# Developing a Flight Delay Prediction Model

## LITERATURE SURVEY

**1. TITLE : Flight Delay Prediction: Data Analysis and Model Development**

YEAR: 2021

AUTHORS : Azib Anees; Wei Huang

DESCRIPTION:

Flight delays in air transportation are a major concern that has adverse effects on the economy, the passengers, and the aviation industry. This matter critically requires an accurate estimation for future flight delays that can be implemented to improve airport operations and customer satisfaction. Having said that, a massive volume of data and an extreme number of parameters have restricted the way to build an accurate model. Many existing flight delay prediction methods are based on small samples and/or are complex to interpret with little or no opportunity for machine learning deployment. This paper develops a prediction model by analysing the data of domestic flights within the United States of America (USA). The proposed model gains insight into factors causing flight delays, cancellations and the relationship between departure and arrival delay using exploratory data analysis. In addition, Random Forest (RF) algorithm is used to train and test the big dataset to help the model development. A web application has also been developed to implement the model and the testing results are presented with the limitations discussed.

## 2. TITLE : Flight Delay Prediction System

**YEAR: 2020**
**AUTHORS :  Yogita Borse , Dhruvin Jain , Shreyash Sharma , Viral Vora, Aakash Zaveri**

DESCRIPTION:

As discussed, considering the standard taxonomy of the flight delay and its problems, one will contemplate the scope of prediction to be one in every of these factors or combination of those factors[3]. The models developed during this system may be applied to predict the incidence of flight delay at airports. Such prognosticative capabilities would facilitate traffic managers and airline dispatchers to organize mitigation methods for reducing traffic disruptions.

This issue can be reduced by developing the flight delay prediction tool which can be developed using following methods.

Statistical analysis

Statistical model requires the use of correlation analysis, parametric and non parametric tests, multivariate analysis and econometric models. Government agencies have invested in these econometric models to understand the relationship between delay and Passenger demand, fare, size of aircraft etc

Probabilistic models

Probabilistic model requires analysis tools that estimates the probability of an event based on the historic data. The estimated outcome is given in form of a distribution function of the probability. The factor of randomness

always makes an impact on the decision or the outcome produced by the probabilistic model.

**3. TITLE : Research on Flight Delay Prediction Based on Random Forest**

**YEAR: 2021**
**AUTHORS : Peng Hu; Jianping Zhang; Ning Li**

DESCRIPTION:

Based on the random forest model, this paper proposes a flight delay prediction model. By analyzing the departure flight data of Guangzhou Baiyun International Airport in June 2020, and selecting the data of ten landing airports, it analyzes the distribution of delayed, punctual, and early arrived. It studies the selection of features that impact on flight delays, and establishes random forest predictions model. Through case study, it researches the mean square error of different leaf sizes when the forest scale is 50 trees. The results show that the optimal leaf size is 5, and the minimum mean square error is 0.1096. And it analyzes the importance of features such as departure flight delay time, scheduled flight time, number of scheduled departure flights on the day, date, and landing airport. The research results also found that, when the forest size is 100 trees and the leaf size is 5, the out-of-bag mean square error is 0.1090, and the accuracy of the prediction model is high, which is close to 90%.

## 4. TITLE: A deep learning approach to flight delay prediction

**YEAR: 2016**
**AUTHORS : Young Jin Kim; Sun Choi; Simon Briceno; Dimitri Mavris**

DESCRIPTION:

Deep learning has achieved significant improvement in various machine learning tasks including image recognition, speech recognition, machine translation and etc. Inspired by the huge success of the paradigm, there have been lots of tries to apply deep learning algorithms to data analytics problems with big data including traffic flow prediction. However, there has been no attempt to apply the deep learning algorithms to the analysis of air traffic data. This paper investigates the effectiveness of the deep learning models in the air traffic delay prediction tasks. By combining multiple models based on the deep learning paradigm, an accurate and robust prediction model has been built which enables an elaborate analysis of the patterns in air traffic delays. In particular, Recurrent Neural Networks (RNN) has shown its great accuracy in modeling sequential data. Day-to-day sequences of the departure and arrival flight delays of an individual airport have been modeled by the Long Short-Term Memory RNN architecture. It has been shown that the accuracy of RNN improves with deeper architectures. In this study, four different ways of building deep RNN architecture are also discussed. Finally, the accuracy of the proposed prediction model was measured, analyzed and compared with previous prediction methods. It shows best accuracy compared with all other methods.

**5.TITLE : The prediction of flight delays based the analysis of Random flight points**

**YEAR: 2015**

**AUTHORS : Fei Rong; Li Qianya; Hu Bo; Zhang Jing; Yang Dongdong**

DESCRIPTION:

The prediction of flight delays is heavily investigated in the last few decades. However, there is a relatively low level of study on Random flight point delays in these important problems. In this paper, we present an influence factor model of random flight points by series analysis on actual airline data, which is combined with BN (Bayesian Network) and GMM-EM(Gaussian mixture model-expectation maximization algorithm) algorithm. The creation of the initial parameters is based on the analysis for the continuous flights fly over the same flight point. The test data is offered by some Air Traffic Management Bureau. And the test result clearly demonstrates the value of Bayesian Network for analyzing the system-level effects arising from micro-level causes.