IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 2, February 2023
Impact Factor: 7.301

Flight Delay Prediction by Machine Learning

Prof. Swati Dhadake[1], Tuljai Kadam[2], Amanoddin Shaikh[3], Sumit Sabale[4], Bhagyashri Shinde[5]
Faculty, Department of Computer Engineering[1]
Students, Department of Computer Engineering[2,3,4,5]
Smt. Kashibai Navale College Of Engineering, Pune, Maharashtra, India

Abstract: *Nowadays, the aviation industry plays a crucial role in the world's transportation sector, and a lot of businesses rely on various airlines to connect them with other parts of the world. But, extreme weather conditions may directly affect the airline services by means of flight delays. To solve this issue, accurately predicting these flight delays allows passengers to be well prepared for the deterrent caused to their journey and enables airlines to respond to the potential causes of the flight delays in advance to diminish the negative impact. The purpose of this project is to build a web app powered by Machine Learning Algorithm to predict flight delays. We look upon implementing Random Forest Regression algorithm for the prediction model. Also the main idea behind it being a web app is to enable the public interactions with the platform for retrieving the predicted delays for their flights.*

Keywords: Flight delay prediction, machine learning, random forest, web app

I. INTRODUCTION

In the modern world, the key components of any transportation system are the passenger planes, cargo planes and air traffic control systems. Since, countries around the world have tried to develop numerous ways to improve the air travel system. This radically changed airline 's operations. Flight delays periodically annoy Hyundai passengers. An estimated 20% of airline flights are canceled or delayed each year, costing 4,444 passengers more than $20 billion in cost and time.

II. PROPOSED SYSTEM

The proposed system for predicting flight delays using machine learning can be based on the following steps: Input Data: Input of data from the user through the web application, including details such as airline, departure location, arrival location, and departure date. Preprocessing: The system retrieves data for the highest delay factor for both origin and destination on a given date. Processing: Data is passed to a machine learning model for processing (using a random forest regression algorithm). Output: Returns the predicted output in HH: MM format.

III. MOTIVATION

The aviation industry plays a crucial role in the world's transportation sector, and a lot of businesses rely on various airlines to connect them with other parts of the world. But, extreme weather conditions may directly affect the airline services by means of flight delays. To solve this issue, accurately predicting these flight delays allows passengers to be well prepared for the deterrent caused to their journey and enables airlines to respond to the potential causes of the flight delays in advance to diminish the negative impact. We are making an attempt to provide a prototype as a solution for the same. Average flight delay is regularly referred to as an indication of airport capacity. Flight delays are common problems worldwide. It is very difficult to explain the reason for the delay.

IV. OBJECTIVE

Increase your flight delay forecast period. Create an easy-to-use web application that allows Internet users to access the predictor. Use the current weather data as input to the machine learning model. Allow users to retrieve expected delays for their flights a week prior to date of departure. The purpose of flight delay prediction is to accurately determine flight delays at airports. This information is used for adaptive cruise control and departure warning.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 2, February 2023
Impact Factor: 7.301

V. LITERATURE REVIEW

1. The main task of the researchers and analysts is to predict the cause of flight's delay. Direct domestic flights from Orlando International Airport.
2. They primarily focused on air travel demand and cyclical changes in the weather at this particular airport.
3. Research motivation is to propose an approach that will improve the performance of the model without interfering with or impacting the planned cost.
4. Created a data mining model that allows to predict flight delays by observing weather conditions. They built a model using WEKA and R, choosing different classifiers and choosing the one that gave the best results. They used various machine learning techniques such as Naive Bayes Analysis and the Linear Discriminant Analysis classifier.
5. Focused on overcoming the effect of data imbalance in the data learning process. They used methods such as decision trees, Ada Boost, and K-Nearest Neighbors to predict individual flight delays. Binary classification was performed by the model to predict planned flight delays.
6. Developed the Detailed Policy Assessment Tool (DPAT) used to encourage minor variations in flight delays due to weather.
7. Conducted sentiment analysis and opinion research, during which he analyzed opinions, people's moods and studied their behavior. The research result is a summary of opinions based on characteristics, which is also called sentiment classification

VI. SYSTEM ANALYSIS

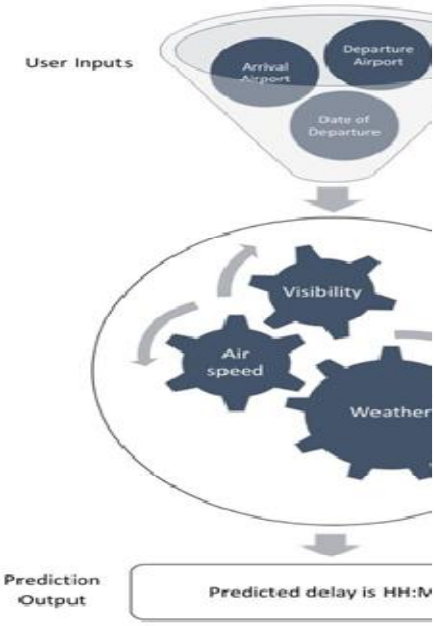

Figure: System Architecture

VII. ALGORITHM

7.1 Random Forest Regression

Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model.

