

# Flight Delay Prediction

Rudranshi Dave<sup>1</sup>, Maulik Mehta<sup>2</sup>, Prof. Ajaykumar T. Shah<sup>3</sup>

<sup>1,2</sup> Dept of Computer Engineering

<sup>3</sup>HOD, Dept of Computer Engineering

<sup>1,2,3</sup> Alpha College of Engineering and Technology, Gandhinagar, Gujarat, India

**Abstract-** Flight delays are quite frequent because 19% of the Indian domestic flights arrive more than 30 minutes late and not only have economic impact but is a major source of frustration for the passengers. Booking of Flight tickets is costly for many of us and delaying or cancellation of flight may affect us all. As delays are a stochastic phenomenon, it is interesting to study their entire probability distribution, instead of looking for an average value. This system proposes model to estimate delay based on method of regression. These are data-driven methods which means that it only consider past observations. Our model has been implemented, optimized and evaluated on a large scale, using several years of records of INDIA's domestic flight delays as well as its affecting aspects as weather and air traffic data too.

**Keywords-** Flight Delay Prediction, Data Analysis, Machine learning.

## I. INTRODUCTION

Imagine having a scheduled meeting in other state or city or planning a tour with having other scheduled bookings or having other emergencies, a delay of flight can ruin your trip and can cause and economic impact on your pocket. But having a knowledge of your flight's delay you can make arrangements according that with booking another flight or can take an insurance of that ticket that now most of insurance companies provides.

The goal of this machine learning model is to predict whether the flight you are booking is likely to arrive on time or not. We would like to give this information to the customers during the search and reservation process, the model will have to give long term prediction with the analysis of data.

In case of flight delay the customer can change the flight as well as cancel it but doing that will cost them a huge amount of money, that's where this system can help the user, like when the user get to buy a ticket with having an idea for the probability of that's flight's delay or cancellation and can make decisions according to that.

Overall most of the previous work has limitations. The traditional approaches can model the delay propagation of air traffic operation, but they lack the ability to analytically analyze the huge volume of traffic and weather dataset. On the other side, the machine learning approaches has the ability to discover the hidden patterns in the data, but they lack the inner relationships between different airports, such as delay propagation. Therefore, a mixed approach including the machine learning and air traffic operation is proposed with feature selection to improve the performance of the prediction. Moreover this new approach can work as a model to provide sequence future delay prediction for individual flights along their scheduled itineraries.

## II. LITERATURE REVIEW

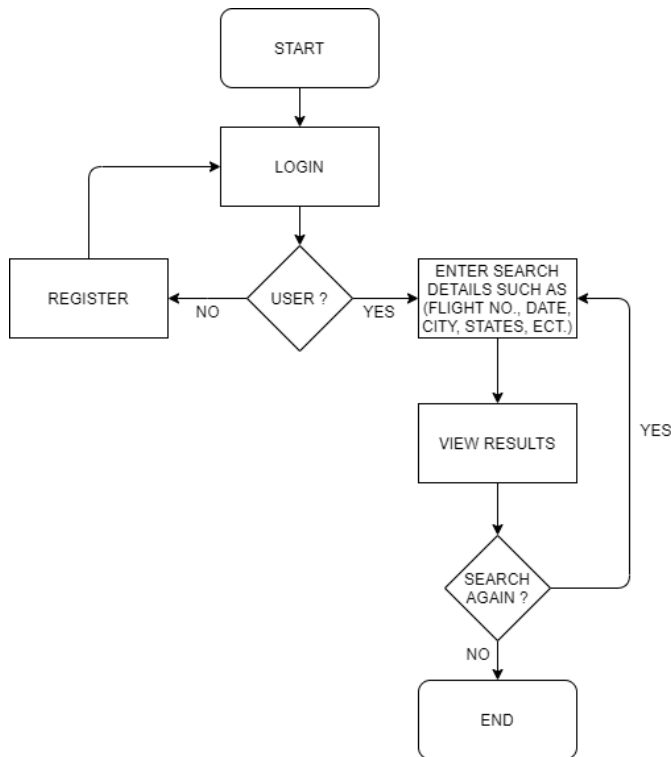
Flight delays affect airlines, airports and passengers. Their prediction is necessary and critical during decision-making process for all the players of commercial aviation. Moreover, the development of accurate prediction models for flight delays became clumsy due to the complex system of air transport system.

