CSCI-B 565

Data Mining

Fall 22

Airline Recommender System

- Himanshu Joshi (hsjoshi)
- Atharv Jangam (ajangam)
- Trishna Patil (tripatil)

Problem Statement

Recognizing user-specific historical behaviors through the analysis of various feedbacks made by them is a crucial component of recommendation systems.

The problem statements that we are trying to solve are-

1) To create a model for airlines that analyzes the sentiment of customer evaluations in order to provide clear feedback about the airlines.

2) Helping other customers choose personally curated airline recommendations based on other customer reviews and feature ratings.

We are targeting to create a highly trained Airline Recommender Model, using various machine learning algorithms, which accurately predicts whether a customer should select any particular airline based on previous customer experiences.

Motivation

The airline business experienced continuous, rapid expansion from 2009 to the COVID-19 era, with the annual number of passengers nearly doubling at that time. Because of this circumstance, airline businesses now compete more fiercely, which has significantly reduced their profitability. With remarkable success, recommender systems have already been implemented in a number of industries, including retail and entertainment. Although, their use in the airline sector is still in its early development stages. As far as we are aware, there are no major machine learning models in place to recommend supplementary services to airline firms.

Here, we use the Kaggle ‘Airline Dataset Mining’ dataset. High accuracy is attained, demonstrating the project's dependability for making future predictions.

We maintain that recommender systems, hold the key to achieving customer centricity because of their capacity to recognize and cater to the demands of the customer at all points of contact throughout the traveler journey.

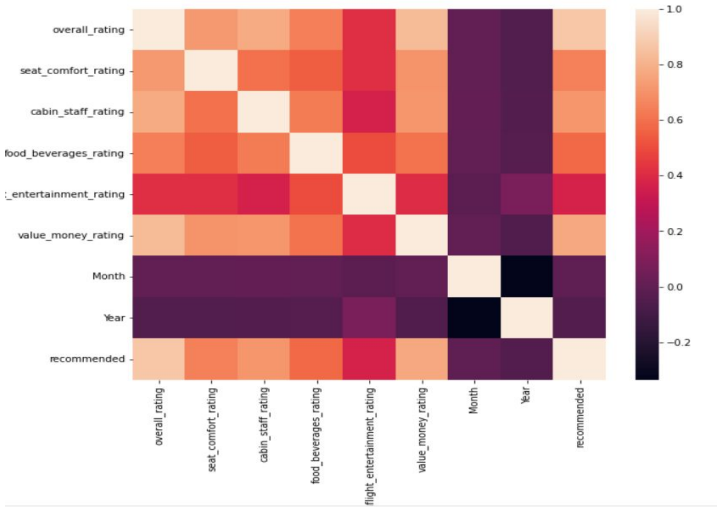
Approaches

- Cleaning the data set - dropped less significant columns, removing outliers
- Performing EDA - analyzing data set to identify the main features
- Training model using Supervised Learning algorithm - Linear Regression, Logistic Regression, KNN, Decision Tree Classifiers, Time Series Classifier
- Association Rules Mining - Apriori Algorithm
- Making a WordCloud for sentiment analysis
- Sentiment Analysis for airlines using reviews provided
- Applying Random Forest and Adaboost to find the accuracy of sentiments.

