**NLP Capstone Project: Airline Sentiment Analysis from Twitter Data**

**🚀 Project Overview**

Analyze tweets posted by passengers of six major US airlines to determine sentiment polarity (Positive, Neutral, Negative). Build a complete NLP pipeline that cleans data, extracts features, builds predictive models, and evaluates sentiment classification performance.

**🎯 Problem Statement**

Identify the sentiments expressed by passengers towards several airlines based on tweets. Classify each tweet into one of three sentiment categories:

* Positive
* Neutral
* Negative

**📊 Dataset**

* **Source:** Twitter data collected from passengers of 6 US airlines
* **Size:** 14,640 tweets
* **Labels:** 3 sentiment classes (Positive, Neutral, Negative)

**🛠️ Project Workflow**

**1. Data Loading and Exploration**

* Load the dataset into a pandas DataFrame
* Explore data distribution, class balance, and tweet lengths
* Visualize sample tweets and class frequencies

**2. Text Cleaning**

* Remove noise such as URLs, mentions, hashtags, special characters
* Convert text to lowercase
* Remove stopwords and punctuation
* Optionally apply stemming or lemmatization

**3. Data Preparation**

* Encode sentiment labels into numeric format (e.g., 0, 1, 2)
* Split the dataset into training and testing sets (e.g., 80:20 split)

**4. Feature Engineering**

* Extract features using **TF-IDF** vectorization
* Consider n-grams (unigrams, bigrams) to capture phrase context
* Optionally explore word embeddings for richer features

**5. Model Building**

* Train baseline models such as **Naive Bayes** and **Logistic Regression**
* Tune hyperparameters with cross-validation
* Compare model performances using accuracy, precision, recall, and F1-score

**6. Pipeline Creation**

* Combine preprocessing, feature extraction, and classification into a single pipeline for easy deployment
* Validate pipeline on test data

**7. Evaluation and Insights**

* Present classification metrics
* Visualize confusion matrix
* Analyze misclassified tweets for improvement insights

**🧩 Deliverables**

* Jupyter Notebook(.ipynb file) and HTML file