Twitter Sentiment Analysis:
Understanding Political Trends and Voter Sentiment

by

Sai Uday Reddy, Bhumireddy

Inamullah, Mohammad

Sai Keerthi Reddy, Mora

Under the Supervision of

Dr. Theodore B. Trafalis

GALLOGLY
COLLEGE OF ENGINEERING
The UNIVERSITY of OKLAHOMA

Data Science & Analytics

DSA 5900, Fall 2024 (4 Credit Hours)

Contents

1 Introduction

Social media has revolutionized the way people communicate, especially in the context of political discussions. Platforms like Twitter provide a space where individuals freely express their opinions on current events, including political candidates and elections. With millions of tweets generated daily, Twitter offers an immense amount of data that can be analyzed to capture the pulse of public opinion.

In this project, we analyzed these conversations using sentiment analysis to classify tweets about political candidates as positive, negative, or neutral. To accomplish this, we fine-tuned several advanced machine learning models, including DistilBERT, BERTweet, RoBERTa, and an Aspect-Based RoBERTa model, using a manually labeled dataset. Alongside these models, we also incorporated VADER, a sentiment analysis tool specifically designed for social media data. Each of these tools was adapted to handle the unique challenges of Twitter, such as informal language, abbreviations, and hashtags. The Aspect-Based RoBERTa model further allowed us to analyze sentiment in specific contexts—such as individual candidates or particular issues—capturing the subtleties of public opinion.

The broader goal of this project is to provide a clearer picture of public sentiment in a highly polarized political climate. By analyzing the emotions and opinions shared on social media, we can identify key issues that matter to people, track emerging trends, and understand how candidates are perceived. We hope these insights can help policymakers, researchers, and the public make more informed decisions and foster healthier discussions. Additionally, the methods and findings from this project can serve as a starting point for future studies on political conversations and public sentiment, encouraging greater civic awareness and engagement.

Our approach is designed to capture the nuances of political sentiment on Twitter, offering a detailed analysis of how various candidates are perceived in online public discourse.

2 Objective

With the 2024 U.S. presidential election approaching, this project focuses on understanding and analyzing public sentiment on Twitter about key political figures, including Donald Trump, Joe Biden, and Kamala Harris. Using advanced Natural Language Processing (NLP) techniques and machine learning models like DistilBERT, BERTweet, RoBERTa, and Aspect-Based RoBERTa, we aim to capture shifts in public opinion over time, particularly in response to significant political events.

The primary goal of this project is to predict the winner of the election based on sentiment analysis. We've collected tweets from critical swing states during the months leading up to the election, ensuring that the dataset reflects the perspectives of voters in key electoral regions. By categorizing tweets as positive, negative, or neutral, we aim to uncover valuable insights into how candidates are perceived and how voter sentiment changes over time.

This project brings together expertise in data collection, preprocessing, and sentiment analysis, applying modern data science techniques to a real-world political challenge. Success will be measured by the accuracy of our sentiment classification and our ability to predict the likely outcome of the election based on the trends and patterns in public opinion observed on social media. Through this work, we aim to provide a data-driven understanding of voter sentiment during this pivotal election period.

3 Data

3.1 Data Source

For our project, we needed to collect a substantial number of tweets based on specific filters, such as location, date, user handle, and hashtags. While the Twitter API offers an official method for retrieving tweet data, we encountered limitations in terms of the available filtering options and access levels. The basic API did not provide the flexibility we required to apply all the necessary filters simultaneously. Moreover, the API's tiered access models posed a significant challenge. Given the volume of tweets we aimed to collect, none of the available tiers were suitable for our data collection needs, and the only tier that met our requirements came at a prohibitively high cost [1].

To overcome these limitations, we explored alternative approaches to gather the data. One potential solution was using Python-based Twitter scrapers. We tried various scrapers, each offering a different range of functionalities. However, every tool we tested had some limitations in terms of filter criteria, either lacking the ability to filter by location, date, or user handle, or being inefficient for the large-scale data collection we intended.

After careful consideration, we decided to use an online Twitter scraping tool provided by Apify. This tool allowed us to overcome the restrictions imposed by the traditional API and Python-based scrapers. The Apify scraper offered more advanced filtering options, enabling us to extract tweets based on the specific criteria we needed. Additionally, it supported the large-scale data collection necessary for our analysis, making it the most viable option for this project.

3.2 Data Collection

For our data collection, we leveraged a news article[2] from U.S. News to identify seven key swing states: Arizona, Georgia, Michigan, Nevada, North Carolina, Pennsylvania, and Wisconsin. These states were considered crucial in the context of the upcoming U.S. presidential election. The data collection spanned a six-month period, from March 15, 2024, to September 15, 2024, to capture sentiment and discussion leading up to the election.

The tweets were selected based on specific user handles related to the primary political figures in the election, such as @JoeBiden, @KamalaHarris, and @realDonaldTrump. Alongside these user handles, we targeted tweets containing broadly discussed hashtags relevant to the election. This approach ensured a representative dataset that reflected public sentiment and engagement during the specified period.

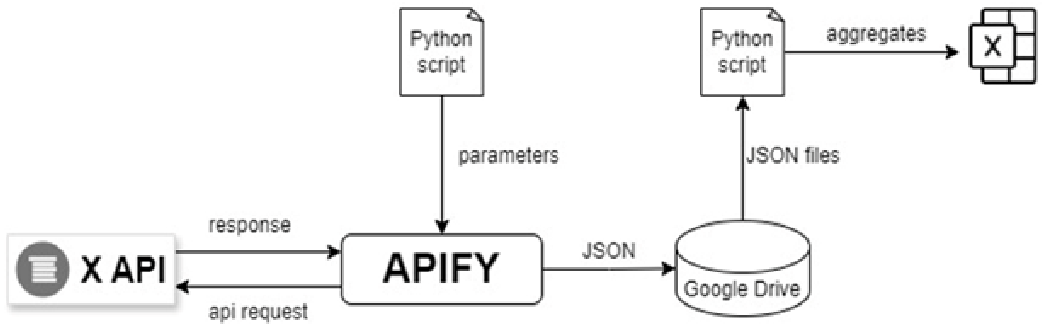

Figure 1: Workflow for data collection and aggregation using Apify

The data collection process, shown in Figure 1, starts with a Python script that triggers the Apify workflow to handle the scraping of Twitter data. Apify sends requests to Twitter's X API with specific filters like location, date, user handles, and hashtags. In response, the X API provides the requested tweet data, which Apify saves as JSON files directly to Google Drive for easy access. Once the data collection is complete, another Python script takes over to process the JSON files, combining them into a single, organized Excel file that includes only the most relevant columns needed for analysis. These fields included:

1. **id** : A unique identifier assigned to each tweet.
2. **tweet_text**: The content of the tweet, representing the textual data posted by the user.
3. **created_at**: The exact timestamp when the tweet was created and posted on Twitter.
4. **retweet_count**: The total number of times the tweet has been retweeted by other users.
5. **reply_count** : The total number of replies that the tweet has received from other users.
6. **like_count** : The total number of likes (or "favorites") the tweet has garnered from users.
7. **view_count** : The total number of views the tweet has accumulated, indicating its reach.
8. **state** : The U.S. state associated with the tweet, reflecting the geographical location from which it was fetched.
9. **Candidate** : The political candidate associated with the user handle or hashtag mentioned in the tweet (Trump, JoeBieden, KamalaHarris).
10. **party** : The political party associated with the user handle or hashtag mentioned in the tweet (e.g., Democratic or Republican).
11. **handle_or_hash** : Specifies whether the tweet pertains to a user handle or a hashtag related to the analysis.

By streamlining the dataset to these essential columns, we ensured a more focused and efficient analysis while preserving the key information required for understanding voter sentiment and engagement across the selected swing states.

