Text Analysis Project
Twitter Sentiment Analysis

Arun Totad
Hithesraja Navarethinam
Leekshitha Resoju
Preethi Sekar

STAT 5870
Big Data Analysis Using Python
Professor Dr. Kevin Lee
December 2024

Problem Statement & Solution:

Twitter is a major platform in which people express opinions on various topics in real-time, making it a valuable source for understanding public sentiment. This project aims to address these challenges by creating an end-to-end Twitter sentiment analysis model for efficient and accurate sentiment classification. This project will develop an unsponsored Twitter sentiment analysis system to classify tweets as positive, negative, or neutral, focusing on the following key areas: Data Collection, Data Preprocessing, Modeling, Evaluation and Tuning.

Description of the Data Set:

The data set will consist of tweets collected using the Twitter API. Tweets will be gathered based on specific keywords and hashtags relevant to the sentiment analysis project. The data will include tweet text, timestamps, user information, and metadata. This real-time data collection will ensure that the sentiment analysis reflects current public opinions.

Description of the Applications:

The sentiment analysis tool will have multiple applications:

- **Brand Monitoring**: Companies can track public sentiment about their products or services in real-time.
- **Customer Service**: Businesses can prioritize and respond to customer feedback promptly.
- **Political Analysis**: Analysts can gauge public opinion on policies, events, and political figures.
- **Market Research**: Researchers can understand consumer trends and opinions to inform decision-making.

Brief Explanation of Methods and Tools:

The project will employ a combination of machine learning and deep learning techniques to classify tweets as positive, negative, or neutral. Key steps include:

- **Data Collection**: Using the Twitter API to gather tweets.
- **Data Preprocessing**: Cleaning and standardizing text data to remove noise and improve quality.
- **Modelling**: Implementing models such as logistic regression, SVM, LSTM, and BERT for sentiment classification.
- **Evaluation and Tuning**: Assessing model performance using metrics like accuracy, precision, recall, and F1-score, and optimizing hyperparameters for better results.

Citation for Data Source: Twitter API documentation (https://developer.twitter.com/en/docs/twitter-api)