**Phase-3 Submission**

**Student Name:** PRIYAN P

**Register Number:** 712523205045

**Institution:** PPG Institute of Technology

**Department:** B.Tech Information Technology

**Date of Submission:** 16/05/2025

**Github Repository Link*:*** *<https://github.com/P-priyan-05/NM_PRIYAN_DS>*

### **1. Problem Statement**

*This project aims to build a multi-class classification model for detecting human emotions from social media text data. Given the growing reliance on sentiment analysis in marketing, customer support, and mental health applications, accurate emotion recognition provides actionable insights.* *This project aims to build a multi-class classification model for detecting human emotions from social media text data. Given the growing reliance on sentiment analysis in marketing, customer support, and mental health applications, accurate emotion recognition provides actionable insights*.

### **2. Abstract**

*We aim to design a system that decodes emotions in textual data from social media. This multi-class classification task leverages Natural Language Processing for text cleaning and feature extraction, followed by model training with various machine learning and deep learning algorithms. We used vectorization methods like TF-IDF and contextual embeddings (e.g., BERT) to represent text data. Models such as Random Forest, SVM, and LSTM were evaluated using classification metrics. The best-performing model was deployed using Streamlit. This solution facilitates real-time emotion analysis for various applications.*

### **3. System Requirements**

***Hardware:***

* *RAM: Minimum 8 GB*
* *Processor: i5 or equivalent (for model training and embeddings)*

***Software:***

* *Python 3.9+*
* *Jupyter Notebook / Google Colab*
* *Libraries: pandas, numpy, scikit-learn, nltk, spacy, seaborn, matplotlib, tensorflow, keras, wordcloud*

### **4. Objectives**

* *Extract and classify emotions from social media text.*
* *Apply NLP techniques for preprocessing and representation.*
* *Compare traditional ML models with deep learning approaches.*
* *Evaluate performance using multi-class classification metrics.*
* *Deploy a user-friendly interface for real-time emotion detection.*

**5. Flowchart of Project Workflow**

*Data Collection → Preprocessing → EDA → Feature Engineering → Modeling → Evaluation → Deployment*

*(Insert flowchart image made using draw.io, Canva, etc.)*

### **6. Dataset Description**

* ***Source:*** *Kaggle / Twitter API / Public NLP Repositories*
* ***Type:*** *Public*
* ***Size:*** *~20,000–50,000 rows*
* ***Structure:***
  + *Input: Social media text*
  + *Output: Emotion label (happy, sad, etc.)*
* ***Sample:***(Insert screenshot of *df.head()*)

### **7. Data Preprocessing**

* *Removed HTML tags, emojis, special characters*
* *Tokenized and lemmatized text (using NLTK, SpaCy)*
* *Removed stopwords*
* *Encoded emotion labels (LabelEncoder)*
* *Text transformed using BoW, TF-IDF, Word2Vec*
* (Include before/after screenshots if available)

### **8. Exploratory Data Analysis (EDA)**

* *Word frequency analysis (word clouds)*
* *Emotion class distribution*
* *Text length analysis*
* *Heatmaps for correlations between emotion categories*
* (Include screenshots of visualizations)

### **9. Feature Engineering**

* *BERT embeddings for deeper semantic understanding*
* *Dimensionality reduction (PCA, T-SNE)*
* *Feature selection via model-based importance (Random Forest)*
* *Added contextual and syntactic features*

### 

### **10. Model Building**

*Trained the following models:*

* *Logistic Regression*
* *Random Forest*
* *SVM*
* *Naive Bayes*
* *LSTM (Keras)*

*Model selection based on:*

* *Performance metrics*
* *Ability to generalize to unseen data*

### **11. Model Evaluation**

* *Accuracy, Precision, Recall, F1-score*
* *Confusion Matrix (multi-class)*
* *ROC-AUC (macro and micro average)*
* *Model comparison table*

