

Sentiment Analysis of Real-World Topics Using Social Media Posts

IDA Final Project - Group 10

Bhuvanesh So Muruganandam
Vignesh Murugan
Vivek Satya Sai Veera Venkata Talluri
Pardhu Burlu

1 Problem Description

Social media platforms, such as Twitter, serve as a rich source of public sentiment, influencing market trends, political outcomes, and societal behaviors. Analyzing these digital interactions offers valuable insights into collective opinions and trends on topics like finance, pandemics, politics, and global events. However, the volume and unstructured nature of social media data make extracting meaningful insights a challenging task.

2 Type of Problem

This project falls under **text mining and sentiment analysis**, focusing on classifying public sentiment into positive, neutral, or negative categories. It also includes elements of trend tracking, entity recognition, and network analysis to explore the relationships between users, topics, and influential entities. The core problem lies in understanding **how public sentiment shifts** in response to major events and how these patterns evolve over time.

3 Initial Thoughts and Approach

3.1 Data Collection

- Use APIs from **Twitter** or **Reddit** to collect posts related to specific domains:
 - **Stocks:** Keywords like “Tesla,” “bull market.”
 - **Pandemics:** Mentions of “COVID-19” or “vaccination.”
 - **Politics:** Sentiment around elections or leadership changes.
 - **Global Events:** Discussions on breaking news or disasters.

3.2 Data Cleaning

- Standardize text by removing noise (URLs, emojis) and normalize the text (case conversion, stemming).

3.3 Sentiment Classification

- Apply **lexicon-based techniques** in R using the Bing, NRC, and AFINN lexicons for sentiment scoring.
- The lexicons are available on Kaggle. You can access them here: [Bing-NRC-AFINN Lexicons on Kaggle](#).

3.4 Visualizations

- Utilize **ggplot2** for sentiment trend analysis.
- Generate **word clouds** to highlight keyword prominence.
- Build **network graphs** to identify influencers and connections between topics.

4 Outcome Expectations

- Discover shifts in public sentiment during key events.
- Identify influential entities and trends across topics.
- Deliver data-driven insights to help policymakers and analysts respond effectively to changing public moods.