Phase-1

Student Name: Umesh M

Register Number: AU410723104090

Institution: Dhanalakshmi College of Engineering

Department: computer science engineering

Date of Submission: 28-Aprl-2025

1.Problem Statement
Decoding emotions through sentiment analysis of social media conversation. Develop a system to analyze social media conversations and accurately decode the emotions expressed by users, considering the nuances of language, context, and cultural differences.

2.Objectives of the Project
- To decode emotions expressed in social media conversations using sentiment analysis techniques.
- To classify the sentiment into positive, negative, or neutral categories.
- To compute and interpret the proportion of positive and negative sentiments from the conversations.

.

3.Scope of the Project

This project focuses on analyzing social media conversations to detect underlying emotions using sentiment analysis techniques.

It aims to classify text into categories such as positive, negative, or neutral sentiments.

The project includes:

- *Preprocessing text data,*
- *Implementing machine learning or deep learning models for sentiment classification,*
- *Visualizing sentiment trends and distributions.*

Limitations:

- *The project is limited to English-language text.*
- *Only public posts from social media platforms (primarily Twitter) will be used.*

4.Data Sources

- ***Dynamic Data:*** *Text data collected from Twitter using the Twitter API, reflecting real-time conversations.*
- ***Static Data:*** *Pre-labeled datasets such as **Sentiment140** from Kaggle, used for training and testing machine learning models.*

Both datasets are publicly available and used strictly for academic and research purposes.

Link: https://search.app/ipxjwx5QpGGrude38

5.High-Level Methodology

● **Data Collection:**

- Collect real-time tweets using the Twitter API with necessary authentication.
- Use the Sentiment140 dataset from Kaggle for model training and validation.

● **Data Cleaning:**

- Remove unwanted elements such as emojis, URLs, hashtags, mentions, special characters, and redundant whitespaces.

- Convert text to lowercase.
- Remove stop words, punctuation, duplicate entries, and null values.

● **Exploratory Data Analysis (EDA):**

- Use libraries like Matplotlib, Seaborn, and WordCloud for data visualization.
- Analyze sentiment distribution, frequent words, hashtags, and correlations among features.

● **Feature Engineering:**

- Perform tokenization and lemmatization.
- Convert text into numerical representations using techniques like TF-IDF or Word2Vec.
- Create additional features such as word count, sentiment keywords, etc.

● **Model Building:**

- Experiment with various machine learning models like:
  - Logistic Regression
  - Naïve Bayes
  - Support Vector Machines (SVM)
  - Random Forest
- Explore deep learning models like:
  - LSTM (Long Short-Term Memory)
  - BERT (Bidirectional Encoder Representations from Transformers)

● **Model Evaluation:**

- Evaluate models based on metrics like Accuracy, Precision, Recall, F1-score, and Confusion Matrix.
- Perform Cross-Validation to avoid overfitting and ensure robust results.

● **Visualization & Interpretation:**

- Present insights using interactive and static graphs with libraries like Plotly, Seaborn, and Matplotlib.
- Visualize sentiment trends, model comparisons, and key findings.

● **Deployment (Optional):**

- If time permits, develop a web application using Flask or Streamlit.
- Allow users to input text and receive real-time sentiment predictions.
- Deploy the app locally or on cloud platforms like Heroku.

.

6.Tools and Technologies

● **Programming Language:**

- Python

● **Notebook/IDE:**

- Google Colab

- Jupyter Notebook

- VS Code (for deployment or local testing)

● **Libraries:**

- **Data Processing:** pandas, numpy, re (regular expressions) , sentimentIntensityAnalyzer.

- **Visualization:** matplotlib, seaborn, wordcloud, plotly

- **Modeling and Machine Learning:** scikit-learn, TensorFlow, Keras, nltk, transformers (for BERT)

● **Optional Tools for Deployment:**

- Streamlit

- Flask

- Heroku (for hosting)

7.Team Members and Roles

- ***Team Leader**- Srikanth T*
  - *Data collection*
  - *Coding*
- ***Team Member**- Umesh M*
  - *Report Writing and Documentation*
  - *Model training and evaluation*
- ***Team Member**- Ravi Kumar B*
  - *Data collection*
  - *Data cleaning and preprocessing*
- ***Team Member**- Yashwanth KS*
  - *EDA and Visualization*
  - *Data collection*