 2022: Voter preferences and topics of discussion on twitter. In *2023 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)* (2023), pp. 724–729.

[13] DULAY, D., HICKEN, A., MENON, A., AND HOLMES, R. Continuity, history, and identity: Why bongbong marcos won the 2022 philippine presidential election. *Pacific Affairs 96*, 1 (2023), 85–104.

[14] GRUJIC, I., BOGDANOVIC-DINIC, S., AND STOIMENOV, L. Collecting and analyzing data from e-government facebook pages. *ICT Innovations* (2014), 86–96.

[15] HOLMES, R. D. The dark side of electoralism: Opinion polls and voting in the 2016 philippine presidential election. *Journal of Current Southeast Asian Affairs 35*, 3 (2016), 15–38.

[16] INTERNATIONAL, A. Five things to know about martial law in the philippines. *Amnesty International* (21 Sep 2022).

[17] JAZEERA, A. Donald trump to face 'hush money' criminal trial: What's it all about? *Al Jazeera* (15 February 2024).

[18] JOHNSON, H., AND HEAD, J. Leni robredo: The woman leading the philippines' 'pink revolution'. *BBC* (7 May 2022).

[19] LEE, C. Kamala harris is using social media to reach young voters. *TIME* (3 Sep 2024).

[20] LIU, B. *Sentiment analysis and opinion mining*. SPRINGER INTERNATIONAL PU, 2012.

[21] MACROHON, J. J. E., VILLAVICENCIO, C. N., INBARAJ, X. A., AND JENG, J.-H. A semi-supervised approach to sentiment analysis of tweets during the 2022 philippine presidential election. *Information 13*, 10 (2022).

[22] MENDOZA, M. E. H. Philippine elections 2022: Tiktok in bongbong marcos' presidential campaign. *Contemporary Southeast Asia: A Journal of International and Strategic Affairs 44*, 3 (2022), 389–395.

[23] PERNIA, R. A., AND PANAO, R. A. L. Electing the dictator's son: the 2022 philippine election in an era of authoritarian nostalgia and democratic decline. *Asian Affairs: An American Review* (2025), 1–22.

[24] SETIAWAN, D., ANANDA, D., AND KARTIKA, T. Media framing of donald trump's 2024 election victory: A case study on international media. *MEDIASI Jurnal Kajian dan Terapan Media, Bahasa, Komunikasi 6*, 1 (2025).

[25] SINPENG, A., GUEORGUIEV, D., AND ARUGAY, A. A. Strong fans, weak campaigns: Social media and duterte in the 2016 philippine election. *Journal of East Asian Studies 20*, 3 (2020), 353–374.

[26] VASSEY, J., DONALDSON, S. I., DORMANESH, A., AND ALLEM, J.-P. Themes in tiktok videos featuring little cigars and cigarillos: Content analysis. *Journal of Medical Internet Research 24*, 11 (2022).