



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

# IJIRSET

International Journal of Innovative Research in  
**SCIENCE | ENGINEERING | TECHNOLOGY**

# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 12, Special Issue 2, April 2023

**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA

9940 572 462

6381 907 438

[ijirset@gmail.com](mailto:ijirset@gmail.com)

[www.ijirset.com](http://www.ijirset.com)

# Developing a Flight Delay Prediction Model Using Machine Learning

Abinesh V<sup>1</sup>, Sneha A<sup>2</sup>, Haripriya M<sup>3</sup>, Rupiya K<sup>4</sup>

U.G. Student, Department of Information Technology, Gojan School of Business and Technology, Chennai, India<sup>123</sup>

Assistant Professor, Department of Information Technology, Gojan School of Business and Technology, Chennai, India<sup>4</sup>

**ABSTRACT:** Over the last twenty years, air travel has been increasingly preferred among travelers, mainly because of its speed and in some cases comfort. This has led to phenomenal growth in air traffic and on the ground. An increase in air traffic growth has also resulted in massive levels of aircraft delays on the ground and in the air. These delays are responsible for large economic and environmental losses. The main objective of the model is to predict flight delays accurately in order to optimize flight operations and minimize delays. Using a machine learning model, we can predict flight arrival delays. The input to our algorithm is rows of feature vector like departure date, departure delay, distance between the two airports, scheduled arrival time etc. We then use decision tree classifier to predict if the flight arrival will be delayed or not. A flight is considered to be delayed when difference between scheduled and actual arrival times is greater than 15 minutes. Furthermore, we compare decision tree classifier with logistic regression and a simple neural network for various figures of merit.

**KEYWORDS:** Flight delay prediction, Airline transport, Machine Learning, Classification Algorithms, Data Analytics.

## I. INTRODUCTION

As additional individuals decide to go via plane, the quantity of flights that neglect to require some investment develops. This extension worsens the congestion at air terminals and causes monetary challenges for the carrier business. Air transportation delays demonstrate an absence of productivity in the flying framework. It comes at an excessive cost for the two carriers and travelers. As per the Complete Defer Effect Study, the all out cost of air transportation postponements to air explorers and the carrier business in 2007 was \$32.9 billion in the US, coming about in a \$4 billion decrease in Gross domestic product [1]. Foreseeing flight deferrals can hence further develop carrier tasks and traveler fulfillment, bringing about a positive monetary effect. In this study, the main goal is to compare the performance of machine learning classification algorithms when predicting flight delays. The airport used in the study is John F. Kennedy International Airport that located in New York City. The information of flights leaving JFK airport between one-year periods was being analyzed. The study made use of several algorithms, and their predictions were evaluated using a number of measures.

The theoretical aspects of selected machine learning models and performance evaluation methods are explained in Section 3. In Section 2, related works by past researchers are discussed. The empirical processes and results of different models are presented and compared in Section 4. The conclusion of the comparative analysis and directions for future research are presented in Section 5.

