# **Sentiment Analysis Project Documentation**

## **1. Introduction**

This document provides comprehensive documentation for a sentiment analysis project. The project aims to analyse the sentiment of textual data using machine learning techniques.

## **2. Data Exploration**

* Dataset:
  + The dataset used for sentiment analysis contains eight samples.
  + It consists of label features, including text content and sentiment labels.
* Structure:
  + The dataset is structured in a [describe the structure].

## **3. Data Preprocessing**

* Text Preprocessing:
  + Lowercasing: Convert text to lowercase.
  + Stop Words Removal: Remove common stop words.
  + Special Characters Handling: Remove or replace special characters.
  + Tokenization: Split text into tokens.
  + Lemmatization: Reduce words to their base form.
* Preprocessed Data: [Describe the preprocessed data format].

## **4. Exploratory Data Analysis (EDA)**

* Sentiment Distribution:
  + Analysed the distribution of sentiment labels.
  + Visualised the distribution using histograms or pie charts.

## **5. Text Vectorization**

* Vectorization Technique:
  + Used TF-IDF (Term Frequency-Inverse Document Frequency) for text vectorization.
* Vectorized Data: [Describe the format of the vectorized data].

## **6. Model Selection**

* Machine Learning Models:
  + Explored Naive Bayes, Support Vector Machines, and LSTM networks for text classification.
* Model Evaluation:
  + Evaluated model performance using accuracy, precision, recall, and F1 score.

## **7. Hyperparameter Tuning**

* Hyperparameter Optimization:
  + Conducted fine-tuning using grid search/random search.
* Optimised Parameters: [List the optimised hyperparameters].

## **8. Cross-Validation**

* Cross-Validation Technique:
  + Implemented [type of cross-validation] to assess model generalisation.
* Cross-Validation Results: [Describe the results and their implications].

## **9. Model Interpretability**

* Feature Importance:
  + Analysed feature importance to understand model predictions.
* Interpretability Techniques:
  + Utilised LIME for local interpretable explanations.

## **10. Evaluation Metrics**

* Metrics Used:
  + Utilised confusion matrix, precision-recall curves, and ROC-AUC for evaluation.
* Evaluation Results: [Provide the results and interpretation].

## **11. Deployment (Optional)**

* Model Deployment:
  + Deployed the trained model for real-time sentiment analysis.
* API Integration: [Describe how the model was integrated into an API or web application].
* User Interface: [Describe the user interface for sentiment analysis].

## **12. Conclusion**

This documentation provides a detailed overview of the sentiment analysis project, covering data preprocessing, model development, evaluation, and deployment. The project aims to analyse sentiment in textual data efficiently and effectively.