3.3 Data Exploration

We collected a total of 110,077 records from seven key swing states over a six-month period. Since Twitter stores user locations as approximate bounding boxes [3] rather than precise coordinates, there is a possibility that some tweets may originate from neighboring states, leading to a slight overlap in the geographical distribution of the data. This potential overlap is a known limitation of geolocation data in social media platforms.

3.3.1 Distribution of Tweets Over Swing States

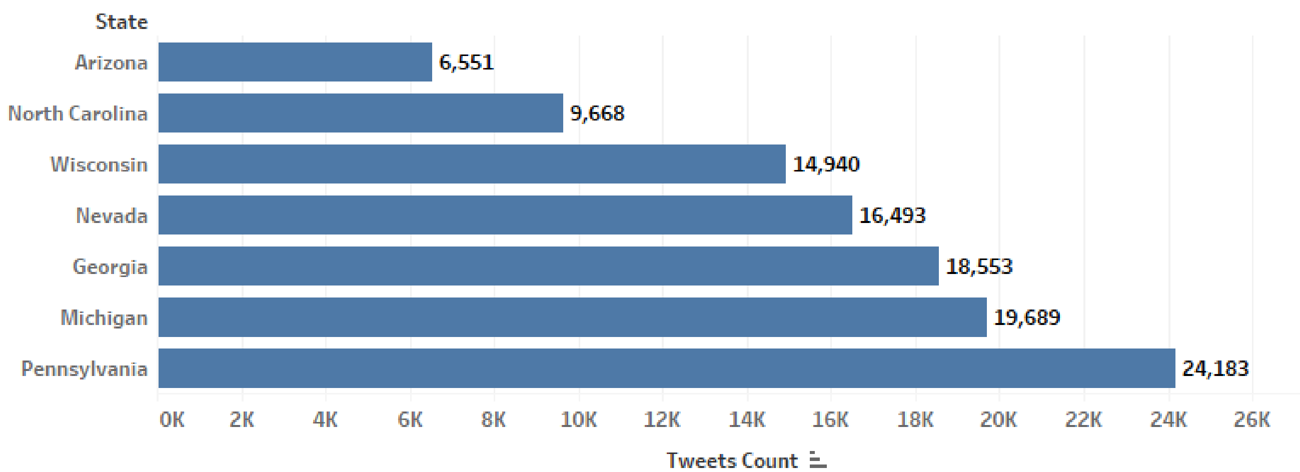

Figure 2: Distribution of Tweets Over Swing States, displaying the tweet counts for each state.

The distribution of tweets over swing states can be seen in Fig. 2. Pennsylvania leads with the highest tweet count at 24,183, followed by Michigan with 19,689 tweets and Georgia with 18,553 tweets. Nevada and Wisconsin also have significant engagement, with 16,493 and 14,940 tweets, respectively. North Carolina and Arizona have comparatively lower tweet counts, at 9,668 and 6,551. The variation in tweet counts across these states highlights differing levels of social media engagement among the swing states.

These differences may reflect varying levels of public interest or political activity across the states. States like Pennsylvania, Michigan, and Georgia generated considerably more Twitter activity, possibly due to higher population sizes, key political events, or active online communities, while Arizona and North Carolina saw lower engagement.

3.3.2 Distribution of Tweets Over Time

The graph, Fig. 3, displays the distribution of tweets over time, segmented by political affiliation: Democratic, Republican, and both. Throughout the observed period, tweet activity maintains a relatively steady flow, punctuated by significant surges during key political events. These surges

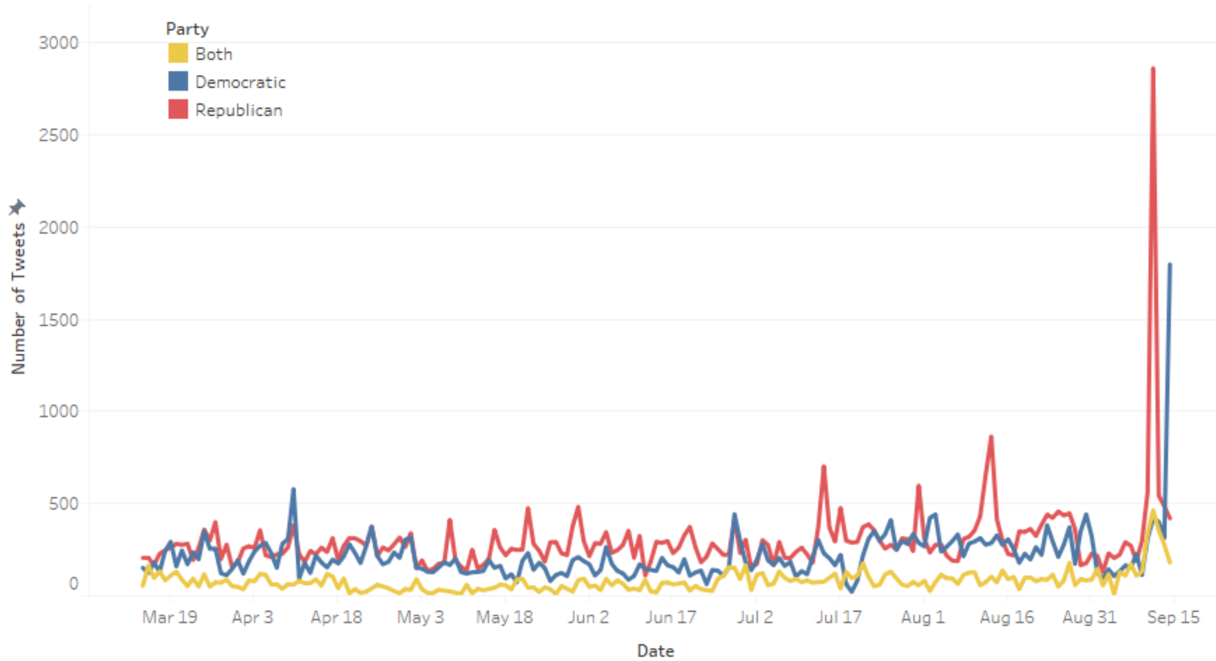

Figure 3: Distribution of Tweets Over Time, displaying the spikes in tweet activity around notable political events.

illustrate the public's tendency to engage more actively on social media during critical moments. For instance, around July 11, a sharp increase in tweets can be seen following the Trump assassination attempt, demonstrating how events of great political consequence can trigger heightened discourse and engagement on platforms like Twitter. Such events not only generate immediate responses but also often spark extended conversations that influence online activity for days or even weeks.

In addition to this event, other significant political occurrences also triggered spikes in tweet activity. Key moments, such as presidential debates, Kamala Harris's entry into the presidential race, and the Trump-Musk interview, show noticeable increases in tweets shortly after they occurred. These patterns demonstrate how pivotal political events influence online discussions and engagement across party affiliations.

3.3.3 Distribution of Tweets by State and Party

The figure, Fig. 4, shows the distribution of tweet counts by political affiliation (Democratic, Republican, and Both) across key swing states. Across these states, Republican candidates, particularly Donald Trump, dominate the conversation, with the highest tweet counts for the Republican party in several regions.

