Phase-1

Student Name: Gagan.P.V

Register Number: 410723106022

Institution: Dhanalakshmi college of engineering

Department: Electronics and Communication Engineering

Date of Submission: 28/04/2025

1.Problem Statement

DECODING EMAIONS THOUGH SENTIMENT ANALYSIS OF SOCIAL MEDIA CONVERSATION

Social media platforms are overflowing with usergenerated content, including conversations about brands, products, services, and public events. However, this data is often unstructured and overwhelming in volume. Businesses and organizations struggle to monitor, understand, and act upon these conversations in real time. Analyzing social media conversations is critical for sentiment analysis, trend detection, crisis management, and improving customer engagement strategies

2.Objectives of the Project

1. Collect and analyze social media conversation data (e.g. tweets, posts).
2. Perform sentiment analysis to classify conversations as positive, neutral, or negative.
3. Identify trending topics and keywords in real-time.
4. Provide visual insights through dashboards to assist in decision-making.

5. Optionally, deploy a simple interface where users can input hashtags or keywords and see real-time analytics

3.Scope of the Project

List the features you plan to analyze or build as part of the project. Also, specify any limitations or constraints, such as restrictions on deployment, use of only certain models, or reliance on specific datasets or tools.

4.Data Sources 1. Describe the dataset you will use

in the project.

2. Mention its source (Kaggle,UCI, APIs, synthetic, etc.), indicate whether it is public, private, or generated by you, and specify whether the dataset is static (downloaded once) or dynamic (updated in real-time).

Data Set Link:

https://www.kaggle.com/datasets/vidyapb/elon-musk-tweets-2015-to-2020?resource=download

5.High-Level Methodology

1. Data Collection

   a. Explain how you will obtain the data (e.g., download, scrape, API access, generate synthetically**)**

2. Data Cleaning

   a. Identify potential issues such as missing values, duplicates, or inconsistent formats, and describe how you plan to address them**.**

3. Exploratory Data Analysis (EDA)

a. Describe the techniques or visualizations you'll use to uncover patterns, trends, and relationships in the data.

4. Feature Engineering

a. Indicate whether you will create new features or transform existing ones to improve model performance**.**

5. Model Building

i. List the types of algorithms or models you plan to experiment with, and mention why they are suitable for your problem.

6. Model Evaluation

i. Specify the metrics or validation strategies you will use to measure model accuracy and ef ectiveness. *Visualization & Interpretation - Explain how you will present the key findingsr insights, or predictions (e.g., chartsr graphs, dashboards).

7. Deployment

i. State whether you will deploy your project (e.g., as a web app, dashboard, or notebook) and briefly describe the method or tools you plan to use, if applicable.]

6.Tools and Technologies

1. **List the tools, programming languages, and libraries you plan to use in your project. Include the following detailsProgramming Language - State the main language you will use (e.g. Python, R).**

2. **Notebook/lDE - Mention the platform or environment you' Il work in (e.g**
3. **Google Colab, Jupyter Notebook, VS Code)**
4. **Libraries - List the key libraries you plan to use for data processing, visualization, and modeling (e.g., pandas, numpy, seaborn, matplotlibr scikit-learn, TensorFlow)**
5. **Optional TOols for Deployment - If applicable, name any tools or frameworks you might use for deployment (e.g., Streamlit, Flask, Gradio,**

7.Team Members and Roles

S.NO	NAMES	ROLES	RESPONSIBILITY
1	Karthik S	Leader	Data Collection, Data Cleaning
2	Santha kumar P	Member	Visualization & Interpretation
3	Joshua Judson J	Member	Exploratory Data Analysis (EDA), Feature Engineering
4	P Venkatesan Gagan	Member	Model Building, Model Evaluation