Flight Delay Prediction

[1]Harshal Bharambe, [2]Rishikesh More [3]Anuja Rothe, [4]Yashita Rajkotia, [5]Varsha Bhole, [6]Amruta Kothavade

[1,2,3,4,5,6]Information Technology A.C.Patil College of EngineeringNavi Mumbai, Kharghar
Lecturer, Dept. of Information Technology, A.C. Patil College of EngineeringNavi Mumbai, Kharghar

Abstract: Flight Planning is one of the demanding situations in commercial world, which faces many unsure conditions. There is such condition in delay occurrence, which stems from various factors and imposes considerable costs on airlines, operators, and travellers. Delays in departure can occur due to bad weather conditions, seasonal and holiday demands, airline policies, technical problems such as problems with airport facilities, baggage handling and mechanical equipment, and the accumulation of delays from previous flights. In This flight delay prediction system based on the Aviation Data, which can result in delays. The system considers the of various parameters. Random Forest (RF), K-Nearest Neighbour (KNN) and Support Vector Machine (SVM) arethe algorithm used in this system.

I. INTRODUCTION

One of the important business issues that airways face are the vital prices that are related to flights getting delayed because of natural occurrences and operational shortcomings that is an upscale affair for airways, causing scheduling and operational problems for end users, thereby causing an unhealthy name and customer dissatisfaction. As we know, we tend not to have a flight delay before departure because the airline's customers as well as the airline's ground staff receive a forecast of the delay from the supporting airline under different conditions. However, we all know that one of the reason for flight delays is the weather and the other is some unpredicted problems like some mechanical issue, which risk safety of passengers. This prompted us to use live weather knowledge combined with various metrics to calculate wing lag before departure. In 2017, Indian state of affairs, in line with the reports by the (DGCA) Directorate General of Civil Aviation, from January to April, close to 5.12 hundred thousand domestic passengers faced problems because of airline corporations not boarding, moreover as flight cancellations and delays. Airline corporations had to pay the passenger's compensations of over 25cr for varied inconveniences throughout the first four months of that year. Hence, the prediction analysis retrieved from this project can contribute within the form of a prototype in serving to identify operational variables that contribute to delays in any state of affairs. The main reason associated with flight delay prediction are known and arranged in taxonomy. It includes the problem that causes the flight to delay, the range of things it affects, and ways that of handling flight delay prediction downside. It considers airline domain options, like problem and solution. Major problem which causes delay in flights can be delay propagation, delay caused during departure, and cancellation of flights. These problems cannot be solved forever, but a latency predictor will allow operators and administrators to take the necessary steps for smooth operation. This delay issue affects airlines, airports and rerouting airspace which are independent entities operating in harmony. Thus, flight delays cause problems in all areas. Machine Learning, Probabilistic models, Statistical analysis or Network Representations are various methods that are used to create a system that predicts the delay in flights.

II. PROBLEM STATEMENT

This proposed system helps Airline passengers to know whether the flight will get delayed or not. To make the system more scalable it is necessary to choose an algorithm which considers all the parameters to be independent. Supervised learning as the name indicates a presence of supervisor as teacher. Essentially supervised learning could be a learning that within which we tend to teach or train the machine exploitation data which is well tagged which means some data is already labelled with correct answer. After that, machine is given new set of examples(data) so supervised learning algorithm rule analyses the coaching knowledge (set of training examples) and produces a correct outcome from tagged data Using supervised machine learning approach, the labelled data gives it authenticity. Apart from this a web application will be developed that will use Open- Source APIs to fetch real time data of flight delay with the reason of flight delay and also provide support contacts and options to the airline users.

III. EXISTING SYSTEM

Flight delay is inevitable and it plays an important role in both profits and loss of the airlines. An accurate estimation of flight delay is critical for airlines because the results can be applied to increase customer satisfaction and income of airline agencies. A fundamental sub-area of artificial intelligence has come into notice, called as Machine Learning, which enables computers to get into a mode of self-learning without being explicitly programmed. With the concept of machine learning, we have been able to apply complex mathematical computations to big data iteratively and automatically, that too with efficient speed, this phenomenon has been encompassing momentum over the last several years. There have been many researches on modelling and predicting flight delays, where most of them have been trying to predict the delay through extracting important characteristics and most related features. However, most of the proposed methods are not accurate enough because of massive volume data, dependencies and extreme number of parameters. It is a machine learning task where the dataset inputs and outputs are clearly recognized and already given, then several types of algorithms are trained using labelled examples.

