Twitter US Airline Sentiment Analysis

05.06.2022

—

Sidhant Mohanty
UCSD ML BootCamp
April 11 2022 Batch

Overview

The airline industry is a very competitive market which has grown rapidly in the past 2 decades. Customer feedback is very crucial to Airline companies as it helps in improving the quality of services and facilities provided to customers. Airline companies resort to traditional customer feedback forms for sentiment analysis which in turn are very tedious and time consuming. The digital connectivity bestows immense power to the customers in terms of vocalizing their thoughts, opinions, and reviews on a brand. The customer views expressed on Twitter, Facebook, and other online forums are forming the base of customer strategy for brands worldwide as they are genuine and much more reliable.

In this capstone project we will be focusing on the sentiment analysis of 6 major US Airlines from Twitter data. Job is to find the problems of each major US airline. For this project we will be using Twitter data (reviews submitted by individuals who traveled through various Airlines) which was scraped from February 2015 and contributors (mostly workers from Crowdflower) were asked to first classify positive, negative and neutral tweets followed by categorizing negative reasons (such as “late flight” or “rude service”)

Goals

Create a Model which on providing the tweets(reviews) to the training should provide the outcome that whether a particular tweet done by an individual is a positive response or a negative one or neutral. Ultimately with this sentiment analysis or opinion mining Airlines can gain better user insights and strategize their brands to provide better service to the customers.

- **Positive Response** : Reviews that contains good experience of traveler with airlines
- **Negative Response** : Reviews that contains difficulty faced by the traveler
- **Neutral Response** : Reviews which are not specific to be considered in positive or negative

Dataset

We will be using Twitter scrapped data from 2015 which can be found on Kaggle

https://www.kaggle.com/datasets/crowdflower/twitter-airline-sentiment?select=Tweets.csv

It is a structured dataset with about 6 columns and 15000 rows

Final Deliverable

As a final deliverable it will be an application deployed as a web service with an API or a simple Web UI where the user can go and enter some text and the application/Model will predict if it's a positive, negative or neutral response.

Approach

We will be using traditional ML approach like Logistic Regression, Random Forest, SVM, Naive Bayes, Decision Tree etc

Specifications

Laptop Specifications (where most of the code and model will be run) :

Processor : 2.6 GHz 6-Core Intel Core i7

Memory : 16 GB 2667 MHz DDR4

Graphics : Intel UHD Graphics 630 1536 MB

On top of this if required we will be using paperspace for additional memory and computing.

Evaluation

We will be evaluating this Model on some the Metrics like :

- Accuracy
- Precision
- Recall
- F1 Score
- Confusion Matrix,ROC and AUC (Area under the Curve,etc