# Predicting Flight Delays with Error Calculation Using Machine Learned Algorithm

[1]Syed Mustafa Ali, [2]Abdul Aziz Akbar Khan, [3]Md Abdul Wadood Khan, [4]Mrs. T. Anita

[1] BE Student, Dept. of Computer Science Engineering, ISL Engineering College

[2] BE Student, Dept. of Computer Science Engineering, ISL Engineering College

[3] BE Student, Dept. of Computer Science Engineering, ISL Engineering College

[4] Assistant Professor, Dept. of Computer Science Engineering, ISL Engineering College

**Article Info**
**Page Number: 1428-1434**
**Publication Issue:**
**Vol. 72 No. 1 (2023)**

**Article History**
**Article Received:** 15 October 2022
**Revised:** 24 November 2022
**Accepted:** 18 December 2022

**ABSTRACT**
Flight delay is a major problem in the aviation sector. During the last two decades, the growth of the aviation sector has caused air traffic congestion, which has caused flight delays. Flight delays result not only in the loss of fortune also negatively impact the environment. Flight delays also cause significant losses for airlines operating commercial flights. Therefore, they do everything possible in the prevention or avoidance of delays and cancellations of flights by taking some measures. In this paper, using machine learning models such as Logistic Regression, Decision Tree Regression, Bayesian Ridge, Random Forest Regression and Gradient Boosting Regression we predict whether the arrival of a particular flight will be delayed or not.

## INTRODUCTION

Flight delay is studied vigorously in various research in recent years. The growing demand for air travel has led to an increase in flight delays. According to the Federal Aviation Administration (FAA), the aviation industry loses more than $3 billion in a year due to flight delays and, as per BTS, in 2016 there were 860,646 arrival delays [1][13].

The reasons for the delay of commercial scheduled flights are air traffic congestion, passengers increasing per year, maintenance and safety problems, adverse weather conditions, the late arrival of plane to be used for next fligh [2][14]. In the United States, the FAA believes that a flight is delayed when the scheduled and actual arrival times differs by more than 15 minutes. Since it becomes a serious problem in the United States, analysis and prediction of flight delays are being studied to reduce large costs [3][15].

## LITERATURE REVIEW

Much research has been done on studying flight delays. The prediction, analysis and cause of flight delays have been a major problem for air traffic control, decision-making by airlines and ground delay response programs. Studies are conducted on the delay and departure delay with meteorological features is encouraged [4]. In the past, researchers have tried to predict flight delays with Machine Learning. Chakrabarty et al used supervised automatic learning algorithms (random forest, Gradient Boosting Classifier, Support Vector Machine and the k-nearest neighbour algorithm) to predict delays in the arrival of operated flights including the five busiest US airports [5][16]. The maximum precision achieved was 79.7% with gradient booster as a classifier with a limited data set. Choi et al. applied machine learning algorithms like decision tree, random forest, AdaBoost and kNearest Neighbours to predict delays on individual flights [6] .Flight schedule data and weather forecasts have been incorporated into the model. Sampling techniques were used to balance the data and it was observed that the accuracy of the classifier trained without sampling was more that of the trained classifier with sampling techniques[7].

Cao et al. used a Bayesian Network model to analyse the turnaround time of a flight and delay prediction. Juan José Rebollo and Hamsa Balakrishnan used a hundred pairs of origin and destination to summarise the result of various regression and classification models[8]. The find outs reveal that among all the methods used, random forest has the highest performance. However, predictability may additionally range because of factors such as the number of origin destination pairs and the forecast horizon [9]. Sruti Oza, Somya Sharma used multiple linear regression to predict weather induced flight delays in flight-data, as well as climatic factors and probabilities due to weather delays. The forecasts were based on some key attributes, such as carrier, departure time, arrival time, origin and destination. Anish M. Kalliguddi and Aera K. Leboulluec predicted both departure and arrival delays using regression models such as Decision Tree Regressor, Multiple Linear Regression and Random Forest Regressor in flight-data. It has been observed that the longer forecast horizon is useful for increasing the accuracy with a minimum forecast error for random forests. Etani J Big Data[10][17].

A supervised model of on-schedule arrival fight is used using weather data and flight data. The relationship between flight data and pressure patterns of Peach Aviation is found. On-Schedule arrival flight is predicted with 77% accuracy using Random Forest as a Classifier [11][20].