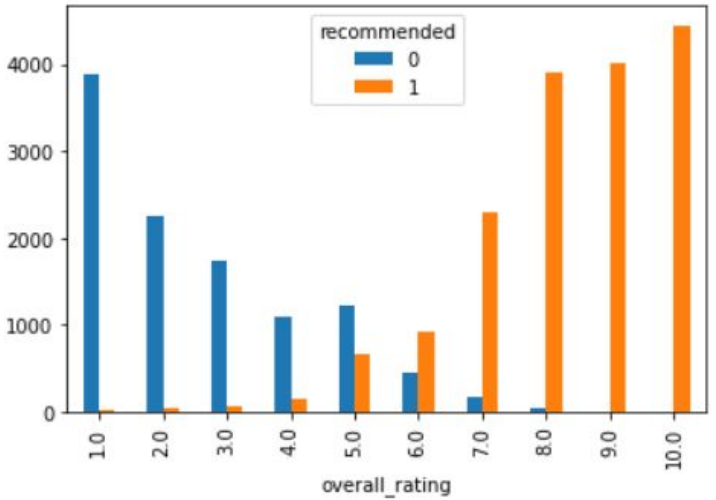
Data Set

- It is available freely on Kaggle. (Click here to access the dataset)
- It has 27,284 rows and 14 columns from 2011 to 2015.
- It has a data of 292 airlines parametrized by seat comfort rating, cabin staff rating, food beverages rating, and Overall Rating .
- A single target column was identified (‘recommended’).
- The dataset is original scraped from Skytrax website (https://www.airlinequality.com/airline-reviews/)

Exploratory Data Analysis (EDA)

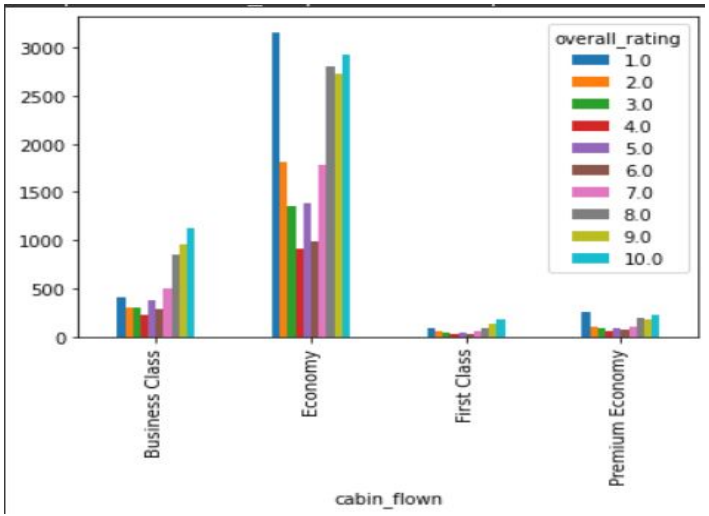


Heat map



Recommended
vs ratings

Exploratory Data Analysis (EDA) - continued



Overall ratings by cabins

Current Results (note* the project is in progress, we are still trying to improve the model accuracy)

	precision	recall	f1-score	support
0	0.94	0.92	0.93	3530
1	0.95	0.96	0.96	5474
accuracy			0.95	9004
macro avg	0.94	0.94	0.94	9004
weighted avg	0.95	0.95	0.95	9004

Linear Regression (Accuracy = 95%)

	precision	recall	f1-score	support
0	0.93	0.93	0.93	3599
1	0.95	0.95	0.95	5405
accuracy			0.94	9004
macro avg	0.94	0.94	0.94	9004
weighted avg	0.94	0.94	0.94	9004

Logistic Regression (Accuracy = 94%)

	precision	recall	f1-score	support
0	0.93	0.92	0.92	3602
1	0.95	0.95	0.95	5402
accuracy			0.94	9004
macro avg	0.94	0.94	0.94	9004
weighted avg	0.94	0.94	0.94	9004

Decision Tree Classifier (Accuracy = 94%)

	precision	recall	f1-score	support
0	0.86	0.76	0.81	2780
1	0.85	0.92	0.88	4041
accuracy			0.85	6821
macro avg	0.85	0.84	0.84	6821
weighted avg	0.85	0.85	0.85	6821

Random Forest on Sentiments (Accuracy = 85%)

Expected Results

To create a highly trained Airline Recommendation model to predict whether any customer would recommend a particular airline based on different airline features ratings.

This would act helpful for other customers while booking an airline of their choice based on other customer ratings, as well as help airlines improve based on the customer feedbacks.

Targeting to create a high accuracy model, demonstrating the project's dependability for making future predictions.