

CS 513 B – KDD PROJECT PROPOSAL: Airline Passenger Satisfaction Prediction

Project Group No:17

Problem Statement:

Airline Industry has a market cap of 1 trillion US dollars, With more than 5000 airline companies to choose from . Airline passenger satisfaction is a crucial factor in determining the success of an airline company. In today's competitive market, understanding customer satisfaction levels can give airlines a competitive edge. However, analyzing the vast amount of data available to identify factors that contribute to passenger satisfaction can be a challenging task. As such, there is a need to develop a machine learning model that can accurately predict passenger satisfaction levels based on a range of factors such as flight time, route, cabin class, in-flight amenities, and more. The aim of this project is to build such a model to help airline companies improve their services and provide better experiences to their passengers. The target variable is satisfaction 1 denotes passenger is satisfied and 0 which is not satisfied.

Dataset:

The dataset comprises of 25 features in form of columns, out of which we may opt to use the essential features only, during implementation. We can use PCA and correlation for feature reduction.

Source of Dataset:

<https://www.kaggle.com/datasets/teejmahal20/airline-passenger-satisfaction>

Implementation Strategy and algorithms used:

We have decided to implement and compare 9 different models among four different group members. We have chosen a few models from our course and few from outside the course. The following are the models selected by us:

1. Decision Trees
2. AdaBoost Classifier
3. Multilayer Perceptron
4. Bagging along with Random Forest
5. K-nearest neighbor with Cross Validation
6. Logistic Regression with Cross Validation
7. Random Forest with Cross Validation
8. Support Vector Machine with Cross Validation
9. Artificial Neural Network

Model metrics and Evaluation:

Evaluation matrices of different machine learning models are used in the project to determine the usability of the algorithm for the classification of the medical dataset based on the input features. The defined metrics to evaluate the algorithmic model are reported as follows:

- Cross-validation score
- Accuracy Score
- R2/F1 score
- Precision and recall
- Confusion matrix

Team Members: Group 17

1. Aditya Lalit Adhduk
2. Dhwanil Rameshbhai Gangani
3. Kishan Hasmukhbhai Kathiriya
4. Harshil Mehta