Phase-1 Submission Template

Student Name: G.Eraiarutselvi

Register Number: 421223121010

Institution: karpagavinayaga college of engineering and technology

Department: B.E Biomedical engineering

Date of Submission: 27.04.2025

1. Problem Statement:

* Social media has become a digital mirror of human emotion—where people openly share opinions, experiences, and feelings. However, with millions of conversations happening daily, it’s nearly impossible to manually understand the emotional pulse of such content. This project focuses on using sentiment analysis to decode the emotions hidden within these social media conversations. By doing so, we aim to uncover trends, moods, and public reactions that can be valuable for businesses, mental health monitoring, and understanding societal behaviors in real time.

2. Objectives of the Project:

* Our main goal is to build a system that can automatically analyze social media posts and determine the underlying emotions—such as happiness, anger, sadness, surprise, or neutrality. By the end of the project, we plan to:

1. Classify sentiments with good accuracy (positive, negative, neutral).
2. Visualize emotional trends over time or during specific events.
3. Offer insights that can help stakeholders make informed decisions based on public sentiment.

3. Scope of the Project:

* We aim to:

Collect social media posts (e.g., from Twitter or Reddit).

Analyze the emotional tone using natural language processing.

Develop visualizations that show emotion patterns.

Limitations:

The project is limited to English-language posts.

We’ll rely on pre-trained models and open-source datasets due to time constraints.

The focus is on offline analysis (no real-time deployment for now).

4. Data Sources:

We'll use publicly available datasets like:

Twitter datasets from Kaggle or APIs (like Tweepy).

Pre-labeled sentiment datasets (e.g., Sentiment140).

The data will be static (downloaded once and used throughout).

5. High-Level Methodology:

Here’s how we plan to tackle the project:

* Data Collection: We'll gather tweets or social media posts using APIs or existing datasets from platforms like Kaggle.
* Data Cleaning: This includes removing irrelevant characters, handling missing data, and standardizing formats.
* Exploratory Data Analysis (EDA): We’ll use graphs and word clouds to understand the frequency of emotions and popular keywords.
* Feature Engineering: We'll experiment with text vectorization (like TF-IDF or word embeddings) to improve model accuracy.
* Model Building: We'll test models like Logistic Regression, Random Forest, and deep learning models such as LSTM or BERT.
* Model Evaluation: Accuracy, F1-score, confusion matrix, and cross-validation will be used to evaluate model performance.
* Visualization & Interpretation: We'll present the findings using bar charts, pie charts, and time-series plots.
* Deployment: For now, deployment will be in the form of a Jupiter Notebook or a basic web interface using Streamlit.

6. Tools and Technologies:

Programming Language: Python

Notebook/IDE: Google Colab / Jupyter Notebook

Libraries: pandas, numpy, matplotlib, seaborn, scikit-learn, nltk, spaCy, TensorFlow/BERT (if needed)

Optional Tools for Deployment: Streamlit, Flask (if time permits).

7. Team Members and Roles:

* [G.ERAIARUT SELVI] – Data Collection & Cleaning
* [K.DEVIKA] – Model Building & Evaluation
* [V.DIVYA DHARSHINI] – Visualization & Reporting
* [K.DANIEL RENO] – Documentation & Presentation