TASK 4

Title: Sentiment Analysis on Twitter Data

Objective

To analyze and visualize **sentiment patterns in social media (Twitter) data** to understand **public opinion and attitudes** toward specific topics, brands, or entities.

Dataset Information

- **Dataset Name:** Twitter Entity Sentiment Analysis
- **Source:** Kaggle
- **URL:** https://www.kaggle.com/datasets/jp797498e/twitter-entity-sentiment-analysis

- **Features in Dataset:**
- Tweet content
- Entity (topic/brand)
- Sentiment (Positive, Neutral, Negative)
- Tweet ID
- Tweet time (optional)

Steps Performed

1. Data Preprocessing

- Cleaned tweet text (removed URLs, mentions, hashtags, punctuation)
- Converted text to lowercase
- Tokenized and removed stop words

- Performed lemmatization

2. Sentiment Label Analysis

- Counted number of tweets under **Positive**, **Neutral**, and **Negative**
- Visualized with **bar chart** and **pie chart**

3. Entity-wise Sentiment Breakdown

- Grouped tweets by Entity
- Calculated sentiment distribution per entity (e.g., brand or topic)
- Visualized using **stacked bar charts**

4. Word Cloud & N-grams

- Created **Word Clouds** for each sentiment class to show most common words

- Generated **bi-grams and tri-grams** to understand phrase patterns

Key Insights

- Most tweets were **Neutral**, followed by Positive, then Negative.
- Certain entities received highly **positive sentiment**, indicating good brand reputation.
- Entities involved in controversy showed a **high proportion of negative tweets**.
- Common positive words: “great,” “love,” “awesome”
- Common negative words: “worst,” “hate,” “disappointed”

Tools & Technologies Used

- **Programming Language:** Python
- **Libraries:** Pandas, NumPy, NLTK, Scikit-learn, Matplotlib, Seaborn, WordCloud
- **Platform:** Jupyter Notebook / Google Colab
- **Visualization Tools:** Seaborn, Matplotlib, WordCloud

Conclusion

Sentiment analysis reveals how users express their opinions on brands/topics. This analysis provides valuable insights into public perception, which can be used by marketers, product teams, and public relations departments for decision-making.

Link to Dataset

🔗 Twitter Sentiment Analysis Dataset – Kaggle