INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN2454-9940

www.ijasem.org

Vol 18, Issue 2, 2024

Flight Delay Prediction Based on Aviation Big Data and Machine Learning

[1] P. SRINIVASA REDDY, [2] CHIRLA LAKSHMI HIMAJA

[1](Assistant Professor), MCA, S.V.K.P & Dr K.S. Raju Arts & Science College

[2]MCA, scholar, S.V.K.P & Dr K.S. Raju Arts & Science College

ABSTRACT

Flight Planning is one of the demanding situations in commercial world, which faces many unsure conditions. There is such condition in delay occurrence, which stems from various factors and imposes considerable costs on airlines, operators, and travellers. Delays in departure can occur due to bad weather conditions, seasonal and holiday demands, airline policies, technical problems such as problems with airport facilities, baggage handling and mechanical equipment, and the accumulation of delays from previous flights. In This flight delay prediction system based on the Aviation Data, which can result in delays. The system considers the of various parameters. Random Forest (RF), K-Nearest Neighbour (KNN) and Support Vector Machine (SVM) are the algorithm used in this system

1.INTRODUCTION

SCOPE:

AIR traffic load has experienced rapid growth in recent years, which brings increasing demands for air traffic surveillance system. Traditional surveillance technology such as primary surveillance radar (PSR) and secondary surveillance radar (SSR) cannot meet requirements of the future dense air traffic.

Therefore, new technologies such as automatic dependent surveillance broadcast (ADS-B) have been proposed, where flights can periodically broadcast their current state information, such as international civil aviation organization (ICAO) identity number, longitude, latitude and speed [1]. Compared with the traditional radar-based schemes, the ADSB- based scheme is low cost, and the corresponding ADS-B receiver (at 1090 MHz or 978 MHz) can be easily connected to personal computers [2]. The received ADS-B message along with other collected data from the Internet can constitute ahuge volumes of aviation data by which data mining can support military, agricultural, and commercial applications.

In the field of civil aviation, the ADS-B can be used to increase precision of aircraft positioning and the reliability of air traffic management (ATM) system [3]. For example, malicious or fake messages can be detected with the use of multi alteration (MLAT) [1], allowing open, free, and secure visibility to all the aircrafts within airspace [2]. Thus, the ADS-B provides opportunity to improve the accuracy of flight delay prediction which contains great commercial value. The flight delay is defined as a flight took off or arrived later than the scheduled time, which occurs in most airlines around the world, costing enormous economic losses for airline company, and bringing huge inconvenience for passenger. According to civil aviation administration of China (CAAC), 47.46% of the delays are caused by severe weather, and 21.14% of the delays are caused by air route problems. Due to the own problem of airline company or technical problems, air traffic control and other reasons account for 2.31% and 29.09%, respectively. Recent studies have been focused on finding a suitable way to predict probability of flight delay or delay time to better apply air traffic flow management (ATFM) [4] to reduce the delay level.

Classification and regression methods are two main ways for modeling the prediction model. Among the classification models, many recent studies applied machine learning methods and obtained promising results [5]– [7]. For instance, L. Hao et al. [8] used a regression model for the three major commercial airports in New York to predict flight delay. However, several reasons are restricting the existing methods from improving the accuracy of the flight delay prediction.

The reasons are summarized as follows: the diversity of causes affecting the flight delay, the complexity of the causes, the relevancy between causes, and the insufficiency of available flight data. In [6], a public dataset named VRA [9] was used to compare the performance of several machine learning models including k-nearest neighbors (K-NN) [10], support vector machines (SVM) [11], naive Bayes classifier, and random forests for predicting flight delay, and achieved the best accuracy of 78.02% among the four schemes. However, the air route information (e.g., traffic flow and size of each route) was not considered in their model, which prevents them from obtaining higher accuracy. In [4], D. A. Pamplona et al. built an artificial neural network with 4 hidden layers, and achieved the highest accuracy of 87%; their proposed model suggested that the day of the week, block hour, and route has great influence on the flight delay. This model did not consider meteorological factors, so there is room for improvement. Y. J. Kim et al. [12] proposed a model with two stage. The first stage is to predict day-to-day delay status of specific airport by using deep RNN model, where the status was defined as an average delay of all flights arrived at each airport.

