

**Prediction in Flight Delays**

A Project

Presented to :

Dr. Joshan Athanesious,

Assistant Professor Senior Grade 2,

School of Computer Science and Engineering (SCOPE),

Vellore Institute of Technology, Chennai – 600127.

Presented By :

Aayushi – 20BCE1791

Literature Review :-

Paper 1

**A Review on Flight Delay Prediction**

March 2017

<https://www.researchgate.net/publication/315382748_A_Review_on_Flight_Delay_Prediction>

This paper presents the use of methods like

1. Statistical Analysis
2. Statistical analysis usually encompasses the use of regression models, correlation analysis, econometric models, para-
3. metric tests, non-parametric tests, and multivariate analysis (MVA). When it comes to regression models, both delay
4. multiplier and recursive models can help airlines to understand delay propagation eﬀects through the network and to
5. estimate the costs of delay

Statistical analysis usually encompasses the use of regression models, correlation analysis, econometric models, para-metric tests, non-parametric tests, and multivariate analysis (MVA). When it comes to regression models, both delay multiplier and recursive models can help airlines to understand delay propagation eﬀects through the network and to estimate the costs of delays.

1. Probabilistic Models
2. Probabilistic Models encompass analysis tools that estimate the probability of an event based on historical data. Tu
3. et al. [112] developed a probabilistic model based on expectation-maximization combined with genetic algorithms to
4. predict the distribution of departure delay at Denver International Airport.

Probabilistic Models encompass analysis tools that estimate the probability of an event based on historical data. developed a probabilistic model based on expectation-maximization combined with genetic algorithms to predict the distribution of departure delay at Denver International Airport.

3. Network Representation

built direct acyclic graphs to model the schedule of an airline (including ﬂight times and resources availability) to

detect disruptions and their impacts on the rest of the network.

Network representation encompasses the study of ﬂight systems according to a graph theory. Abdelghani et al. built direct acyclic graphs to model the schedule of an airline (including ﬂight times and resources availability) to detect disruptions and their impacts on the rest of the network.

4. Operational Research

This includes advanced operational methods like optimizations, simulations and query theory to help key players make better decisions.

Paper 2

**Flight delay prediction based on deep learning and Levenberg-Marquart algorithm**

* November 2020
* [Journal of Big Data](https://www.researchgate.net/journal/Journal-of-Big-Data-2196-1115) 7
* DOI:[10.1186/s40537-020-00380-z](https://journalofbigdata.springeropen.com/articles/10.1186/s40537-020-00380-z)

<https://www.researchgate.net/publication/346539839_Flight_delay_prediction_based_on_deep_learning_and_Levenberg-Marquart_algorithm>

Nowadays, service quality plays an important role in attracting customers. Flight delay has negative economic eﬀects on the passenger, agencies and airports. 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 dedicated to 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

1. Statistical Methods,
2. Probability methods,
3. network-based methods ,
4. operational methods and
5. machine learning methods

This paper displays the idea of using Machine learning methods to predict the delay. The researcher has presented a model based on support vector regressor (SVR) algorithm to predict ﬂight delay in U.S.A airports. Due to the large amount of data, the data was grouped and sampled by month. The results have shown SVR has the best prediction result for the ﬂight delay time with the best accuracy value was 80.44%. Also, the time characteristics had a large impact on the mode performance.

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.

Paper 3

**Predicting flight delay based on multiple linear regression**

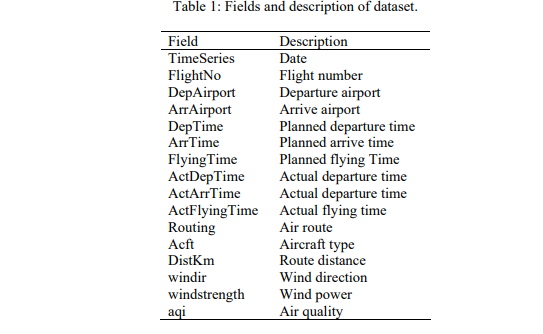
Published under licence by IOP Publishing Ltd

<https://iopscience.iop.org/article/10.1088/1755-1315/81/1/012198>

Data Used: - [www.umetrip.com](http://www.umetrip.com)

There are 25 fields in the dataset, and the main fields are shown in Table 1.

Table 1: Fields and description of dataset.

