### **📄 Final Report: Twitter Sentiment Analysis Project**

#### **Project Objective**

The primary objective of this project was to analyze Twitter data and classify user sentiments (Positive or Negative) using Natural Language Processing (NLP) techniques and machine learning models. The analysis provides insights into general public opinion on social media, which can be valuable for brands, policymakers, and researchers.

#### **Dataset Overview**

* **Size**: 1.6 million tweets.
* **Classes**: Binary (0 = Negative, 1 = Positive after label transformation).

#### **Key Steps**

1. **Data Cleaning and Preprocessing**
   * Removed URLs, mentions, hashtags, emojis, and special characters.
   * Tokenized and lemmatized the tweets.
   * Final preprocessed text stored in final\_text.
2. **Feature Engineering**
   * Applied TF-IDF vectorization with max\_features=5000.
   * Resulting matrix shape: (1,600,000, 5000)
3. **Modeling**
   * Applied and evaluated:  
     + Logistic Regression (tuned with RandomizedSearchCV)
     + Multinomial Naive Bayes
     + LinearSVC (tuned)
     + XGBoost
   * Logistic Regression achieved highest performance with:  
     + Accuracy: **77.44%**
     + F1 Score: **0.78**
4. **Evaluation**
   * All models evaluated using Accuracy, Precision, Recall, F1-score, and Confusion Matrix.
   * Hyperparameter tuning marginally improved model performance.
5. **Visualization**
   * WordClouds for positive and negative tweets.
   * Sentiment distribution plots.
6. **Deployment**
   * Built a simple Streamlit app for real-time sentiment prediction.
   * Users can input a tweet and receive a classification instantly.
   * Saved logistic\_model.pkl and tfidf\_vectorizer.pkl to serve the app.

#### **Conclusion**

This project successfully demonstrated end-to-end sentiment analysis on Twitter data, covering data engineering, model building, tuning, and deployment. The final model performs well and can be extended for multi-class sentiment, multilingual tweets, or even topic modeling in future iterations.