**ELECTION SENTIMENT ANALYSIS**

**CS6611 – CREATIVE AND INNOVATIVE PROJECT**

***Submitted by***

**ARAVIND KRISHNAN S 2021103509**

**NITHISH KUMAR R A 2021103548**

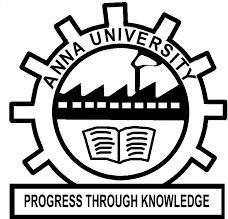
**PAVISHEK T V 2021103721**

***in partial fulfilment of the requirements for the award of the degree of***

**BACHELOR OF ENGINEERING**

**in**

**COMPUTER SCIENCE AND ENGINEERING**



**College of Engineering Guindy**

**ANNA UNIVERSITY, CHENNAI - 600 025**

**MAY 2024**

**ANNA UNIVERSITY, CHENNAI - 600 025**

**BONAFIDE CERTIFICATE**

Certificate that this project request titled **ELECTION SENTIMENT ANALYSIS** is the bonafide work of **ARAVIND KRISHNAN S (2021103509) NITHISH KUMAR R A (2021103548)**, **PAVISHEK T V (2021103721)** who carried out the project work under my supervision, for the fulfilment of the requirements as part of the CS6611 – Creative and Innovative Project.

|  |  |  |
| --- | --- | --- |
| **Dr. S. VALLI**  **Professor**  **HEAD OF THE DEPARTMENT**  Department of  Computer Science  and Engineering,  Anna University,  Chennai – 600025. | **Dr. P. UMA MAHESHWARI**  **Professor**  **SUPERVISOR**  Department of  Computer Science  and Engineering,  Anna University,  Chennai – 600025. | **M.S. KARTHIKA DEVI**  **Assistant Professor**      **SUPERVISOR**  Department of  Computer Science  and Engineering,  Anna University,  Chennai – 600025. |

## 

## 

## ABSTRACT

In the dynamic landscape of modern politics, comprehending the sentiments of voters has emerged as a pivotal element for effective campaigning and policy formulation. This abstract encapsulates an in-depth exploration of election sentiment analysis, employing cutting-edge data analytics methodologies to unravel the multifaceted dimensions of voter perspectives. Leveraging the power of natural language processing, sentiment analysis, and machine learning algorithms, this study endeavours to decode sentiments expressed across diverse platforms, offering profound insights into the intricate tapestry of voter behaviours, preferences, and concerns. Through meticulous scrutiny of textual data extracted from sprawling social media platforms, this research enriches the discourse surrounding electoral dynamics. By facilitating a granular understanding of sentiment patterns, it equips political stakeholders with invaluable tools to craft astute strategies and cultivate a nuanced comprehension of democratic processes.

## ACKNOWLEDGEMENT

Foremost, we would like to express our sincere gratitude to our project guide, **Dr. P. UMA MAHESHWARI**, Professor and **M.S. KARTHIKA DEVI,** Assistant Professor , Department of Computer Science and Engineering, College of Engineering Guindy, Chennai for their constant source of inspiration. We thank them for the continuous support and guidance which was instrumental in taking the project to successful completion.

We are grateful to **Dr. S. Valli**, Professor and Head, Department of Computer Science and Engineering, College of Engineering Guindy, Chennai for her support and for providing necessary facilities to carry out for the project.

We would also like to thank our friends and family for their encouragement and continued support. We would also like to thank the Almighty for giving us the moral strength to accomplish our task.