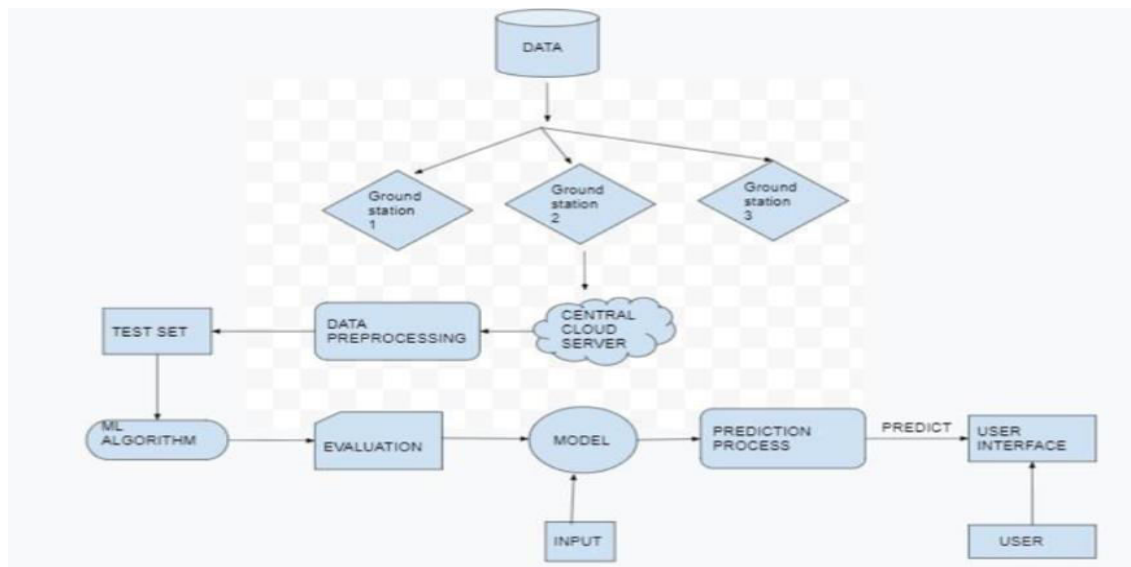
## II. RELATED WORK

The demand for airline transportation is increasing dramatically these days. As a result, flight delay analysis has become a popular research topic. To conduct the investigation, various researchers used various machine learning and data mining techniques. They were interested in a variety of factors, including airport facility location, weather conditions, and airport capacity. The use of machine learning enables researchers to handle large amounts of flight data for storage and processing. Khaksar and Sheikholeslami conducted a study that used Bayesian modelling, decision trees, cluster classification, and random forest to estimate the occurrences and magnitude of delay in the US and Iranian airline networks [1]. They determined that the main parameters affecting the US airline network are visibility, wind, and departure time, while those affecting Iran's airline network are fleet. The procedure to do this project is to install required libraries. Data Collection is to collect the dataset or create the dataset. Data Pre processing is to import the Libraries, Importing the

dataset, Understanding Data Type and Summary of features, Take care of missing data, Data Visualization, Drop the column from Data Frame & replace the missing value, Splitting the Dataset into Dependent and Independent variables, Splitting Data into Train and Test. Model Building is the process of training and testing , evaluation and saving the Model. Application Building is the process of create an HTML file and build a Python Code Finally the user interface will show the dashboard of the flask app. This paper is divided into four sections (i)Data Collection (ii) Comparison of machine learning models on collected data (iii) Training of system on most promising model (iv) Testing. The training data set is now supplied to machine learning model, on the basis of this data set the model is trained. Every new applicant details filled at the time of application form acts as a test data set. After the operation of testing, model predict whether the new applicant is a fit case for approval of the loan or not based upon the inference it conclude on the basis of the training data sets.