This study analyzes high-dimensional data from Beijing International Airport and presents a practical flight delay prediction model. Following a multifactor approach, a novel deep belief network method is employed to mine the inner patterns of flight delays. Support vector regression is embedded in the developed model to perform a supervised fine-tuning within the presented predictive architecture. The proposed method has proven to be highly capable of handling the challenges of large datasets and capturing the key factors influencing delays. This ultimately enables connected airports to collectively alleviate delay propagation within their network through collaborative efforts (e.g., delay prediction synchronization)[12][18].

Flight delays hurt airlines, airports, and passengers. Their prediction is crucial during the decision-making process for all players of commercial aviation. Moreover, the development of accurate prediction models for flight delays became cumbersome due to the complexity of air transportation system, the number of methods for prediction, and the deluge of flight data. In this context, this paper presents a thorough literature review of approaches used to build flight delay prediction models from the Data Science perspective. We propose a taxonomy and summarize the initiatives used to address the flight delay prediction problem, according to scope, data, and computational methods, giving particular attention to an increased usage of machine learning methods. Besides, we also present a timeline of significant works that depicts relationships between flight delay prediction problems and research trend[12][19].

## SYSTEM ARCHITECTURE

Django is based on MVT (Model-View- Template) architecture. MVT is a software design pattern for developing a web application.

MVT Structure has the following three parts:

Model: Model is going to act as the interface of your data. It is responsible for maintaining data. It is the logical data structure behind the entire application and is represented by a database (generally relational databases such as MySql, Postgres). To check more, visit – Django Models.

View: The View is the user interface — what you see in your browser when you render a website. It is represented by HTML/CSS/Javascript and Jinja files. To check more, visit – Django Views.

Template: A template consists of static parts of the desired HTML output as well as some special syntax describing how dynamic content will be inserted. To check more, visit – Django Templates.

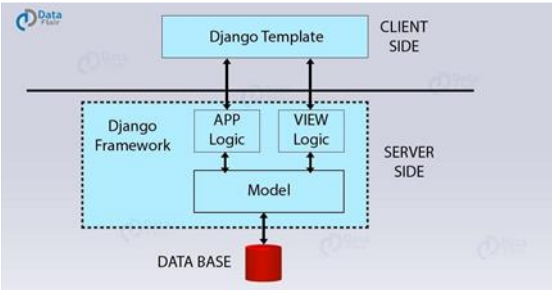



**Fig 1.** System Architecture

## PROPOSED ALGORITHM

"Boosting" in machine learning is a way of combining multiple simple models into a single composite model. This is also why boosting is known as an additive model, since simple models (also known as weak learners) are added one at a time, while keeping existing trees in the model unchanged. As we combine more and more simple models, the complete final model becomes a stronger predictor. The term "gradient" in "gradient boosting" comes from the fact that the algorithm uses gradient descent to minimize the loss.

When gradient boost is used to predict a continuous value – like age, weight, or cost – we're using gradient boost for regression. This is not the same as using linear regression. This is slightly different than the configuration used for classification, so we'll stick to regression in this article.

Decision trees are used as the weak learners in gradient boosting. Decision Tree solves the problem of machine learning by transforming the data into tree representation. Each internal node of the tree representation denotes an attribute and each leaf node denotes a class label.

The loss function is generally the squared error (particularly for regression problems). The loss function needs to be differentiable.

Also like linear regression we have concepts of residuals in Gradient Boosting Regression as well. Gradient boosting Regression calculates the difference between the current prediction and the known correct target value.

This difference is called residual. After that Gradient boosting Regression trains a weak model that maps features to that residual. This residual predicted by a weak model is added to the existing model input and thus this process nudges the model towards the correct target. Repeating this step again and again improves the overall model prediction.

Also it should be noted that Gradient boosting regression is used to predict continuous values like house price , while Gradient Boosting Classification is used for predicting classes like whether a patient has a particular disease or not.

## IMPLEMENTATION

Machine Learning algorithms are trained with data to create a model. New input data is then used to make predictions. If the accuracy is acceptable, the algorithm is deployed; otherwise, it is retrained with augmented data. The goal is to build a predictive model to solve a given problem.

In the Machine Learning process, the first step is to define the objective of the problem, such as predicting rain based on weather conditions. Understanding the required data and the approach to solve the problem is crucial.

