Developing a Flight Delay Prediction Model using Machine Learning

Project Discription:

The main objective of the model is to predict flight delays accurately in order to optimize flight operations and minimize delays.

Using a machine learning model, we can predict flight arrival delays. The input to our algorithm is rows of feature vector like departure date, departure delay, distance between the two airports, scheduled arrival time etc. We then usedecision tree classifier to predict if the flight arrival will be delayed or not. A flight is considered to be delayed when difference between scheduled and actualarrival times is greater than 15 minutes.

Base papers related to our project:

Title :	Application of Machine Learning Algorithms to Predict Flight Arrival Delays	Flight delay predictions and the study of its casual factors using machine learning algorithm	Airline delay prediction by machine learning algorithms	An Approach of Applyin Learning Model in Flight Prediction- A Comparati
Methodology used :	Decision Tree, Logistic Regression and Neural Networks	K-nearest neighbour, Random Forest, Naïve Bayes, Decision tree, Artificial Neural Networks (ANNs)	Decision Tree, Random Forest, Bayesian classifycation, K-means clustering, Hybrid approach	Naive Bayes, CART, Classification and Regression Trees, Decision trees, XGB Classifier, Random Forest
Advantages :	Decision tree classifier performs better at predicting on-time flights whereas neural network performs better at predicting delayed flights with a overall accuracy of 91%	The Artificial Neural Network provides better accuracy of 83% in predicting the flight delay due to climate factor	The accuracy levels of the hybrid approach were 71.39% and 76.44% in predicting delay occurrence and 70.16% and 75.93% in predicting delay is provided by hybrid approach	The overall accuracy is 98% out of which the majority accuracy is of CART with the accuracy of 99.15%

Disadvantages :	Deeper neural network and more training data to increase the accuracy	A dataset with longer duration might aid in further increase in accuracy as some data is missing due to cancellation of flight causing reduction in accuracy.	Combing the hybrid method with robust flight may increase the accuracy	Limited dataset due to hardware restriction resulting in decreased accuracy

Problem statement :

To develop the flight delay prediction model with the use of deeper neural network, more training data, dataset with more duration to increase the accuracy and to compare decision tree classifier with logistic regression and neural network for identifying the difference in actual and scheduled arrival inorder tp predict the flight delay.

Team Number – 06

Team Members:

Arsah A

Karolin Kiruba R

Kishan I

Neekita C

Sri jane A