**Phase-2**

**Decoding emotions through sentiment analysis of social media conversations**

**Github Repository link : https://github.com/ganapathyram373/week-7.git**

**Student Name:** GANAPATHY RAM S

**Register Number:** 513523106009

**Institution:** AMCET **Department:**ECE **Date of Submission:** 05/05/2025

1. **Problem Statement :**

**Decoding emotions through sentiment analysis of social media conversations can help understand public opinions and sentiments towards this initiative. The problem statement is to develop a sentiment analysis model that can accurately classify emotions and opinions expressed on social media platforms. This will enable policymakers to identify areas of improvement, measure the impact of the initiative, and make informed decisions. Sentiment analysis can help uncover the emotional tone behind the text, whether it's positive, negative, or neutral, providing valuable insights into public perception.**

1. **Project Objectives:**

**The objectives of this project are to develop a sentiment analysis model that can accurately classify emotions and opinions expressed on social media platforms towards . The project aims to identify trends and patterns in public sentiment, providing insights into the strengths and weaknesses of the initiative. The objectives also include developing a user-friendly interface to visualize the sentiment analysis results, enabling policymakers to track the impact of the initiative over time.**

1. **Flowchart of the Project Workflow:**

**The workflow includes data collection from social media platforms, data preprocessing to clean and normalize the text data, feature extraction using techniques like bag-of-words or word embeddings, model building using machine learning or deep learning algorithms, model evaluation using metrics like accuracy and F1-score, and visualization of results using plots and charts.**

1. **Data Description:**

**The dataset consists of social media posts and comments related to the initiative, collected from platforms like Twitter and Facebook. The dataset includes text data, timestamps, and user information. The text data may include opinions, emotions, and sentiments expressed towards the initiative.**

1. **Data Preprocessing:**

**Data preprocessing involves cleaning and normalizing the text data. Techniques used include tokenization, stopword removal, stemming, and lemmatization. Tokenization breaks down the text into individual words or tokens. Stopword removal eliminates common words like "the" and "and". Stemming and lemmatization reduce words to their base form.**

1. **Exploratory Data Analysis (EDA):**

**EDA involves analyzing the sentiment distribution, word frequencies, and topic modeling to understand the data. Sentiment distribution analysis helps identify the overall sentiment towards the initiative. Word frequencies analysis identifies the most commonly used words. Topic modeling helps identify underlying themes and topics in the text data.**

1. **Feature Engineering:**

**Feature engineering involves extracting relevant features from the text data. Techniques used include bag-of-words, TF-IDF, and word embeddings. Bag-of-words represents text data as a bag of words. TF-IDF weights words based on their importance. Word embeddings like Word2Vec and GloVe represent words as vectors in a high-dimensional space.**

1. **Model Building:**

**Model building involves developing a machine learning or deep learning model to classify sentiments. Algorithms used include support vector machines, random forests, and recurrent neural networks. The model is trained on the preprocessed data and evaluated using metrics like accuracy and F1-score.**

1. **Visualization of Results&Model Insights:**

**# Import necessary libraries**

**import matplotlib.pyplot as plt**

**import seaborn as sns**

**from wordcloud import WordCloud**

**import plotly.express as px**

**# Sentiment Distribution Plot**

**sns.countplot(x='sentiment', data=df)**

**plt.title('Sentiment Distribution')**

**plt.show()**

**# Word Cloud for Positive Sentiments**

**positive\_text = ' '.join(df[df['sentiment'] == 'positive']['text'])**

**wordcloud = WordCloud(width=800, height=400).generate(positive\_text)**

**plt.figure(figsize=(10, 5))**

**plt.imshow(wordcloud, interpolation='bilinear')**

**plt.axis('off')**

**plt.title('Positive Sentiments')**

**plt.show()**

**# Confusion Matrix**

**from sklearn.metrics import confusion\_matrix**

**cm = confusion\_matrix(y\_test, y\_pred)**

**plt.figure(figsize=(8, 6))**

**sns.heatmap(cm, annot=True, cmap='Blues')**

**plt.xlabel('Predicted Labels')**

**plt.ylabel('True Labels')**

**plt.show()**

**# Model Performance Metrics**

**from sklearn.metrics import accuracy\_score, classification\_report**

**print('Accuracy:', accuracy\_score(y\_test, y\_pred))**

**print('Classification Report:')**

**print(classification\_report(y\_test, y\_pred))**

**# Interactive Visualization using Plotly**

**fig = px.bar(df['sentiment'].value\_counts(), title='Sentiment Distribution')**

**fig.show()**

**This code snippet provides various visualizations to gain insights into the results:**

**1. Sentiment distribution plot: A bar chart showing the count of each sentiment label.**

**2. Word cloud: A visual representation of words in positive sentiments.**

**3. Confusion matrix: A heatmap showing the true and predicted labels.**

**4. Model performance metrics: Accuracy and classification report.**

**5. Interactive visualization: An interactive bar chart showing sentiment distribution using Plotly.**

**These visualizations help understand the sentiment distribution, model performance, and insights from the analysis.**

**10. Tools and Technologies Used:**

**Tools and technologies used include Python, NLTK, spaCy, TensorFlow, and PyTorch. Python is the programming language used for data preprocessing, feature extraction, model building, and visualization. NLTK and spaCy are libraries used for natural language processing. TensorFlow and PyTorch are deep learning frameworks used for building and training models.**

**11. Team Members and Contributions:**

**DILLIRAJA A - DATA ANALSYIST**

**DINESHBABU V – CONTENT CREATER**

**DINESHKUMAR D - TEAM LEADER**

**GANAPATHYRAM S – EXECUTER**

**GOPI E - TEAM CO-ORDINATOR**