Sentiment Analysis Project Report

# Project Title

Sentiment Analysis on Textual Data Using Natural Language Processing

# Objective

The primary goal of this project is to perform sentiment analysis on unstructured textual data (like tweets or reviews) to determine the emotional tone behind a body of text.

# Tools and Technologies Used

- Programming Language: Python  
- Libraries: NLTK, Scikit-learn, Pandas, NumPy, Matplotlib, Seaborn, WordCloud  
- Model: Logistic Regression / Naive Bayes  
- Dataset: Sample Tweets/Reviews dataset (labeled with positive, neutral, and negative sentiments)

# Methodology

Step 1: Data Collection  
A dataset containing textual reviews or tweets along with their corresponding sentiment labels (positive, neutral, negative).  
  
Step 2: Data Preprocessing  
- Removal of noise: URLs, mentions, hashtags, punctuation.  
- Tokenization: Splitting text into individual words.  
- Stopword Removal: Common words like 'is', 'the', etc., were removed.  
- Lemmatization: Reducing words to their base form.  
  
Step 3: Exploratory Data Analysis (EDA)  
- Visualizations like word clouds and bar graphs were used.  
- Class distribution was checked.  
  
Step 4: Feature Extraction  
- TF-IDF Vectorization used for converting text into numerical format.  
  
Step 5: Model Building  
- Logistic Regression or Naive Bayes trained using TF-IDF features.  
  
Step 6: Model Evaluation  
- Metrics: Accuracy, Precision, Recall, F1 Score.  
- Confusion matrix and ROC curve evaluated.

# Results and Insights

- Accuracy of around 85%.  
- Most common positive words: love, great, amazing.  
- Most common negative words: bad, hate, terrible.  
- Neutral detection had lower accuracy, suggesting room for improvement.

# Conclusion

Demonstrated effective sentiment analysis using NLP. Enables businesses to gauge public opinion for decision-making.

# Future Enhancements

- Use deep learning models like LSTM/BERT  
- Real-time API integration  
- Multilingual analysis

# Deliverables

- Python Notebook with preprocessing, EDA, model implementation, and results.