Student Name: Sudharsun M

Register Number: 622623205052

Institution: Shreenivasa Engineering College

Department: B.Tech. Information technology

Date of Submission: (Insert Date)

Github Repository Link: https://github.com/Sudharsun2410/-Ammu.git

1. Introduction

In the digital age, social media platforms have become a rich source of human expression, offering insights into public sentiment and emotional states.

This project aims to explore the application of sentiment analysis on social media conversations to decode underlying emotions, providing valuable insights for various domains such as marketing, mental health, and public opinion analysis.

2. Problem Statement

Despite the vast amount of data generated on social media, understanding the emotional tone behind user-generated Traditional methods often struggle context, and the informal nature of tant remains a significant challenge. ces such as sarcasm, cultural guage.
This project seeks to address these challenges bu implementina advanced sentiment analysis

3. Objectives

Objective 1: To collect and preprocess social media data for sentiment analysis.

Objective 2: To apply sentiment analysis models to classify emotions in social media conversations.

Objective 3: To evaluate the performance of different sentiment analysis techniques.

Objective 4: To interpret the results and provide insights into public sentiment trends.

4. Literature Review

Recent studies have highlighted the potential of sentiment analysis in understanding public emotions. For instance. Aim Technologies discusses various techniques for sentiment analysis, including machine learning approaches such as Naive Bayes and Support Vector Machines, and lexicon-based methods. These methods have been employed to analyze social media data. providing insights into user sentiments and opinions.

5. Methodology

5.1 Data Collection

Social media data will be collected using APIs from platforms such as Twitter and Reddit. The data will include text posts, comments, and metadata to provide a comprehensive dataset for analysis.

5.2 Data Preprocessing

The collected data will undergo preprocessing steps, including:

Tokenization: Breaking down text into individual words or tokens.

Normalization: Converting text to lowercase and removing punctuation.

Stopword Removal: Eliminating common words that do not contribute to sentiment.

Lemmatization: Reducing words to their base or root form.

5.3 Sentiment Analysis Models

Various sentiment analysis models will be applied, including:

Alm Technologies +4

Machine Learning Models: Naive Bayes. Support Vector Machines, and Recurrent Neural Networks.

Lexicon-Based Approaches: Using predefined sentiment lexicons to classify emotions.

Hybrid Models: Combining machine learning and lexicon-based methods for improved accuracy. Aim Technologles

5.4 Evaluation Metrics

The performance of the models will be evaluated using metrics such as accuracy, precision, recall, and F1-score. These metrics will provide insights into the effectiveness of each model in classifying emotions.

6. Expected Outcomes

The project aims to achieve the following outcomes:

Outcome 1: Development of a robust sentiment analysis pipeline for social media data.

Outcome 2: Identification of prevalent emotions in social media conversations.

Outcome 3: Insights into public sentiment trends over time.

Outcome 4: Recommendations for improving sentiment analysis models to handle social media nuances.

7. Resources Required

Software: Python, Jupyter Notebook, Scikit-learn, NLTK. TensorFlow.

Hardware: Personal computers with internet access.

Data Sources: APIs from Twitter, Reddit, and other social media platforms. newyorker.com +4