**Dataset:**

Airline sentiment analysis dataset

**Model:**

XGBoost classifier

**Training Steps:**

* Taken same number of records for each of the classes.
* Preprocessed the data by removing unwanted punctuations and stop words.
* Converted the sentences to vectors using CountVectorizer of Scikit Learn.
* Then Split the data into training and test set, where test set is 20% of the whole data set.
* Then trained the model using XGBoost classifier.
* Saved the corpus and the classifier using joblib in order use in our application.
* Used grid search to find the best hyperparameters, but there was no obvious improvement seen.

**API Implementation Steps:**

* Used pyramid framework to implement the API.
* Created different module to predict the sentiment.
* Created HTML pages to test.
* Provided CORS permission to allow cross origin resource share.
* Tested using postman.

**Future Scope:**

* We can use pretrained google news model to improve the accuracy.
* While stratifying the data into same number of records, we prioritize the sentences and choose instead of random selection.