DEVELOPING A FLIGHT DELAY PREDICTION MODEL USING MACHINE LEARNING

SUBMITTED BY

PRIYANKA S	113319205036
PRIYADHARSHINI U	113319205035
HARITHA K	113319205015
KOHILA K	113319205022

LITERATURE SURVEY				
TITLE	AUTHUR	ALGORITHM	ADVANTAGES	DISADVANTAGES
FLIGHT DELAY PREDICTION BASED ON DEEP LEARNING AND LEVENBERG-MARQUART ALGORITHM	Maryam Farshchian Yazdi, Seyed Reza Kamel , Seyyed Javad Mahdavi Chabok and Maryam Kheirabadi	DEEP LEARNING	Accuracy of the proposed prediction model analysed and compared to previous prediction method. results of three models on both imbalanced and balanced datasets shows that precision, accuracy, sensitivity, recall and F-measure of SDA-LM model with imbalanced and balanced dataset is improvement than SAE-LM and SDA models. The results also show that accuracy of the proposed model in forecasting flight delay on imbalanced and balanced dataset respectively has greater than previous model called RNN	However, this model has drawbacks of over fitting, that researchers have solved that through typical data dropout technique for each step of repeated training procedure. Moreover, application of this method decreases the computation time and memory space during the training. The next drawback is the noise of input data. However, the researcher neglects the noise during prediction.
FLIGHT DELAY PREDICTION USING MACHINE LEARNING ALGORITHM XGBOOST	Nathalie Kuhn and Navaneeth Jamadagni	MACHINE LEARNING	Different ML algorithms to predict if a flight will be delayed or not before it is even announced on the departure boards. So it will not be aiming to get the highest accuracy possible, because it would be quite easy by adding a series of features that will biased the model in terms of predictive power. So this information was looked at as part of the Exploratory Data Analysis (EDA).	The authors compare various machine learning algorithms to predict flight delays, but Failed to consider simple neural networks and decision tree classifiers. So simple machine learning algorithms like decision tree and simple neural networks to be implemented to predict flight delays, and investigate if we can predict flight delay with fewer feature-set accurately.
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	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 (Na N) 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

			extremes, visibility, air pressure, humidity and wind	input time to standard date time format. Finally, the data is analysed for distribution, converting and pre-processing. 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 Map Reduce 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 analyses 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..