There are several work in the literature that focuses on the air traffic, airlines, airports and weather datasets. As well as the approaches to reach the accurate prediction of flight's delay which are machine learning and data analysis. An important area that is extensively studied is finding and measuring factors affecting aircraft delays on the ground and in the air and develop machine learning algorithm to optimize airline and airport operations based on the factor responsible for the flight delay.

## III. DESIGN AND DEVELOPMENT

Our project's main goal is to create a system that provides user a knowledge of probability of flight's delay or cancellation to avoid the not having an idea of the future circumstances that can affect the scheduled plans.

First of all, we created a basic flowchart to get a clear idea about how we want our application to work and perform its functions.



**Fig 2: Flight Delay Prediction Flowchart**

### III. A: DATASET AND FEATURES

To train and test our models, we used a publicly available Kaggle dataset for INDIAN domestic air traffic and generated some of our own on weather too. The dataset has historic data for many years and its features categorized as follows:

- Information about flight (day, day of the week, airline, flight)
- Information about origin and destination (origin airport, destination airport)
- Information about the departure (scheduled departure, departure time, departure delay)
- Information about the flight journey
- Information about arrival
- Information about diversion, cancellation and reason of delay (airline delay, weather delay)

### III. B: METHODOLOGY

By applying the models of decision tree and Logistic Regression.

### IV. STUDY FINDINGS

After studying about the problem and finding solutions about it, we can say that the flight delay or cancellation has really a frustrated impact on all the players of

the aviation system that includes airlines, airports and passengers.

And to have a solution about it can improve this system as well as will provide a very good help to users with bookings and cancellations.

Although there will always be something to improve and to work on because the technology as well as enhancing experiences can help discover new problems and their solutions.

### V. CONCLUSION

Using Machine learning and data analysis can give more accurate results and working with dataset, the system can find patterns that can actually lead the system to generate accuracy in result and will be very much to the users of aviation system.

### VI. ACKNOWLEDGEMENT

This project is a part of the Department of Computer Engineering, Alpha College of Engineering, Khatraj, Gandhinagar. The authors would like to thank valuable professors and friends at ACET for providing their helpful support and invaluable advice.

### REFERENCES

- [1] Nathalie Kuhn and Navaneeth Jamadagni, CSS299:AUTUMN2017 <http://cs229.stanford.edu/proj2017>
- [2] Jun Chen (San Diego State University, California, 92182), Meng Li (Purdue University, Indiana, 47906) [https://www.researchgate.net/publication/330185077\\_Chained\\_Predictions\\_of\\_Flight\\_Delay\\_Using\\_Machine\\_Learning#:~:text=Given%20the%20initial%20departure%20delay,along%20the%20same%20aircraft's%20itinerary.&text=Our%20application%20results%20clearly%20demonstrate,traffic%20delay%20in%20daily%20operation.](https://www.researchgate.net/publication/330185077_Chained_Predictions_of_Flight_Delay_Using_Machine_Learning#:~:text=Given%20the%20initial%20departure%20delay,along%20the%20same%20aircraft's%20itinerary.&text=Our%20application%20results%20clearly%20demonstrate,traffic%20delay%20in%20daily%20operation.)
- [3] <https://www.kaggle.com/datasets>
- [4] Hang Zhou and Xinxin Jiang, College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China Volume 2014, Article ID 851202, 13 pages <http://dx.doi.org/10.1155/2014/851202>
- [5] [https://en.wikipedia.org/wiki/Flight\\_cancellation\\_and\\_delay](https://en.wikipedia.org/wiki/Flight_cancellation_and_delay)
- [6] <https://docs.microsoft.com/en-us/learn/modules/predict-flight-delays-with-python/>

- [7] Chetan Khobragade, Shashant Pandit, Devyani Shrikundwar, Abhishek Dahikar, Prof. Manisha Pise, International Journal for Scientific & Development/ Vol. 8, Issue 3,2020, ISSN: 2321-0613