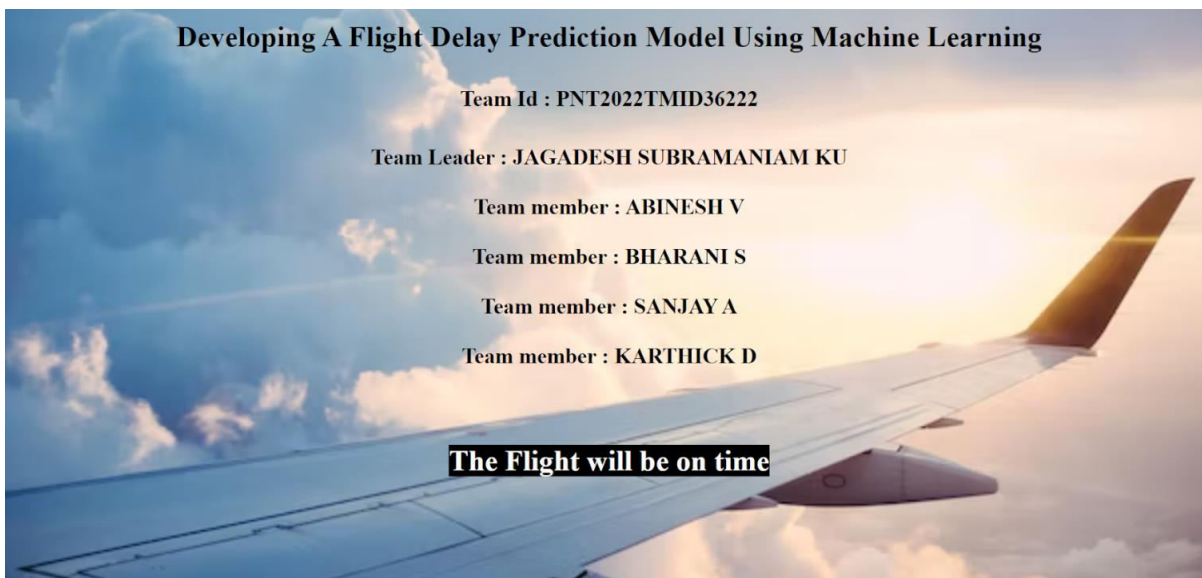
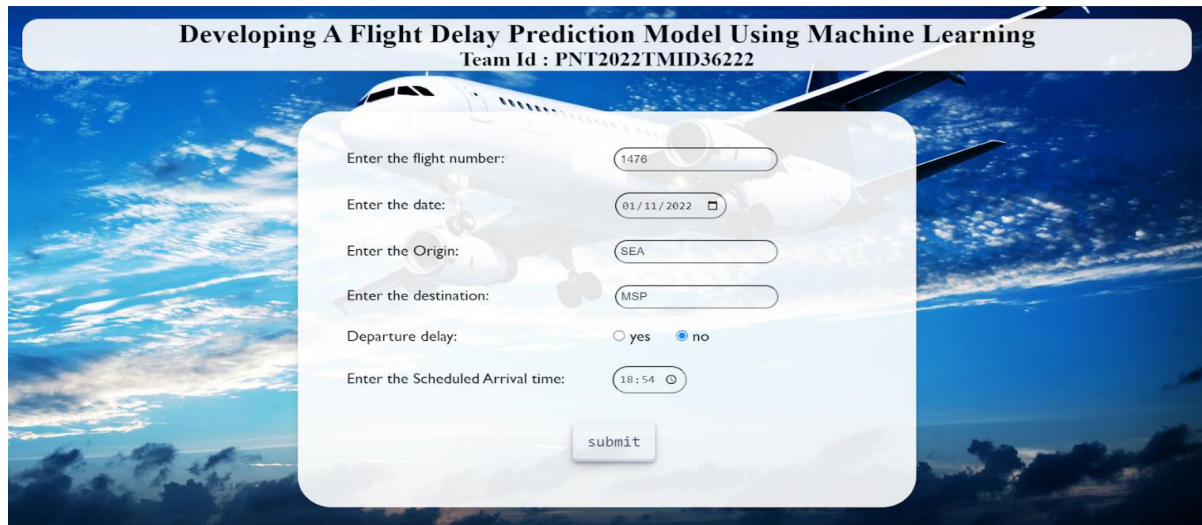
### III.METHODOLOGY

Machine learning is the designation of algorithms that enable the computer to analyze the data, obtain potential patterns, and then use them to predict. Learning algorithms can give insight into the relative difficulty of learning in different environments [8]. Machine learning algorithms are divided into several categories, and the two most common types are supervised learning and unsupervised learning. Algorithms of supervised learning generated a function that translates inputs to desired outputs. The primary forms of supervised learning algorithms include regression and classification. Unsupervised learning models a collection of inputs in the absence of labeled examples.



