**PROJECT TITLE**:DECODING EMOTIONS THROUGH SENTIMENT ANALYSIS OF SOCIAL MEDIA CONVERSATIONS

**PHASE-1**

**STUDENT NAME : MOULITHARAN G**

**REGISTER NUMBER: 620523243036**

**INSTITUTION : CMS COLLEGE OF**

**ENGINEERING ,**

**NAMAKKAL .637003**

**DEPARTMENT : B.TECH- ARTIFICIAL**

**INTELLIGENCE AND**

**DATA SCIENCE**

**DATE OF SUBMISSION:**

**GITHUB REPOSITORY**

**LINK :**

1.**ProblemStatement**:

In the digital age, social media platforms have become primary spaces for individuals to expresstheir emotions, opinions, and experiences. These unstructured conversations offer valuable insightsinto public sentiment and emotional trends. This project focuses on leveraging natural languageprocessing techniques to analyze and classify emotions in social media content. By decodingemotional patterns in real time, organizations can make informed decisions, improve customerexperiences, and detect early signs of mental health issues or societal trends.

2.**Objectives of the Project:**

- To develop a machine learning model that can classify the emotional tone of social media posts.

- To identify and extract dominant emotions and sentiments in conversations.

- To uncover emotional trends over time or across topics.

- To provide actionable insights through visual dashboards or tools.

3.**Scope of the Project**:

Features to Analyze:

- Textual data from social media platforms (e.g., tweets, Reddit posts)

- Linguistic features (e.g., word choice, emoji usage, punctuation)

- Metadata (e.g., timestamp, user location if available)

Constraints and Limitations:

- Access limited to publicly available posts via APIs or datasets.

- Emotion detection is language-specific (focus on English).

- Posts may contain sarcasm, slang, or code-switching which challenges emotion detection.

- Model performance dependent on data quality and annotation.

4.**Data Sources**:

- Dataset: Emotion-labeled datasets (e.g., GoEmotions, Sentiment140, Twitter datasets)

- Source: Google Dataset Search, Kaggle, HuggingFace Datasets

- Type: Public

- Nature: Static or collected via API

5.**High-Level Methodology:**

Data Collection:

- Use Twitter API or Reddit scraping tools to gather recent posts.

- Download existing labeled datasets from repositories.

Data Cleaning:

- Remove irrelevant text (URLs, mentions, hashtags).

- Handle emojis, special characters, and non-English content.

- Normalize text: lowercase, lemmatization, etc.

Exploratory Data Analysis (EDA):

- Analyze distribution of emotions (happy, sad, angry, etc.)

- Visualize word frequency and sentiment patterns.- Group by date/time to find emotion trends.

Feature Engineering:

- Convert text to numeric features (TF-IDF, Word2Vec, BERT embeddings)

- Include additional features like emoji presence, sentiment scores.

Model Building:

- Use NLP models: Naive Bayes, SVM, LSTM, BERT

- Experiment with multi-class classification for emotion categories.

Model Evaluation:

- Metrics: Accuracy, Precision, Recall, F1-Score

- Confusion matrix for emotion categories

- Cross-validation and testing with unseen data

Visualization & Interpretation:

- Emotion frequency charts, timelines

- Word clouds and heatmaps

- Interactive dashboards

Deployment:

- Build a basic interface using Streamlit or Gradio

- Allow users to input text and see detected emotions.

6.**Tools and Technologies:**

- Programming Language: Python

- Notebook/IDE: Google Colab, Jupyter Notebook

- Libraries:

- Data Handling: pandas, numpy

- NLP: nltk, spaCy, transformers (HuggingFace)

- Modeling: scikit-learn, TensorFlow, PyTorch

- Visualization: matplotlib, seaborn, plotly

- Deployment Tools: Streamlit, Gradio, Flask