**Aravind Krishnan S Nithish Kumar R A Pavishek T V**

## 

**TABLE OF CONTENTS**

|  |  |  |  |
| --- | --- | --- | --- |
| **CHAPTER NO** | **TITLE** | | **PAGE NO** |
|  | **ABSTRACT** | | **5** |
|  | **ACKNOWLEDGEMENT** | | **6** |
| **1.** | **INTRODUCTION** | | **11** |
|  | 1.1 OBJECTIVE | | **12** |
|  | 1.2 PROBLEM STATEMENT | | **12** |
|  | 1.3 NEED FOR THE SYSTEM | | **12** |
|  | 1.4 SIGNIFICANCE OF THE STUDY | | **13** |
|  | 1.5 SCOPE OF THE PROJECT | | **14** |
| **2.** | **LITERATURE SURVEY** | | **15** |
|  | 2.1 LITERATURE REVIEW | | **15** |
|  | 2.2 SUMMARY OF LITERATURES | | **17** |
| **3.** | **SYSTEM DESIGN** | | **18** |
|  | | **3.1 SYSTEM OVERVIEW** | **18** |
|  | **3.2** **MODULE DESCRIPTION** | | **20** |
|  | **3.2.1 DATA COLLECTION AND**  **PRE-PROCESSING** | | **20** |
|  | 3.2.1.1 DATA SOURCES IDENTIFICATION | | **20** |
|  | 3.2.1.2 DATA FILTERING | | **20** |
|  | 3.2.1.3 DATA CLEANING | | **20** |
|  | 3.2.1.4 NORMALISATION | | **20** |
|  | **3.2.2** **SENTIMENT ANALYSIS** | | **21** |
|  | 3.2.2.1 TEXT TOKENIZATION | | **21** |
|  | 3.2.2.2 SENTIMENT LEXICON USAGE | | **21** |
|  | 3.2.2.3 SENTIMENT CLASSIFICATION | | **21** |
|  | 3.2.2.4 CONTEXTUAL ANALYSIS | | **21** |
|  | 3.2.2.5 VISUALIZATION | | **21** |
|  | **3.2.3** **POLARITY MAPPING** | | **22** |
|  | 3.2.3.1 TOPIC IDENTIFICATION | | **22** |
|  | 3.2.3.2 SENTIMENT SCORE ASSIGNMENT | | **22** |
|  | 3.2.3.3 POLARITY MAPPING | | **22** |
|  | 3.2.3.4 VISUALIZATION | | **22** |
|  | **3.2.4 PREDICTION** | | **23** |
|  | 3.2.4.1 MODEL TRAINING | | **23** |
|  | 3.2.4.2 VISUALIZATION AND REPORTING | | **23** |
| **4.** | **IMPLEMENTATION AND RESULT**  **4.1 DATASET DESCRIPTION**    **4.2 SYSTEM REQUIREMENTS**  4.2.1 HARDWARE REQUIREMENT  4.2.2 SOFTWARE REQUIREMENT  **4.3 IMPLEMENTATION DETAILS** | | **24**  **24**  **24**  **24**  **25**  **25** |
|  | 4.3.1 IMPORT DATASETS | | **25** |
|  | 4.3.2 INSPECTING DATASETS | | **27** |
|  | 4.3.3 FETCHING SENTIMENT SCORES | | **28** |
|  | 4.3.4 GENERATING POLARITY | | **29** |
|  | 4.3.5 CATEGORIZING BASED ON POLARITY | | **32** |
|  | 4.3.6 REMOVING NEUTRAL POLARITY | | **34** |
|  | 4.3.7 DROPPING NEUTRAL POLARITY TWEETS | | **35** |
|  | 4.3.8 FINAL DATASET | | **36** |
|  | 4.3.9 PREDICTION | | **36** |
|  | **4.4 TEST CASES AND PERFORMANCE METRICS** | | **38** |
| **5.** | **CONCLUSION** | | **41** |
|  | 5.1 CONCLUSION | | **41** |
|  | 5.2 FUTURE WORK | | **41** |
|  | **REFERENCES** | | **42** |
|  |  | |  |

## 

**TABLE OF FIGURE**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **FIGURE NO** | | **TITLE** | | **PAGE NO** |
| **3.1** | | **BLOCK DIAGRAM** | | **18** |
| **4.1** | | **IMPORT DATASETS** | | **26** |
| **4.2** | | **INSPECTING DATASET** | | **27** |
| **4.3** | | **SENTIMENT SCORE** | | **29** |
| **4.4** | | **GENERATING POLARITY** | | **29** |
| **4.5** | | **CATEGORIZING TWEETS** | | **32** |
| **4.6** | | **REMOVING NEUTRAL TWEETS** | | **34** |
| **4.7** | | **DROPPING NEUTRAL TWEETS** | | **35** |
| **4.8** | | **FINAL DATASET** | | **36** |
| **4.9** | | **PREDICTION** | | **37** |
| **4.10** | | **TEST CASE 1** | | **39** |
| **4.11** | | **TEST CASE 2** | | **40** |
|  |
|  | |  |
|  | |  |

**CHAPTER 1**

# INTRODUCTION

### 1.1 OVERVIEW

In the ever-evolving landscape of political discourse, the ability to gauge public sentiment accurately has become increasingly vital, especially during elections. Traditional methods such as polls often struggle to capture the real-time pulse of public opinion, while social media platforms offer a vast reservoir of data that reflects the thoughts and feelings of millions. However, harnessing this data efficiently poses a significant challenge due to its sheer volume, complexity, and inherent biases.

This research endeavours to bridge this gap by leveraging advanced Natural Language Processing (NLP) techniques, particularly sentiment analysis, to automatically classify and analyse opinions expressed in election-related tweets. Sentiment analysis, a subset of NLP, holds the promise of automatically categorizing opinions as positive, negative, or neutral, thereby providing valuable insights into the collective mood of the electorate.

By harnessing the power of NLP, this study aims to unlock the potential of social media as a rich source of real-time, granular data on public sentiment during elections. Through the systematic analysis of vast amounts of textual data, this research seeks to offer policymakers, analysts, and the public a nuanced understanding of the prevailing sentiments, enabling informed decision-making and a deeper comprehension of the electoral landscape.

### 

* 1. **OBJECTIVE**
* Understanding public opinion
* Predicting election outcomes
* Enhancing democratic processes
* To develop robust sentiment analysis techniques to revolutionise our understanding of public opinion and its role in elections .

### 1.3 PROBLEM STATEMENT

Traditional methods like polls struggle to capture the real time pulse of public opinion during elections .Social media offers a massive dataset , but analysing it efficiently is difficult due to volume , complexity and bias.

The proposal is to develop and apply advanced NLP techniques like sentiment analysis to automatically classify and analyse opinions expressed in election-related tweets. Sentiment analysis can automatically classify opinion (positive,negative,neutral) expressed in election-related text, offering valuable insights.

