Developing a Flight Delay Prediction Model using Machine Learning

Introduction

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

Literature Survey

[1] Study of flight Departure Delay and Casual Factors Using Spatial Analysis

Assuming delay as a spatially dependent variable, finds delay distribution pattern to predict delay.

Advantages: Considers spatial factors, people, day types and time ranges of a day to contribute to the prediction.

Disadvantages: Some of the attributes considered cannot be obtained on large scale in real time.

[2] Flight delay forecasting and analysis of direct and indirect factors

LSTM network with attention mechanism to remember spatial dependencies.

Advantages: Direct and Indirect causing factors are weighted differently.

Disadvantages: Air interaction of flights not taken into account.

[3] Flight delay prediction based on aviation big data and machine learning

Comparision of LSTM and Random forest; Uses ADS-B data for improved accuracy.

Advantages: Use of ADS-B can be seen promising. Showed that LSTM suffers from overfitting on test set.

Disadvantages: Deployment of ADS-B is hectic. More data handing takes place.

[4] Prediction of weather-induced airline delays based on machine learning algorithms

Experimented predicting delay using supervised machine learning algorithms. Uses SMOTE for weaker class sampling.

Advantages: Found weather causes to be amounting to a significant percent of delay.

Disadvantages: Not taken spatial dependencies into account. Amount of delay could have been found.

[5] Flight delay prediction from spatial and temporal perspective

ST-Random Forest for flight delay prediction using spatial features of aviation network and temporal correlation of weather condition and airport crowdedness on flight delays.

Advantages: A real-time, highly accurate prediction system that guaranties the influence of the air traffic network in the prediction.

Disadvantages: Overfitting might occur due to LSTM.

[6] Airline Flight Delay Prediction Using Machine Learning Models

Comparison among 7 classification machine learning algorithms.

Advantages: Among the considered alongs using 4 performance indicators decision tree was found to be the best in predicting flight delays.

Disadvantages: The data imbalance issue even through handled through weighted evaluation methods does had a significant effect on performance on the algorithms.

[7]Predicting flight delay based on multiple linear regression

A multiple linear regression algorithm to predict delay.

Advantages: Both airline and weather features are taken into consideration. The methodology used in this gives better results compared to Naïve-Bayes and C4.5 approach.

Disadvantages: Predicts only the flights which are delayed above 30 minutes.

[8]Flight Delay Prediction System

Supervised Machine Learning algorithm using Naïve Bayes.

Advantages: Considers independence among the predictors making the system scalable. Good for real time prediction.

Disadvantages: Does not take into account the impact of unprecedented reasons such as major calamities in flight delays.

[9]A Deep Learning approach to flight delay prediction

A deep RNN and LSTM approach to prediction; uses limited data attributes

Advantages: Predicting two sections namely day prediction and flight prediction seems more reasonable and can give more insights for the airport managers to make necessary arrangements.

Disadvantages: Air traffic/flight interaction doesn't play great roles. Biased towards weather attributes.

References

1. Shaowu Cheng, Yaping Zhang, Siqi IIao, Ruiwei Liu, Xiao Luo, Qian Luo, "Study of Flight Departure Delay and Casual Factors Using Spatial Analysis", Journal of Advanced Transportation, vol.2019, Article ID 3525912, 11 pages, 2019. https://doi.org/10.1155/2019/3525912
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). https://doi.org/10.1049/itr2.12183
3. Gui , G., Liu, F., Sun, J., Zhou, X., Flight delay prediction based on aviation big data and machine learning. *IEEE Transactions on Vehicular Technology, 69(1), 140-150.*

4. Choi, S., Kim, Y. J., Briccno, S., & Mavris, D.(2016, September). Prediction of weather-induced airline delays based on machine learning algorithms. *In 2016 IEEE/AIAA 35th Digital Avionics System Conference (DASC) (pp. 1-6). IEEE.*

5. Li, Q., & Jing, R. (2022). *Flight Delay Prediction from Spatial and Temporal Perspective. Expert Systems with Applications, 117662.*

6. 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).*

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. Borse, Y., Jain, D., Sharma, S., Vora, V.,& Zaveri, A. (2020). Flight Delay Prediction System. *Int. J. Eng. Res. Techno, 9(3), 88-92.*

9. 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 System Conference (DASC) (pp. 1-6). IEEE.*