IV. LITERATURE SURVEY

With increasingly tight flight schedules, the prediction of aviation resources is developing rapidly. The differences in the current research are mainly in the prediction methods and the input factors considered. Prediction methods are either based on statistics (Stats) or based on machine learning (ML) or deep learning (DL). The influencing factors considered are mainly divided into direct and indirect factors. As mentioned earlier, the direct influencing factors are those that have nothing to do with the time series, which will not be accumulated. However, the indirect factors are related to the time series, these factors will accumulate

over time, and finally affect the delay of a flight. Much literature addresses the statistical analysis. Tu et al. used a genetic algorithm to fit delay data and study long- and short-term flight departure trends [1]. The model included seasonal influences, daily trends, and random trends, enabling users to grasp general delay characteristics. Hsiao and Hansen considered the influence of arrival queues, passenger flow, weather and other factors on flight delays [2]. Through econometric analysis of the contribution rates of various factors to delays, the model explained 72– 73% of the variation in the average delay. Hao et al. used econometric and simulation models to calculate and decompose delays, considering direct factors such as quarter-hourly data on throughput, demand, and arrival rates [3]. Rodriguez-Sanza et al. [4] used a Bayesian network and time-series features to model randomness and time variation of flight delays. However, the prediction results consisted of statistical guidance rather than a tactical operation. ML and DL are developing rapidly. Rebollo and Balakrishnan divided flight delay data into temporal data (e.g., day of week, month of year, and time of day) and spatial data (e.g., delay state and type of delay) [5]. They used a random forest model to predict departure delays in the next two hours, with an average error of about 21 min on the test dataset. However, they only conducted a sensitivity analysis of influencing factors and not the primary factors (or time points) of delays. Manna et al. used the gradient boosted decision tree model (GBDT) to predict the delay of a flight with six directed factors [6]. Their model can also be used by passengers or airline agencies. Kim et al. placed delays at different levels for prediction, using a recurrent neural network (RNN) to consider time and other direct factors such as weather and visibility [7]. Based on a single airport, McCarthy et al. studied the delays of multiple airports with a long short-term memory (LSTM) algorithm, using the time series of the past 24 h to predict delays[8].The method was shown to be accurate and robust for low-cost airlines in Europe. The analysis of causes on the whole network is helpful to gain an overall understanding of critical factors. Qiang uses the Random Forest (RF) algorithm to predict the delay of a single airport [9], and the method was alidated by
U.S. domestic flights. However, it cannot explain the delay for one flight. Zhen and Bin et al. combine the RF and the maximal information coefficient to analyse the flight delay of PEK, but there is no analysis of factors of delay in detail [10].

V. WORKING

1. The working is separated into three main stages: Initial, Middle, Last stage.
2. The Initial stage is identified with Data Exploration, Data Cleaning and Data Transformation.
3. The centre stage comprises of data modelling.
4. The final stage comprises of data analysis using multiple models.
5. Data exploration is similar to initial data analysis, visual exploration to understand what is in a dataset and the characteristics of the data, rather than through traditional data management systems.
6. Data Cleaning is the process of detecting and correcting (or removing) corrupt or inaccurate records from a record set, table, or database and refers to identifying incomplete, incorrect, inaccurate or irrelevant parts of the data and then replacing, modifying, or deleting the dirty or coarse data.
7. Data transformation is the process of converting data from one format to another, typically from the format of a source system into the required format of a destination system.
8. Once the first stage is cleared then we move to data modelling. Data modelling is the process of producing a descriptive diagram of relationships between various types
of information that are to be stored in a database. One of the goals of data modelling is to create the most efficient method of storing information while still providing for complete access and reporting.
9. After this the data is processed using algorithms andresults are obtained.
10. This result are the test results generated by training the models on the train dataset.
11. Once the dataset is processed then we can make use of the real time data.
12. A web app that shows flight delays with specific reasons for delay in flight on real time basis with the help of Open- Source APIs is designed.
13. Also, contact information of officials of the respectiveairports and airlines is added for passengers' convenience.