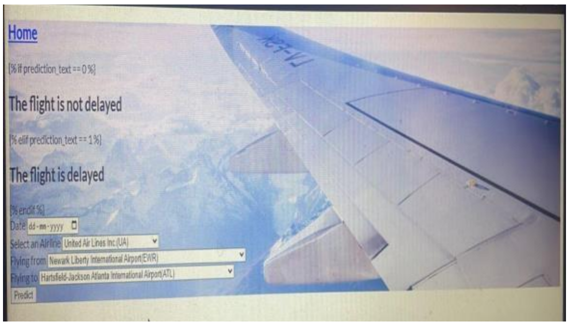
Next, in the data gathering stage, you determine the type of data needed and its availability. Data can be obtained through manual collection or web scraping. For beginners, ready-to-use datasets from online resources are convenient options. For weather forecasting, data like humidity level, temperature, pressure, locality, and other relevant measures must be collected and stored.

After data collection, the data preparation stage addresses inconsistencies like missing values, redundant variables, and duplicates. These inconsistencies need to be resolved to ensure accurate computations and predictions. By scanning the dataset and rectifying issues, data is prepared for further analysis.

## SNAP SHOTS


**Fig 2:** Home Screen

**Fig 3:** Input Screen



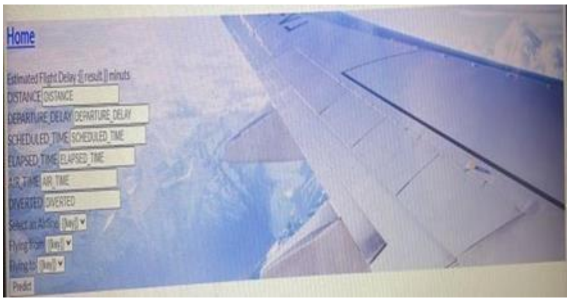




**Fig 4:** Output Screen

## CONCLUSION AND FUTURE ENHANCEMENT

Machine learning algorithms were applied progressively and successively to predict flight arrival & delay. We built five models out of this. We saw for each evaluation metric considered the values of the models and compared them. We found out that: - In Departure Delay, Random Forest Regressor was observed as the best model with Mean Squared Error 2261.8 and Mean Absolute Error 24.1, which are the minimum value found in these respective metrics. In Arrival Delay, Random Forest Regressor was the best model observed with Mean Squared Error 3019.3 and Mean Absolute Error 30.8, which are the minimum value found in these respective metrics. In the rest of the metrics, the value of the error of Random Forest Regressor although is not minimum but still gives a low value comparatively. In maximum metrics, we found out that Random Forest Regressor gives us the best value and thus should be the model selected.

The future scope of this paper can include the application of more advanced, modern and innovative preprocessing techniques, automated hybrid learning and sampling algorithms, and deep learning models adjusted to achieve better performance. To evolve a predictive model, additional variables can be introduced. e.g., a model where meteorological statistics are utilized in developing error-free models for flight delays. In this paper we used data from the US only, therefore in future, the model can be trained with data from other countries as well. With the use of models that are complex and hybrid of many other models provided with appropriate processing power and with the use of larger detailed datasets, more accurate predictive models can be developed. Additionally, the model can be configured for other airports to predict their flight delays as well and for that data from these airports would be required to incorporate into this research.

**REFERENCES**

1. N. G. Rupp, "Further Investigation into the Causes of Flight Delays," in Department of Economics, East Carolina University, 2007.
2. "Bureau of Transportation Statistics (BTS) Databases and Statistics," [Online]. Available: http://www.transtats.bts.gov
3. "Airports Council International, World Airport Traffic Report," 2015,2016.
4. A. M. Kalliguddi, Area K., Leboulluec, "Predictive Modelling of Aircraft Flight Delay," Universal Journal of Management, pp. 485 - 491, 2017.
5. E. Cinar, F. Aybek, A. Caycar, C. Cetek, "Capacity and delay analysis for airport manoeuvring areas using simulation," Aircraft Engineering and Aerospace Technology, vol. 86, no. No. 1,pp. 43-55, 2013.
6. Navoneel, et al., Chakrabarty, "Flight Arrival Delay Prediction Using Gradient Boosting Classifier," in Emerging Technologies in Data Mining and Information Security, Singapore, 2019.
7. Y. J. Kim, S. Briceno, D. Mavris, Sun Choi, "Prediction of weatherinduced airline delays based on machine learning algorithms," in 35th Digital Avionics Systems Conference (DASC), 2016.
8. W.-d. Cao. a. X.-y. Lin, "Flight turnaround time analysis and delay prediction based on Bayesian Network," Computer Engineering and Design, vol. 5, pp. 1770-1772, 2011.
9. J. J. Robollo, Hamsa, Balakrishnan, "Characterization and Prediction of Air Traffic Delays".
10. S. Sharma, H. Sangoi, R. Raut, V. C. Kotak, S. Oza, "Flight Delay Prediction System Using Weighted Multiple Linear Regression," International Journal of Engineering and Computer Science, vol. 4, no. 4, pp. 11668 - 11677, April 2015.
11. Noriko, Etani, "Development of a predictive model for on-time arrival fight of airliner by discovering correlation between fight and
12. 1822-b.tech-it-batchno-331.pdf (sathyabama.ac.in)
13. M.A.Bari & Shahanawaj Ahamad, “Code Cloning: The Analysis, Detection and Removal”, in International Journal of Computer Applications(IJCA),ISSN:0975-887,ISBN:978-93-80749-183,Vol:20,No:7,pp:34.,NewYork,U.S.A.,April 2011