### 

### 1.4 NEED FOR THE SYSTEM

The need for an election sentiment analysis system is paramount in today's digital age where understanding public opinion is crucial in the electoral process. With communication predominantly occurring online, platforms like social media, news outlets, and public forums serve as avenues for expressing political sentiments. Such a system is vital for several reasons. Firstly, it enables informed decision-making for political campaigns and policymakers by providing insights into public sentiment, ensuring alignment with voter preferences. Moreover, the system allows for real-time monitoring of sentiment, facilitating timely adjustments to campaign strategies in response to changing public opinion.

By analyzing sentiment across various topics and themes, it helps identify key issues that resonate with voters, enabling candidates to prioritize agendas effectively. Additionally, the system aids in mitigating risks associated with negative sentiment or controversies, safeguarding a candidate's reputation and campaign success. Furthermore, by fostering meaningful dialogue and addressing voter concerns, it enhances voter engagement and participation, crucial for a healthy democratic process. Ensuring fairness and transparency in elections is also facilitated by understanding overall sentiment towards the electoral process. Lastly, the system serves as a valuable tool for researchers, academics, and political analysts to study trends and patterns in public opinion, contributing to a deeper understanding of political dynamics over time. Overall, an election sentiment analysis system plays a pivotal role in enhancing democratic processes, facilitating informed decision-making, and ensuring responsiveness to public sentiment in the electoral arena.

### 

### 1.5 SIGNIFICANCE OF THE STUDY

The election sentiment analysis system holds immense significance across various domains. Firstly, it enables political stakeholders to make informed decisions by providing real-time insights into voter sentiment, thereby shaping campaign strategies, messaging, and policy priorities. Additionally, the system promotes responsive campaigning by allowing candidates and parties to adapt their approaches dynamically, addressing emerging issues and concerns among the electorate. Moreover, by enhancing voter engagement and participation, it strengthens democratic governance, fostering transparency and accountability in the electoral process.

Furthermore, the system aids in risk mitigation by identifying and addressing negative sentiment or controversies early on, safeguarding the reputations and electoral prospects of political actors. Beyond electoral contexts, the data generated by the system contributes to academic research, policy analysis, and understanding broader societal trends, thus advancing political science and public policy. Lastly, understanding sentiments internationally can also be crucial for diplomacy and international relations, aiding governments in navigating diplomatic relations and public perceptions abroad. Overall, the election sentiment analysis system democratizes access to public opinion, shapes responsive governance, and deepens understanding of electoral dynamics and political discourse.

**1.6 SCOPE OF THE PROJECT**

* The scope of this project encompasses several crucial components aimed at analysing voter sentiments during electoral events comprehensively.
* It involves gathering election-related textual data from diverse sources, including social media platforms, news articles, and public forums.
* The collected data undergoes rigorous pre-processing to ensure consistency and readiness for sentiment analysis.
* Utilizing natural language processing and machine learning techniques, sentiment analysis is then performed to decipher the sentiments expressed in the textual data.
* The project involves identifying key topics or themes within the data through topic modelling, providing insights into the issues driving voter sentiment.
* The findings of sentiment analysis are presented visually using various visualization techniques such as word clouds, sentiment timelines, and thematic maps.
* Through the analysis, actionable insights are derived to inform political strategies and decision-making.
* The accuracy and reliability of the sentiment analysis results are evaluated and validated using validation techniques and comparison with ground truth data.
* Overall, the project aims to contribute to the understanding of electoral dynamics and assist political stakeholders in making well-informed decisions based on voter sentiments.

**CHAPTER 2**

**LITERATURE SURVEY**

**2.1 LITERATURE REVIEW**

**SENTIMENT ANALYSIS**

Li et al,2020[4]. introduce a sentiment information based network model (SINM) for Chinese text sentiment analysis. Their approach incorporates transformer encoder and LSTM components. However, the effectiveness of the model heavily relies on the quality and quantity of the training data, potentially impacting accuracy if the data doesn't capture the nuances of Chinese sentiment expression accurately.

Woldemariam, Y,2022[2]. presents a cross-media analysis framework for sentiment analysis. The pipeline includes components like a chat room cleaner, NLP, and sentiment analyzer. Leveraging the Apache Hadoop framework and Stanford CoreNLP library with the Recursive Neural Tensor Network (RNTN) model, their approach acknowledges the diverse expressions of sentiment across different media. For instance, while a frowning emoji in text typically conveys negativity, the same expression in a video might indicate concentration or contemplation, highlighting the complexity of cross-media sentiment analysis.

Ilmania, A et al,2023[4] introduce a deep neural network approach for aspect detection and sentiment classification in Indonesian text. Their methodology includes two approaches for sentiment classification: one utilizing word embedding, sentiment lexicon, and POS tags with bi-GRU topology, and the other employing an aspect matrix to rescale word embedding vectors with CNN topology. However, effective ABSA models necessitate a substantial amount of accurately labelled Indonesian training data, encompassing both aspects and sentiment polarities, posing a challenge for model development.