VI. MODULES

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. 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, sizeof 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.

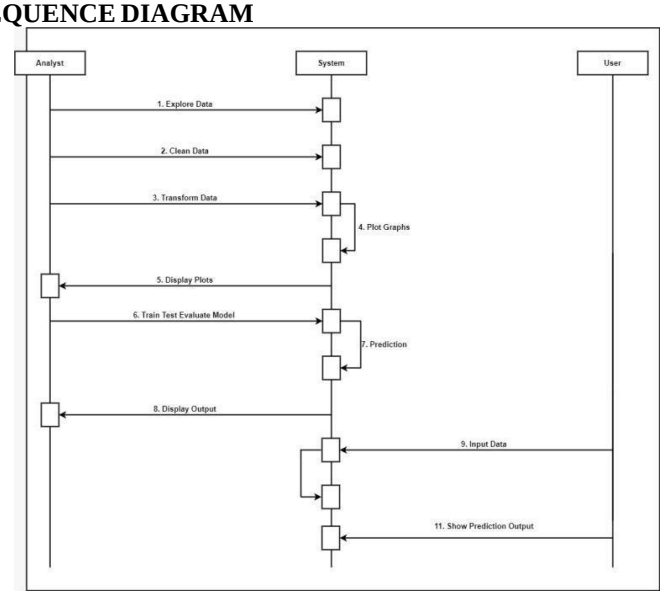
Machine Learning

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.

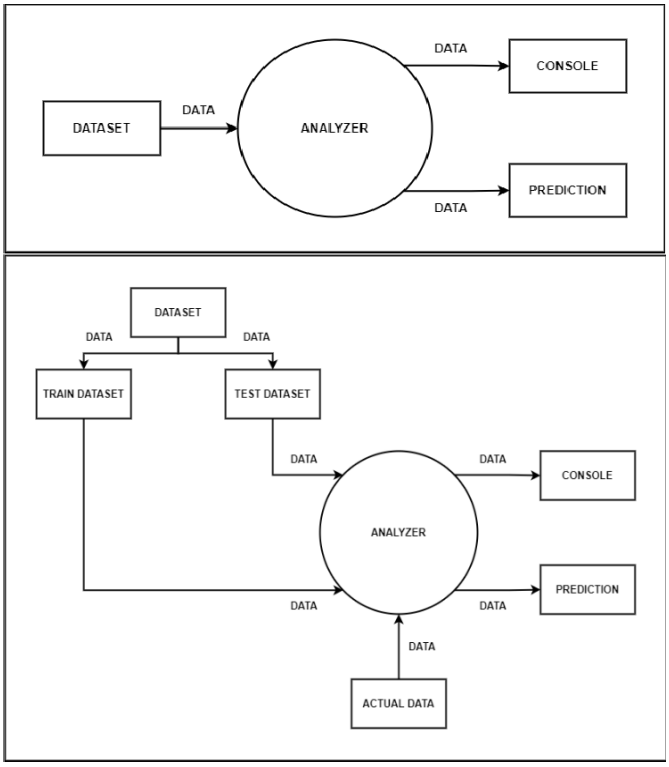
Web Application

Here we plan to design a web app that shows flight delays with specific reasons for delay in flight on real time basis with the help of Open-Source APIs. Also, we would beadding contact information of officials of the respective airports and airlines.