### IV. EXPERIMENTAL RESULTS

The data set used for analysis contains data about flights leaving from JFK airport between one year from November 2019 to December 2020. It was obtained from open data at <https://www.kaggle.com/deepankurk/flight-take-off-data-jfk-airport>. It includes 28820 lines of individual flight information with 23 columns. Other researchers have used this data set to conduct delay prediction using the numerical value of the variable “TAXI\_OUT”. In this study, the prediction of flight delays was conducted on the variable “DEP\_DELAY” by binary classification.



**Developing A Flight Delay Prediction Model Using Machine Learning**  
Team Id : PNT2022TMID36222

Team Leader : JAGADESH SUBRAMANIAM KU

Team member : ABINESH V

Team member : BHARANI S

Team member : SANJAY A

Team member : KARTHICK D

**The Flight will be on time**

## V. CONCLUSION

Although the flight delay are predicted in advance based on record, some natural causes such as climate change, machine malfunction, etc. These causes will stand in our way and are quite problematic. The developed ML model will predicts the flight delay with at most accuracy. The Decision Tree algorithm is used to boost up the performance of the model. This algorithm will enhance the prediction and give a accurateresults to the users.

## REFERENCES

- [1] Shaowu Cheng, Yaping Zhang, Siqi Hao, Ruiwei Liu, Xiao Luo, Qian Luo, "Study of Flight Departure Delay and Causal Factor Using Spatial Analysis", Journal of Advanced Transportation, vol. 2019, Article ID 3525912.
- [2] Wang, F., Bi, J., Xie, D., Zhao,X.: Flight delay forecasting and analysis of direct and indirect factors. IET Intell. Transp. Syst. 16, 890–907(2022).
- [3] Gui, G., Liu, F., Sun, J., Yang, J., Zhou, Z., & Zhao, D. (2019). Flighdelayprediction based on aviation big data and machine learning. IEEE Transactions on Vehicular Technology, 69(1), 140- 150.



**Organized by**

Department of CSE, Velammal Engineering College, Chennai, India

- [4] Choi, S., Kim, Y. J., Briceno, S., & Mavris, D. (2016, September). Prediction of weather-induced airline delays based on machine learning algorithms. In 2016 IEEE/AIAA 35th Digital Avionics Systems Conference (DASC)(pp. 1-6). IEEE.
- [5] Kim, Y. J., Choi, S., Briceno, S., & Mavris, D. (2016, September). A deep learning approach to flight delay prediction. In 2016 IEEE/AIAA 35th Digital Avionics Systems Conference (DASC)(pp. 1-6). IEEE.
- [6] Borse, Y., Jain, D., Sharma, S., Vora, V., & Zaveri, A. (2020). Flight Delay Prediction System. *Int. J. Eng. Res. Technol*, 9(3), 88-92.
- [7] Ding, Y. (2017, August). Predicting flight delay based on multiple linear regression. In *IOP Conference Series: Earth and Environmental Science*(Vol. 81, No. 1, p. 012198). IOP Publishing.
- [8] Tang, Y. (2021, October). Airline Flight Delay Prediction Using Machine Learning Models. In 2021 5th International Conference on E-Business and Internet (pp. 151-154).
- [9] Li, Q., & Jing, R. (2022). Flight Delay Prediction from Spatial and Temporal Perspective. *Expert Systems with Applications*, 117662.
- [10] ATLIOĞLU, M. C., BOLAT, M., ŞAHİN, M., TUNALI, V., & KILINÇ, D. (2020). Supervised learning approaches to flight delay prediction. *Sakarya University Journal of Science*. <https://doi.org/10.16984/saufenbilder.710107>.
- [11] Yu, B., Guo, Z., Asian, S., Wang, H., & Chen, G. (2019). Flight delay prediction for commercial air transport: A deep learning approach. *Transportation Research Part E: Logistics and Transportation Review*, 125, 203–221. <https://doi.org/10.1016/j.tre.2019.03.013>.



**SJIF: 8.423**



INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INTERNATIONAL CENTRE



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 **9940 572 462**  **6381 907 438**  **ijirset@gmail.com**



[www.ijirset.com](http://www.ijirset.com)

Scan to save the contact details