As per UGC guidelines an electronic bar code is provided to seure your paper

International Journal for Modern Trends in Science and Technology
Volume 9, Issue 11, pages 01-04.
ISSN: 2455-3778 online
Available online at: http://www.ijmtst.com/vol9issue10.html
DOI: https://doi.org/10.46501/IJMTST0911001

Flight Delay Prediction using Aviation Big Data and Machine Learning

Sahil Khalkar[1] | Rushikesh Nimbhore[1] | Atharva Pardeshi[1] | Sanket Kanade[1] | V.K.Barbudhe[2]

[1]Department of Information Technology, Sandip Institute of Technology and Research Centre, Nashik, Maharashtra, India
[2]Assistant Professor, Department of Information Technology, Sandip Institute of Technology and Research Centre, Nashik, Maharashtra, India

To Cite this Article
Sahil Khalkar, Rushikesh Nimbhore, Atharva Pardeshi, Sanket Kanade and V.K.Barbudhe. IOT Based Data Monitoring in Secured Block Chain Architecture, International Journal for Modern Trends in Science and Technology, 2023, 9(11), pages. 01-04.https://doi.org/10.46501/IJMTST0911001

Article Info
Received: 15 October 2023; Accepted: 07 November 2023; Published: 10 November 2023.

ABSTRACT

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 incomes of airline agencies. There have been many researches on modeling 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. This paper proposes a model for predicting flight delay based on Machine Learning (ML). ML is one of the newest methods employed in solving problems with high level of complexity and massive amount of data. Moreover, ML is capable to automatically extract the important features from data. Furthermore, due to the fact that most of flight delay data are noisy, a technique based on stack denoising autoencoder is designed and added to the proposed model. Also, Various algorithms like Random Forest, Decision Tree, MLP Classifier are applied to find weight and bias proper values, and finally the output has been optimized to produce high accurate results.

KEYWORDS: Flight delay, machine learning, random forest, decision tree, MLP classifier.

1. INTRODUCTION

Flight delay prediction is a critical area of research in aviation management. Delays in the aviation industry can result from various factors, including weather conditions, air traffic congestion, technical issues, and scheduling problems. Early research focused on deterministic models, but with the advent of big data and machine learning, the focus shifted towards data-driven predictive models. As the air travels have a significant role in economy of agencies and airports, it is necessary for them to increase quality of their services. One of the important modern life challenges of airports and airline agencies is flight delay. Delay in flight is inevitable, which has too much negative economic effects on passengers, agencies and airport. Furthermore, delay can damage the environment through fuel consumption increment and also leads to emission of pollutant gases. In addition, the delay affects the trade, because goods' transport is highly dependent to customer trust, which can increase or decrease the ticket sales, so that on time flight leads to customer confidence. So that, flight prediction can cause a skillful decision and operation for agencies and airports, and also a good passenger information system can relatively satisfy the customer. In this paper, we performed an aviation data analytic and apply machine learning techniques to realistic aviation dataset for flight arrival delay prediction. Instead of blindly testing machine learning models, we leverage the strength of data visualization to discover potential patterns of flight delay for a better understanding of the explored data and reasonable factors selection before building prediction models

2. Literature Review

A. Traditional Approaches to Flight Delay Prediction:

Early studies in flight delay prediction utilized statistical methods and simple regression models. These approaches, while informative, lacked the accuracy and complexity to handle the diverse and dynamic factors contributing to flight delays. Some other research papers proposed a regression-based model considering historical delays and weather conditions but faced limitations in handling non-linear relationships. Several researchers have emphasized the importance of historical data analysis in flight delay forecasting. By examining past flight records, the researchers identified delay patterns, trends and causes. Factors such as weather, congestion, and flight-specific data have been extensively studied to develop predictive models. [1]

B. Statistical analysis

Government agencies have invested in econometric models that incorporate correlation analysis, both parametric and non-parametric tests, multivariate analysis, and various statistical techniques. These models are utilized to comprehend the connections between factors such as delay, passenger demand, fare, and aircraft size. [2]

C. Probabilistic model and Machine Learning Techniques

A probabilistic model utilizes analysis tools to calculate the likelihood of an event occurring, relying on historical data. The model provides an estimated result in the form of a probability distribution function. The element of randomness significantly influences the decisions or outcomes generated by the probabilistic model.

