# Airline Sentiment Analysis Project

This project is focused on building a sentiment analysis system for airline tweets using Natural Language Processing (NLP) techniques and Transformer-based deep learning models. The goal is to classify tweets into three sentiment categories: negative, neutral, and positive.

## 1. Data Loading and Label Encoding

The dataset containing tweets and their sentiment labels was loaded. The sentiment labels (negative, neutral, positive) were mapped to numerical values (0, 1, 2) to make them compatible with machine learning models.

## 2. Text Preprocessing

Tweets were cleaned using several preprocessing steps:  
- Removal of URLs, hashtags, mentions, and special characters.  
- Lowercasing all text.  
- Tokenization of text into words.  
- Removal of stopwords (common uninformative words).  
- Lemmatization to reduce words to their base form.  
This ensures that the text data is clean, consistent, and meaningful for training.

## 3. Dataset Preparation

The cleaned dataset was split into training and testing sets using an 80-20 split. The BERT tokenizer was applied to convert text into input IDs and attention masks. These encoded datasets were then wrapped into Hugging Face Dataset objects for model training.

## 4. Model Building

A pre-trained BERT model ('bert-base-uncased') was loaded using Hugging Face Transformers. It was configured for sequence classification with three output labels corresponding to the sentiment classes.

## 5. Metrics Function

A custom metrics function was implemented to evaluate the model's predictions. It computes accuracy, precision, recall, and F1-score, which provide a detailed picture of model performance.

## 6. Model Training

The model was trained using Hugging Face’s Trainer API with the following settings:  
- Evaluation and saving after each epoch.  
- 20 training epochs.  
- Small batch sizes for both training and evaluation.  
- Logging enabled for monitoring progress.  
A subset of the dataset was used to ensure faster training while testing the workflow.

## 7. Model Evaluation

After training, the model was evaluated on the test dataset. The evaluation results included accuracy, precision, recall, and F1-score, providing insights into how well the model generalizes to unseen data.

## 8. Prompt Demonstration

To showcase the model’s inference capabilities, a few prompt-based tests were conducted. Different prompt variations were tested to classify the sentiment of a sample tweet about flight delays. The model consistently predicted the correct sentiment (negative in this case).

## 9. Ouput

