Phase-1

Decoding Emotions through Sentiment Analysis of Social Media Conversations

Student Name: Durga.M

Register Number:620123106024

Institution:AVS ENGINEERING COLLEGE

Department:ECE

Date of submission:30/04/2025

1. Why This Project is Chosen & Its Importance

Why: Social media is a rich source of public emotion and opinion. Understanding sentiments helps businesses, governments, and researchers respond effectively.

Importance:

- - Track public mood on events/policies.
- - Improve customer service and brand management.

- Detect trends, threats, or opportunities in real time.

2. Objectives of the Project

- - Extract and analyze emotional tone from social media posts.
- - Classify conversations into sentiments/emotions (e.g., happy, sad, angry).
- - Build a predictive model to automate sentiment detection.
- - Visualize trends and insights.

3. Scope of the Project

- - Focused on one or more platforms (e.g., Twitter, Reddit).
- - Language: Primarily English.
- - Text-based content only (no images/videos).
- - Real-time or historical analysis.

4. Data Sources

- - Twitter API (via Tweepy or Twitter Developer Portal)
- - Reddit API (PRAW)
- - Open Datasets (Kaggle, CrowdFlower, Sentiment140)

5. High-Level Methodology

a. Data Collection:

- - Use APIs to extract social media text (tweets, posts, comments).

b. Data Cleaning:

- - Remove stopwords, URLs, hashtags, mentions, emojis.
- - Normalize text (lowercase, stemming/lemmatization).

c. Exploratory Data Analysis (EDA):

- - Word clouds, frequency plots, sentiment distribution.
- - Time-based trends and correlations.

d. Feature Engineering:

- - Text vectorization (TF-IDF, Word2Vec, BERT embeddings).
- - Add meta-features (post length, number of emojis, etc.).

e. Model Building:

- - Classification models: Logistic Regression, Naive Bayes, SVM, LSTM, or Transformers (BERT).

f. Model Evaluation:

- - Accuracy, Precision, Recall, F1-score, Confusion Matrix.

g. Visualization & Interpretation:

- - Dashboards (Plotly, Seaborn, Matplotlib).
- - Sentiment over time, word importance, etc.

h. Deployment (Optional):

- - Build web app (Streamlit/Flask).
- - Host on Heroku, AWS, or Hugging Face Spaces.

6. Tools and Technologies

- M.Durga – Data Collection Learn
- "My role involved gathering and preprocessing data from social media platforms such as Twitter and Reddit. I used APIs and scraping tools to extract posts and comments, ensuring that the data was relevant and ethically sourced. I also helped clean and prepare the dataset for analysis, removing noise, duplicates, and irrelevant content.
- V.Harinisha– NLP & Sentiment Analysis Developer
- "I focused on applying natural language processing techniques to analyze the collected data. I used tools like NLTK and TextBlob to perform sentiment analysis and classified the emotions into categories like happiness, anger, sadness, and surprise. I also experimented with different machine learning models to improve accuracy.
- S.Divya– Model Training & Evaluation
- "I handled the training and evaluation of sentiment classification models. I used deep learning models such as LSTM and BERT to understand complex emotions in conversations. I also worked on fine-tuning hyperparameters and validating the model using various metrics like precision, recall, and F1 score."