Machine learning algorithms, particularly those based on artificial neural networks, decision trees, and support vector machines, have gained prominence in flight delay prediction. Researchers have explored the application of these techniques to effectively capture complex relationships among various influencing factors. Deep learning models, such as recurrent neural networks (RNNs) and long short-term memory networks (LSTMs), have also been employed to handle sequential data and improve prediction accuracy. [3]

D. Real-time data integration

Integrating real-time data sources including flight data, airport crowd data, and even social media data has been a research interest By integrating these dynamic data, researchers aim to increase the timeliness and accuracy of delay forecasting, enabling airlines to respond quickly to probability concerns [4]

E. Hybrid models and cluster methods

To strengthen the forecasts, researchers have developed hybrid models that combine the strengths of different forecasting approaches. Ensemble techniques such as bagging and boosting have been used to create an integrated model that outperforms individual algorithms. These hybrid methods aim to reduce the limitations of specific methods and increase the overall prediction performance. [5]

3. Methodology

1. The primary task of researchers and analysts is to identify the causes of flight delays.
2. Particular attention was paid to the demands of air travel and the cyclical changes in the weather at this particular airport.
3. The motivation of the study is to propose a method to improve the performance of the model without interfering with or affecting the planned costs.
4. Developed a data mining model that can predict flight delays based on weather conditions. WEKA and R were used to model, identify classifiers and select those that gave the best results. Machine learning techniques such as Random forest, K-Nearest neighbor, and multilayer perceptron Analysis were used.

5. Focused on eliminating the impact of data imbalances in the data learning process. Methods such as decision trees, Ada boost, and K-nearest neighbors were used to predict individual flight delays. The model performed binary classification to identify scheduled flight delays.
6. Developed a Detailed Policy Assessment Tool (DPAT) to promote minor changes to flight delays due to weather.
7. He conducted emotional analysisand psychoanalysis, analyzing the minds and thoughts of people and studying their behavior. The result of the analysis is a collection of symptom-based concepts, also known as perceptual classification.

A. Random Forest Algorithm

Random Forest is an ensemble learning method that combines multiple decision trees to make predictions. It is effective for handling large datasets with numerous features and capturing complex relationships within the data. Random forest is a popular machine learning algorithm that includes supervised learning methods. It can be used for Classification and Regression problems in ML. It is based on the concept of group learning, which is the process of combining multiple classifiers to solve a complex problem and improve the performance of the model. Random forest is a Supervised Machine Learning Algorithm that is used widely in Classification and Regression problems. It builds decision trees on different samples and takes their majority vote for classification and average in case of regression.

B. Decision Tree Algorithm

Decision trees are versatile algorithms that can handle classification and regression tasks. Predictions are made by dividing the data set into smaller units based on different features. Decision trees are easy to interpret, making them useful for gaining insights into flight delay factors. Decision trees (DTs) are a supervised nonparametric learning method used for classification and regression. The goal is to build a model that predicts the value of an objective variable by learning simple decision rules from different data segments.

C. K-Nearest Neighbor

K-Nearest Neighbors (KNN) is a versatile machine learning algorithm that can be used for both classification and regression tasks. In the context of flight delay prediction, KNN can be applied as a regression algorithm to predict continuous numerical values, such as the delay time of flights, based on relevant input features from aviation big data.

- Train the KNN regression model on the preprocessed aviation big data. The algorithm calculates the distance between the input data point and its K nearest neighbors in the feature space.

- When making predictions for a new data point, the algorithm averages the target values of its K nearest neighbors to predict the flight delay time.

D. Multilayer Perceptron

Multilayer Perceptron (MLP) is a type of artificial neural network that can be used for flight delay prediction using aviation big data and machine learning. MLPs are particularly powerful for capturing complex, nonlinear relationships in the data.

-Split the preprocessed data into training and testing sets.

-Train the MLP model on the training data using backpropagation and an optimization algorithm like stochastic gradient descent (SGD) or Adam optimizer.

-Monitor the model's performance on the validation set during training to prevent overfitting. You can use metrics like Mean Absolute Error (MAE) or Root Mean Squared Error (RMSE) to assess the model's accuracy.