Fujihira, et al,2019[3] propose a multilingual sentiment analysis method relying on word-to-word translation using a sentiment dictionary from any native language. They conducted sentiment classification experiments on tweets in English, German, French, and Spanish. However, they caution that languages convey sentiment nuances differently, and direct word-to-word translation may overlook subtleties like sarcasm, irony, and cultural references, potentially resulting in inaccurate sentiment classification.

Wang, Z et al,2023[8] present a social media analytics engine capable of fine-grained sentiment analysis and emotion sensing. Their method utilizes a social adaptive fuzzy similarity-based classification approach to automatically categorize text messages into sentiment categories (positive, negative, neutral, and mixed), while also identifying prevailing emotion categories (e.g., satisfaction, happiness, excitement, anger, sadness, and anxiety). However, they acknowledge the challenge posed by ambiguous language, slang, sarcasm, and context-specific expressions commonly found in social media posts, which can hinder the accurate interpretation of sentiment by analysis algorithms.

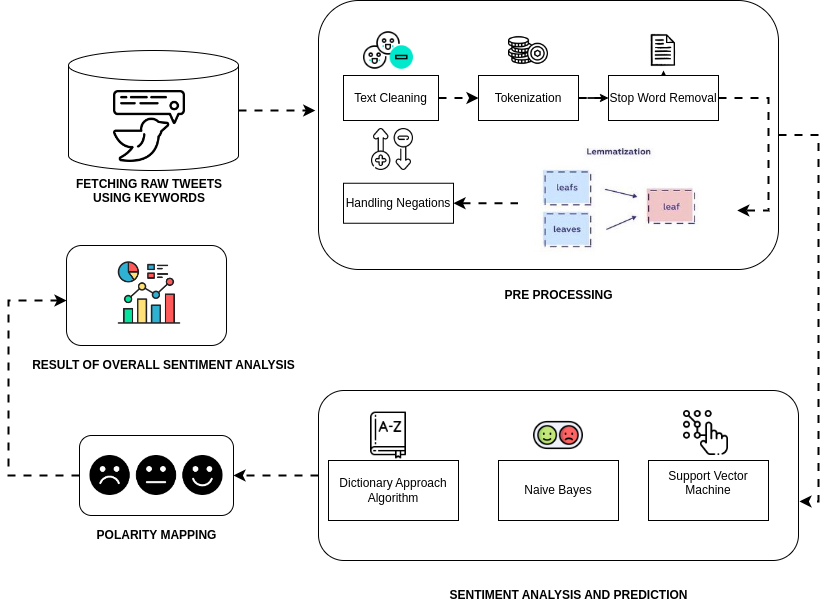
**2.2 SUMMARY OF LITERATURES**

* A summary of literature related to election sentiment analysis reveals a growing body of research focused on understanding public opinion dynamics in the context of electoral events.
* Studies in this area utilize a variety of methodologies, including natural language processing, machine learning, and social network analysis, to analyse sentiment expressed in textual data from sources such as social media, news articles, and public forums.
* Key themes explored in the literature include the impact of sentiment on electoral outcomes, the role of social media in shaping public opinion, and the challenges associated with sentiment analysis in diverse linguistic and cultural contexts.
* Additionally, researchers investigate the effectiveness of sentiment analysis techniquesin predicting election results, identifying key issues driving voter sentiment, and informing political strategies.
* Overall, the literature underscores the importance of election sentiment analysis in enhancing our understanding of voter behaviour, informing political decision-making, and fostering transparency and accountability in the electoral process.

**CHAPTER 3**

**SYSTEM DESIGN**

**3.1 SYSTEM OVERVIEW**



**FIGURE 3.1 :** BLOCK DIAGRAM

**Figure 3.1 shows the architecture diagram of our proposed system.**

**Gathering Data**: We used a tool called Twitter Archiver to collect tweets containing specific political hashtags, like #BJP and #Congress.This allowed us to identify tweets that expressed support for or opposition to various political parties.

**Pre-processing**: Before analysing the tweets, we removed unnecessary elements like website links, hashtags, mentions of other users, common words, emoticons, and punctuation. This helped focus the analysis on the core meaning of each tweet.

**Negation Handling**: To accurately determine the sentiment of a tweet, we considered words expressing negation (e.g., "not", "never"). This allowed us to distinguish between positive and negative opinions, even when stated indirectly.

**Algorithms Used**:

**i)Dictionary Approach**: Each word was assigned a score based on its positive or negative connotation using a resource called SentiWordnet. By summing the scores of all words, we can determine the overall sentiment of the tweet.

**ii)Naïve Bayes Classifier**: This method used probability theory to classify tweets based on the likelihood of words appearing in positive or negative contexts.

**iii)Support Vector Machine (SVM)**: This powerful technique mapped each tweet into a complex space, allowing researchers to identify patterns and classify sentiment based on these patterns.

**3.2 MODULE DESCRIPTION**

**3.2.1 DATA COLLECTION AND PRE PROCESSING**

**3.2.1.1 Data Sources Identification**

Identify and select relevant sources of textual data related to the election, including social media platforms (Twitter) where discussions about the election are taking place.