Democratic engagement is also substantial, particularly in states like Pennsylvania and Georgia, though generally lower than Republican tweet counts across most regions. The Both category,

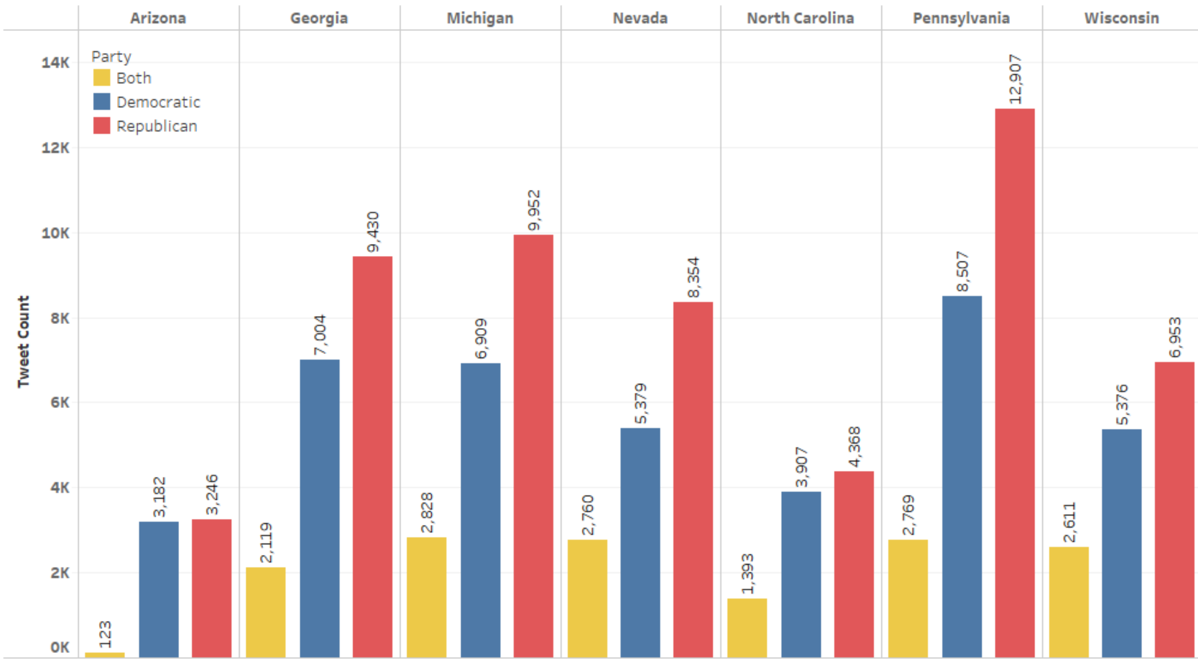

Figure 4: Distribution of Tweets by State and Party, showing how often each party is mentioned.

representing tweets mentioning multiple candidates, has the lowest engagement overall.

While this figure reflects the volume of tweets, it is important to note that the number of tweets does not necessarily indicate positive sentiment. The tweets may include both supportive and critical discourse, as social media activity around political figures often represents a broad spectrum of opinions. Therefore, high tweet counts for a particular candidate or party may reflect a mix of both praise and criticism.

3.3.4 Engagement Metrics Over Time

The figure, Fig.5, presents the evolving patterns of Twitter engagement over time, showcasing the dynamic interaction between users and content through various metrics. Engagement levels rise and fall, with certain moments marked by distinct peaks, reflecting periods of heightened public interest and interaction. These surges in activity often align with key political or social events, sparking increased visibility and user responses.

As engagement intensifies, users contribute through a variety of actions, including liking, sharing, and commenting on tweets. The data reveals moments when content resonates more broadly, drawing widespread attention and sparking conversations. During these peaks, users not only view the content but are also compelled to share their own perspectives, adding to the discourse and amplifying the reach of specific tweets.

The recurring spikes in activity suggest that social media engagement is closely tied to external events, with moments of high interaction providing a glimpse into the public's collective reaction.

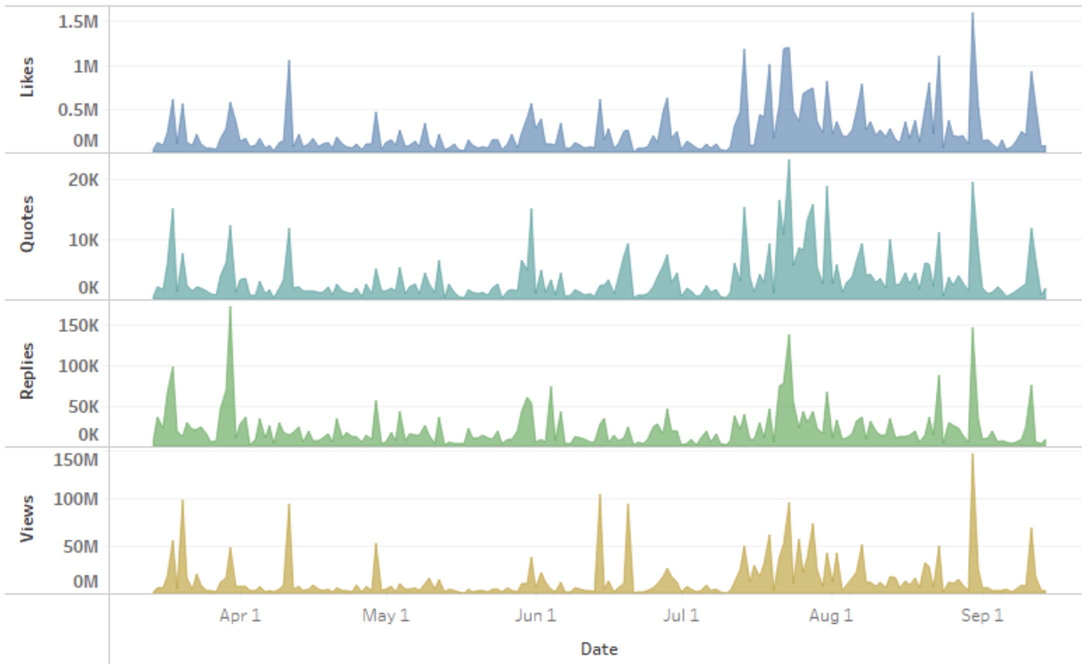

Figure 5: Distribution of Engagement Metrics over Time, showing how user interactions evolve through likes, quotes, replies, and views.

These patterns demonstrate how Twitter serves as a platform for real-time conversations, where significant moments generate a ripple effect of engagement, driving likes, shares, and discussions across the platform.

3.3.5 Average Engagement Metrics by Party

Fig. 6 provides interesting observations about how user interactions vary between Democratic, Republican, and multi-party (Both) tweets. Notably, Democratic-affiliated tweets tend to see the highest engagement in multiple metrics, such as retweets, replies, and quotes. This indicates that Democratic content is not only widely viewed but also actively shared and discussed by users. The higher quote counts suggest that users often feel compelled to add their opinions or commentary when sharing Democratic-affiliated tweets, driving deeper conversations.

On the other hand, while Republican-affiliated tweets show slightly lower average engagement in several categories, they still maintain significant interaction levels, particularly in likes and views. This suggests that although users may engage less frequently through quotes and replies, they still express their opinions through simpler interactions, such as likes. This might reflect a different mode of engagement among the audience, where Republican content is consumed more passively, with fewer instances of users adding their commentary or entering into discussions.

