**Sentiment Analysis**

I analyzed customer feedback on airline services using the **Twitter US Airline Sentiment Dataset (**[**Twitter US Airline Sentiment**](https://www.kaggle.com/datasets/crowdflower/twitter-airline-sentiment)**)**. Tweets were classified into three categories: **Negative**, **Neutral**, and **Positive**. The goal was to accurately capture the sentiment expressed in these tweets which can be used to gain insights into customer satisfaction and highlight areas where airlines can improve. I have performed data preprocessing, tuning a pre-trained BERT model and evaluating its performance to assess its effectiveness in sentiment classification.

**Dataset Overview:**

Number of records: 14640

Negative: 9178

Neutral: 3099

Positive: 2363

A bar graph with different colored squares

Description automatically generated

**EDA:**

1. **Word Cloud for Negative Tweets**

**A close up of words

Description automatically generated**

The words "flight," "united," "USAirways," and “customer service” appear alot meaning that complaints are often associated with flight issues, specific airlines and customer service problems. Terms like "delayed," "hold," and "luggage" shows dissatisfaction related to delays and baggage handling.

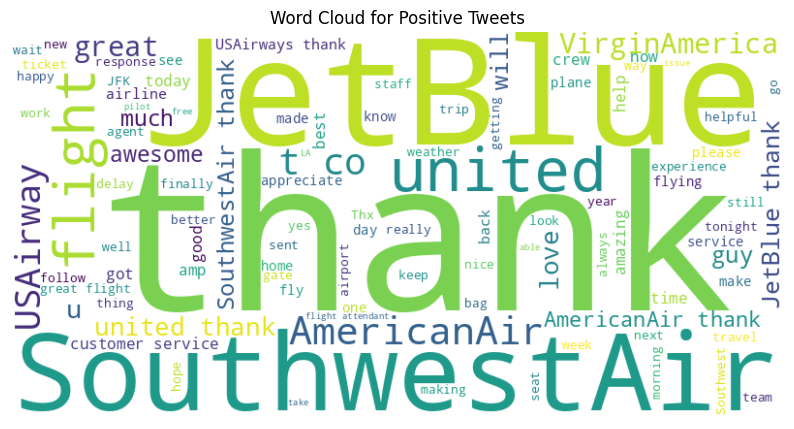
1. **Word Cloud for Neutral Tweets**

**A close up of words

Description automatically generated**

Words like "united," "JetBlue," and "flight" are common in neutral tweets which mention the airline without expressing a strong opinion. Words like "thank," "need," and "time" appear frequently showing that these tweets may involve general inquiries, updates or non-critical feedback.

1. **Word Cloud for Positive Tweets**



Positive tweets highlights words like "thank," "JetBlue," and "SouthwestAir," suggesting that these airlines receive a praise. Common words like "great," "awesome," and "love" indicate customer satisfaction. The usage of words like "helpful" and "amazing" shows positive experiences with airline staff and services.

1. Sentiment Distribution by Airline

A graph of different colored bars

Description automatically generated

Distribution of sentiments (negative, neutral, and positive) for airlines.

* **United Airlines,** **US Airways** and **American** have the highest count of negative sentiments showing higher dissatisfaction.
* **JetBlue** and **Southwest Airlines** have a higher proportion of positive sentiments.
* **Virgin America** has balanced sentiment distribution skewed towards positive feedback.

**Data Preprocessing:**

* Removed URLs, special characters, and converted text to lowercase.
* Mapped sentiment labels to numerical values:
  + Negative → 0
  + Neutral → 1
  + Positive → 2

**BERT Model:**

Data was Split in Train and Valid in ratio of 70% and 30%.

Training size: 10248; Validation size: 4392

A screenshot of a computer

Description automatically generated

The model was fine-tuned for 3 epochs and final **validation accuracy of 85%**. Because of increase in the **validation loss** during the last epoch may be because of overfitting. To improve performance, **early stopping** and hyperparameter tuning could have been explored (which I Didn’t for this time).

I tested it on a few sample tweets to assess its sentiment classification capabilities.

A screenshot of a computer

Description automatically generated

These results show that the model performs well in identifying positive and negative sentiments. But prediction of "It was an okay flight, nothing special" as **Positive** instead of **Neutral** suggests that the model could benefit from further fine-tuning.

**Model Evaluation:**

**A blue squares with white text

Description automatically generated**

**Negative:** The model correctly classified 2,550 negative tweets, with only 180 misclassified as neutral and 84 as positive.

**Neutral:** Out of 884 neutral tweets, the model correctly predicted 608, while 185 were misclassified as negative and 91 as positive.

**Positive:** The model performed reasonably well on positive tweets, with 571 correct predictions. 45 positive tweets were misclassified as negative and 78 as neutral.

**A screenshot of a graph

Description automatically generated**

**Accuracy:** 85% means model classifies tweets correctly most of the time.

**Precision, Recall, F1-Score:**

Negative: highest performance with a precision and recall of 0.92 and 0.91.

Neutral: model struggles with neutral tweets [lower precision (0.70) and recall (0.69)]

Positive: model performs well with a precision of 0.77 and recall of 0.82.

**A red green and blue rectangular bars

Description automatically generated**

**Conclusion:** The bar chart shows that the model predominantly predicts tweets as Negative (9,083), with fewer classifications for Neutral (3,079) and Positive (2,478) sentiments. This distribution shows that the model might be biased towards negative classifications due to the dataset's class imbalance. Further fine-tuning and addressing the imbalance could help improve the model's ability to recognize neutral and positive sentiments more accurately.