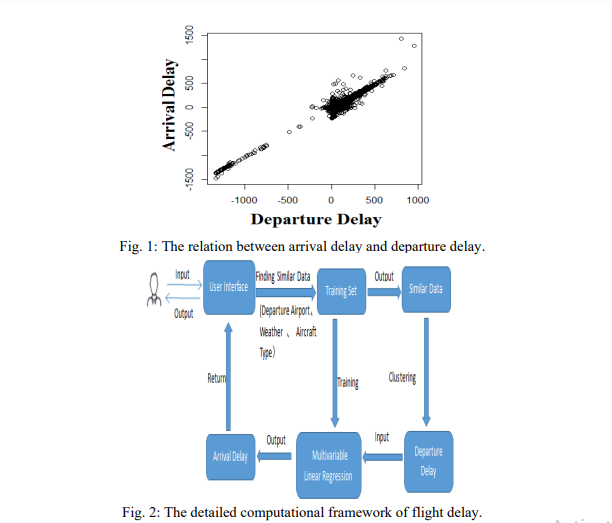


Flight delay has become the shackles of the development of the aviation industry. Today, flight delay has not only become a problem for the majority of travellers, but also the world's civil aviation industry problems. . However, flight delay was caused by many reasons. The main factor is that the capacity of airport and airspace is insufficient. For other reasons, such as the weather, the airport scheduling, the company plan, passengers and luggage etc. may cause flight delay.

This paper has depicted the use of a multiple linear regression model using departure delay and route distance to predict arrival delay, and presented the design and implementation of a flight delay prediction system.

Principle of multiple linear regressions:-

In statistics, linear regression is an approach for modelling the relationship between a scalar dependent variable y and one or more explanatory variables (or independent variables) denoted X. The case of one explanatory variable is called simple linear regression. For more than one explanatory variable, the process is called multiple linear regressions. In linear regression, the relationships are modelled using linear predictor functions whose unknown model parameters are estimated from the data. Such models are called linear models.



Paper 4

**Flight Delay Prediction System**

Vol. 9 Issues 03, March-2020

<https://www.ijert.org/research/flight-delay-prediction-system-IJERTV9IS030148.pdf>

This paper presents that prediction and reduction in flight delays can be brought up by using following methods:-

1. 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.

1. Probabilistic Model

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.

1. Machine learning

Supervised Machine learning could be a task where the dataset input and also the output are recognized, then many algorithms are used to analyse this data to map new examples. Here in this case is that the prediction of delay in flight.

1. Logistic Regression

Logistic regression an algorithm that performs classification using,

ℎ𝜃 (𝑥) = (𝜃 𝑇𝑥) = 1 1 + 𝑒 −𝜃𝑇𝑥

Which in turn represents the maximum likelihood of estimation and gradient ascent Logistic regression is the applicable multivariate analysis to conduct once the variable is divided (binary). Like all regression analyses, the logistical regression is a predictive analysis. Logistic regression is employed to explain data and to explain the relationship between one dependent binary variable and one or additional nominal, ordinal, interval or ratio-level independent variables.

1. Neural Networks

Neural Network is made by stacking along multiple neurons in layers to provide a final output. Initial layer is that the input layer and therefore the last is that the output layer.

This paper presented the need to develop a system to predict the delay in flights along with its methodology. The paper gives details about the range of different methodology that is used or can be used to find out the delay in flights.

Paper 6

**Airline Flight Delay Prediction Using Machine Learning Models**

October 2021

Computing methodologies → Machine learning;

The paper performed a prediction of the occurrence of flight delays by adapting it into a machine learning problem. A supervised machine learning approach in the form of binary classification was used for the prediction. Seven algorithms were used for delay prediction, and four measures were used for algorithms performance evaluation. Machine learning is the designation of algorithms that enable the computer to analyse the data, obtain potential patterns, and then use them to predict.

This study, classification models were selected and trained using seven algorithms: Logistic Regression, K-Nearest Neighbour (KNN), Gaussian Naïve Bayes, Decision Tree, Support Vector Machine (SVM), Random Forest, and Gradient Boosted Tree.

Due to the imbalanced nature of the data set, evaluation measures were weighted to eliminate the dominant effect of non-delayed flights over delayed flights. After applying classifiers to the delay prediction, the values of their four measures were compared to evaluate the performance of each model.

The result shows that the highest values of accuracy, precision, recall, and f1-score are generated by the Decision Tree model (accuracy: 0.9778; precision: 0.9777; recall: 0.9778; f1-score: 0.9778). Such high values indicate that the Decision Tree performs well when predicting flight delays in the data set.