

OPINION MINING FOR POLITICAL DISCOURSE

Group Members : 1. Reethika.G : 2303A51190

2. Vyshnavi.M : 2303A51205

3. Vyshnavi.R : 2303A51216

4. Sreeja.T : 2303A51220

Course Name : Applications of Data Mining

Instructor : Dr.R.S.Dubey (Dean R&D)

Submission Date : 25th April 2025

Opinion Mining for Political Discourse

# 1. Introduction

**Project Overview:**  
This project explores the application of sentiment analysis and opinion mining techniques to analyze political discourse on social media. By processing tweets related to political figures and topics, we aim to identify public sentiments and understand their trends over time. The goal is to utilize these findings to shed light on public perspectives regarding policies, political candidates, and controversial issues.  
  
**Importance of Opinion Mining:**  
Opinion mining is crucial in today’s digital world where public opinion is largely shaped and shared on platforms like Twitter. Understanding these opinions helps policymakers, political strategists, and researchers gauge the mood of the electorate and the impact of political events. Analyzing this data can help detect early signs of political change, measure support for public policies, and assess public response to political campaigns or decisions.  
 **Problem Statement:**  
How can sentiment and opinion mining be utilized to effectively gauge public sentiment towards political issues and candidates using data from social media? This project seeks to address this question by leveraging sentiment analysis tools to classify and interpret public reactions shared on Twitter.

# 2. Objectives

- Collect political tweets from Kaggle dataset.  
- Clean and preprocess the data for analysis.  
- Perform sentiment analysis using TextBlob to classify sentiments as positive, negative, or neutral.  
- Visualize the distribution of public opinion and sentiment trends.  
- Identify significant patterns and shifts in sentiment related to political discourse.

# 3. Dataset Description

The dataset used in this project is sourced from Kaggle and contains tweets posted by political figures, particularly focusing on Donald Trump. The dataset includes fields such as tweet content, timestamp, and other metadata. It provides a comprehensive view of political communication and public engagement on Twitter, making it ideal for sentiment analysis tasks.

# 4. Tools and Libraries

- Python: Primary programming language used for scripting and data processing.  
- Pandas: Used for data manipulation, cleaning, and transformation.  
- Matplotlib & Seaborn: Libraries for creating visual representations of data.  
- TextBlob: NLP library used for basic sentiment analysis.

# 5. Data Processing

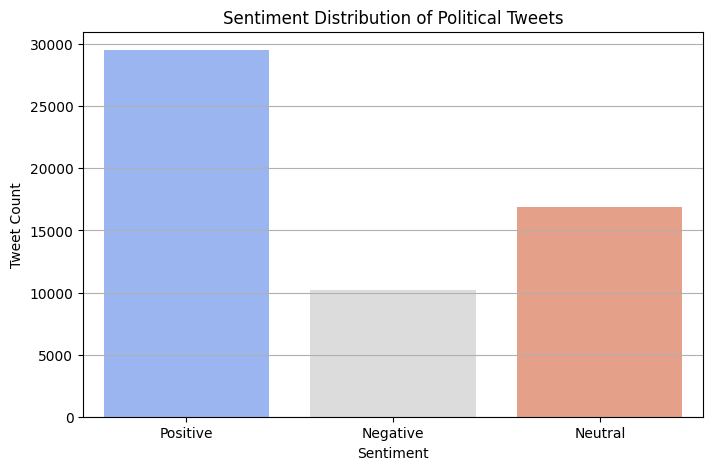
Python code used to process the dataset:

import pandas as pd  
import matplotlib.pyplot as plt  
import seaborn as sns  
from textblob import TextBlob  
  
df = pd.read\_csv("/content/trump\_tweets.csv")  
print("Column names in dataset:", df.columns)  
  
if 'content' in df.columns:  
 text\_column = 'content'  
elif 'text' in df.columns:  
 text\_column = 'text'  
elif 'Tweet' in df.columns:  
 text\_column = 'Tweet'  
else:  
 raise ValueError("Tweet content column not found. Check column names in your CSV.")  
  
df['clean\_text'] = df[text\_column].astype(str).str.replace(r"http\S+|www\S+|https\S+", '', regex=True)  
df['clean\_text'] = df['clean\_text'].str.replace(r"[^a-zA-Z\s]", '', regex=True)  
df['clean\_text'] = df['clean\_text'].str.lower()  
  
def get\_sentiment(text):  
 analysis = TextBlob(text)  
 polarity = analysis.sentiment.polarity  
 if polarity > 0:  
 return 'Positive'  
 elif polarity < 0:  
 return 'Negative'  
 else:  
 return 'Neutral'  
  
df['Sentiment'] = df['clean\_text'].apply(get\_sentiment)  
df.to\_csv("processed\_tweets.csv", index=False)

# 6. Exploratory Data Analysis

The dataset was examined to understand the distribution of sentiments among political tweets. This helps in identifying the overall tone of discourse and the frequency of varying sentiments.  
Graph:  
A bar plot was generated to visualize the distribution of positive, negative, and neutral tweets:

plt.figure(figsize=(8, 5))  
sns.countplot(data=df, x='Sentiment', palette='coolwarm')  
plt.title('Sentiment Distribution of Political Tweets')  
plt.xlabel('Sentiment')  
plt.ylabel('Tweet Count')  
plt.grid(axis='y')  
plt.show()



# 7. Model Building

In this project, we utilized the pre-built TextBlob library which implements a rule-based approach for sentiment analysis. It doesn't require model training and provides sentiment scores directly, making it suitable for straightforward analysis tasks without the need for deep learning or machine learning models.

# 8. Evaluation

Since a pre-trained sentiment analyzer was used, the evaluation was performed by manually reviewing classified tweets to validate their sentiment labels. For future work, incorporating manually labeled datasets would allow for more rigorous evaluation using metrics like accuracy, precision, recall, and F1-score.

# 9. Conclusion

Opinion mining on social media can provide deep insights into public sentiment about political issues and personalities. This project demonstrated how tools like TextBlob can help automate sentiment classification and visualize public opinion trends. It confirmed that a significant amount of political discourse on Twitter could be effectively categorized using simple sentiment analysis methods.

# 10. Future Enhancements

- Use a labeled dataset to train and evaluate a custom machine learning model.  
- Incorporate more advanced NLP techniques such as BERT, which could improve accuracy.  
- Analyze sentiments across different time frames to observe changes related to major political events.  
- Expand the dataset to include tweets from multiple politicians and compare their public perception.

# 11. Appendix

- Kaggle Dataset: "Trump Twitter Archive" (https://www.kaggle.com/datasets/rohanrao/trump-tweets)  
- Python Libraries: Pandas, Matplotlib, Seaborn, TextBlob  
- Sentiment Analyzer: TextBlob polarity-based scoring system