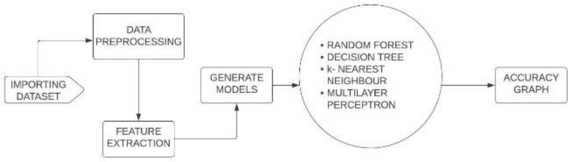

Figure 1: Architecture of System

4. Conclusion

After using these two models to predict whether the flight should be delayed and by how much it is expected to be delayed, we found that the following aspects are important: week, month, flight of use, scheduled time (within the flight duration), both departure and departure point. destination Distance to destination, scheduled flight departure time, departure airport code and vehicle Check-in and check-out time. Using our model, it is possible to predict whether the flight will be delayed and, more importantly, how much it is expected to be delayed, based on the data collected. However, our model has some limitations, firstly, due to our power calculation, our model only includes one year of data, and making predictions will be easier than several years of data. In addition, some important information such as aircraft type is not included, as is information about specific weather conditions at the airport. Therefore, researchers may try to collect more data and use better computing power to build better models. This paper presents a method to estimate the total flight delay at an airport by analyzing the learning process. This way, we can predict the delay of new flights without needing months of data to build a forecast model. The next step will be to extend the model to international flights, or at least leverage more data to create more accurate predictions. Finally, the most interesting step will be the integration of these models into the flight booking tool to provide passengers with an estimate of future delay; however, given the impact this will have on the booking, confidence must be placed in the information provided.

Acknowledgement

We are very grateful to Prof. V. Barbudhe (Guide and Assistant Professor, Department of Information Technology, Sandip Institute of Technology and Research Centre, Nashik) For his expert guidance and continuous encouragement throughout the project.

At last, we must express our sincere, heartfelt gratitude to all staff members and students of the Department of Information technology who helped us directly or indirectly during this course of work.

Conflict of interest statement

Authors declare that they do not have any conflict of interest.

References

[1] Chakrabarty N. A data mining approach to flight arrival delay prediction for American airlines. In 2019 9th Annual Information Technology, Electromechanical Engineering and Microelectronics Conference (IEMECON). New York: IEEE; 2019.

[2] Mrs. Yogita Borse, Dhruvin Jain, Shreyash Sharma, Viral Vora, Aakash Zaveri, Assistant Professor, Department of Information Technology, K.J Somaiya College of Engineering, Mumbai, India. Flight Delay Prediction System in International Journal of Engineering Research & Technology (IJERT) Vol. 9 Issue 03, March-2020

[3] Carvalho, A. Sternberg, L. Maia Goncalves, A. Beatriz Cruz, J.A. Soares, D. Brandao, D. Car- ˜ valho, e E. Ogasawara, 2020, On the relevance of data science for flight delay research: a systematic review, Transport Reviews.

[4] Yuemin Tang. 2021. Airline Flight Delay Prediction Using Machine Learning Models. In 2021 5th International Conference on E-Business and Internet (ICEBI 2021), October 15- 17, 2021, Singapore, Singapore. ACM, New York, NY, USA, 7 Page

[5] Vishrut Raj, Viran Raj, Satyam Singh, Adityanath Mishra, Rajkumar Goel Institute of Technology/AKTU – "Flight Delay Prediction" in IJIRT | Volume 8 Issue 1 | ISSN: 2349-6002, June 2021.

[6] F. Azadian, A. E. Murat, and R. B. Chinnam. Dynamic routing of time-sensitive air cargo using real-time information. Transportation Research Part E: Logistics and Transportation Review, 48(1):355–372, Jan. 2012. ISSN 1366-5545.

[7] E. Balaban, I. Roychoudhury, L. Spirkovska, S. Sankararaman, C. Kulkarni, and T. Arnon. Dynamic routing of aircraft in the presence of adverse weather using a POMDP framework. In 17th AIAA Aviation Technology, Integration, and Operations Conference, 2017, 2017.

[8] Gopalakrishnan K, Balakrishnan H. A comparative analysis of models for predicting delays in air traffic networks. ATM Seminar. 2017.

[9] U.S. Department of Transportation Bureau of Transportation Statistics. Airline On-Time Statistics. https://www.transtats.bts.gov/ONTIME/. 2019.

[10] Saadat MN, Moniruzzaman M. Enhancing airlines delay prediction by implementing classification based deep learning algorithms. In International Conference on Ubiquitous Information Management and Communication. Berlin: Springer; 2019

[11] Yu B, et al. Flight delay prediction for commercial air transport: A deep learning approach. Transport Part E Logits Transport Rev. 2019; 125:203–21.

[12] Vandal T, et al. Prediction and uncertainty quantification of daily airport fight delays. In International Conference on Predictive Applications and APIs. 2018.

[13] Esmaeilzadeh E, Mokhtarimousavi S. Machine learning approach for fight departure delay prediction and analysis. Transport Res Rec, 2020.