JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

# Prediction of Delay in Flights using Machine Learning Techniques: A Review

**[1]Maswood Alam, [2]Dr. Sandeep Dubey**

[1]Research Scholar, [2]Associate professor
Department of Computer Science and Engineering
SAM College of Engineering and Technology, Bhopal, India

***Abstract :*** Predicting flight delays is crucial for the aviation industry to improve operational efficiency and enhance passenger experience. Machine learning techniques have emerged as powerful tools for forecasting flight delays by leveraging historical data and various features. This review provides an overview of the prediction of delay in flights using machine learning techniques. The review highlights the importance of data quality in achieving accurate predictions. Comprehensive and reliable datasets, encompassing factors such as historical flight data, weather conditions, airport congestion, and aircraft information, are essential for robust models. Effective feature engineering is another crucial aspect, as it enables capturing relevant indicators such as departure/arrival time, airline, airport, weather conditions, previous delays, and holidays.

***Index Terms*** **– Machine Learning, Flight, Delay, Prediction.**

## I. INTRODUCTION

Predicting flight delays using machine learning techniques is an important and challenging problem in the aviation industry. By leveraging historical data and various features, machine learning models can provide valuable insights and predictions to help airlines, passengers, and other stakeholders anticipate and manage delays.

Prediction of delay in flights is a critical task in the aviation industry, aiming to forecast the likelihood and duration of delays for improved operational planning and passenger satisfaction. Several factors contribute to flight delays, including weather conditions, air traffic congestion, aircraft maintenance issues, and crew availability. Machine learning techniques have been widely employed to predict flight delays, leveraging historical data and relevant features.

To predict flight delays, various machine learning algorithms can be utilized, such as decision trees, random forests, support vector machines (SVM), gradient boosting, and neural networks. These algorithms analyze historical flight data and associated features to identify patterns and relationships that can be used to estimate the probability of delays. By considering features like departure/arrival time, airline, airport, weather conditions, historical delay patterns, and other contextual factors, machine learning models can provide accurate predictions.

Data quality and availability are crucial for accurate predictions. Comprehensive and up-to-date datasets that incorporate factors like historical flight records, weather data, and airport information are essential for training reliable models. Feature engineering plays a vital role in capturing the relevant information that affects flight delays. Domain knowledge and understanding of the aviation industry help in selecting and engineering meaningful features.

Model evaluation is an important step in assessing the performance of flight delay prediction models. Common evaluation metrics include accuracy, precision, recall, F1 score, and area under the receiver operating characteristic curve (AUC-ROC). These metrics provide insights into the model's ability to correctly classify delayed and non-delayed flights.

Real-time prediction of flight delays presents additional challenges due to the need for up-to-the-minute data and the dynamic nature of flight operations. Integration with live data feeds and continuous model updates are necessary to ensure accurate and timely predictions. Additionally, interpretability of the models is crucial for understanding the factors contributing to delays, allowing stakeholders to make informed decisions.

The successful implementation of flight delay prediction models using machine learning techniques can benefit various stakeholders, including airlines, airport authorities, and passengers. Airlines can optimize their operations by proactively managing delays, adjusting schedules, and allocating resources efficiently. Airports can anticipate congestion and optimize capacity utilization, while passengers can plan their travel better, reducing inconvenience and potential disruptions.

## II. BACKGROUND

S. Addu et al.,[1] Because of the rise of the aviation business, there has been an increase in the amount of air traffic, which has led to delays in flight times. Flight delays have repercussions not just in terms of the economy but also the environment. The job of monitoring and controlling air traffic is becoming ever more challenging. Flight delays may be caused by a number of different things, including safety concerns, technical failures, adverse weather conditions, crowded airports, and so on. In order to solve these issues, the authors of this study recommend implementing machine learning techniques such as Random Forest, Decision Tree, MLP Classifier, Naive Bayes, and KNN classifier.

B. T. L. S. S et al.,[2] a variety of machine learning and data analysis techniques in order to identify abnormalities concealed within the data. We used both supervised and unsupervised algorithms appropriate to the data set because of their strengths in assessing the aviation operational data, which is often unlabeled. The reason for this is because unsupervised algorithms are more likely to provide accurate results. We used the K means clustering technique, the K-Nearest Neighbors (KNN) algorithm, the support vector machine (SVM) program, and the XGBoost algorithm on the data in order to identify which model was the most effective in predicting the delay in the flight.

M. T. Vo et al.,[3] A flight delay is an unforeseen occurrence that may occur in the aviation industry specifically and the transportation industry in general. The ability of an airline to anticipate the potential of a flight delay or cancellation is critical for both the proactive scheduling of flights by the airline and the enhancement of the company's reputation among its customers. A real-time flight delay prediction system has been implemented, and this paper describes its findings. The whole system is created utilizing big data technologies to guarantee that it is both practical and scalable. The flight data is sent over Apache Kafka to trained machine learning models that are integrated inside Apache Spark to provide real-time prediction results.