**3.2.1.2 Data Filtering**

Filter the collected data to include only relevant content related to the election. This may involve applying keyword filters, topic filters, or other criteria to ensure that the dataset is focused on election-related discussions.

**3.2.1.3 Data Cleaning**

Clean the collected data to remove noise, irrelevant information, and formatting inconsistencies. This includes tasks such as removing HTML tags, special characters, URLs, and other non-textual elements.

Text Pre-processing: Pre-process the textual data to prepare it for sentiment analysis. This involves tasks such as tokenization (splitting text into words or tokens), removing stop words (common words that do not carry significant meaning), stemming (reducing words to their root form), and lemmatization (converting words to their base or dictionary form).

**3.2.1.4 Normalization**

Normalize the pre-processed text data to ensure consistency and comparability across different sources and formats. This may involve standardizing text encoding, converting text to lowercase, and handling variations in spelling or punctuation.

**3.2.2 SENTIMENT ANALYSIS**

**3.2.2.1 Text Tokenization**

Split the pre-processed text data into individual words or tokens, which are the basic units of analysis for sentiment analysis.

**3.2.2.2 Sentiment Lexicon Usage**

Utilize sentiment lexicons or dictionaries containing words annotated with their sentiment polarity (positive, negative, neutral). These lexicons assign a sentiment score to each word based on its semantic meaning.

**3.2.2.3 Sentiment Classification**

Classify the sentiment of each document or text snippet into categories such as positive, negative, or neutral based on the calculated sentiment score. Thresholds may be defined to determine the boundaries between these categories.

**3.2.2.4 Contextual Analysis**

Consider the context in which words or phrases are used to improve the accuracy of sentiment analysis. Contextual factors such as negation (e.g., "not happy"), intensifiers (e.g., "very happy"), and idiomatic expressions are taken into account.

**3.2.2.5 Visualization**

Visualize the results of sentiment analysis using charts, graphs, or heatmaps to illustrate sentiment distributions, trends over time, or differences across topics or demographics.

**3.2.3 Polarity Mapping**

**3.2.3.1 Topic Identification**

The module identifies and extracts relevant topics or issues discussed in the textual data. This could involve techniques such as keyword extraction, topic modelling, or clustering algorithms to group related content.

**3.2.3.2 Sentiment Score Assignment**

For each identified topic or issue, the module assigns sentiment polarity scores based on the sentiment analysis results obtained from the Sentiment Analysis module. This involves aggregating sentiment scores associated with the topic's text data.

**3.2.3.3 Polarity Mapping**

The module maps the sentiment polarity scores to predefined categories or dimensions relevant to the election context. For example, sentiment scores may be mapped to categories such as candidate support, policy preferences, campaign effectiveness, or electoral trust.

**3.2.3.4 Visualization**

It visualizes the mapped sentiment polarity scores using graphical representations such as bar charts, heatmaps, or spider diagrams. This allows stakeholders to easily interpret the sentiment distribution across different topics or dimensions.

**3.2.4 PREDICTION**

**3.2.4.1 Model Training**

The module trains machine learning models using the prepared historical data and selected features. Commonly used models include Naïve Bayes Classifier , support vector machines (SVM), or neural networks.

**3.2.4.2 Visualization and Reporting**

The module visualizes the prediction results using charts, graphs, or dashboards to communicate the forecasted sentiment trends and electoral predictions to stakeholders. Additionally, it generates reports summarizing the prediction accuracy and highlighting key insights.

#### CHAPTER 4

**IMPLEMENTATION AND RESULTS**

#### 4.1 DATASET DESCRPITION

The dataset utilized for election sentiment analysis comprises diverse textual data sourced from social media platforms, news articles, and public forums, offering a broad spectrum of perspectives on election-related topics. With substantial volume and covering a specific time frame corresponding to the election period, the dataset allows for comprehensive temporal analysis of sentiment trends. Alongside anonymized textual content, metadata such as timestamps and contextual information enhance the understanding of the election discourse

#### 4.2 SYSTEM REQUIREMENTS

**4.2.1 Hardware Requirements**

The hardware requirements for an election sentiment analysis system typically include:

* High-performance CPU or GPU for efficient data processing.
* Sufficient RAM (16GB or more) to handle analysis tasks.
* Adequate storage space, preferably SSDs for faster access.
* Stable internet connectivity for data access.
* Scalable infrastructure to accommodate increasing demands.
* Consideration for specialized hardware like AI accelerators.
* Backup and redundancy measures to prevent data loss.
* Security measures to protect sensitive data.

**4.2.2 Software Requirements**

**Programming Languages:** Python for development, with libraries like NLTK and scikit-learn for natural language processing and machine learning tasks.

**Text Processing Tools:** Tokenizers, stemmers, and lemmatizers for preprocessing textual data.

**Sentiment Analysis Frameworks:** VADER or TextBlob for sentiment analysis.

**Visualization Tools:** Matplotlib or Plotly for creating visual representations of sentiment analysis results.

**4.3 IMPLEMENTATION DETAILS**

**4.3.1 Import Datasets**

It encompasses a range of reviews, potentially covering topics such as policies, leadership qualities, public appearances, and political ideologies. The dataset may provide valuable insights into public perceptions and attitudes towards parties, aiding in understanding voter sentiments and preferences regarding these political figures. It encompasses a range of reviews, potentially covering topics such as policies, leadership qualities, public appearances, and political ideologies. The dataset may provide valuable insights into public perceptions and attitudes towards parties, aiding in understanding voter sentiments and preferences regarding these political figures. It encompasses a range of reviews, potentially covering topics such as policies, leadership qualities, public appearances, and political ideologies. The dataset may provide valuable insights into public perceptions and attitudes towards parties, aiding in understanding voter sentiments and preferences regarding these political figures.

