PHASE-2

STUDENT NAME:

Sivanesh. B

REGISTER NUMBER:
422223106031

INSTITUTION:
Surya Group Of Institutions

DEPARTMENT:
B.E “Electronics And Communication Engineering”

DATE OF SUBMISSION:
08/05/2025
GitHub link: https://github.com/bsivanesh/Emotions-sentiment-/tree/main

Project : DECODING EMOTIONS THROUGH SENTIMENT ANALYSIS OF SOCIAL MEDIA CONVERSATIONS

1. Problem Statement

This project focuses on identifying and interpreting human emotions through sentiment analysis of social media conversations. It aims to classify emotions such as happiness, anger, sadness, fear, and surprise based on textual content shared online. Solving this problem can provide valuable insights for mental health monitoring, customer sentiment tracking, and public opinion analysis.

2. Project Objectives

- Build a sentiment analysis model that detects basic emotions with high accuracy.
- Use natural language processing (NLP) to process and analyze social media data.
- Implement multi-class classification for emotion detection.
- Apply the model in contexts such as customer feedback analysis and mental well-being assessment.

3. Flowchart of the Project Workflow

Data Collection → Preprocessing → Feature Extraction → Model Training → Evaluation → Emotion Prediction

4. Data Description

Dataset Name: GoEmotions
Origin: Google Research - Reddit Comments
Type: Textual, labeled data with 27 emotions + neutral
Size: 58,000 labeled entries
Dataset Link: https://github.com/google-research/google-research/tree/master/goemotions

5. Data Preprocessing

- Removed URLs, mentions, and hashtags.
- Tokenized text using NLTK.
- Removed stopwords and applied lemmatization.
- Encoded emotion labels using one-hot encoding.

6. Exploratory Data Analysis (EDA)

- Plotted emotion distribution.
- Analyzed word frequency using word clouds.
- Visualized correlations using heatmaps.

7. Feature Engineering

- Used TF-IDF vectorization for feature extraction.
- Explored word embeddings using Word2Vec.
- Considered n-gram features (bigrams and trigrams).

8. Model Building

- Implemented Logistic Regression and Random Forest for baseline.
- Used LSTM for sequential emotion classification.
- Evaluated with accuracy, precision, recall, and F1-score.

9. Visualization of Results & Model Insights

- Confusion matrix for emotion classification.
- Bar plots comparing model performances.
- Feature importance analysis from tree-based models.

10. Tools and Technologies Used

- Programming Language: Python
- Libraries: pandas, numpy, nltk, sklearn, tensorflow, seaborn, matplotlib
- IDE: Google Colab / Jupyter Notebook

11. Team Members and Contributions

- Data Collection & Cleaning: B. Sivanesh
- Feature Engineering: S. Vasanth
- Model Development & Reporting:

S. Vishnu