

Phase-1 Submission

Student Name: HARIHARAN R

Register Number: 712523205024

Institution: PPG Institute Of Technology

Department: Information Technology

Date of Submission: 09/05/2025

1.Problem Statement

Social media platforms like Twitter and Facebook are flooded with emotional expressions daily, making it difficult to extract meaningful emotional insights from the vast data. Traditional sentiment analysis tools can detect basic sentiment but often miss complex emotions like anger or joy. This project aims to build an advanced sentiment analysis model to detect a broader range of emotions in real time, helping businesses and organizations track emotional trends and make better data-driven decisions.

2.Objectives of the Project

- *Develop a model that can accurately classify social media posts into different emotional categories, such as happy, sad, angry, excited, etc.*
- *Build a system that can process and analyze social media data in real-time to capture emotional trends and patterns as they emerge.*
- *Identify and visualize how emotional sentiments evolve over time, especially during major events like elections, product launches, or crises.*
- *Predict future trends based on historical sentiment data to help businesses and organizations take proactive actions.*

3.Scope of the Project

- *We will gather real-time data from social media platforms (e.g., Twitter API) using specific keywords, hashtags, or accounts related to a particular topic or event. The data will be preprocessed to remove noise, such as irrelevant posts, ads, or spam.*
- *The primary focus will be on leveraging Natural Language Processing (NLP) techniques and machine learning models like BERT, RoBERTa, or GPT-based models for sentiment analysis and emotion classification.*
- *Using data visualization tools (Python libraries like Matplotlib and Seaborn), we will create dashboards to track emotional trends over time and correlate these trends with major events.*

4.Data Sources

1. *Twitter API / X API: <https://developer.x.com/en/docs/x-api>*
2. *Kaggle Emotion Datasets: <https://www.kaggle.com/datasets/markmedhat/twitter>*

5.High-Level Methodology

- *Use static datasets like GoEmotions and Kaggle emotion datasets, along with dynamic data from Twitter and Reddit APIs using Python scripts.*
- *Remove duplicates, non-English text, special characters, URLs, and normalize text by lowercasing, removing noise, and handling emojis.*
- *Use visual tools (e.g., word clouds, bar charts, time-series plots) to explore emotion distribution and identify patterns*
- *Train models such as Logistic Regression, SVM, LSTM, and fine-tuned BERT for multi-class emotion classification.*
- *Create charts and graphs to show emotion trends and predictions using Matplotlib, Seaborn, or Streamlit dashboards.*

1. Tools and Technologies

- *Python*
- *Notebook/IDE*
- *Google Colab*
- *Jupyter Notebook*
- *VS Code*

Team Members and Roles

Team Member	Role	Responsibility
Danish P	<i>Coordinator</i>	<i>Assist in sentiment analysis and review code</i>
Abishek S	<i>Backend Support</i>	<i>Responsible for assisting in the development and maintenance of backend systems,</i>
Hariharan R	<i>Data and Research Specialist</i>	<i>Collect datasets and support data preprocessing</i>
Atsaya D	<i>Frontend and Documentation Support</i>	<i>Assist in building the UI and Documenting the process</i>

<i>Karthika V</i>	<i>Tester and Reviewer</i>	<i>Responsible for conducting thorough testing, reporting bugs, providing suggestions for improvements, and contributing to overall product quality.</i>
-------------------	----------------------------	--