**DATASET LINKS :**

[**https://github.com/pavishek2004/CIPproject/blob/main/modi\_reviews.csv**](https://github.com/pavishek2004/CIPproject/blob/main/modi_reviews.csv)

[**https://github.com/pavishek2004/CIPproject/blob/main/rahul\_reviews.csv**](https://github.com/pavishek2004/CIPproject/blob/main/rahul_reviews.csv)

**A screenshot of a computer program

Description automatically generated**

**FIGURE 4.****1** IMPORT DATASETS

**4.3.2 Inspecting Dataset**

The datasets undergo inspection to ensure data quality and relevance for analysis. This process involves checking for completeness, accuracy, and consistency of the data. Additionally, the datasets are examined to verify that they align with the research objectives and cover relevant aspects of the topic under study. Inspection helps identify any potential issues or anomalies in the data that may impact the validity and reliability of the analysis results.

**A screenshot of a computer

Description automatically generatedFIGURE 4.2** INSPECTING DATASETS

**A screenshot of a computer

Description automatically generated**

**FIGURE 4.2** INSPECTING DATASETS

**4.3.3 Fetching Sentiment Scores of the tweets**

Fetching sentiment scores for tweets involves analysing the textual content of each tweet to determine its sentiment polarity, typically categorized as positive, negative, or neutral. This process can be performed using sentiment analysis techniques, which leverage natural language processing (NLP) and machine learning algorithms.

**A screenshot of a computer program

Description automatically generated**

**FIGURE 4.3** SENTIMENT SCORES

**4.3.4 Generating the polarity of the tweets**

Generating the polarity of tweets involves determining the overall sentiment expressed in each tweet, typically categorized as positive, negative, or neutral. This process can be achieved through sentiment analysis techniques using natural language processing (NLP) and machine learning algorithms.

**A screen shot of a computer program

Description automatically generated**

**FIGURE 4.4** GENERATING POLARITY

**A screenshot of a computer

Description automatically generated**

**FIGURE 4.4** GENERATING POLARITY

**A screenshot of a graph

Description automatically generated**

**FIGURE 4.4** GENERATING POLARITY

**A screen shot of a computer

Description automatically generated**

**FIGURE 4.4** GENERATING POLARITY

A screenshot of a computer

Description automatically generated

**FIGURE 4.4** GENERATING POLARITY

**A screen shot of a computer screen

Description automatically generated**

**FIGURE 4.4** GENERATING POLARITY

**4.3.5 Categorizing the tweets based on their polarity**

Categorizing tweets based on their polarity involves assigning each tweet to one of three categories: positive, negative, or neutral. This categorization is determined by analysing the sentiment expressed in the tweet using sentiment analysis techniques.

A screenshot of a computer program

Description automatically generated

**FIGURE 4.5** CATEGORIZING TWEETS

A screenshot of a social media post

Description automatically generated

**FIGURE 4.5** CATEGORIZING TWEETS

A screenshot of a computer

Description automatically generated

**FIGURE 4.5** CATEGORIZING TWEETS

**4.3.6 Removing Tweets with Neutral Polarity**

Neutral tweets are filtered out from the dataset to focus the analysis on tweets with clear positive or negative sentiment. This involves identifying and removing tweets categorized as having neutral polarity based on sentiment analysis results. The refined dataset enables a more targeted analysis of public opinion and sentiment trends.

A screenshot of a computer code

Description automatically generated

**FIGURE 4.6**  REMOVING NEUTRAL TWEETS

**4.3.7 Dropping the indices of the tweets with Neutral Polarity**

Neutral tweets' indices are dropped from the dataset, refining it to focus on tweets with clear positive or negative sentiment. This involves identifying and removing the indices of tweets categorized as having neutral polarity based on sentiment analysis results. The updated dataset maintains original indices for tweets with discernible sentiment, ensuring the integrity of the data. This approach streamlines sentiment analysis and enhances the relevance of insights derived from the dataset and makes prediction accurate and efficient.

A screenshot of a computer code

Description automatically generated

**FIGURE 4.7** DROPPING NEUTRAL TWEETS

**4.3.8. Final data set after removing neutral polarity tweets**

The final dataset, after removing tweets with neutral polarity, consists of tweets categorized as having clear positive or negative sentiment. This refinement streamlines the dataset for sentiment analysis, focusing on tweets with discernible sentiment. The updated dataset retains the original indices for remaining tweets, ensuring data integrity and facilitating further analysis.