14. Jain, V., Beram, S. M., Talukdar, V., Patil, T., Dhabliya, D., & Gupta, A. (2022). Accuracy enhancement in machine learning during blockchain based transaction classification. Paper presented at the PDGC 2022 - 2022 7th International Conference on Parallel, Distributed and Grid Computing, 536-540. doi:10.1109/PDGC56933.2022.10053213 Retrieved from www.scopus.com
15. Kathole, A. B., Katti, J., Dhabliya, D., Deshpande, V., Rajawat, A. S., Goyal, S. B., . . . Suciu, G. (2022). Energy-aware UAV based on blockchain model using IoE application in 6G network-driven cybertwin. Energies, 15(21) doi:10.3390/en15218304
16. Keerthi, R. S., Dhabliya, D., Elangovan, P., Borodin, K., Parmar, J., & Patel, S. K. (2021). Tunable high-gain and multiband microstrip antenna based on liquid/copper split-ring resonator superstrates for C/X band communication. Physica B: Condensed Matter, 618 doi:10.1016/j.physb.2021.413203
17. Kothandaraman, D., Praveena, N., Varadarajkumar, K., Madhav Rao, B., Dhabliya, D., Satla, S., & Abera, W. (2022). Intelligent forecasting of air quality and pollution prediction using machine learning. Adsorption Science and Technology, 2022 doi:10.1155/2022/5086622
18. Dr. M.A.Bari, "EffectiveIDS To Mitigate The Packet Dropping Nodes From Manet ", JACE, Vol -6,Issue -6,June 2019.
19. Syed Shehriyar Ali, Mohammed Sarfaraz Shaikh, Syed Safi Uddin, Dr. Mohammed Abdul Bari, "Saas Product Comparison and Reviews Using Nlp", Journal of Engineering Science (JES), ISSN NO:0377-9254, Vol 13, Issue 05, MAY/2022
20. Hafsa Fatima, Shayesta Nazneen, Maryam Banu, Dr. Mohammed Abdul Bar," Tensorflow-Based Automatic Personality Recognition Used in Asynchronous Video Interviews", Journal of Engineering Science (JES), ISSN NO:0377-9254, Vol 13, Issue 05,MAY/2022
21. Mohammed Shoeb, Mohammed Akram Ali, Mohammed Shadeel, Dr. Mohammed Abdul Bari, "Self-Driving Car: Using Opencv2 and Machine Learning", The International journal of analytical and experimental modal analysis (IJAEMA), ISSN NO: 0886-9367, Volume XIV, Issue V, May/2022
22. Mr. Pathan Ahmed Khan, Dr. M.A Bari, Impact Of Emergence With Robotics At Educational Institution And Emerging Challenges", International Journal of Multidisciplinary Engineering in Current Research(IJMEC), ISSN: 2456-4265, Volume 6, Issue 12, December 2021,Page43-
23. Shahanawaj Ahamad, Mohammed Abdul Bari, *Big Data Processing Model for Smart City Design: A Systematic Review* ", VOL 2021: ISSUE 08 IS SN : 0011-9342 ;Design Engineering (Toronto) Elsevier SCI Oct
24. Ijteba Sultana, Mohd Abdul Bari and Sanjay," *Impact of Intermediate Bottleneck Nodes on the QoS Provision in Wireless Infrastructure less Networks*", Journal of Physics: Conference Series, Conf. Ser. 1998 012029 , CONSILIO Aug 2021