

Airline Delay Prediction

Neha Jain¹ Poornima Harjai² Ms Narinder Kaur³

^{1,2,3}Department of Information Technology

^{1,2,3}Maharaja Agrasen Institute of Technology, Rohini, India

Abstract— The project deals with airline delay problem. The airport delay can be understood by delay distribution of airport. The assessment of an airport's schedule performance is also discussed. Various factors that lead to flight delay are analyzed with the help of functions and graphs. The outcomes conclude the dependence of originate and arrival delay. The airport arrival delay is found to show monthly and weekly patterns, which is related to the schedule performance.

Key words: Airline Delay Problem, Delay Prediction

I. INTRODUCTION

Due to increasing demand, there has been a constant increase in the air traffic. Major airports around the world face significant delay problems because of imbalance between capacity and demand. It leads to flight delay. Here demand is the total number of flights that are scheduled to arrive or depart in a given time frame whereas capacity is the maximum number of flight arrival and departures in given time frame. Capacity demand imbalance occurs when the demand of resources exceed the capacity that can be afforded. It is also the reason behind airport congestion. The only way of handing various activities efficiently so as to avoid delay is done by Flight scheduling. It is necessary for planning and executing functionalities. But flight schedule are subjected to irregularity. Flight delays can occur due to various reasons like weather, arrival delay, etc. Sometimes delay of a flight is propagated to successive connecting flights just like chain reaction. This is known as reactionary delay. In such cases, delays are spread to various other flights. But there exist some patterns in flight delays due to schedule performance which can be further analyzed with the help of certain factors so that flight delays can be avoided.

II. AIM OF THE STUDY

- 1) Analyzing flight delay using R programming. Representing dataset with help of functions and graphs
- 2) Predicting flight delay using python.

III. PROBLEM

There are three issues related to flight delay

A. Delay propagation

It deals with the spreading of delays to many flights. Here delay of a flight is propagated to the successive Connecting flights like chain reaction thus increasing the problem.

B. Root delay

Root delay is basically concerned about the origin and causes of a delay. Proper prediction of delay so as to analyze when and where it occurred.

C. Cancellation

Flights get cancelled due to various reasons. Like for example: delays or whether issues.

IV. NEED OF THE PROJECT

Delays can prove to be a headache for the passengers. People who are on their business trips might miss out their conferences or meetings. People with some medical condition also have to face a lot of challenges while waiting for their flights to arrive or depart. Troubles would also be faced by customers who have connecting flights. They need to reschedule their plans. If the passengers would be alerted hours ago, they would be able to have a comfortable experience with the airline. It would save resources as well as time of the passengers. It would lead to a convenient journey.

Its beneficial for the passengers as they don't have to waste time waiting for the delayed flight. They can get warned and accordingly schedule their activities. They would get enough time to have their plans executed properly rather than sitting around, wasting their time in an uncomfortable airport.

V. DATASET DESCRIPTION

The data was collected from US Department of transportation ie Bureau of Transportation Statistics. The dataset comprise of list of airports and carriers. It included the year, month, arrival delay, departure delay, elapsed time, origin and destination of various unique carriers. Code system was used to indicate the origin and destination airports.

The entire process starts from the collection of various factors in the form of dataset from the web site. These datasets are further checked for any missing values. This data is analyzed by the help of R programming and we get the desired results in the form of graphs.

The factors are as follows:

- Year
- Month
- Day of Month
- Day of Week
- Unique Carrier
- Origin Airport ID
- Destination Airport ID
- Departure Time - It stands for the local time at which the plane was scheduled to depart
- Arrival Time - It stands for the local time at which the plane was scheduled to arrive.
- Arrival Delay - It shows how early/late the plane was at its final destination in minutes: our target variable
- Distance - It tells how far the plane traveled for the route.
- Elapsed Time - It stands for the scheduled difference between departure and arrival

VI. RESEARCH METHODOLOGY

- 1) The data is first fetched into R Studio.
- 2) The semi-structured format is converted into structured format.