Tweets mentioning both parties show a balanced level of engagement, with metrics like views and replies falling between those of Democratic and Republican tweets. This suggests that such tweets

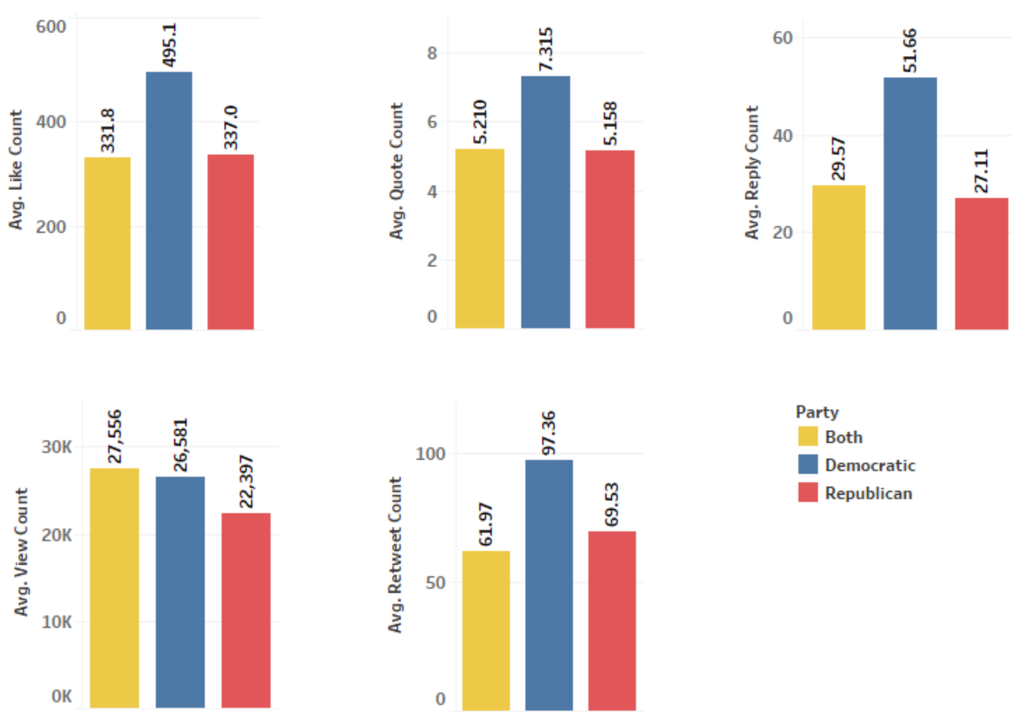

Figure 6: Average Engagement Metrics by Party, illustrating the variation in average engagement for each party.

attract a broader audience, encouraging interaction from users with different affiliations. These tweets may spark cross-party discussions, leading to a more evenly distributed engagement pattern. Overall, the figure highlights how content and political alignment influence user interaction on social media.

3.3.6 Tweets Across User Handles or Hashtags

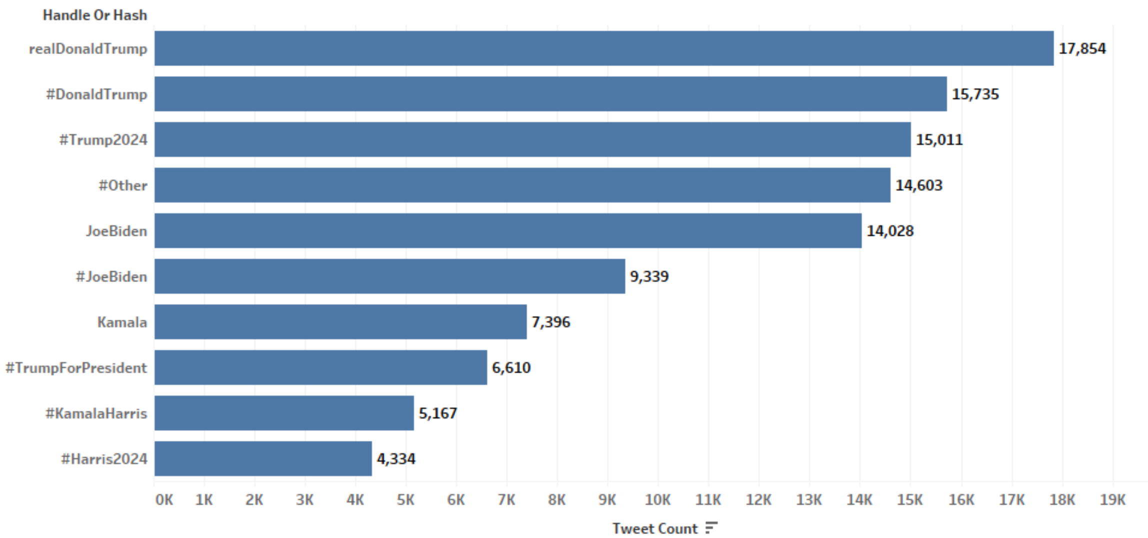

Figure 7: Breakdown of Tweets by User Handles and Hashtags, highlighting the most discussed figures and topics on Twitter.

The plot, Fig. 7, provides a breakdown of tweet counts by user handles and hashtags, revealing the most frequently used handles and hashtags in discussions. realDonaldTrump leads with the highest mentions at 17,854 tweets, followed by #DonaldTrump with 15,735 tweets. Other Trump-related hashtags, such as #Trump2024 and #TrumpForPresident, also rank prominently, reflecting strong engagement surrounding his political activities.

Mentions of Joe Biden and #JoeBiden are also significant, though they trail behind Trump-related mentions, with 9,339 and 7,396 tweets, respectively. Kamala Harris and associated hashtags, like #KamalaHarris and #Harris2024, show fewer mentions, with 5,167 and 4,334 tweets, indicating a relatively lower level of Twitter engagement.

It is worth noting that the #Other category, which accounts for 14,603 tweets, includes important election-related hashtags such as #USElection, #TrumpVsBiden, and #TrumpHarrisDebate, capturing broader discussions around the 2024 election. This breakdown emphasizes that Trump-related content dominates the social media landscape, followed by Biden, with Harris receiving less attention during this period.

3.4 Data Preparation

Preparing the dataset for sentiment analysis involved a carefully designed preprocessing pipeline to ensure data quality and relevance. Tweets labeled as "both," indicating no clear association with a specific party or candidate, were excluded, reducing the dataset to approximately 90,000 records for analysis. To enhance model accuracy, a subset of 1,300 tweets was manually labeled, providing high-quality training data that added depth to the dataset. This manual effort ensured the models could learn from accurately categorized examples, improving their performance on real-world data.

Unlike conventional cleaning methods that strip away elements like hashtags and user handles, we retained these components for their contextual value. Hashtags often indicated political alignment or key topics, while user handles frequently referenced specific candidates, making them essential for understanding the nuances of political discourse. Other steps included removing URLs and non-alphabetic characters to minimize noise. By adopting this tailored approach, which preserved meaningful context while eliminating irrelevant data, the dataset was optimized for sentiment analysis, providing a robust foundation for training and evaluating the models.

4 Methodology

For this project, we aim to predict which candidate has the highest chance of winning the upcoming presidential elections based on sentiment analysis of Twitter data. We use Large Language Models (LLMs) and VADER, a sentiment analysis tool, to classify the sentiment in tweets, providing insight into public opinion toward various candidates. Since our data is unlabeled, we rely on LLMs and fine-tuning techniques.

we use transformer-based models from the Hugging Face library and state-of-the- art Large Language Models (LLMs) such as DistilBERT, RoBERTa and BERTweet to perform sentiment analysis on Twitter data related to the upcoming presidential election. These models, pre-trained on large text corpora, are fine-tuned to classify the sentiment of tweets as positive, negative, or neutral, helping us understand public opinion on various political candidates[4].

4.1 VADER Sentiment Analysis Model: Limitations and Transition

Traditional machine learning (ML) models face several limitations when applied to sentiment analysis tasks. One major challenge is their dependence on labeled data, which can be both time-consuming and costly to acquire. Additionally, these models rely heavily on manual feature engineering to extract relevant features, requiring significant effort and domain expertise. Traditional ML models are also better suited for structured data and often struggle when dealing with unstructured text such as social media posts. Another significant drawback is their inability to capture nuanced contexts, including sarcasm or ambiguous language. Furthermore, these models can be computationally intensive, requiring substantial resources for training and inference.

