**Phase-1 Submission Template Student Name:** ARUNKUMAR.P

**Register Number:** 412723104009

**Institution:** TAGORE ENGINEERING COLLEGE

**Department:** COMPUTERSCIENCE AND ENGINEERING

**Date of Submission: 23/04/2025**

# Problem Statement

In today’s digital era, social media has become a key platform where people express opinions and emotions. Understanding these emotions can provide valuable insights for businesses, mental health monitoring, and public sentiment analysis. This project focuses on decoding human emotions using sentiment analysis techniques applied to social media conversations. By using NLP and AI models, the goal is to categorize text into emotional states (like happy, sad, angry, etc.) and understand sentiment trends.

# Objectives of the Project

 Collect and preprocess social media text data (e.g., tweets, Instagram captions)

 Perform sentiment analysis using NLP techniques.

 Classify emotions such as joy, anger, sadness, etc.

 Visualize emotion trends across conversations or time periods.

Compare different machine learning models (Logistic Regression, Naive Bayes, LSTM, etc.).

# Scope of the Project

1. Focused on analyzing English-language posts from platforms like Twitter or Reddit.
2. Emotion detection is limited to text-based data (no image/video analysis).
3. Emotions considered: joy, anger, fear, surprise, sadness, etc.
4. Limited to historical and static datasets (no real-time deployment in Phase-1).
5. Final output will be a model and dashboard/report for emotion classification

# Data Sources

1. source : Kaggle datasets (e.g., Tweet Emotion Dataset, Sentiment140)
2. Twitter API (if access is granted for tweet extraction)
3. Other public datasets for emotion/sentiment classification

# High-Level Methodology

* + **Data Collection** – Download datasets from Kaggle or fetch tweets via Twitter API.
  + **Data Cleaning** *Remove stopwords, emojis, links, and perform tokenization.*
  + **Exploratory Data Analysis (EDA)** –– Understand emotion distribution, word frequency, etc.
  + **Feature Engineering** –– Convert text into vectors using TF-IDF, Word2Vec, or embeddings
  + **Model Building** – Train models like Logistic Regression, Naive Bayes, or LSTM for emotion classification.
  + **Model Evaluation** – Use metrics like Accuracy, F1-Score, and Confusion Matrix.
  + **Visualization & Interpretation** – Pie charts, bar plots, and word clouds for emotions and trends.
  + **Deployment** – – Possible deployment as a Streamlit app to test input text live.

# Tools and Technologies

## **● Programming Language:**

## Python

## **● Notebook/IDE:**

## Google Colab / Jupyter Notebook

## **● Libraries:**

## pandas, numpy – Data processing

## matplotlib, seaborn – Visualization

## scikit-learn – traditional ML models

## tensorflow.keras – other deep learning models

## nltk, spacy – Text preprocessing

## tweepy, snscrape – Social media scraping

# 7.Team Members and Roles

|  |  |
| --- | --- |
| **NAME** | **ROLE** |
| AVINASH N | Model Building & PRESENTATION |
| ARUNAGIRI D | DATA COLLECTION |
| ARUNKUMAR P | Report & EVALUATION |