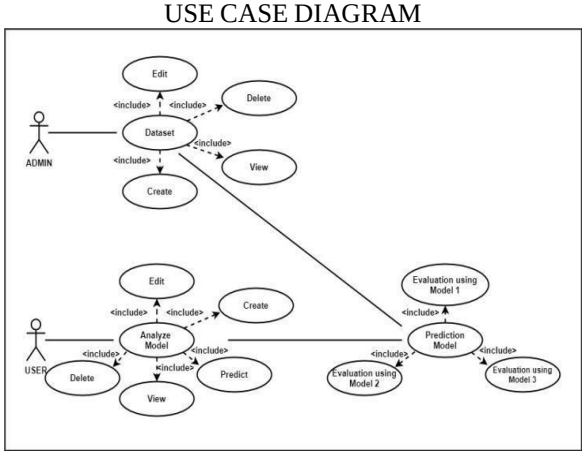
VII. DESIGN DETAILS SEQUENCE DIAGRAM



DATA FLOW DIAGRAM



USE CASE DIAGRAM



VIII. SCOPE

The system proposed has a lot of scope as it will solve various issues related to Flight Delays. The system not only analyses flight delays with respect to machine learning based on past history but also shows real-time flight delay with actual reason of flight delay and support options.

IX. CONCLUSION

With the rapid development of the air transport industry, in countries with less aviation airspace and high population density, such as China, flight delays increasingly affect the improvement of the air transport efficiency and travel experience. Among the policies and measures to alleviate flight delays, accurate prediction plays a critical role. The prediction of flight delays can provide a decision-making reference for airline crew configuration and ground guarantee resource allocation. To the extent possible, we considered all influencing factors related to flight delays. These are of two types: one is directly related to flight attributes and arrival times, which are not subject to direct intervention, and the other consists of indirect influencing factors associated with the pre-order flight and airport state. The indirect influencing factors allow human intervention before flight delay occurs. The result will help ease passenger tension and anxiety, and provide a reference for passenger time management.

X. ACKNOWLEDGEMENT

I thank my college Principal Dr. V. N. Pawar sir for providing the required resources for the development of the project. I would also like to thank HOD Dr. V. Y. Bhole for suggesting such a great project topic for departmental purposes. My sincere thanks to my Project Guide Prof._____for helping, suggesting new ideas and guiding me throughout the semester. I am also grateful to all the faculty members for their support.

REFERENCES

1. Tu, Y., Ball, M.O., Jank, W.S.: Estimating flight departure delay distributions - A statistical approach with long-term trend and short-term pattern. *J. Am. Stat. Assoc.* 103(481), 112– 125 (2008)
2. Hsiao, C., Hansen, M.: Econometric analysis of USairline flight delays with time-of-day effects. *Transp. Res. Rec.* 1951, 104– 112 (2006)
3. 6Hao, L., Hansen, M., Zhang, Y., Post, J.: New York: Two ways of estimating the delay impact of New Yorkairports. *Transp. Res. Part E Logist. Transp. Rev.* 70: 245– 260 (2014)
4. Rodriguez-Sanza, A., Gomez Comendador, F., Arnaldo Valdes, R., Perez-Castan, J., Barragan Montes, R., Camara Serrano, S.: Assessment of airport arrival congestion anddelay: Prediction and reliability. *Transp. Res. Part C Emerging Technol.* 98, 255– 283 (2019)
5. Rebollo, J.J., Balakrishnan, H.: Characterization andprediction of air traffic delays. *Transp. Res. Part C Emerging Technol.* 44, 231– 241 (2014)
6. Manna, S., Biswas, S., Kundu, R., Rakshit, S., Gupta, P., Barman, S.: A statistical approach to predict flight delay using gradient boosted decision tree. In: 2017 International Conference on Computational Intelligence in Data Science (ICCIDS). Chennai, Tamil Nadu (2017)
7. Y.J. Kim, S. Choi, S. Briceno, D. Mavris, (eds.): *A Deep Learning Approach to Flight Delay Prediction*. IEEE, Sacramento, CA (2016)
8. McCarthy, N., Karzand, M., Lecue, F.: Amsterdam to dublin eventually delayed? LSTM and transfer learning for predicting delays of low cost airlines. *Proc. AAAI Conf. Artif. Intell.* 33(01), 9541– 9546 (2019)
9. Li, Q., Jing, R.: Generation and prediction of flight delaysin air transport. *IET Intell. Transp. Syst.* 15(6), 740– 753 (2021)
10. Guo, Z., Yu, B., Hao, M., Wang, W., Jiang, Y., Zong, F.: A novel hybrid method for flight departure delay prediction using Random Forest Regression and Maximal Information Coefficient. *Aerosp. Sci.Technol.* 116(106822), 1– 10 (2021)