R. A. Sugara et al.,[4] a flight delay prediction is modeled, and the modeling process is carried out with the help of the algorithms Decision Tree, Random Forest, Gradient Boosted Tree, and XGBoost Tree. In addition to using and merging the data on meteorological characteristics, this research has also utilized and combined the data on airport operating flights. Many different sample strategies were used in order to prepare for the unbalanced class.

M. Guimarães et al.,[5] analyze the elements that contribute additively to the expected result for each decision horizon, and we use historical data on flights and passengers to forecast missed aircraft connections in the hub airport of an airline. Our data is high-dimensional, diverse, unbalanced, and noisy; in addition, it does not provide any information on the arrival and departure times of passengers. Boosting, data balancing using Gaussian mixture models, and probabilistic encoding of categorical classes are some of the methods that we use.

S. Sharan et al.,[6] it seems that everyone is becoming more and more busy, which forces them to keep particularly close track of the time. Customers are often dissatisfied with the commercial aviation industry due in large part to the fact that flight delays are a common occurrence. Hence, accurate forecasting of flight delays is of critical importance to both the comfort of passengers and the reduction of the economic losses suffered by airlines.

C. Y. Yiu et al.,[7] During the course of the last several years, the global civil aviation sector has seen a period of fast development. Airports are experiencing challenges related to saturation as a direct result of the increased demand for air travel. It is anticipated that there will be a lot of traffic and a lengthy line of people waiting to take off and land. As a result, the issue of having an increasing number of flight delays has been amplified by the physical limits. But, if the delay continues to build up, it will be detrimental to both the operating efficiency and the reputation of the airport. It is also possible that there may be additional costs. In this study, many different methods of machine learning were used to forecast flight delays. These methods included the decision tree, the random forest, the k-nearest neighbor, the Naive Bayes algorithm, and artificial neural networks.

R. Hendrickx et al.,[8] presents a methodical strategy for dealing with unbalanced data in order to solve classification difficulties, taking into consideration the preferences of airport planners. A wide variety of practicable imbalance ratios, in addition to a number of classification algorithms and sampling strategies, are taken into consideration here. With consideration given to the applicable performance criteria, an ideal imbalance ratio has been determined. The methodology is shown by carrying out a binary classification of aircraft cancellations and delays at a major airport in Europe.

J. Huo et al.,[9] Making accurate forecasts of flight delays via the use of machine learning approaches is the primary focus of this endeavor. Using data taken from the Hong Kong International Airport, the outcomes of various different machine learning algorithms are compared with one another and subjected to an in-depth analysis. The aviation and insurance sectors might benefit significantly from the insights and suggestions that come from this investigation. By accurate prediction of flight delays, better design of the airport system may be achieved.

M. Bardach et al.,[10] the technologies that are now used by air traffic control do not make adequate use of the vast amount of data that is accessible for the purpose of the early identification of impending congestion and, as a consequence, flight delays. As a result, flight plans are not altered quickly enough in air traffic circumstances that have a significant potential for delay. In the work that we are doing, our goal is to determine the risk class of an air traffic scenario by basing it on the anticipated cost of the delays and taking into account information about the circumstances of the environment and the events that occur outside of the scenario.

## III. PROPOSED STRATEGY

The overall quality and effectiveness of a prediction model for flight delays depend on several factors, including the quality and quantity of the data used, the choice of features, the selection of appropriate machine learning algorithms, and the evaluation metrics employed. Here's a general review of prediction of delay in flights using machine learning techniques:

Data Quality: The quality and reliability of the input data play a crucial role in the accuracy of the predictions. Factors such as missing data, inconsistencies, or errors can impact the model's performance. It is essential to ensure that the dataset is comprehensive, up-to-date, and contains relevant information, such as historical flight data, weather conditions, airport congestion, and aircraft information.

Feature Selection: Effective feature engineering is critical to capture the important aspects that contribute to flight delays. Features like departure/arrival time, airline, airport, weather conditions, previous delays, and holidays can be relevant indicators. The selection of features should be based on domain knowledge and an understanding of the factors that influence flight delays.

Algorithm Selection: Various machine learning algorithms can be employed for flight delay prediction, including decision trees, random forests, and support vector machines (SVM), gradient boosting, and neural networks. The choice of algorithm depends on the specific characteristics of the data, the size of the dataset, and the trade-off between accuracy and computational complexity. Ensemble methods, such as random forests or gradient boosting, often yield good results due to their ability to handle complex relationships and handle noise in the data.

