Phase-2

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 :10.05.2025

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

1. Problem Statement

With the rise of social media platforms, individuals frequently express emotions, opinions, and feedback online. This data, while valuable, is unstructured and massive in volume, making manual analysis impractical. The problem lies in effectively decoding and understanding human emotions from these social media conversations using sentiment analysis techniques.

2. Project Objective

The objective of this project is to analyze social media text data to identify and classify

user sentiments and emotions. By building a robust sentiment analysis system, we aim to derive actionable insights and emotional trends from social media conversations.

3. Flow Chart of the Project Workflow

1. Data Collection
2. Data Cleaning & Preprocessing
3. Exploratory Data Analysis (EDA)
4. Feature Engineering
5. Model Building
6. Model Evaluation
7. Visualization and Insights

4. Data Description

The dataset includes:
- Social media posts or comments (e.g., tweets)
- Time of posting
- Anonymized user ID
- Sentiment/Emotion labels (if available)
- Meta-data such as likes, shares (optional)

5. Data Processing

Preprocessing steps include:
- Removing URLs, emojis, and special characters
- Lowercasing text
- Removing stopwords
- Tokenization and Lemmatization
- Handling null and duplicate values

6. Exploratory Data Analysis

EDA involves analyzing the distribution of sentiments, identifying common keywords, creating word clouds, and observing time-based trends in sentiment variation.


7. Feature Engineering

Feature extraction techniques used include:

- Bag of Words (BoW)
- TF-IDF Vectorization
- Word Embeddings (e.g., GloVe, Word2Vec)
- N-grams and POS tagging

8. Model Building

Models used for sentiment classification:
- Logistic Regression
- Support Vector Machine (SVM)
- Random Forest
- Deep Learning models (RNN, LSTM)
- Transformer-based models like BERT

9. Visualization of Results and Model Insights

Visual outputs include:
- Confusion Matrix
- Accuracy, Precision, Recall, F1 Score graphs
- Word Clouds and Sentiment Distribution
- Feature importance visualization
- Time-series trends of sentiments

10. Tools and Technologies Used

- Python (Pandas, NumPy, NLTK, SpaCy, Scikit-learn, TensorFlow, Keras)
- Jupyter Notebook / Google Colab
- Matplotlib, Seaborn, Plotly
- Git & GitHub for version control
- Tableau / Power BI for dashboard creation

11. Team Members and Contributions

Name	Role	Contributions
M.Durga	Project leader and developer	Data collection and cleaning

V.Harinisha	Data Analyst	Exploratory Data Analysis and Visualization
S.Divya	Modal building and Deployment	Model building and evaluation