PHASE-1

**Student Name:** **DHANUSH M**

**Register Number:** **422623104024**

**Institution:** **University College of Engineering Panruti**

**Department:** **Computer Science and Engineering**

**Date Of Submission: 29.04.2025**

**🧩 1. Problem Statement**

* Social media has become a powerful medium where people express their emotions.
* Unstructured data from platforms like Twitter can reveal public sentiment about sensitive topics like mental health.
* There is a growing need to automatically detect and understand these emotional signals at s

**🎯 2. Objectives of the Project**

* To collect real-time tweets related to mental health using Twitter API v2.
* To preprocess and clean tweet data for analysis.
* To classify tweets into Positive, Negative, or Neutral emotions using sentiment analysis.
* To visualize sentiment trends and provide insights.
* To lay the groundwork for potential deployment in mental health monitoring tools.

**🔭 3. Scope of the Project**

* Focused on English-language tweets related to *mental health*.
* Uses keyword-based tweet extraction (e.g., "mental health").
* Employs basic NLP techniques with TextBlob for sentiment classification.
* Short-term scope: sentiment detection.
* Long-term potential: real-time emotional monitoring, mental health dashboards, integration with chatbots, etc.

**🗂️ 4. Data Resources**

* **Primary Source:** Twitter API v2
* **Query Used:** "mental health -is:retweet lang:en"
* **Volume:** 50 most recent tweets
* **Fields Collected:** Text content of tweets
* tweets

**🛠️ 5. High-Level Methodology**

**📥 Data Collection**

* Twitter API v2 via Tweepy.
* Fetch tweets related to mental health (non-retweets, English).

**🧹 Data Cleaning**

* Remove URLs, mentions, hashtags, special characters.
* Convert text to lowercase for normalization.

**📊 Exploratory Data Analysis (EDA)**

* Distribution of sentiment categories (positive, negative, neutral).
* Word frequency analysis (optional improvement).

**🧬 Feature Engineering**

* Not applicable in basic TextBlob sentiment analysis (already has built-in features).
* Future improvements: TF-IDF, word embeddings.

**🧠 Model Building**

* TextBlob for rule-based sentiment scoring using polarity.

**✅ Model Evaluation**

* Limited as it's rule-based (no train/test split).
* Could add manual evaluation or benchmark against labeled datasets.

**📈 Visualization & Interpretation**

* Count plot of sentiment categories using Seaborn.
* First few tweets and their associated sentiments displayed in a table.

**🚀 Deployment (Optional/Future Scope)**

* Could deploy as a Flask API, Streamlit app, or integrate with dashboards.
* Real-time sentiment tracking for public awareness or organization use.

**🧰 6. Tools and Technologies**

**💻 Programming Language**

* **Python**: powerful for data manipulation, NLP, and visualization.

**📒 Notebook/IDE**

* **Google Colab**: cloud-based notebook for coding, visualization, and sharing.

**📚 Libraries**

* tweepy – to interact with Twitter API.
* textblob – for sentiment analysis.
* pandas – for data manipulation.
* matplotlib & seaborn – for visualizing data.

**⚙️ Optional Tools for Deployment**

**Flask / Streamlit** – for creating web apps.

**Heroku / Render** – for deploying apps online.

**Docker** – for containerization (advanced use case).

**📊 7. Results & Visualization**

* Bar chart showing the number of tweets per sentiment category.
* Example tweet table with sentiment classification.

**🧠 8. Conclusion & Future Work**

* Successfully demonstrated emotion detection in social media data.
* Shows potential for tracking public sentiment on mental health.
* Future improvements:
  + Use of deep learning models like BERT.
  + Larger datasets for training & testing.
  + Real-time dashboard or alert system.

**🤝 9. Team Members and Roles**

**Team Member Roles**

**MADHUMATHI S Project Lead, Twitter API Integration.**

**SIVA JENANI S Sentiment analysis logic, Data cleaning.**

**KAVIYA P Exploratory data analysis, Model building.**

**DHANUSH M Presentation Design, Deployment Research.**

**RAGAVAN R Report Writing, Visualization.**