Phase-1 Submission

Student Name: Dharshini V

Register Number: 712523205301

Institution: PPG Institute of Technology

Department: B.Tech Information Technology

Date of Submission:

1.Problem Statement

Social media platforms are a rich source of real-time, unfiltered public opinions and emotions. However, manually understanding the emotional tone of vast conversations is not feasible. This project aims to automate the process of decoding emotion from social media.

2.Objectives of the Project

1. Classify social media conversations into different sentiment categories (e.g., positive, negative, neutral).

2. Detect emotions such as joy, anger, sadness, fear, and surprise.

3. Identify patterns in emotional shifts over time or in response to events.

4. Visualize the distribution of sentiments and emotions for easier interpretation.

3. Scope of the Project

Features:

- *Sentiment polarity (positive/negative/neutral)*
- *Emotion classification (joy, anger, sadness, etc.)*

Constraints:

- *Focus on English language posts only*
- *Analyze only public tweets/posts (due to privacy/legal issues)*

4. Data Sources

- ***Sentiment140*** *(Twitter-based dataset with pre-labeled sentiments)*
- ***Kaggle datasets*** *related to Twitter sentiment and emotion*
- *Optionally, data collected via **Twitter API** (subject to rate limits and account permissions)*

5. High-Level Methodology

Data Collection:*Use pre-existing datasets from Kaggle and possibly collect live tweets via the Twitter API.*

Data Cleaning:*Remove stopwords, emojis, special characters, and URLs.*

Exploratory Data Analysis (EDA):*Visualize sentiment distribution, word clouds, hashtag analysis.*

Feature Engineering:*Convert text to numerical features using TF-IDF, word embeddings (Word2Vec, BERT).*

Model Building:*Experiment with Logistic Regression, Naive Bayes, SVM for sentiment classification.*

Model Evaluation:*Use accuracy, precision, recall, F1-score, and confusion matrix.*

6. Tools and Technologies

Programming Language: *Python*

Notebook/IDE: *Google Colab / Jupyter Notebook*

Libraries:

- *pandas, numpy – data processing*
- *matplotlib, seaborn, wordcloud, plotly – visualization*
- *scikit-learn, NLTK, TextBlob, transformers – NLP and ML*
- *tweepy – Twitter API (if applicable)*

7. Team Members and Roles

S.NO	NAME	ROLE	DESCRIPTION
1	**DHARSHINI V**	*Data Collection, Cleaning*	**Data Collection:** *Social media data is sourced from public datasets and optionally via the Twitter API.*
2	**VASANTHA PRIYAN E**	*EDA*	*EDA explores sentiment and emotion .*
3	**KRISHNAMOORTHI M**	*Model Development, Evaluation*	*Sentiment and emotion models are built and assessed .*
4	**PRIYAN P**	*Visualization,Interpretation*	*A visualization is agraphical representation of data or concepts.*
5	**BALAGANESH V**	*Documentation, Deployment*	*Documentation provides detailed information .*