To address some of these challenges, the project explored the use of VADER (Valence Aware Dictionary for Sentiment Reasoning), a pre-trained sentiment analysis tool from the NLTK module. One of VADER's key advantages is that it does not require labeled data, making it a more efficient choice for scenarios where annotated datasets are unavailable. Moreover, VADER is optimized for short, informal text, such as tweets, making it particularly well-suited for social media sentiment analysis [5].

Despite its advantages, VADER is not without limitations. Similar to traditional models, it struggles with understanding sarcasm and nuanced language [6], leading to occasional inaccuracies in sentiment predictions. Additionally, VADER does not support aspect-based sentiment analysis, meaning it cannot distinguish sentiment toward specific entities or aspects within a text. While effective for general sentiment analysis, these limitations highlight the need for more advanced models in complex scenarios.

Tweet Text	**Candidate**	**Actual Sentiment**	**Predicted Sentiment**
@JoeBiden Insighted an assassination attempt. He needs to face justice.	JoeBiden	Negative	Positive
#JoeBiden is stepping out of office. He is too old to run this country, and #DonaldTrump should not run for president either. He is also too old and a liar who only cares about himself. If I were president, I would give free insurance to all Americans and drop gas prices to $1 a gallon.	Trump	Negative	Positive

Table 1: Examples of VADER Sentiment for Tweets..

VADER is useful for sentiment analysis of short, informal text but struggles with context, sarcasm, and aspect-based sentiment analysis, making it less effective for complex tasks. It is a good starting point, but transformer-based models offer better performance for nuanced sentiment analysis.

4.2 Transformers Overview

The Transformer model proposed by Vaswani is a state-of-the-art architecture for sequence transduction tasks, revolutionizing the field of natural language processing (NLP). Unlike traditional models that rely on recurrent or convolutional layers, the Transformer uses an entirely new mechanism called self-attention [4]. This approach enables the model to process input sequences in parallel rather than sequentially, offering significant improvements in training efficiency. Self-attention allows each word in a sentence to directly attend to all other words, regardless of their position, capturing long-range dependencies more effectively than previous models.

At the core of the Transformer is the multi-head self-attention mechanism, which allows the model to learn multiple attention patterns at once. In this setup, the input tokens are projected into three vectors: Query (Q), Key (K), and Value (V). These vectors are used to compute attention scores, which determine the relevance of one word to another in the context of a given sequence. The multi-head design enables the model to capture various relationships within the input data simultaneously, improving its ability to model complex dependencies.

The Transformer architecture includes a position-wise feedforward network (FFN) after each attention layer, consisting of two linear transformations with a ReLU activation. This helps the model learn complex patterns. To maintain word order, positional encodings are added to the input embeddings, using sine and cosine functions. Transformers also use residual connections and layer normalization to prevent issues like vanishing gradients and improve training stability.

The model follows an encoder-decoder structure, with both parts containing layers of multi-head self-attention and feedforward networks [4]. These layers are stacked multiple times to capture higher-level representations. The Transformer's parallel processing capabilities make it faster and more scalable than previous models, such as RNNs, handling longer sequences and large datasets efficiently.

Due to these innovations, the Transformer has become the foundation for many modern NLP models, including BERT, RoBERTa, and BERTweet, which achieve top performance on various language tasks.

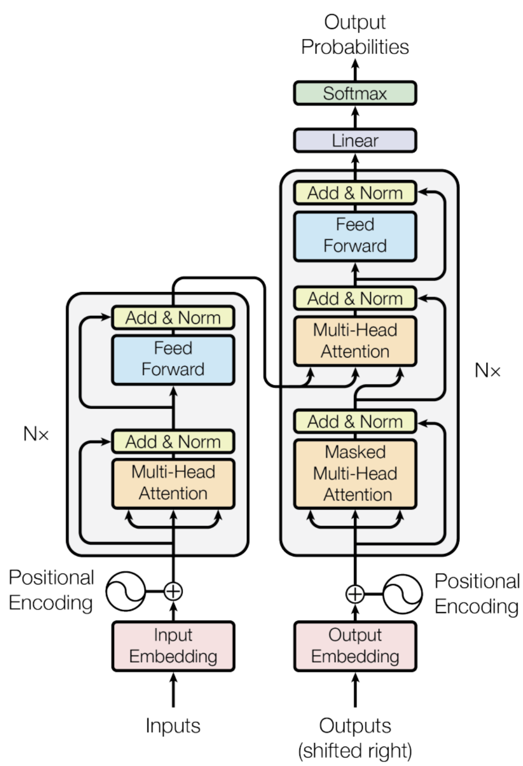

Figure 8: Transformer model architecture.

4.3 BERTweet

BERTweet is a transformer-based model specifically pre-trained on a large corpus of Twitter data, enabling it to effectively capture the unique linguistic characteristics of tweets. Tweets often contain informal language, slang, emojis, hashtags, and abbreviations, all of which are challenging for generic models. BERTweet's specialization in this domain makes it particularly well-suited for sentiment analysis and similar tasks on social media platforms [7].

The specific model we used, finiteautomata/bertweet-base-sentiment-analysis, is available on Hugging Face [8] and has been pre-trained for sentiment analysis. However, its pretraining does not include aspect-based sentiment analysis, which required us to fine-tune the model further to meet our project's requirements.

Before fine-tuning, the model achieved an accuracy of 36% on our dataset. This was likely due to the model's inability to handle specific nuances or domain-specific requirements in the unaltered dataset. After fine-tuning the model on our labeled data, the performance significantly improved, achieving an accuracy of 58%. This enhancement demonstrates the effectiveness of adapting pre-trained models to specific tasks through fine-tuning.

Tweet Text	**Candidate**	**Actual Sentiment**	**Predicted Sentiment**
@JoeBiden Insighted an assassination attempt. He needs to face justice.	JoeBiden	Negative	Negative
#JoeBiden is stepping out of office. He is too old to run this country, and #DonaldTrump should not run for president either. He is also too old and a liar who only cares about himself. If I were president, I would give free insurance to all Americans and drop gas prices to $1 a gallon.	Trump	Negative	Negative
The jobs are going away, the salaries aren't keeping up with record inflation. #maga #Trump2024 #Kamala caused chaos in the economy. Her advice was catastrophic.	Trump	Positive	Negative

Table 2: Examples of BERTweet Sentiment Analysis for Tweets.

While Examples 1 and 2 were correctly predicted, Example 3 highlights a limitation. Although the overall tweet sentiment is negative, the sentiment with respect to Trump is actually positive due to the supportive hashtags and statements. This issue arises because BERTweet is not explicitly pre-trained for aspect-based sentiment analysis, which requires focusing on specific entities within a text

BERTweet, while powerful, has some limitations. Its context window length of 128 tokens can truncate longer tweets or combined texts, leading to a loss of important information. The model is not pre-trained for aspect-based sentiment analysis, requiring additional fine-tuning to handle sentiments specific to entities or aspects. It can also face errors in aspect recognition, misclassifying sentiments when the overall sentiment of a tweet differs from the sentiment toward a particular entity. Additionally, BERTweet is resource-intensive, demanding more computational power than lighter models like VADER, making it less practical for very large datasets.

Despite these challenges, BERTweet outperforms lexicon-based models like VADER in capturing nuanced sentiments in tweets. Fine-tuning allowed us to tailor the model to our needs, achieving better accuracy and relevance. These limitations, however, highlight the importance of adapting models to fit specific tasks effectively.

4.4 RoBERTa (Robustly Optimized BERT Approach)

RoBERTa (Robustly Optimized BERT Pretraining Approach) is a transformer-based model that builds upon BERT's architecture but is trained with improved optimization techniques. It was designed to achieve better performance by removing certain training constraints such as the Next Sentence Prediction (NSP) task and using larger mini-batches and more training data [9]. RoBERTa excels at understanding contextual nuances in text, making it an ideal candidate for tasks like sentiment analysis.

