**Decoding emotions through sentiment analysis of**

**social media conversations**

**Phase-3**

**Student Name:** [A.NIVETHITHA]  
**Register Number:** [621123104048]  
**Institution:** [IDHAYA ENGINEERING COLLEGE FOR WOMEN]  
**Department:** [COMPUTER SCIENCE AND ENGINEERING]  
**Date of Submission:** [06\05\2025]

**1. Problem Statement**

In today's digital era, social media has become a powerful platform for expressing public opinion and emotions. However, the sheer volume and unstructured nature of social media content make it challenging to analyze and extract meaningful insights. This project focuses on decoding emotions embedded in social media conversations using sentiment analysis. The aim is to classify posts into various emotional categories (e.g., joy, anger, sadness, fear) to help brands, policymakers, or mental health professionals understand public sentiment.  
**Problem Type:** Multi-class Classification

**2. Abstract**

This project applies natural language processing (NLP) techniques to analyze and decode emotions from social media conversations. The main objective is to build a machine learning model that classifies textual posts into emotional categories such as joy, anger, sadness, and fear. We collected a dataset from public social media sources, cleaned and preprocessed the data, and applied exploratory data analysis to uncover trends. Several classification models were evaluated, including logistic regression, random forest, and deep learning-based LSTM. The best-performing model was deployed using Streamlit, providing a user-friendly interface to predict emotions from text input.

**3. System Requirements**

* **Hardware:**
  + Minimum 4GB RAM
  + Dual-core processor (i5 or above for training deep learning models)
* **Software:**
  + Python 3.8+
  + Libraries: pandas, numpy, scikit-learn, nltk, matplotlib, seaborn, tensorflow, streamlit
  + IDE: Google Colab / Jupyter Notebook / VS Code

**4. Objectives**

* Build a model to classify social media posts into emotional
* Visualize emotional trends over time or topics.
* categories.Help organizations gain emotional insights from online discussions.
* Deploy the model via an accessible web app for real-time use.

**5. Flowchart of Project Workflow**

*(Create a visual using draw.io or Canva, and insert here)*  
Example steps:

java

CopyEdit

Data Collection → Data Cleaning → Tokenization & Lemmatization → EDA → Feature Extraction (TF-IDF or Embeddings) → Model Training (ML/DL) → Evaluation → Streamlit Deployment

**6. Dataset Description**

* **Source:** Kaggle / Twitter API / Public emotion dataset (e.g., Emotion Dataset by Saravia et al.)
* **Type:** Public
* **Size & Structure:** 20,000+ rows, 2 columns (text, emotion)
* **Sample Screenshot:**  
  *(Insert a screenshot of df.head())*

**7. Data Preprocessing**

* Removed URLs, mentions, hashtags, and special characters.
* Tokenized text, removed stopwords, and applied lemmatization.
* Encoded emotion labels.
* Transformed text using TF-IDF and Word Embeddings.
* *(Insert before/after screenshots or sample code)*

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

* Plotted distribution of emotions.
* Word clouds for each emotion.
* Heatmap showing correlation between length of text and emotion.
* Key insights:
  + "Joy" was the most common emotion.
  + "Fear" and "Sadness" often occurred in longer posts.
* *(Insert histograms, boxplots, etc.)*

**9. Feature Engineering**

* Created new features like word count and sentiment score.
* Used TF-IDF and GloVe embeddings.
* Feature selection based on model performance and interpretability.

**10. Model Building**

* Models tried:
  + Logistic Regression (Baseline)
  + Random Forest
  + LSTM Neural Network (Best Performance)
* Chose LSTM for capturing sequence patterns in text.
* Hyperparameter tuning done via GridSearchCV for ML models.
* *(Insert training screenshots)*

**11. Model Evaluation**

* Metrics:
  + Accuracy: 85%
  + F1-Score: 84% (macro average)
  + AUC-ROC: Not applicable (multi-class)
* Confusion matrix and classification report used.
* *(Insert confusion matrix and model comparison table)*

**12. Deployment**

* **Platform:** Streamlit Cloud
* **Public Link:** [Insert live link]
* **UI Screenshot:** *(Insert image)*
* **Sample Output:**
  + Input: *“I just got a promotion!”* → Output: *Emotion: Joy*

**13. Source Code**

All source code is available on GitHub, including notebooks, preprocessing scripts, models, and deployment code.  
[Insert GitHub link]

**14. Future Scope**

* Expand emotion categories using a larger, more diverse dataset.
* Integrate speech-to-text for analyzing voice inputs.
* Deploy as an API service for integration with third-party apps.

**15. Team Members and Roles**

| **Name** | **Role** | **Responsibility** |
| --- | --- | --- |
| **R. Parameswari** | **Team Lead** | **Model building, deployment, final integration** |
| **A. Nivethitha** | **Data Engineer** | **Data collection, preprocessing, feature engineering** |
| **R. Thavaselvi** | **Analyst** | **EDA, visualization, insights extraction** |
| **S. Swathi** | **Documentation Lead** | **Report writing, screenshots, presentation prep** |