As the name suggests, "Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset." Instead of relying on onedecision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 2, February 2023
Impact Factor: 7.301

VIII. MATHEMATICAL MODEL

We first give the definition of weather vector w and flight schedule vector s as follows:

w = [w1,w2,w3,w4,w5,w6]

where w1=visibility at departure airport,
w2=visibility at destination airport,
w3=wind direction at departure airport,
w4=wind direction at destination airport,
w5=wind speed at departure airport and
w6=wind speed at destination airport.

s = [s1,s2]

where s1=scheduled time of departure and
s2=scheduled time of arrival.

Input vector is given by,

I = [d, p1, p2]

where d=date of departure,
p1=departure point and
p2=destination point

Processing of the input vector (I) returns the vectors w and s. These vectors are used in the input vector of the ML model. The ML model input vector (x) is given by,

x = [w,s,n]

where w=weather vector,
s=schedule vector and
n=airline ID

On processing returns y,

y = {model, x}
y = [minutes]
y is further represented in the HH: MM format at the user-end

IX. CONCLUSION

In this paper, random forest-based and architectures have been implemented to predict individual flight delay. The experimental results show that the random forest based method can obtain good performance for the binary classification task and there are still room for improving the multi-categories classification tasks. In summary, the random forest-based architecture presented better adaptation at a cost of the training accuracy when handling the limited dataset. The future work would include training the model on latest aviation data for the selected airport. Testing the model on live data fed from aviation data servers. Inclusion of other factors responsible for delays in evaluating the expected delays.

ACKNOWLEDGMENT

Our sincere thanks to Prof. Swati Dhadake for her invaluable guidance and support throughout the course. We also thank the Department of Computer Engineering, Smt. Kashibai Navale College of Engineering, Pune for it's support and provision of the necessary resources to complete this project. We thank to everyone who contributed to the success of this project.

REFERENCES

[1]. M. Leonardi, "Intrusion Detection with Ads-b Anomaly and Sensor Surveillance Tracking," IEEE Trans. Aerosp. former. system, in press, doi:10.1109/TAES.2018.2886616.

[2]. Yu. A. Nijsour, G. Caddum, G. Gagnon, F. Gagnon, K. Yuen and R. Ma hapatra, "Adaptive air-to-ground communication systems based on ads-b and global multilateration", IEEE Trans. Milestone Technologies, Vol.65,5, p. 3150–3165, 2015.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 2, February 2023
Impact Factor: 7.301

[3]. J.A. F. Zuluaga, J.F.V. Bonya, J.D. O. Pabon and K. M. S. Rios, "Radar Error Calculation and Correction System Based on ads-b and Smart Business Tools", Proc. International Qanahan Conference. security technology., seconds. 1–5, IEEE, 2018.

[4]. D.A. Pamplona, L. Weigang, A.G. (1999). Wright, E.H. Shiguemori, и c. J.P. Alves, "Supervised Neural Networks with Multi-Input Layers for Air Traffic Delay Prediction", Proc. International Jt. Meeting Neural Networks , pp. 1–6, IEEE, 2018.

[5]. S. Manna, S. Biswas, R. Kundu, S. Rakshit, P. Gupta, and S. Barman, "Statistical Approach to Flight Delay Prediction Using the Gradient Boosted Decision Tree", Proc. International Conference Intel Account. data science , str. 101-1 1–5, IEEE, 2017.

[6]. L. Moreira, C. Dantas, L.D., et al. Oliveira, J. Soares and E. Ogasawara, "A Evaluation of Preprocessing Techniques in Machine Learning Models for Flight Delay", Proc. International Jt. Meeting Neural Networks, pp. 1–8, IEEE, 2018.

[7]. J. J. Rebollo and H. Balakrishnan, "Characterization and Prediction of Air Traffic Delay", Transp. Resolution Part C Technology, Vol. 44, s. 231–241, 2014.

[8]. L. Hao, M. Hansen, Y. Zhang and J. Post, "New York, New York: Two Methods to Assess the Impact of New York Airport Delays", Transp. Part of resolution ELogist. Transportation Obr., Vol. 70, p. 245–260, 2014.

[9]. ANAC, “National Civil Aviation Authority of Brazil”. anac.gov, 2017. [Online] Available at: http://www.anac.gov.br/.

[10]. S. Zhang, X. Li, M.Zong, X. Zhu and R. Wang, "Efficient knnclassification with different number of nearest neighbors", IEEE Trans.Neural Networks. study.sist., vol. 29, no. 5, p. 1774–1785, 2017.

BIOGRAPHY

- Tuljai Kadam is currently pursuing Bachelor’s Degree in Computer Engineering at Smt. Kashibai Navale College of Engineering, Pune.
- Amanoddin Shaikh is currently pursuing Bachelor’s Degree in Computer Engineering at Smt. Kashibai Navale College of Engineering, Pune.
- Sumit Sabale is currently pursuing Bachelor’s Degree in Computer Engineering at Smt. Kashibai Navale College of Engineering, Pune.
- Bhagyashri Shinde is currently pursuing Bachelor’s Degree in Computer Engineering at Smt. Kashibai Navale College of Engineering, Pune.