For this project, we used the pretrained RoBERTa model, specifically cardiffnlp/twitter-roberta-base-sentiment, which was fine-tuned for sentiment analysis tasks on Twitter data [10]. This model is adept at handling the informal language, slang, emojis, and hashtags commonly found in tweets. However, as with BERTweet, the model was not originally pre-trained for aspect-based sentiment analysis, which required additional fine-tuning for our task.

Before fine-tuning, we observed that the RoBERTa model performed reasonably well but lacked the precision needed for our specific needs, achieving a baseline accuracy of 40%. After fine-tuning the model with our labeled data, the performance improved to 62%, demonstrating that fine-tuning effectively enhanced the model's ability to capture and predict the sentiment accurately in the context of specific candidates.

Tweet Text	**Candidate**	**Actual Sentiment**	**Predicted Sentiment**
@JoeBiden Insighted an assassination attempt. He needs to face justice.	JoeBiden	Negative	Negative
#JoeBiden is stepping out of office. He is too old to run this country, and #DonaldTrump should not run for president either. He is also too old and a liar who only cares about himself. If I were president, I would give free insurance to all Americans and drop gas prices to $1 a gallon.	Trump	Negative	Negative
The jobs are going away, the salaries aren't keeping up with record inflation. #maga #Trump2024 #Kamala caused chaos in the economy. Her advice was catastrophic.	Trump	Positive	Negative

Table 3: Examples of RoBERTa Sentiment Analysis for Tweets.

RoBERTa, a powerful model for natural language processing, demonstrates notable strengths in handling context and informal language. However, it does have several limitations that can impact its performance in specific tasks, particularly in sentiment analysis for social media platforms like Twitter.

One limitation is its context window length. RoBERTa, like other transformer models, has a fixed input size (usually 512 tokens) [9]. This can cause truncation of longer tweets or concatenated texts, which may result in the loss of important contextual information that could influence sentiment analysis outcomes. Furthermore, RoBERTa is not pre-trained for aspect-based sentiment analysis. While it can capture overall sentiment, it struggles to differentiate between sentiments directed at different entities or aspects in a text. Fine-tuning is required to address this issue and improve its ability to analyze sentiment toward specific candidates or topics accurately.

Additionally, aspect recognition remains a challenge for RoBERTa. In some instances, such as the example involving Trump in the results section, the model may misinterpret the sentiment directed at the candidate, even if the overall sentiment of the tweet is negative. This highlights the model's difficulty in accurately identifying sentiments related to specific entities when the broader context may suggest a different interpretation.

Another limitation is its computational intensity. RoBERTa, being a large transformer model, is computationally expensive compared to lightweight models like VADER. As a result, it requires substantial resources to perform real-time or large-scale sentiment analysis, making it less efficient for scenarios with limited computational capacity.

In conclusion, RoBERTa represents a significant advancement in sentiment analysis, offering superior performance over traditional models. Its ability to understand context and handle informal language, especially in the form of tweets, makes it an excellent choice for many sentiment analysis tasks. The model's performance can be significantly improved through fine-tuning, increasing its accuracy from 40% to 60%. However, its limitations in aspect-based sentiment analysis and its resource-intensive nature highlight the need for careful adaptation and optimization for domain-specific tasks. Despite these challenges, RoBERTa remains a powerful tool in the field of natural language processing, capable of delivering valuable insights when properly fine-tuned and applied.

4.5 DistilBERT

DistilBERT is a smaller and faster version of the BERT model, designed to address challenges associated with deploying large-scale pre-trained models, especially in scenarios with constrained computational resources or limited budgets. Unlike traditional approaches that focus on task-specific distillation, DistilBERT leverages knowledge distillation during the pre-training phase to create a general-purpose language representation model. This innovative approach enables DistilBERT to retain 97% of BERT's performance while reducing its size by 40% and operating 60% faster [11].

For this project, we started with the pre-trained DistilBERT model (`distilbert-base-uncased`) [12] and customized it for our needs by fine-tuning it to handle multi-class sentiment classification. While DistilBERT is originally designed for binary classification (positive and negative), we expanded its capabilities to include a third category: neutral. This adjustment was crucial for capturing the subtle and varied sentiments often found in political discussions on Twitter, where opinions are not always polarized. Fine-tuning the model helped it better understand and classify the complex range of sentiments expressed in the tweets we analyzed.

When we started with the pre-trained DistilBERT model, it was only set up for binary classification, so we couldn't directly measure its performance against our goal of classifying tweets into three categories: positive, negative, and neutral. The lack of a neutral label meant the initial accuracy didn't fully reflect how well the model could handle the complexity of our task. After fine-tuning the model with our labeled dataset to include the neutral category, we saw a significant improvement, achieving an accuracy of 71%. This demonstrated how tailoring the model to our specific needs made it much better at understanding and classifying the diverse range of sentiments expressed in political tweets.

Tweet Text	**Candidate**	**Actual Sentiment**	**Predicted Sentiment**
@JoeBiden Insighted an assassination attempt. He needs to face justice.	JoeBiden	Negative	Negative
#JoeBiden his stepping out office his to dam old to run this country and #DonaldTrump should not run for president his also to old and a lier he only cares about himself if I was president I would give free insurance to all the American people drop the prices gas to $1 a gallon.	Trump	Negative	Negative
The jobs are going away, the salaries aren't keeping up with record inflation. #maga #Trump2024 #Kamala caused chaos in the economy. Her advice was catastrophic	Trump	Positive	Positive

Table 4: Examples of DistilBERT Sentiment for Tweets.

While DistilBERT is efficient and effective, it does have a few limitations. Its fixed input size of 512 tokens means longer tweets or threads might get cut off, potentially losing important context. Additionally, the model's performance heavily depends on the quality and diversity of the fine-tuning dataset. Biases or gaps in the training data can lead to errors, particularly with less common or nuanced sentiments.

DistilBERT proved to be a fast and reliable choice for analyzing political tweets, especially after fine-tuning it to include a neutral sentiment category. Its ability to balance strong performance with computational efficiency makes it ideal for large-scale analysis. While it has some limitations, such as handling longer inputs and reliance on quality training data, these challenges are manageable. Overall, it's a powerful tool for capturing public sentiment in real-world scenarios.

4.6 RoBERTa ABSA

For this project, the pre-trained "cardiffnlp/twitter-roberta-base-sentiment" model from Hugging Face was adapted to perform Aspect-Based Sentiment Analysis (ABSA). This approach enabled sentiment analysis at a granular level, identifying sentiments toward specific aspects (handle_or_hash, Hashtages or handles that are related to respective political Candidates) within tweets[13].

The chosen model, RoBERTa, was optimized for sentiment analysis on Twitter data, capable of handling the informal language, abbreviations, and hashtags often found in tweets. To adapt the model for ABSA, the input format was modified to incorporate both the tweet content and the associated aspect in the form : tweet_text [ASPECT] handle_or_hash This adjustment allowed the model to focus on sentiment specific to the aspect while maintaining the context provided by the entire tweet.

Following the successful fine-tuning of the model, It is fed to predict sentiment with input combined with the tweet text and its aspect, which was tokenized and passed to the model and the results were mapped back to the original sentiment labels (Negative, Neutral, Positive).

Tweet Text	**Candidate**	**Actual Sentiment**	**Predicted Sentiment**
@JoeBiden Insighted an assassination attempt. He needs to face justice.	JoeBiden	Negative	Negative
#JoeBiden his stepping out office his to dam old to run this country and #DonaldTrump should not run for president his also to old and a lier he only cares about himself if I was president I would give free insurance to all the American people drop the prices gas to $1 a gallon.	Trump	Negative	Negative
The jobs are going away, the salaries aren't keeping up with record inflation. #maga #Trump2024 #Kamala caused chaos in the economy. Her advice was catastrophic	Trump	Positive	Positive

Table 5: Examples of Roberta ABSA Sentiment for Tweets.

