Developing a Flight Delay Prediction Model using Machine Learning

SUBMITTED BY

YASHWANTH R	113319205047
RAGHUL R	113319205038
NARESH D	113319205026
DIVYASHREE A	113319205010

BACHELOR OF TECHNOLOGY IN INFORMATION TECHNOLOGY

LITERATURE SURVEY:

TITLE	AUTHOR	ALGORITHM	ADVANTAGES	DISADVANTAGES
Flight Delay Prediction Using a Hybrid Deep Learning Method	Warittorn Cheevachaipimol Bhudharhita Teinwan and Parames Chutima	Deep learning	Accurately predicting flight delay has been a challenging issue for researchers and practitioners for decades. However, recent studies have demonstrated the applicability of using state-of-the-art computer-based approaches such as big data, machine learning, and deep learning to achieve better flight delay prediction results relative to conventional statistical approaches.	.

| Flight Delay Prediction Based on Gradient Boosting Ensemble Models | Rahemeen Khan and Tooba Zahid | Machine Learning | applied stacking based algorithms including Random forest, Naïve bayes, KNN, Logistic regression, Decision tree by incorporating the SMOTE to process imbalance dataset with feature selection technique for flight delay prediction. For predicting the flight delay, various techniques such a machine learning models, deep learning models and big data analytics techniques utilized in the past. | In machine learning and deep learning models to compare the performance with the traditional statistical models on air traffic dataset for flight delay prediction.established as system to avoid delays in flight with the help of data mining and machine learning techniques covering top 5 busiest airport of US. Gradient boosting classifier attained the 85.73% prediction results of flight delay. |

| BIG DATA IN FLIGHT DELAY PREDICTION FOR MAINTAINING DATABASE OF THE FLIGHTS DEPARTING THE PORT | Yushan Jiang Yongxin Liu | Big Data | The data source and pre-processing steps including the data merging and cleansing will be introduced. As each dataset is untidy with messy redundant records and missing values.The data used in this paper consists of two parts, Airline On-Time Performance Data (AOTP) and Quality Controlled Local Climatological Data (QCLCD) in the year 2016.It contains basic hourly airport weather data including temperature extremes, visibility, air pressure, humidity and wind | Data is thoroughly examined for integrity criteria as well. Since expected model is to work with all the forms like offline, near line and online data, the irrelevant and unnecessary parameters that could overburden the dataset is reduced.Dropped the null values and assigned zero to Not a Number (NaN) values as one of the data cleansing activities. The data types of time factors such as scheduled time, airtime etc., are found to be in float point and proper conversion of input time to standard date time format. Finally, the data is analyzed for distribution , converting and preprocessing. Then different datasets such the airline, flight, airports and weather datasets are integrated and normalized to identify the correlating factors that affect the flight cancellations |

| PREDICTION OF DELAY OF FLIGHT USING DATA MINING | L.BeLCASTRO Fabrizio marozo Domencia Talia | Data Mining | . Two open datasets of airline flights and weather observations have been collected and exploratory data analysis has been performed to discover initial insights, evaluate the quality of data, and identify potentially interesting subsets The data preparation and mining tasks have been implemented as MapReduce programs. Other than providing the necessary computing resources for our experiments, the Cloud makes the proposed process more general. If the amount of data increases (e.g., by extending the analysis to many years of flight and weather data), the Cloud can provide the required resources with a high level of elasticity, reliability, and scalability. | The air time and flight distance would also have a greater impact on on-time performance of specific flight; Different carriers and specific aircraft would also have a slight influence of on time performance. Accuracy of this model is low because detailed weather and aircraft data could not be collected.

A research analyzes flight information operated by American Airlines, predicting possible arrival delay of the flight using Data Mining . Due to the imbalanced data, Over-Sampling technique, Randomized SMOTE was applied for Data Balancing. The Gradient Boosting Classifier Model was deployed by training and then Grid Search on Gradient Boosting Classifier Model on flight data, caused hyper-parameter tuned and achieving a maximum accuracy of 85.73%. Result showed that deleting some features affected the value of accuracy and reduced it.. |