**Phase-1**

**Student Name:** ANUSREE.M,ARUNAETHI.M.S, ARUNIKA.E,CHELLA,VETHAVALLI .J

**Register Number:**613023243006, 613023243007, 613023243008, 613023243009, 613023243060.

**Institution:**VIVEKANANDHA COLLEGE OF TECHNOLOGY FOR WOMEN.

**Department:**B.TECH-AI&DS

**Date of Submission:**25-04-2025

**1.Problem Statement:**

In the age of digital communication, social media platforms have become a primary space for individuals to express their thoughts, feelings, and opinions. However, the vast volume and unstructured nature of this data make it challenging to understand and interpret the emotional undertones in these conversations. The problem is to develop an effective system that can decode human emotions by performing sentiment analysis on social media conversations. This involves identifying and categorizing emotional states—such as happiness, anger, sadness, fear, or surprise—expressed in textual data from platforms like Twitter, Facebook, or Instagram . The goal is to enable deeper insights into public sentiment and emotional trends, which can be leveraged for applications in marketing, public health, crisis management, and more.

**2.Objectives of the Project:**

1. To identify and extract emotional content from social media conversations using natural language processing (NLP) techniques.

2. To classify sentiments expressed in posts, comments, and tweets into categories such as positive, negative, neutral, or specific emotions (e.g., joy, anger, sadness, fear, surprise).

3. To develop or utilize sentiment analysis models that can accurately interpret informal, slang-rich, and context-dependent language typical of social media.

4. To analyze patterns and trends in emotional expression across different timeframes, events, or topics.

5. To assess the impact of influential posts or users on the overall sentiment of social media discussions.

6. To visualize sentiment dynamics through dashboards or graphs that help in understanding public opinion or emotional shifts over time.

7. To support decision-making for businesses, policymakers, or mental health researchers based on insights gained from emotional sentiment analysis.

**3.Scope of the Project:**

**Features to Include**:

1. Data Collection Module:Integration with APIs (e.g., Twitter API, Reddit API) to fetch real-time or historical social media data.

2. Emotion and Sentiment Classification:Use of pre-trained models or fine-tuned classifiers to detect sentiments (positive, negative, neutral).

3. Visualization Dashboard:Graphs and charts to display emotional distribution, sentiment polarity, and trends over time.

4. Report Generation:Auto-generation of summary reports with visual and textual insights.

**Limitations and Constraints:**

1. Limited accuracy in emotion detection.

2.Lack of contextual understanding.

3.Privacy and ethical consideration.

**4.Data Source:**

**Dataset Source**: Google AI Blog / GitHub.

**Type**: Reddit comments labeled with 27 emotion categories.

**Nature:**Multilabel emotional classification,Balanced and cleaned dataset,Useful for training and fine-tuning emotion detection models.

**5.High-Level Methodology:**

**1. Data Collection:**Extract data from open-source datasets (e.g., Sentiment140, GoEmotions, Kaggle) and via APIs (e.g., Twitter, Reddit).

**2. Data Cleaning:**Remove duplicates, irrelevant data, and stop words.Normalize text: lowercase, correct spelling, expand contractions. (e.g., URLs, mentions, hashtags).

**3. Exploratory Data Analysis (EDA):**Generate word clouds and frequency plots to identify common terms and sentiment-bearing phrases.Analyze trends across time, topics, or sources.

**4. Feature Engineering:**WordEmbeddings such as Word2Vec, GloVe, or BERT embeddings for deep learning models**.**Incorporateemoji analysis and sentiment scores as features.

**5. Model Building:**Transformer-based models like BERT or RoBERTa for contextual understanding**.**Traditional models (e.g., Logistic Regression, SVM) for baseline comparison.

**6. Model Evaluation:**Use metrics such as accuracy, precision, recall, F1-score, and confusion matrix.Conduct user-based evaluation if applicable, e.g., rating of model output relevance or emotional accuracy.

**7. Visualization &Interpretation:**UseMatplotlib, Seaborn, and Plotly to visualize Sentiment/emotion distribution,Model performance metrics,Emotion trends over time or topics.

**8. Deployment:**Deploy the final sentiment analysis tool as a web app using Streamlit or Flask.

**6.Tools and Technologies:**

**1.Programming Language** –*Python,R(optional),Java script(optional).*

**2.Notebook/IDE-***Google Colab, JupyterNotebook,kagglekernels,pycharm.*

**3.Libraries** –*pandas, numpy, seaborn, matplotlib,scikit-learn,tensor flow.*

**4.Optional Tools for Deployment** –*Streamlit, Flask,Github,Jupiterlab.*

**7.Team Members and Roles:**

* **Anusree M**

**Responsibilities:**Coordinate overall project progress and deadlines.Allocate tasks and manage collaboration tools (e.g., Git, Trello, etc.).Ensure timely communication among team members.Prepare documentation and final project report/presentation.

* **Arunaethi M S**

**Responsibilities:**Source data from APIs or datasets (Twitter, Reddit, GoEmotions, etc.).Clean and preprocess text (tokenization, stop-word removal, etc.).Handle data augmentation if needed .Ensure data is ready for analysis/modelling.

* **Arunika E**

**Responsibilities:**Perform sentiment/emotion-specific feature extraction (TF-IDF, word embeddings, etc.).Handle emotion lexicons or use libraries like NLTK, spaCy, or Hugging Face.Explore and implement techniques for emotion classification (e.g., keyword-based, BERT).

* **Chella**

**Responsibilities:**Design and train ML/DL models (e.g., SVM, LSTM, BERT, etc.).Tune hyperparameters and evaluate models (accuracy, F1-score, etc.).Implement cross-validation and test for robustness.Visualize and interpret model predictions.

* **VethavalliJ**

**Responsibilities:**Analyze model outcomes and emotional trends from the data.Create dashboards or visualizations using matplotlib, seaborn, or Plotly.Interpret insights for real-world applications (e.g., crisis detection, brand sentiment).Help with the final presentation design and storytelling.