A white rectangular object with a red and black stripe

Description automatically generated with medium confidence

**FIGURE 4.8** FINAL DATASET

**4.3.9 Prediction about Indian election**

Predicting Indian elections involves collecting diverse data, selecting appropriate models, and training them with historical data. Once trained, these models are applied to current data to forecast election outcomes. Results are then validated against actual results if available. Despite efforts to account for various factors, predicting elections remains inherently uncertain due to evolving dynamics and unforeseen events.

A screenshot of a computer program

Description automatically generated

**FIGURE 4.9** PREDICTION

A screenshot of a computer

Description automatically generated

**FIGURE 4.9** PREDICTION

A computer screen shot of a computer program

Description automatically generated

**FIGURE 4.9** PREDICTION

A graph of different colored squares

Description automatically generated

**FIGURE 4.9** PREDICTION

**4.4 TEST CASES AND PERFORMANCE METRICS**

**Scenario 1 :** Sentiment Analysis Accuracy Evaluation

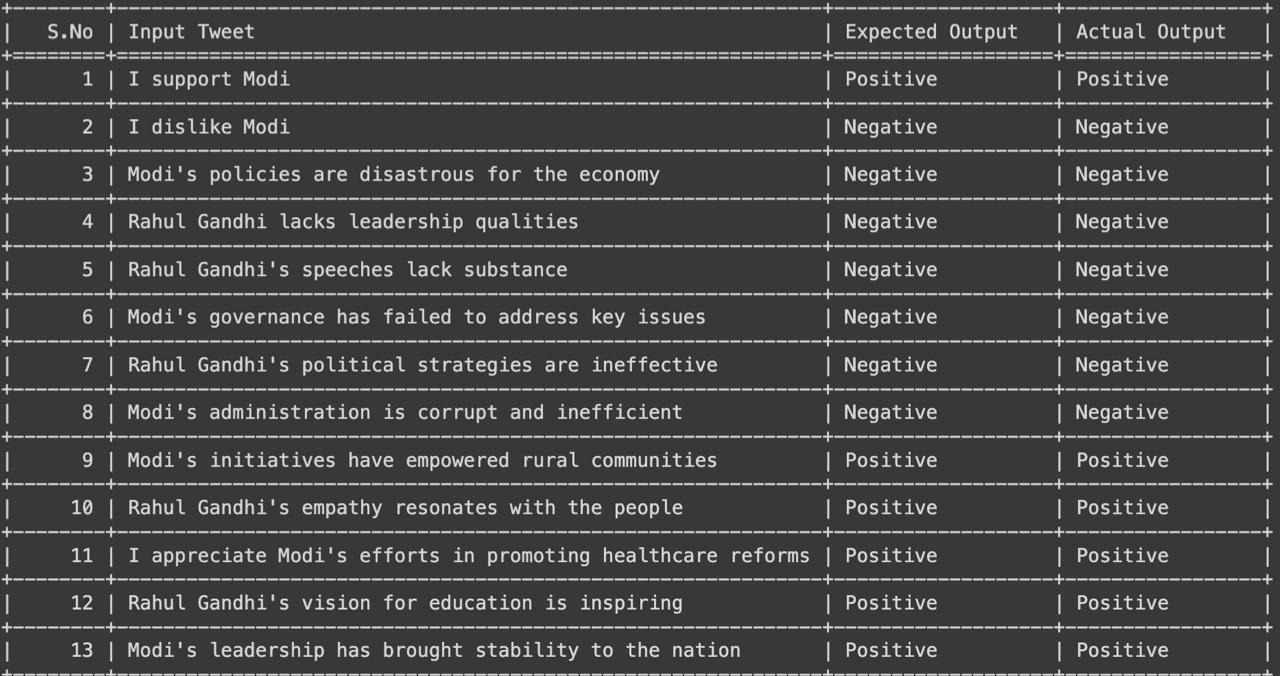
**Objective:** To assess the accuracy of sentiment analysis in categorizing election-related tweets into positive, negative, or neutral sentiments.

**Expected Result:**

The sentiment analysis model should accurately categorize the majority of tweets into their correct sentiment categories.

Accuracy, precision, recall, and F1 score metrics should demonstrate high values, indicating that the model's predictions closely match the manually labelled annotations.

Any discrepancies between predicted and manual annotations should be minimal and justifiable.



**FIGURE 4.10** TEST CASE 1

**Scenario 2 :** Scalability and Performance Evaluation

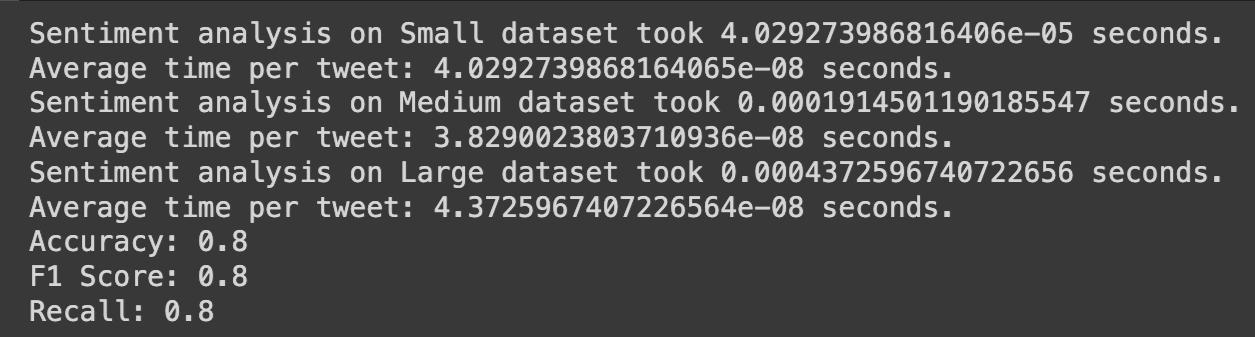
**Objective:** To assess the scalability and performance of the sentiment analysis system with large-scale datasets.