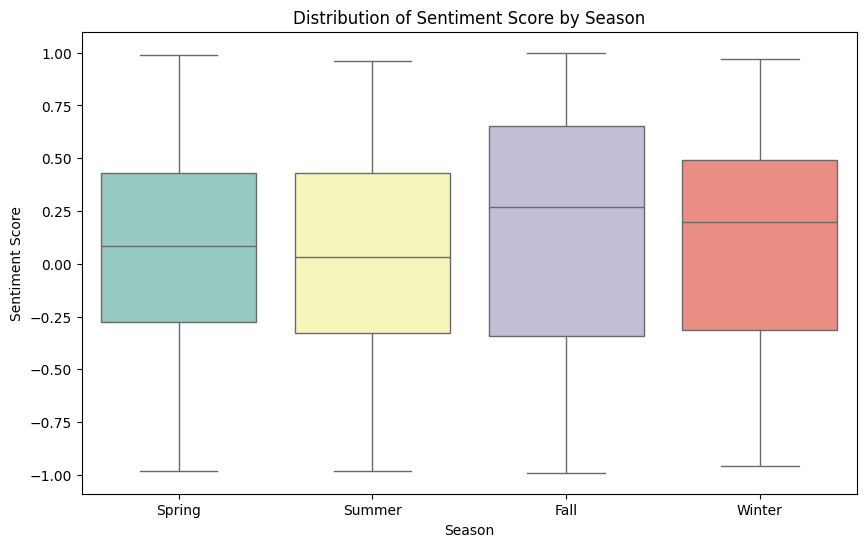
### **12. Deployment**

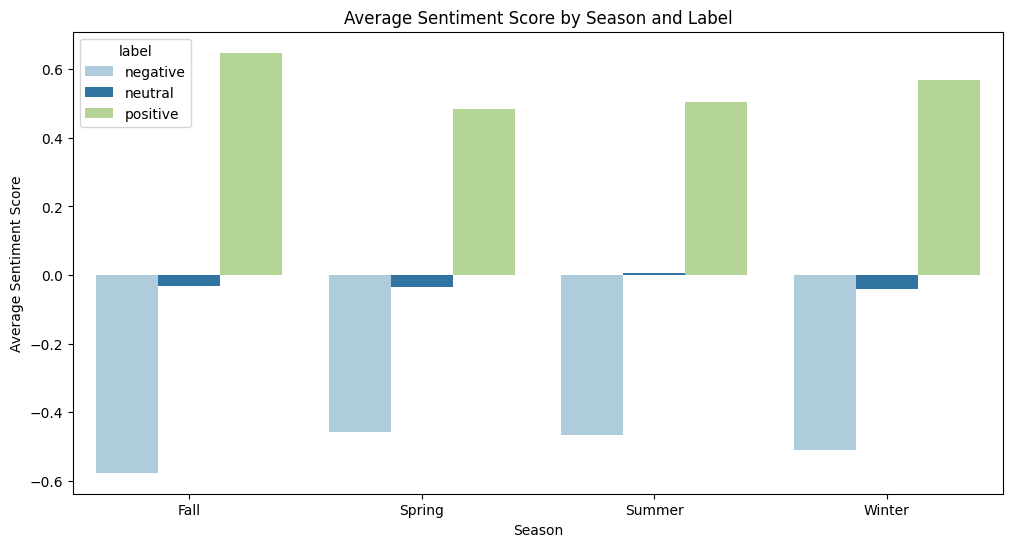
***Platform:*** *Streamlit Cloud****Deployment Link:*** *[Insert live link]****UI Screenshot:***(Insert UI screenshot)***Sample Prediction Output:*** *Input: “I feel so lonely today.”  
Output:* ***Emotion: Sad***

**13. Source code**

*Complete source code is available in the GitHub repository:* [*https://github.com/P-priyan-05/NM\_PRIYAN\_DS*](https://github.com/P-priyan-05/NM_PRIYAN_DS)

*Output:*

**

**

**14. Future scope**

* *Integrate multimodal sentiment analysis (text + images)*
* *Expand emotion categories using hierarchical classification*
* *Real-time social media monitoring dashboard*
* *Feedback loop with user correction to improve model over time*

**15. Team Members and Roles**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| | ***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* |