Phase-2 Submission

Student Name: Dharshini V

Register Number: 712523205301

Institution: PPG INSTITUTE OF TECHNOLOGY

Department: B.TECH INFORMATION TECHNOLOGY

Date of Submission: 09/05/2025

Github Repository Link: https://github.com/Dharshini-0905/NM_DHARSHINI_DS

---

1. Problem Statement

*The project focuses on designing and implementing a robust model capable of decoding human emotions expressed in social media text data. This task falls under the domain of **multi-class classification**, wherein each input text is classified into one of the predefined emotional categories, such as happy, sad, angry, fearful, surprised, or neutral. The aim is to improve sentiment detection by leveraging Natural Language Processing (NLP) and machine learning techniques.*

2. Project Objectives

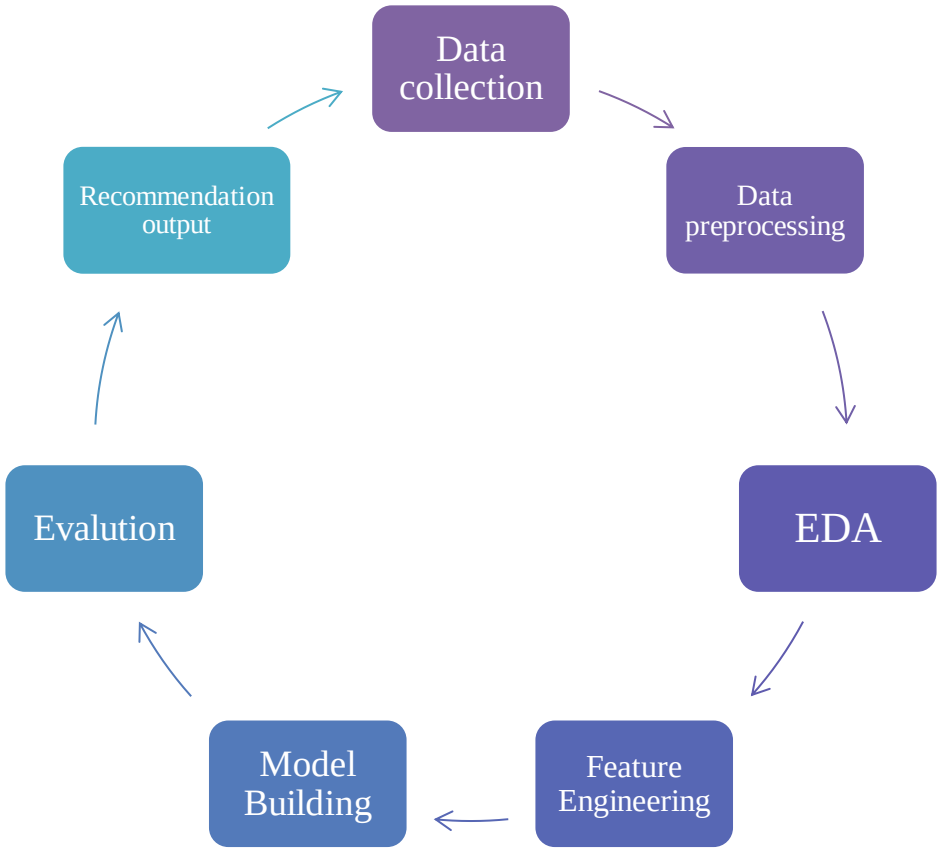
To extract and classify emotional cues from user-generated text content on social media platforms.

To employ advanced NLP techniques for effective preprocessing and feature representation.

To develop, train, and compare both machine learning and deep learning models for emotion classification.

To optimize model performance through evaluation metrics and hyperparameter tuning.

3. Flowchart of the Project Workflow



4. Data Description

- ***Dataset Name: Emotion Detection from Text***
- ***Source: Kaggle / Twitter API / Public NLP repositories***
- ***Data Type: Unstructured (text data)***
- ***Size: Approximately 20,000 to 50,000 social media posts/comments***

- ***Features:***

- ***Input: Text***
- ***Output: Emotion Label***
  - ***Target Variable: Emotion class (happy, sad, angry, fear, etc.)***
  - ***Dataset Type: Static***

5. Data Preprocessing

Performed the following steps to prepare the data:

- *Handled missing values by removing or imputing records.*
- *Identified and removed duplicate records.*
- *Treated outliers where applicable.*
- *Converted data types and ensured uniform formatting.*
- *Encoded categorical variables using Label Encoding and One-Hot Encoding.*
- *Normalized and standardized numerical features when required.*
- *Documented all transformation steps in code with explanatory markdowns.*

6. Exploratory Data Analysis (EDA)

- *Removed HTML tags, special characters, and emojis from text.*
- *Converted all text to lowercase for uniformity.*
- *Applied tokenization and lemmatization using NLTK and SpaCy.*

- *Removed stopwords to reduce noise in the dataset.*
- *Encoded emotion labels using LabelEncoder.*
- *Transformed text using vectorization techniques such as:*
  - *Bag of Words (BoW)*
  - *TF-IDF (Term Frequency-Inverse Document Frequency)*
  - *Word Embeddings (Word2Vec / GloVe)*

7. Feature Engineering

Reiterated preprocessing steps and enhanced feature representation through:

- *Contextual embedding methods for deeper understanding of sentiment (e.g., BERT-based embeddings).*
- *Dimensionality reduction using PCA or T-SNE, where applicable.*

8. Model Building

The following models were trained and evaluated for performance:

- *Logistic Regression*
- *Random Forest Classifier*
- *Multinomial Naive Bayes*
- *Support Vector Machine (SVM)*
- *Long Short-Term Memory (LSTM) Neural Network*

9. Visualization of Results & Model Insights

- ***Confusion Matrix:*** *Assessed performance across emotion categories.*
- ***ROC Curve:*** *Visualized multi-class ROC-AUC scores.*

- ***Feature Importance:*** *Interpreted influential features (particularly for Random Forest & Logistic Regression).*
- ***Word Clouds:*** *Generated for the most frequent terms per emotion class.*

10. Tools and Technologies Used

- *Programming Language: Python*
- *Development Environment: Jupyter Notebook / Google Colab*
- *Libraries & Frameworks:*
  - *Pandas, NumPy*
  - *NLTK, SpaCy*
  - *Scikit-learn*
  - *TensorFlow / Keras*
  - *Matplotlib, Seaborn, WordCloud*

11. Team Members and Contributions

S.NO	***NAME***	***ROLE***
1	***DHARSHINI V***	*Data Collection, Cleaning*
2	***VASANTHA PRIYAN E***	*EDA*
3	***KRISHNAMOORTHI M***	*Model Development, Evaluation*
4	***PRIYAN P***	*Visualization, Interpretation*
5	***BALAGANESH V***	*Documentation, Deployment*