Airline Sentiment Analysis API

Step 1 : Text - Preprocessing -

It was done by firstly **removing all the numerical data , punctuations, stop words and emojis** from the text , then the text was converted into lower case. **Regex Tokenization and Lemmatization** was done next.

Step 2 : Feature Extraction -

The text data was converted into integer matrix using **TF IDF** and the imbalance data was handled using **SMOTE** .

Step 3 : Classifier Model -

The model was trained using **Random Forest Classifier** . And the model performance is as follows

Confusion matrix :

[[2591 160]
[147 2609]]

Precision : 0.9422174070061394

Recall : 0.9466618287373004

F1 Score : 0.9444343891402714

ROC - AUC : 0.99