5 Results & Analysis

5.1 Performance Comparison of Models

Model	**Accuracy**	**Precision**	**Recall**	**F1 Score**
RoBERTa (ABSA)	0.78	0.783	0.766	0.766
DistilBERT	0.71	0.703	0.695	0.692
RoBERTa	0.612	0.587	0.608	0.584
BERTweet	0.58	0.575	0.584	0.555

Table 6: Performance Comparison of Models on Sentiment Analysis.

Comparative performance of four NLP models RoBERTa ABSA, RoBERTa, BERTweet, and DistilBERT on sentiment analysis tasks, evaluated using metrics such as Accuracy, Precision, Recall, and F1 Score. The results highlight significant differences in the models' ability to handle the task effectively.

Among the models, RoBERTa ABSA demonstrates the best performance across all metrics, achieving an accuracy of 0.78 and an F1 score of 0.766. These results indicate its robustness and suitability for sentiment analysis tasks. DistilBERT also performs well, with an accuracy of 0.71 and an F1 score of 0.69, showcasing its balance between computational efficiency and performance.

In comparison, RoBERTa achieves moderate performance, with an accuracy of 0.61 and an F1 score of 0.584. This suggests that while it is effective, it does not reach the level of specialization seen in RoBERTa ABSA. On the other hand, BERTweet performs the lowest, with an accuracy of 0.58 and an F1 score of 0.555, indicating its limitations for this particular task despite its focus on social media text.

Overall, the results underline the importance of using task-specific models like RoBERTa ABSA for achieving higher performance in sentiment analysis tasks. Models such as BERTweet, while useful in other contexts, may not be as effective in this domain.

5.2 Distribution Of sentiment by Party:

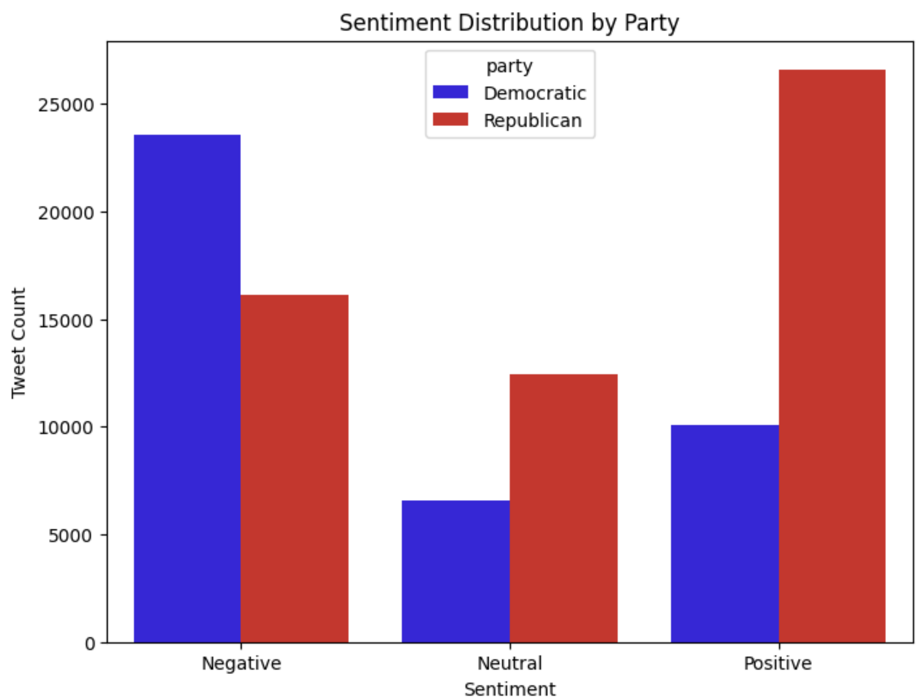

Figure 9: Sentiment Distribution by Party

Figure 9 provides a clear visualization of the results obtained from the fine-tuned RoBERTa (ABSA) model. It highlights the sentiment distribution across tweets mentioning the Democratic and Republican parties, categorized into three sentiment classes: negative, neutral, and positive.

It is evident that the Democratic Party received a significantly higher proportion of negative sentiments compared to the Republican Party. This indicates that a considerable portion of tweets mentioning Democratic-related topics expressed criticism or dissatisfaction. On the other hand, the Republican Party is associated with a noticeably larger volume of both positive and neutral sentiments. This suggests a more favorable or balanced tone toward Republican-related aspects, reflecting comparatively less critical public discourse.

5.3 Distribution Of sentiment over time

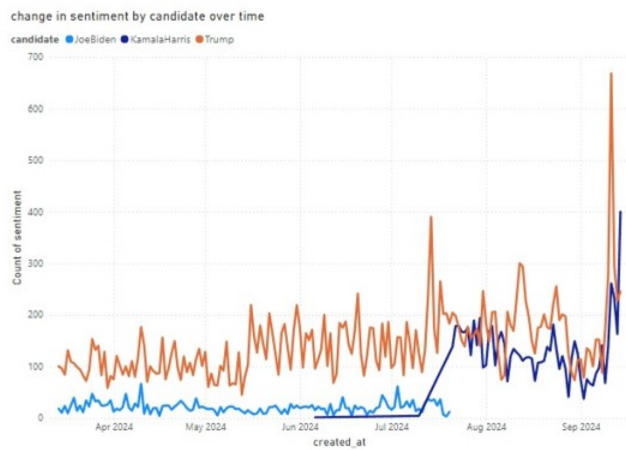

Figure 10: Candidates Positivity Trend Over Time.

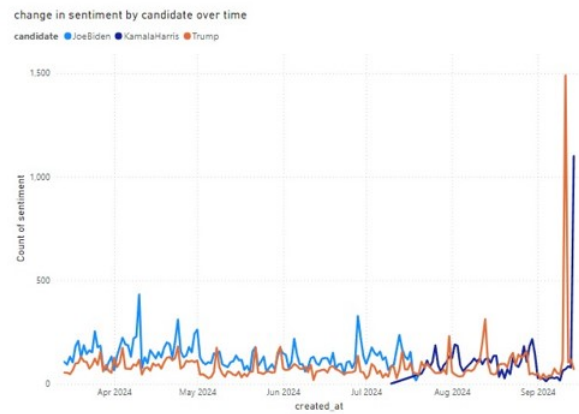

Figure 11: Candidates Negativity Trend Over Time.

The positive sentiment trends over time reveal significant patterns in public opinion regarding the candidates Joe Biden, Kamala Harris, and Donald Trump. Joe Biden consistently maintained lower levels of positive sentiment throughout the observed period compared to Trump. A noticeable shift occurred when Biden announced Kamala Harris as the new presidential candidate. Following the announcement, Harris managed to generate more positive sentiment than Biden. However, her positive sentiment levels still fell short when compared to Trump, who consistently dominated in positive sentiment trends.

A particularly sharp spike in Trump's positive sentiment was recorded around July 10th, coinciding with an attempted assassination event. This incident appears to have drawn substantial public attention and may have temporarily boosted positive sentiment for Trump. This trend underscores the dynamic nature of public sentiment, which can be significantly influenced by major events and candidate announcements.

The negative sentiment trends over time present a contrasting narrative to the positive sentiment trends. Joe Biden consistently received more negative sentiment compared to Donald Trump, who maintained relatively lower levels of negative sentiment. This pattern highlights Biden's greater challenges in managing public opinion.

