Project Title: Hotel Reviews Sentiment Analysis

Business Understanding:

Problem Statement:

Build a sentiment analysis model to analyse customer reviews and feedback for a hospitality company. The model should be able to classify reviews as positive, negative, or neutral, and identify key topics or areas of concern.

Objective:

To develop a machine learning model that can accurately predict the sentiment of hotel reviews as Positive, Negative, or Neutral.

Stage 1: Data Collection and Understanding.

Data has been downloaded from Kaggle
website: https://www.kaggle.com/datasets/andrewmvd/trip-advisor-hotel-reviews

Features:

- review: Text of the review
- sentiment: Sentiment label (0: Negative, 1: Neutral, 2: Positive)

Data Understanding

- I described the dataset columns and their meanings.
- Checked for missing values and duplicated records, but there no missing and duplicate values in my data.
- exploratory data analysis (EDA) to gain insights into the dataset's characteristics, including statistical summaries and visualizations.

Stage 2: Data Preparation

- Tokenization
- Converted emojis into text and converted text into lowercase
- Removal of stop words and punctuation
- Lemmatization
- Count Vectorization

Stage 3: Model Training:

1. Split data into training and test sets.

2.Train multiple models: Naive Bayes, Logistic Regression, XGBoost.
3.Evaluate models using accuracy, precision, recall, and F1-score

Stage 4: Model Selection:

- Naive Bayes: Accuracy 83.36%
- Logistic Regression: Accuracy 80.09%
- XGBoost: Accuracy 79.85%

- **Selected Model:** Considering the performance and model size, Logistic Regression might be the best choice for deployment due to its balance between accuracy and handling of class 2. It also has the smallest model size, which could be beneficial for server deployment.

Stage 5: Deployment and Results

Deployment:

- **Platform:** Streamlit
- **Steps:**
    1. Saved trained model ("C:\Users\madas\Downloads\Prudent_Hackathon\logistic_regression1.pkl")
    2. Created a Streamlit app for user interaction and prediction.
    3. Host the application on Streamlit Cloud.

Streamlit App Features:

- **User Input:** Text box for users to input hotel reviews.
- **Prediction Output:** Display predicted sentiment (Positive, Negative, Neutral)

Usage Example:

1. **Review Prediction:**
    o User enters: "The hotel was fantastic with great service!"
    o Predicted Sentiment: Positive

Future Enhancements:

- Improve performance on Neutral class by tuning hyperparameters.
- Explore additional models and feature engineering techniques.
- Enhance the user interface of the Streamlit app for better user experience.

Access the Deployed Application:

You can access the deployed sentiment analysis application using the following link:

[Sentiment Analysis Web App](#)