**Capstone Project Submission**

| **Team Member’s Name, Email and Contribution:** |
| --- |
| 1. Prashant Gour ([gourprashant787@gmail.com](mailto:gourprashant787@gmail.com))  * Importing Libraries * Loading Dataset * Data Analysis using EDA * Model Development * Logistic Regression Model * Decision Tree Model * Random Forest Model * Support Vector Machine Model * K-Nearest-Neighbour Model * Naïve Bayes * Model Selection * Conclusions  1. Jyotsana Gour ([gourjyotsana@gmail.com](mailto:gourjyotsana@gmail.com))  * Importing Libraries * Loading datasets * Feature Engineering * Model Development * Logistic Regression Model * Decision Tree Model * Random Forest Model * Support Vector Machine Model * K-Nearest-Neighbour Model * Naïve Bayes * Model Selection * Conclusions  1. Anuj Gupta ( [anujgupta19899@gmail.com)](about:blank)  * Importing Libraries * Loading datasets * Feature Engineering * Model Development * Logistic Regression Model * Decision Tree Model * Random Forest Model * Support Vector Machine Model * K-Nearest-Neighbour Model * Naïve Bayes * Model Selection * Conclusions  1. Ashutosh Sharma ([)](about:blank)  * Importing Libraries * Loading datasets * Feature Engineering * Model Development * Logistic Regression Model * Decision Tree Model * Random Forest Model * Support Vector Machine Model * K-Nearest-Neighbour Model * Naïve Bayes * Model Selection * Conclusions |
| **Please paste the GitHub Repo link.** |
| Github Link:- <https://github.com/GourPrahasnt/Airline_passenger_referal_prediction> |
| **Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)**   | Data includes airline reviews from 2006 to 2019 for popular airlines around the world with multiple choice and free text questions. Data is scraped in Spring 2019. The main objectives are to predict whether passengers will refer the airline to their friends.  We have Started with data loading and we have done EDA, feature engineering, data cleaning, target encoding feature selection and then model building.  So, we have used this models:   * Logistic Regression Model * Decision Tree Model * Random Forest Model * Support Vector Machine Model * K-Nearest-Neighbour Model * Naïve Bayes * We performed hyperparameter tuning for decision tree models, random forest models, K-Nearest Neighbors, Support Vector Machines, and Naive Bayes using the GridSearch CV method. This is done to improve accuracy and avoid overfitting criteria. After that, we completed the Gradient Boosting model by fine-tuning the hyperparameters. * Based on an understanding of the business and problem use cases. The classification metrics for Recall are given first priority, Accuray second priority, and ROC AUC third priority. * We created classifier models using 6 different classifier types, all of which provided over 90% accuracy. We can conclude that LogisticRegression gives the best model. * Comparing the model evaluation metrics, we can see that the SVM is the most accurate model with a very small margin, and performs the best among the experimental models on the given dataset. * The most important features were overall rating and value for money, which helped the model predict whether passengers would recommend a particular airline to their friends. * The developed classifier models can be used to predict passenger recommendations as they enable airlines to identify influential passengers who can help generate more revenue. * Therefore, in order to grow or grow their business, our customers must provide outstanding cabin service, ground handling, entertaining food, beverages and comfortable seating.   **Thank You** | | --- | |