Model Evaluation: Evaluating the performance of the prediction model is crucial to assess its accuracy and generalization ability. Common evaluation metrics include accuracy, precision, recall, F1 score, and area under the receiver operating characteristic curve (AUC-ROC). Additionally, techniques like cross-validation and train-test splits can help estimate the model's performance on unseen data.

Interpretability: While accuracy is important, the interpretability of the model is also significant, especially in critical applications such as aviation. It is crucial to understand how the model arrives at its predictions and identify the key features driving the predictions. Decision tree-based algorithms provide interpretability through feature importance rankings, while complex models like neural networks may offer less interpretability.

## IV. CONCLUSION

The prediction of flight delays using machine learning techniques provides valuable insights to airlines, passengers, and stakeholders. By considering data quality, feature selection, algorithm choice, evaluation metrics, interpretability, and addressing industry-specific challenges, these models can significantly enhance operational efficiency and improve the travel experience for all involved parties.

## REFERENCES


1. S. Addu, P. R. Ambati, S. R. Kondakalla, H. Kunchakuri and M. Thottempudi, "Predicting Delay in Flights using Machine Learning," 2022 International Conference on Applied Artificial Intelligence and Computing (ICAAIC), Salem, India, 2022, pp. 374-379, doi: 10.1109/ICAAIC53929.2022.9793243.
2. B. T. L. S. S., H. Al Ali, A. A. A. M. Majid, O. A. A. A. Alhammadi, A. M. Y. M. Aljassmy and Z. Mukandavire, "Analysis of Flight Delay Data Using Different Machine Learning Algorithms," 2022 New Trends in Civil Aviation (NTCA), Prague, Czech Republic, 2022, pp. 57-62, doi: 10.23919/NTCA55899.2022.9934398.
3. M. -T. Vo, T. -V. Tran, D. -T. Pham and T. -H. Do, "A Practical Real-Time Flight Delay Prediction System using Big Data Technology," 2022 IEEE International Conference on Communication, Networks and Satellite (COMNETSAT), Solo, Indonesia, 2022, pp. 160-167, doi: 10.1109/COMNETSAT56033.2022.9994427.
4. R. A. Sugara and D. Purwitasari, "Flight Delay Prediction for Mitigation of Airport Commercial Revenue Losses Using Machine Learning on Imbalanced Dataset," 2022 International Conference on Computer Engineering, Network, and Intelligent Multimedia (CENIM), Surabaya, Indonesia, 2022, pp. 1-8, doi: 10.1109/CENIM56801.2022.10037369.
5. M. Guimarães, C. Soares and R. Ventura, "Decision Support Models for Predicting and Explaining Airport Passenger Connectivity From Data," in IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 9, pp. 16005-16015, Sept. 2022, doi: 10.1109/TITS.2022.3147155.
6. S. Sharan, M. Sriniketh, H. Vardhan and D. Jayanth, "State-Of-Art Machine Learning Techniques to Predict Airlines Delay," 2021 International Conference on Forensics, Analytics, Big Data, Security (FABS), Bengaluru, India, 2021, pp. 1-6, doi: 10.1109/FABS52071.2021.9702590.
7. C. Y. Yiu, K. K. H. Ng, K. C. Kwok, W. Tung Lee and H. T. Mo, "Flight delay predictions and the study of its causal factors using machine learning algorithms," 2021 IEEE 3rd International Conference on Civil Aviation Safety and Information Technology (ICCASIT), Changsha, China, 2021, pp. 179-183, doi: 10.1109/ICCASIT53235.2021.9633571.
8. R. Hendrickx, M. Zoutendijk, M. Mitici and J. Schäfer, "Considering Airport Planners' Preferences and Imbalanced Datasets when Predicting Flight Delays and Cancellations," 2021 IEEE/AIAA 40th Digital Avionics Systems Conference (DASC), San Antonio, TX, USA, 2021, pp. 1-10, doi: 10.1109/DASC52595.2021.9594367.

9. J. Huo, K. L. Keung, C. K. M. Lee, K. K. H. Ng and K. C. Li, "The Prediction of Flight Delay: Big Data-driven Machine Learning Approach," 2020 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), Singapore, Singapore, 2020, pp. 190-194, doi: 10.1109/IEEM45057.2020.9309919.
10. M. Bardach, E. Gringinger, M. Schrefl and C. G. Schuetz, "Predicting Flight Delay Risk Using a Random Forest Classifier Based on Air Traffic Scenarios and Environmental Conditions," 2020 AIAA/IEEE 39th Digital Avionics Systems Conference (DASC), San Antonio, TX, USA, 2020, pp. 1-8, doi: 10.1109/DASC50938.2020.9256474.