# Sentimental Analysis API

**Dhruv Viradiya**

**IIT Jodhpur**

## Overview

This project is about creating an API for a machine learning model. This includes applying NLP to the airline sentimental dataset, implementing various models like Random Forest, Decision Tree, XGBoost, LightGBM, and Logistic Regression, and comparing results of all the models, generating API for the best model using fast API.

## Approach

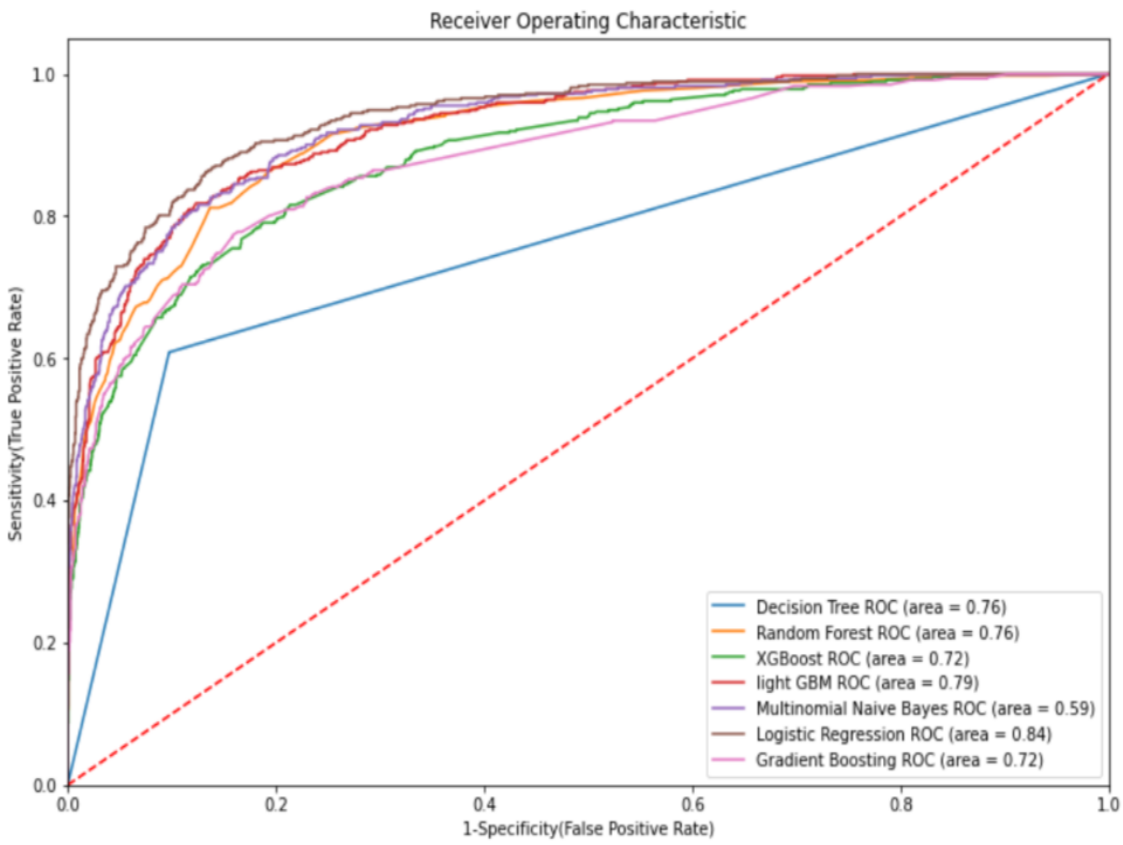
- Cleaning the dataset:
  - Removing special characters, emojis, and extra words.
  - Removing stop words
  - lemmetizing the words
- Training:
  - Using vectorizor for making training dataset
  - Applying PCA for reducing dimensions
  - Train different models and compare their results.
- Generating API
  - Using fastapi for creating API for the generated model.

## Models Used:

1) Decision Tree
2) Random Forest
3) XGBoost
4) Multinomial NB
5) Light GBM
6) Gradient Boosting
7) Logistic Regression

## Accuracy score for different models

| Model | Accuracy |
|---|---|
| Decision Tree | 0.8423 |
| Random Forest | 0.8847 |
| XGBoost | 0.87 |
| Multinomial NB | 0.8310 |
| Light GBM | 0.8930 |
| Gradient Boosting | 0.8739 |
| Logistic Regression | 0.9073 |




## Best Model

**Logistic Regression**