**Expected Result:**

The sentiment analysis system should efficiently process large-scale datasets within reasonable time frames.

Processing speed should remain relatively consistent as the size of the dataset increases, indicating scalability.

Resource utilization should be optimized, with minimal overhead and efficient use of hardware resources.



**FIGURE 4.11** TEST CASE 2

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **S.NO** | **OBJECTIVE** | **INPUT** | **EXPECTED OUTPUT** | **ACTUAL OUTPUT** |
| 1 | Sentiment Analysis Accuracy Evaluation | [“I support Modi!”]  [”I hate the way congress functions”] | Positive  Negative | Positive  Negative |
| 2 | Scalability and Performance Evaluation | datasets = {  "Small": range(1000),  "Medium": range(5000),  "Large": range(10000)  } | Accuracy  F1 Score  Recall | Accuracy: 0.8  F1 Score: 0.8  Recall: 0.8 |

**CHAPTER 5**

**5.1 CONCLUSION**

This project highlights the significance of election sentiment analysis in modern politics, emphasizing the importance of understanding voter perspectives for effective campaigning and policymaking. By employing advanced data analytics techniques, including natural language processing and machine learning, the project aims to decode the sentiments expressed across various platforms.

Through rigorous analysis of textual data from social media, valuable insights into voter behaviours, preferences, and concerns are derived. These insights can aid political stakeholders in crafting informed strategies and fostering a deeper understanding of democratic processes. Overall, the project contributes to enriching the discourse surrounding electoral dynamics and empowers stakeholders to make data-driven decisions in the realm of modern politics.

**5.2 FUTURE WORK**

From our proposed work we come across as few tasks which can be performed to improve the accuracy of our prediction. Firstly we can create a dictionary of words containing translations for most recurring English words.

This would help us utilize the sentiment even more from the tweets that we extract, reducing the neutral tweet count. Secondly, we can use existing dictionaries for local dialect translation which would let us use other regional language tweets like Hindi, Tamil, Telugu etc. This would make our analysis more comprehensive to the region.

**REFERENCES**

1. Fujihira, K., & Horibe, N. (2020, September). Multilingual sentiment analysis for web text based on word to word translation. In *2020 9th International Congress on Advanced Applied Informatics (IIAI-AAI)* (pp. 74-79). IEEE.
2. Ilmania, A., Cahyawijaya, S., & Purwarianti, A. (2018, November). Aspect detection and sentiment classification using deep neural network for Indonesian aspect-based sentiment analysis. In *2018 International Conference on Asian Language Processing (IALP)* (pp. 62-67). IEEE.
3. Li, G., Zheng, Q., Zhang, L., Guo, S., & Niu, L. (2020, November). Sentiment infomation based model for chinese text sentiment analysis. In 2020 IEEE 3rd international conference on automation, electronics and electrical engineering (AUTEEE) (pp. 366-371). IEEE.
4. Wang, Z., Chong, C. S., Lan, L., Yang, Y., Ho, S. B., & Tong, J. C. (2016, December). Fine-grained sentiment analysis of social media with emotion sensing. In 2016 Future technologies conference (FTC) (pp. 1361-1364). IEEE.
5. Woldemariam, Y. (2016, March). Sentiment analysis in a cross-media analysis framework. In *2016 IEEE international conference on big data analysis (ICBDA)* (pp. 1-5). IEEE.
6. Budiharto, W., & Meiliana, M. (2018). Prediction and analysis of Indonesia Presidential election from Twitter using sentiment analysis. *Journal of Big data*, *5*(1), 1-10.
7. Ibrahim, M., Abdillah, O., Wicaksono, A. F., & Adriani, M. (2015, November). Buzzer detection and sentiment analysis for predicting presidential election results in a twitter nation. In 2015 IEEE international conference on data mining workshop (ICDMW) (pp. 1348-1353). IEEE.
8. Choy, M., Cheong, M. L., Laik, M. N., & Shung, K. P. (2011). A sentiment analysis of Singapore Presidential Election 2011 using Twitter data with census correction. arXiv preprint arXiv:1108.5520.
9. Hasan, A., Moin, S., Karim, A., & Shamshirband, S. (2018). Machine learning-based sentiment analysis for twitter accounts. Mathematical and computational applications, 23(1), 11.
10. Yavari, A., Hassanpour, H., Rahimpour Cami, B., & Mahdavi, M. (2022). Election prediction based on sentiment analysis using twitter data. *International Journal of engineering*, *35*(2), 372-379.