

# **LITERATURE REVIEW ON FLIGHT DELAY PREDICTION**

## **MODEL USING MACHINE LEARNING**

### **ABSTRACT**

The current and the existing circumstances due to the traffic congestion causing flight delays these flight delays not only causing economic impact but also have harmful environment effects and degrading the passenger quality of service and fuel consumption and gas consumption the airline management had becoming the increasingly challenging to overcome this issues. By using the factors causing the airline delay we carry out the predictive analysis and machine learning algorithms to find the causes of flight delays.

To resolve this situation, supervised machine learning models were implemented to predict flight delays. Seven algorithms (Logistic Regression, K-Nearest Neighbor, Gaussian Naïve Bayes, Decision Tree, Support Vector Machine, Random Forest, and Gradient Boosted Tree) were trained and tested to complete the binary classification of flight delays. The evaluation of algorithms was fulfilled by comparing the values of four measures: accuracy, precision, recall, and f1- score. These measures were weighted to adjust the imbalance of the selected data set.

### **INTRODUCTION**

Flight delays could always be annoying, especially in the case when the period of delay was so long that there was even a danger to miss the next flight. However, if there was a way to predict whether there would be a delay or even better – how long the delay could be, then people could make earlier preparation to reschedule following flights in an earlier manner. For that consideration, we adopted a dataset containing airline delayed time and other air liner information provided by Kaggle to building a model, mainly aiming to solve the following questions. Whether there would be a delay with certain publicly

reachable resources; and 2. How long delayed time one could expect with the same information given. We deployed python sklearn and pandas library to build our model, and evaluate our model based on R-Square for linear regression and accuracy rate for logistic regression. As a brief result of our project, we found, it would be helpful to use the following factors in evaluating our model: week, month, airline carrier reference, planned elapsed time (in air time), distance between two departure and destinations, flight planned departure time, departure airport code, and taxi-in and taxi out time.(1) The continuous increase of storage capacities and computational power is currently pulling the development of data analytics. Indeed, companies (and especially IT-intensive ones) are collecting massive volume of data (often referred as Big Data), such as web logs, customer information, production and sales tracking, etc. Analyzing these datasets, with data mining algorithms for example, allows the extraction of information that can help a company to gain knowledge (for example on customs' behaviors) or to use the information as a basis for new products or services.

## **LITERATURE REVIEW**

Since flight delays cause multiple problems across the world, there has been a significant improvement in delay prediction models right from the 1990s. The quantity of the delay decreased the quality of marketing strategies. A delay in the departure or arrival of a domestic flight affects the operation of an international flight. A small amount of change in the delay value can be a massive amount of success for airport sectors. The models developed during this system can be applied to predict the incidence of flight delay at airports. Such prognosticative capabilities would help traffic managers and airline dispatchers to organize mitigation strategies for lowering traffic disruptions.

Nowadays, service quality plays an important role in attracting customers. Among these, air travels have their special customers and the most important matter in these travels is the flight time, on time arrival at destination for passengers such those who have an important

meeting, that has been leading to high expenses for the passenger until they get to their destination on time. Flight delay has negative economic effects on the passenger, agencies and airport. Therefore, any reduction of these effect requires decreasing postponed flight price, so that prediction or estimation has a great significance and numerous studies has been to dedicated this subject. Correspondingly, all the scientists have tried to design a model that understands effective factors and computes effect of each factor and their relation. Overall, the prediction methods are classified into five groups including Statistical Methods, Probability methods, network-based methods, operational methods and machine learning methods .

In one of the best studies that has been performed based on statistics delay time has been considered to be reduced. Their study has investigated important factors before fly and those which occur on the ground. In the next step, it has predicted the delay at destination based on factors that occur in the vicinity of arrival time at destination. Eventually, results have shown that whenever, the delay is correctly predicted, passenger disaffection and fuel consumption decrease and consequently number of flight increases. Moreover, it is possible to increase the agencies benefits through reducing number of passengers who wrongly selected their routes or specifying the probabilities for some flights and optimizing delay time prediction.

Another prominent investigation based on Probability has been done and the author believes that huge storm in U.S.A has highly affected the flight delay. This study has been devoted to predict delay based on mathematical calculations and through considering delay time duration of the flights that had been engaged to storm in the same day. Metrological reports have shown the effect of storm one hour before and after event cause ephemeral climate at the region. In the next step, Monte-Carlo simulation has been used to estimate the airport runway capacity, so that traffic of each runway would have been estimated. As the research has employed only one factor, the model has not enough accuracy, but it is possible to increase region air capacity path structure.

## **CONCLUSION**

Hence, Machine learning and deep learning algorithms can be used to predict flight delays. The purpose of doing the above classification and analysis is to gauge the delay to not only suffice the various purposes of humankind, but also analyze factors affecting delay. The Proposed System is build using Support Vector Machine, Random Forest and KNN. The overall precision, recall and accuracy is calculated by using above techniques As seen from the articles and papers These methodologies offer virtually identical accuracy however we want an algorithmic rule that is good with real world prediction and analysis and thus: naïve-Bayes. Except being clever with real time prediction algorithmic rule that considers or assumes independence among predictions that makes the system scalable. As other independent attribute may be superimposed up to the algorithmic rule for computation of the delay, the anticipated postpone can thus facilitate the floor personnel for implementing accurate and easy operation plans and consequently the facts, if sent to the passengers will benefits the airways and also the passengers.

## **REFERENCES**

1. Airline Flight Delay Prediction Using Machine Learning Models - Yuemin Tang - University of Southern California, USA  
<https://dl.acm.org/doi/pdf/10.1145/3497701.3497725>
2. Yi Ding 2017 IOP Conf. Ser.: Earth Environ. Sci. 81 012198  
<https://iopscience.iop.org/article/10.1088/1755-1315/81/1/012198/pdf>
3. FLIGHT DELAY PREDICTION BASED ON AVIATION BIG DATA AND MACHINE LEARNING By Rahul Garg, Soham Gosavi, Tejas Choulwar, Onkar Vallal, Prof. Sarika Aundhakar  
[https://www.irjmets.com/uploadedfiles/paper//issue\\_5\\_may\\_2022/22985/final/fin\\_irjmets1652551445.pdf](https://www.irjmets.com/uploadedfiles/paper//issue_5_may_2022/22985/final/fin_irjmets1652551445.pdf)