Flight Delay Analysis Using Machine Learning

SK. Ahmed Mohiddin[1] , G. Komali Priya[2] , D. Karthik Sai[3] , P. Lalitha Sree[4] , P. Sriram[5] ,

[1]Professor, [2,3,4,5] Students
Department of Computer Science And Engineering
Sri Vasavi Institute Of Engineering and Technology, Pedana, A.P, India

ABSTRACT:

The prediction of flight delays is heavily investigated in the last few decades. Flight delays hurt airlines, airports, and passengers. The development of accurate prediction models for flight delays became cumbersome due to the complexity of air transportation system, the number of methods for prediction, and the deluge of flight data. The flight delay analysis is based on scheduled arrival, departure and actual time. In this context, this paper presents a thorough literature review of approaches used to build flight delay prediction models. We propose a taxonomy and summarize the initiatives used to address the flight delay prediction problem, according to scope, data, and computational methods, giving particular attention to an increased usage of machine learning methods. Besides, then we will check the accuracy metrics for flight delay prediction.

INTRODUCTION:

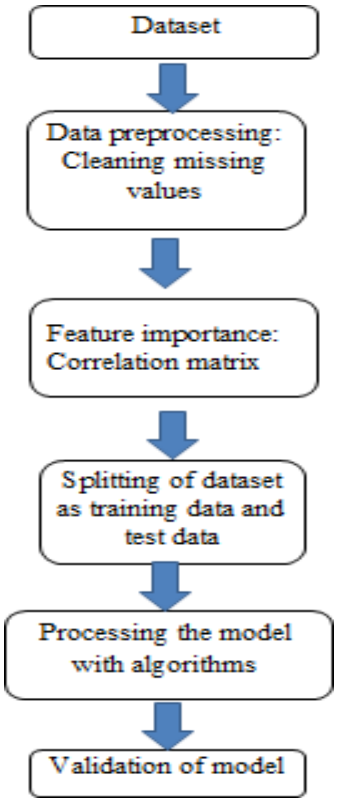
Air transportation plays a vital role in the transportation infrastructure as well as contributes significantly to the economy. Airports are known for their capability to increase business activities near them and hence result in economic development. The International Air Transport Association showed that the demand for air travel increased by 6.3 percent in the year 2016 as compared to the year 2015.There are several causes of an aircraft being delayed such as weather changes, problems in maintenance, previous delays being propagated down the line, traffic congestion and many more. A delay of an aircraft can be problematic for the travelling passengers as it prevents them from fulfilling their commitments and attending preplanned events. This can result in the passenger losing a lot of money as well as make him or her frustrated and angry.

PROPOSED SYSTEM:

The Flight delays will be calculated based on the scheduled time i.e arrival time of flight, departure time of flight and actual time of the flight. Based on the scheduled time will calculate the difference in time and make it as a target variable. We considered the datatset for flight delay analysis, where we can start the analytics preprossessing of the dataset in order to make it feasible to machine learning format. The flight delay analysis is a regression problem, then will use regression based models like linear regression and logistic regression etc. If the data has collinearity or interdependencies we will go for lasso or ridge regression then we will check the accuracy metrics, like Rmse for validating our model.

DOI : 10.36893.DRSR.2021.V09I01N01.841-844

ARCHITECTURE:



ALGORITHMS:

Elastic net Regression:

Linear regression refers to a model that assumes a linear relationship between input variables and the target variable.With a single input variable, this relationship is a line, and with higher dimensions, this relationship can be thought of as a hyperplane that connects the input variables to the target variable. The coefficients of the model are found via an optimization process that seeks to minimize the sum squared error between the predictions (yhat) and the expected target values (y).

LASSO Regression:

DOI : 10.36893.DRSR.2021.V09I01N01.841-844

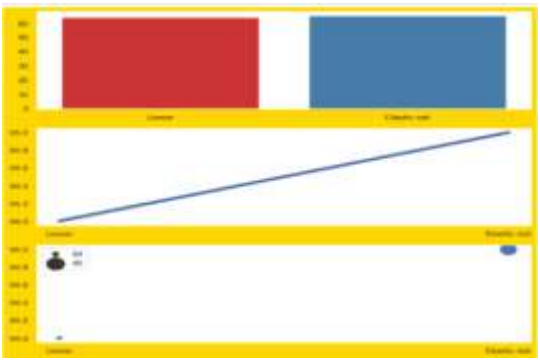
Lasso regression is a type of linear Regression that uses shrinkage. Shrinkage is where data values are shrunk towards a central point, like the mean. The lasso procedure encourages simple, sparse models (i.e. models with fewer parameters). This particular type of regression is well-suited for models showing high levels of multicollinearity or when you want to automate certain parts of model selection, like variable selection/parameter elimination.

The acronym "LASSO" stands for **L**east **A**bsolute **S**hrinkage and **S**election **O**perator. A tuning parameter, λ controls the strength of the L1 penalty. λ is basically the of amount shrinkage.

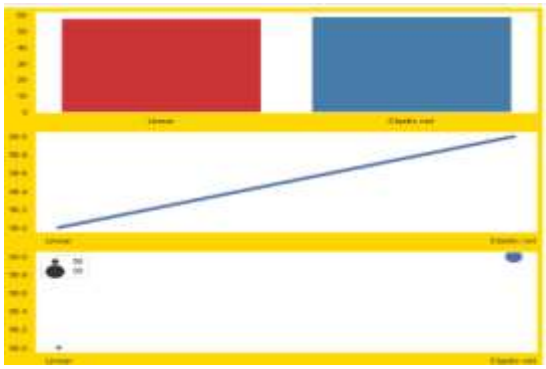
Results and Analysis:

Results for Sliptting ratio 90:10

Results for Sliptting ratio 80:20



Results for Sliptting ratio 70:30



Results for Splitting ratio 60:40

DOI : 10.36893.DRSR.2021.V09I01N01.841-844

CONCLUSION:

Overall, our models are only of limited utility since none were capable of correctly predicting flight delays with both precision and recall greater than 50%. This seemingly low performance is likely due to the many causes of flight delays being outside the scope of our data. It is unclear if it is even possible to predict whether or not a flight will be delayed so far in advance, as we have set up the problem, because so many of the causes of delays (e.g. mechanical issues and weather) cannot be known in advance. Despite this, we were successful in creating models that outperform baseline models, and perform at least about as well as prior work, even when we often use less information, and generalize to more airports. Although imperfect, this model still makes potentially useful predictions about which flights are more or less likely to be delayed.

REFERENCES :

[1] Yufeng Tu, Michael Ball, Wolfgang Jank. Estimating Flight Departure Delay Distributions-A Statistical Approach with Long-term Trend and Short-term Pattern. 2006

[2] Pernkopf, F. and D. Bouchaffra. A genetic-based em algorithm for learning Gaussian mixture models. IEEE Transactions on Pattern Analysis and Machine Intelligence 27, 1344–1348. (2005)

[3] Mueller, Eric R., and Gano B. Chatterji. "Analysis of aircraft arrival and departure delay characteristics." AIAA aircraft technology, integration and operations (ATIO) conference. 2002.

[4] Beatty, Roger, et al. "Preliminary evaluation of flight delay propagation through an airline schedule." Air Traffic Control Quarterly 7.4 (1999): 259-270.

[5] Sternberg A, Soares J, Carvalho D, Ogasawara E. A Review on Flight Delay Prediction. arXiv preprint arXiv:1703.06118. 2017 Mar 1.

DOI : 10.36893.DRSR.2021.V09I01N01.841-844