Phase-3

Decoding Emotions through Sentiment Analysis of Social Media Conversations

Student name : M.Durga

Register number : 620123106024

Institution : AVS Engineering college

Department : Electronics and communication engineering

Date of submission :17.05.2025

 Github repository link :
http://https://github.com/durgaqueen16/Durga-M-naan-mudhalvan-project-.git

1. Problem Statement
Social media platforms generate vast amounts of user-generated content, reflecting public opinions and emotions. However, decoding these emotions in a structured manner remains a challenge. The project aims to build a sentiment analysis system to decode and analyze emotions from social media conversations effectively.

2. Abstract
This project presents a sentiment analysis system that decodes emotions from social media conversations using natural language processing and machine

learning techniques. It involves collecting data from platforms like Twitter, processing the text, and building a model to classify sentiments such as positive, negative, and neutral. The system aims to provide insights into public opinion trends and can be used in various domains such as marketing, politics, and public health.

3. System Requirements

- Programming Language: Python
- Libraries: pandas, numpy, sklearn, nltk, matplotlib, seaborn, Flask
- Tools: Jupyter Notebook, Anaconda, Git, Heroku/Streamlit
- Dataset Source: Twitter API or pre-collected datasets (e.g., Sentiment140)
- Operating System: Windows/Linux/Mac

4. Objectives

- To collect and preprocess social media text data.
- To perform exploratory data analysis to understand sentiment distribution.
- To build and evaluate machine learning models for sentiment classification.
- To deploy the model as a web application.

5. Flowchart of Project Workflow

1. Data Collection
2. Data Preprocessing
3. Exploratory Data Analysis
4. Feature Engineering
5. Model Building
6. Model Evaluation
7. Deployment

6. Dataset Description

The dataset used for this project will be collected from Twitter API or pre-collected datasets like Sentiment140. The dataset will contain text data labeled as positive, negative, or neutral.

7. Data Processing

- Data cleaning: removing special characters, punctuation, and stop words.
- Tokenization: splitting text into individual words or tokens.
- Stemming or Lemmatization: reducing words to their base form.

- Sentiment distribution: analyzing the distribution of positive, negative, and neutral sentiments in the dataset.
- Word frequency: analyzing the most common words in the dataset.

9. Feature Engineering

- Bag-of-words: representing text data as a bag-of-words.
- TF-IDF: representing text data using term frequency-inverse document frequency.

10. Model Building

- Supervised learning algorithms: using algorithms like Naive Bayes, Logistic Regression, and Support Vector Machines to classify sentiments.

11. Model Evaluation

- Accuracy: evaluating the model's performance using accuracy metrics.
- Precision: evaluating the model's performance using precision metrics.
- Recall: evaluating the model's performance using recall metrics.
- F1-score: evaluating the model's performance using F1-score metrics.

12. Deployment

- Deploying the model as a web application using Flask or Streamlit.

13. Source Code

The source code for this project will be written in Python and will utilize libraries like pandas, numpy, sklearn, and nltk.

14. Future Scope

- Improving the model's performance using deep learning algorithms.
- Expanding the dataset to include more diverse text data.
- Integrating the model with other natural language processing tasks.

15. Team Members and Roles

- M.Durga: Project Lead and Developer
- V.Harinisha: Data Analyst
- S.Divya: Model Building and Deployment