Toward the end of the timeline, a significant spike in negative sentiment was observed for both

Kamala Harris and Donald Trump. This surge corresponds to the Trump vs. Harris debate, an event that generated considerable public attention and polarization. The debate not only amplified negative sentiment but also contributed to increased positive sentiment, as reflected in earlier trends, showcasing the polarizing impact of high-stakes political events.

5.4 Comparative Analysis — Predictive Sentiment vs. Election Outcomes

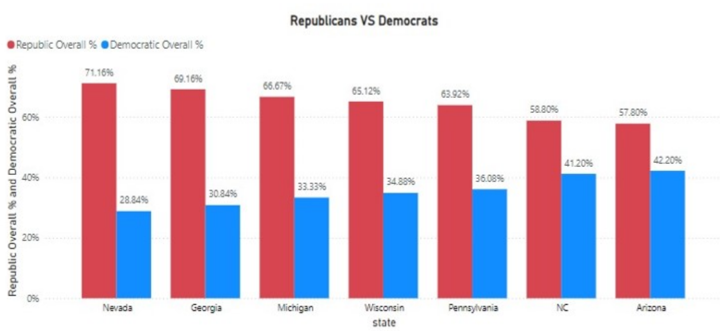

Figure 12: Overall Predicted Sentiment.

State	Republican	Democratic
Nevada	51	47
Georgia	51	48
Michigan	50	48
Wisconsin	49.6	48.8
Pennsylvania	50	49
NC	51	48
Arizona	52	47

Figure 13: Election Outcomes Percentage.

The comparison between the predicted sentiment by the model and the actual election outcomes reveals interesting insights. The predicted sentiment, as seen in the plot, reflects the general trend of public opinion as captured from social media data. It incorporates an assumption that negative sentiment toward one party translates to positive sentiment toward the opposing party. This approach stems from the manual labeling process, where tweets are categorized based on the hashtags or handles through which they were extracted.

For instance, in Nevada, the predicted sentiment for the Republican Party is significantly higher than for the Democratic Party, with Republicans showing a sentiment share of 71.16% compared to the Democrats' 28.84%. This aligns with the actual election outcomes, where Republicans received 51% of the votes compared to 47% for the Democrats. Similarly, states like Georgia, Michigan, and Arizona demonstrate a consistent trend, where the model's sentiment prediction correlates with the actual voting percentages, albeit with varying magnitudes.

The methodology accounts for the observed trend during manual annotation, where negative sentiments directed at a particular party often reflected positive attitudes toward the rival party. This is particularly relevant for swing states like Wisconsin and Pennsylvania, where close sentiment proportions and election results highlight the competitive nature of the political landscape.

6 Deliverables

The deliverables for this project include a comprehensive GitHub repository that houses all the necessary code and resources for data collection, preprocessing, and fine-tuning. The repository features implementations of the four machine learning models—DistilBERT, BERT, RoBERTa, and aspect-based RoBERTa—used for sentiment analysis, along with scripts for collecting and cleaning Twitter data. Everything is organized to make it easy for others to use and build upon.

The repository also includes the labeled dataset used for fine-tuning, as well as the final versions of the fine-tuned models. Clear and detailed documentation is provided to guide users through the entire process, from data ingestion and preparation to model training and evaluation. It also explains the reasoning behind the methods and tools chosen, ensuring transparency and replicability.

Another key outcome of this project is this report, which outlines the entire workflow and highlights the challenges faced and the solutions developed along the way. It serves not just as a summary of the work done but as a valuable reference for anyone interested in political sentiment analysis. Together, these deliverables create a strong foundation for future research and provide a practical framework for understanding public sentiment through social media.

7 References

[1] *Twitter x developer home*, Accessed: 2024-10-08. [Online]. Available: `https://developer.x.com/en`.

[2] E. D. Jr., "7 states that could sway the 2024 presidential election," *U.S. News & World Report*, Oct. 2, 2024. [Online]. Available: `https://www.usnews.com/news/elections/articles/7-swing-states-that-could-decide-the-2024-presidential-election`.

[3] *Place object model*, Accessed: 2024-10-08. [Online]. Available: `https://developer.x.com/en/docs/x-api/data-dictionary/object-model/place`.

[4] A. Vaswani *et al.*, *Attention is all you need*, 2023. arXiv: `1706.03762 [cs.CL]`. [Online]. Available: `https://arxiv.org/abs/1706.03762`.

[5] C. Hutto and E. Gilbert, "Vader: A parsimonious rule-based model for sentiment analysis of social media text," *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 8, no. 1, 2014. [Online]. Available: `https://doi.org/10.1609/icwsm.v8i1.14550`.

[6] A. Indian, P. Manethia, G. Meena, and K. Mohbey, "Decoding emotions: Unveiling sentiments and sarcasm through text analysis," in *The Future of Artificial Intelligence and Robotics. ICDLAIR 2023*, D. Pastor-Escuredo, I. Brigui, N. Kesswani, S. Bordoloi, and A. Ray, Eds., ser. Lecture Notes in Networks and Systems, vol. 1001, Springer, Cham, 2024, pp. 762–774. DOI: `10.1007/978-3-031-60935-0_62`.

[7] D. Q. Nguyen, T. Vu, and A. T. Nguyen, "Bertweet: A pre-trained language model for english tweets," *arXiv preprint arXiv:2005.10200*, 2020. [Online]. Available: `https://arxiv.org/abs/2005.10200`.

[8] HuggingFace, *Finiteautomata/bertweet-base-sentiment-analysis*, 2021. [Online]. Available: `https://huggingface.co/finiteautomata/bertweet-base-sentiment-analysis`.

[9] Y. Liu *et al.*, "Roberta: A robustly optimized bert pretraining approach," *arXiv preprint arXiv:1907.11692*, 2019. [Online]. Available: `https://arxiv.org/abs/1907.11692`.

[10] HuggingFace, *Cardiffnlp/twitter-roberta-base-sentiment*, 2021. [Online]. Available: `https://huggingface.co/cardiffnlp/twitter-roberta-base-sentiment`.

[11] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, "Distilbert, a distilled version of bert: Smaller, faster, cheaper and lighter," *arXiv preprint arXiv:1910.01108*, 2020, Version 4, 1 March 2020. [Online]. Available: `https://arxiv.org/abs/1910.01108`.

[12] HuggingFace, *Distilbert-base-uncased*, 2019. [Online]. Available: `https://huggingface.co/distilbert/distilbert-base-uncased`.

[13] S. Onalaja, E. Romero, and B. Yun, "Aspect-based sentiment analysis of movie reviews," *SMU Scholar*, 2021. [Online]. Available: `https://scholar.smu.edu/cgi/viewcontent.cgi?article=1205&context=datasciencereview`.

8 Self-Assessment

- Gained hands-on experience in sentiment analysis using NLP models like RoBERTa, BERTweet, and DistilBERT.
- Learnend how to analyze political discourse on social media and predict sentiment trends related to election outcomes using Twitter data.
- Improved our skills in data preprocessing, including handling unstructured text, managing large datasets for sentiment analysis.
- We have applied performance metrics which were taught to us to evaluate model performance using metrics such as accuracy, precision, recall, and F1 score.
- Gained experience in fine-tuning pre-trained models, specifically RoBERTa ABSA, to perform aspect-based sentiment analysis for political content.
- We independently learned how to collect and process political tweets using the Twitter API, ensuring relevant data was extracted for sentiment analysis.
- We have used various data analysis techniques, using PowerBI and Python libraries like Matplotlib and Seaborn to visualize engagement metrics, sentiment trends, and election results.
- Developed an understanding of predictive analytics, correlating sentiment trends with actual election outcomes to gauge the accuracy of sentiment-based predictions.
- Gained experience in visualizing Twitter engagement metrics over time, including likes, shares, replies, and views, which helped us better understand social media trends and their impact on public opinion.