Capstone Project Submission

Team Member's Name, Email, and Contribution:

Name: Shahfaissal I Dharwad
email id: shahfaissal21@gmail.com
Contribution:
   • Feature Engineering
   • NLP
   • Model Building
 	
Name: Ajinkya Satish Jumde
email id: jumde.ajinkya@gmail.com
Contribution:
   • Data Wrangling
   • Handling Missing and duplicate values
   • Exploratory Data Analysis
   Model Building

GitHub Repo link.
GitHub Link:- https://github.com/shahfaissal/capstone_project_3
Please write a summary of your Capstone project and its components. Describe the problem statement, your approaches, and your conclusions. (200-400 words)
The data shown includes airline reviews from 2006 to 2019 for popular airlines around the world with multiple choice and free text questions. The data is discarded Spring 2019. The main goal is to predict whether travelers will refer to airline to your friends. We tested the data and conducted survey data analysis to create machine learning models to predict the dependent factor which is the recommendation of airlines by a passenger to a friend. Our main goal is to use previous passenger data and predict recommendations based on that passenger. Data has various factors influencing the recommended part of the dataset, such as an airline and its aircraft carrier, an overall rating of 10, customer name, review date, review content, cabin, route and more other relative factors. We get mixed data types which also included some of the outliers and almost half the zero values. Most rating categories are numerical while the review that has a big influence on the recommendation is categorical data. So we did an exploratory data analysis to find out the main influencing factors anticipated passenger recommendations. We performed the function engineering followed by EDA to test the important features and release less important. After feature engineering, we made natural language processing (NLP) review review and content then one-time coding for categorical data. After preparing the final data set that is ready to be fed to ML, we split it into training and testing data and used a logistic regression model, Decision tree, random forest, cv lattice search random forest, support vector Machines (SVM), K-Nearest Neighbor (KNN) model, XG Boost and compared a different model with its evaluation matrix.