- 3) The structured data set is analyzed to find out desired outcomes.
- 4) Regression testing is used to predict delay using python.

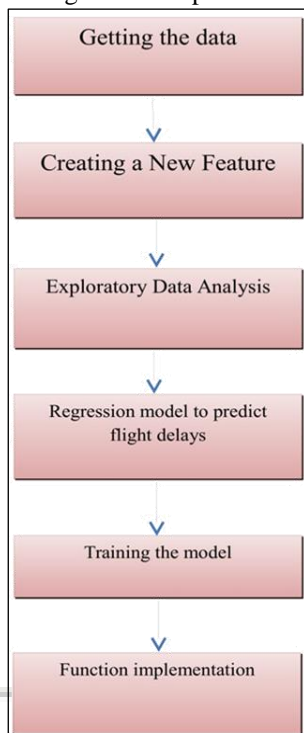


Fig. : Research methodology

VII. REGRESSION ANALYSIS

It is basically a process for defining relationship between a dependent variable and two or more independent variables. Here we can predict the value of one variable by the other one. Independent variable is used to analyze the value of dependent variable.

In every situation, regression function is generated. Regression function is the function of independent variables. Regression model can scale many training examples. Once the model has been trained, it can be effectively used to analyze flight delays. Inputting one training example at a time greatly helps in reducing the memory needed while running the model. Factors like arrival delay, departure delay, elapsed time, origin, destination, day of the week, day of the month, etc helps in analyzing and predicting various flight delays.

VIII. MODEL CREATION

Logistic regression technique is used for prediction of delay. The dataset is split into two categories namely train set and test set. We use stratified KFold (K) cross validation technique so as to avoid overfitting the model. It leads to K folds, K-1 train sets and the remaining are the test sets.

Finally logistic regression confusion matrix is created which depicts the relationship between true labels and the predicted labels.

IX. CONCLUSION

The results of the research show that the arrival delay is highly related to the originate delay. The airport arrival delay shows monthly and weekly patterns, which is further closely

associated with schedule performance. The factors contributing to flight delay are distance, day of the week, day of the month, season, holiday, arrival time, departure time and space between two successive flights. This research also helps us examine flight delay levels which further depict arrival delay patterns. Characteristics of flights and their effect on flight delay is also taken in consideration. The patterns of delay from the flight level in which delays occur are analyzed, and the significant reasons of delay are given out. It was also concluded that delay in flight depends on the subsequent flight delay and is not confined to a certain time period. If there is delay in the arrival of an aircraft then this delay will be further passed to the next connecting flight.

Its beneficial for the passengers as they don't have to waste time waiting for the delayed flight. They can get warned and accordingly schedule their activities. They would get enough time to have their plans executed properly rather than sitting around, wasting their time in an uncomfortable airport

ACKNOWLEDGMENT

Every successful venture has the blessings and the support of many behind it. It would be unjust if every such support is not revealed and thanked in person as well as in the published report.

First and foremost, we would like to thank Ms. Narinder Kaur for giving us an opportunity to work under her as our mentor and her constant guidance throughout the project. Also I would like to thank Dr. M.L Sharma (HOD) for his support and guidance.

I would even like to thank my parents and friends to support me throughout this internship period.

I perceive this opportunity as a big milestone in my career development.

REFERENCES

- [1] Jesse Steinweg-Woods, Ph.D. <https://jessesw.com/Air-Delays/> March 16, 2015
- [2] Joanna Medina, <http://www.flightbucks.com/blog/how-scientists-are-predicting-flight-delays> December 14th, 2016
- [3] Qi Wang, <https://hortonworks.com/tutorial/predicting-airline-delays-using-sparkr/> Aug 31, 2016
- [4] Alice Sternberg, Jorge Soares, Diego Carvalho, Eduardo Ogasawara, A Review on Flight Delay Prediction, Nov 6, 2017