LITERATURE REVIEW OF DEVELOPING A FLIGHT DELAY PREDICTION MODEL USING MACHINE LEARNING

INTRODUCTION

Airports are significant nodes of air transportation. The number of airport flight delays has been on increase in recent years. Delayed flights are defined by the Federal Aviation Administration when they arrive or depart more than 15 minutes later than scheduled. In 2019, the arrival delay rate is 19.2% and the departure delay rate is 18.18% in the United States. Flight delays can cause many negative effects, such as passengers' inconvenience, increased airport pressure, and airline losses. Effective flight delay prediction could provide support for flight plan and emergency plan formulation, reduce the economic loss, and alleviate the negative impact (1). Hence, a delay prediction turns out very important. International Civil Aviation Organization (ICAO) has enabled a program called Air Traffic Flow Management (ATFM) with the objective of ensuring that the traffic volume is compatible with the capacities declared by aviation authorities in order to reduce ground and en-route delays. Another application of traffic management is the Free Route Airspace (FRA) concept which consists of using the shortest routes possible in order to reduce flight time, CO_2 emissions, and fuel waste. Moreover, several other models have been developed to solve delays problem based on probability, statistics, graph and network representations, operational research studies, and so on (2). The Related Work gives the recent works and research on this topic.

RELATED WORKS

The principle of the Stacking classification algorithm is introduced, the SMOTE algorithm is selected to process imbalanced datasets, and the Boruta algorithm is utilized for feature selection. There are five supervised machine learning algorithms in the first-level learner of Stacking including KNN, Random Forest, Logistic Regression, Decision Tree, and Gaussian Naive Bayes. The second-level learner is Logistic Regression. To verify the effectiveness of the proposed method, comparative experiments are carried out based on Boston Logan International Airport flight datasets from January to December 2019. Multiple indexes are used to comprehensively evaluate the prediction results, such as Accuracy, Precision, Recall, F1 Score, ROC curve, and AUC Score. The results show that the Stacking algorithm not only could improve the prediction accuracy but also maintains great stability (1). Multi Layer Perceptron (MLP) to train and test data applied. The neural network MLP was able to predict flight arrival delay with a coefficient of determination $R2$ of 0.9048, and the selective procedure achieved a time saving and a better $R2$ score of 0.9560. To enhance the reliability of the proposed method, the performance of the MLP was compared with that of Gradient Boosting (GB) and Decision Trees (DT). The result is that the MLP outperformed all existing benchmark methods (2). Here present the first datadriven systemic study of air transport delays in China, of their evolution and causes, based on 11 million flights between 2016 and 2018. A significant fraction of the delays can be explained by a few variables, e.g., weather conditions and traffic levels, the most important factors being the presence of thunderstorms and the season of the year. Remaining delays can often be explained by en-route weather phenomena or by reactionary delays. This study contributes towards a better understanding of delays and their prediction through a data-driven methodology, leveraging on statistics and data mining concepts (3). Recognize useful patterns of the flight delay from aviation data and perform accurate delay prediction. The best result for flight delay prediction (five classes) using machine learning models is 89.07% (Multilayer Perceptron). A Convolution neural network model is also built which is enlightened by the idea of pattern recognition and success of neural network method, showing a slightly better result with 89.32% prediction accuracy (4). Explores a

broader scope of factors which may potentially influence the flight delay, and compares several machine learningbased models in designed generalized flight delay prediction tasks. Compared with the previous schemes, the proposed random forest-based model can obtain higher prediction accuracy (90.2% for the binary classification) and can overcome the overfitting problem (5). The model demonstrated to reduce by 30% the take-off time prediction errors of the current system one hour before the time that flight is scheduled to depart from the parking position and presents an extension of the model, which overcomes this look-ahead time constraint and allows to improve take-off time predictions as early as the initial flight plan is received. In addition, a subset of the original set of input features has been meticulously selected to facilitate the implementation of the solution in an operational air traffic flow and capacity management system, while minimising the loss of predictive power. Finally, the importance and interactions of the input features are thoroughly analysed with additive feature attribution methods (6). The designed prediction tasks contain different classification tasks and a regression task. Experimental results show that long short-term memory (LSTM) is capable of handling the obtained aviation sequence data, but overfitting problem occurs in our limited dataset. Compared with the previous schemes, the proposed random forest-based model can obtain higher prediction accuracy (90.2% for the binary classification) and can overcome the overfitting problem (7).

References

1. A Multilayer Perceptron Neural Network with Selective-Data. **Hajar Alla, Lahcen Moumoun , and Youssef Balouki**. [ed.] Jianping Gou. Settat, Morocco : Hindawi, 2021. Hindawi Scientific Programming. p. 12.

2. Flight Delay Classification Prediction Based on. **Jia Yi, 1 Honghai Zhang** ,. [ed.] Chi-Hua Chen. Honghai Zhang; zhh0913@163.com : Wiley, 2021. p. 10.

3. FLIGHT DELAY PREDICTION USING MACHINE LEARNING. **Sarah Ajmeria, Srushti V,Prof. Kavitha S Patil.** Bangalore, India : IJIREEICE, 2022. DOI: 10.17148/IJIREEICE.2022.10584. p. 5.

4. Applying Machine Learning to Aviation Big Data. **Yushan Jiang, Yongxin Liu,Dahai Liu,Houbing Song**. 2020. 2020 IEEE Intl Conf on Dependable,Intl Conf on Cloud and Big Data Computing. p. 8.

5. Characterization and Prediction of Air Transport. **Massimiliano Zanin 1, * , Yanbo Zhu 2,3, Ran Yan 3.** 2020, MDPI Journals, p. 15.

6. An explainable machine learning approach to improve take-off time. **Ramon Dalmau, Franck Ballerini,Herbert Naessens,Seddik Belkoura**. 2021, Journal of Air Transport Management, p. 12.

7. Flight Delay Prediction Based on Aviation Big Data. **Guan Gui, Senior Member, IEEE, Fan Liu, Student Member, IEEE, Jinlong Sun, Member, IEEE,**. 2020, IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, VOL. XX, NO. XX, XXX 2015, p. 11.