The second stage is a layered neuron network model to predict the delay of each individual flight using the day-to-day delay status from the first stage and other information. The two stages of the model achieved accuracies of 85% and 87.42%, respectively. This study suggested that the deep learning model requires a great volumes of data. Otherwise, the model is likely to end up with poor performance or overfitting [13]. To address the problems in ATM, the received ADS-B messages can be coupled with weather information, traffic flow information, and other information to constitute an aviation data lake, which provides an opportunity to find a better approach to accurately predict the flight delay. Meanwhile, machine learning have made great progress and have obtain amazing performance in many domains, such as internet of things [14], heterogeneous network traffic control [15], autonomous driving [16], unmanned aerial vehicle [17]–[21], wireless communications [22]–[28], and cognitive radio [29]–[31]. The above successes motivate us to apply machine learning in the field of air traffic data analytic applications [12], [32]. Compared with the scenarios in wireless communications, the air traffic also faces dynamic environment and can be affected by many dynamic factors. Therefore, it is worthy to apply machine learning models for the flight delay prediction by making full use of the aviation data lake. By combining the advantages of all the available different data, we can feed the entire dataset into specific deep learning models, which allows us to find optimal solution in a larger and finer solution space and gain higher prediction accuracy of the flight delay. Our work benefits from considering as many factors as possible that may potentially influence the flight delay. For instance, airports information, weather of airports, traffic flow of airports, traffic flow of routes. The contributions of this paper can be summarized as follows: We explore a broader scope of factors which may potentially influence the flight delay and quantize those selected factors. Thus we obtain an integrated aviation dataset. Our experimental results indicate that the multiple factors can be effectively used to predict whether a flight will delay. Several machine learning based-network architectures are proposed and are matched with the established aviation dataset. Traditional flight prediction problem is a binary classification task. To comprehensively evaluate the performance of the architectures, several prediction tasks covering classification and regression are designed. Conventional schemes mostly focused on a single route or a single airport [4], [6], [12]. However, our work covers all routes and airports which are within our ADSB platform.

2.LITERATURE SURVEY

TITLE:

Flight Delay Prediction Based On Aviation Big Data And Machine Learning

DESCRIPTION:

With the development of the BIG DATA aviation transportation industry in recent years, the volume of civIL aviation transportation has increased rapidly. Increased carrier costs and reduced airport operating efficiency caused by flight delays have become issues that need to be addressed.How to improve the accuracy of predicting flight arrival delay time is of great significance for improving airport transportation efficiency, rationally scheduling flights and improving passenger comfort. In this paper, the RANDOM BASED FOREST model is utilized on the U.S Domestic airline on-time performance data from U.S. Transportation Administration, combined with the characteristics of the model to determine the influencing factors, and to predict the arrival delaysof flights within the United States.

The accuracy;precision and some other criterion of the model are given to evaluate the performance on the data. A better effect is obtained: the accuracy reach 80.44% in this case. Finally, the specific delay time is predicted, we found that the support vector machine has the best prediction result for the flight delay time, the average prediction error is 9.733 min, which has a certain reference value for flight operation and airport scheduling.

ALGORITHM:

In the context of predicting flight delays, several machine learning algorithms are commonly employed for binary classification. Let's explore some of these algorithms:

->**Logistic Regression:** A widely used algorithm that models the probability of a binary outcome (such as flight delay) based on input features. It's particularly useful for understanding the impact of individual features.

->**K-Nearest Neighbor (KNN):** KNN classifies data points based on their proximity to other data points. In the context of flight delays, it considers the similarity of a given flight to its neighboring flights.

->**Gaussian Naïve Bayes**: This probabilistic algorithm assumes that features are conditionally independent given the class label. It's efficient and works well with categorical data.

->**Decision Tree:** Decision trees recursively split data based on feature values to create a tree-like structure. Each leaf node represents a class label (e.g., delayed or not delayed).

->**Support Vector Machine (SVM):** SVM aims to find a hyperplane that best separates data points into different classes. It's effective for both linear and non-linear classification tasks.

->**Random Forest:** An ensemble method that combines multiple decision trees. It reduces overfitting and provides robust predictions by aggregating results from individual trees.

->**Gradient Boosted Tree:** Another ensemble technique that builds decision trees sequentially, with each tree correcting the errors of the previous ones. It often performs well in practice.

In a study comparing these algorithms for flight delay prediction using data from John F. Kennedy International Airport, the Decision Tree algorithm demonstrated the best performance with an accuracy of 0.9777 1. Keep in mind that the choice of algorithm depends on the specific dataset, features, and problem requirements.

Predicting flight delays is crucial for improving airline operations and passenger satisfaction, ultimately benefiting the economy. 🛫✈️

ethods, namely k-nearest neighbors (k-NN), support vector machine (SVM), and linear regression (LR).

3. EXISTING SYSTEM

->Nowadays, aircrafts have become a necessity because they easy life. They are efficient in carrying goods and passengers around the world. It also supplies emergencies in warfare and takes a vital role in carrying medical necessities. Hence, advent of airplanes is considered important. Delays in aircrafts can affect thousands of people across the globe either directly or indirectly. There are a lot of reasons of delays in aircrafts such as critical weather, security issues, traffic and many more.

->There are several methods implemented in the existing system to predict the flight delays but due to various complexities of the ATFM and the huge datasets involved, it has become very difficult to find an accurate solution for this complication. Many algorithms have been implemented to forecast flight delays. We are using Python in Visual Studio Code. We implement Binary Classification to prepare a model that can predict the delays.

ADVANTAGES:

->Proposed methods implementing ADS-B Message Based Aviation Big Data Platform which is more effective and fast.

->ADS-B system is a communication and surveillance integrated system for air traffic management (ATM) where flights periodically broadcast location and other information on the same frequency band.

3.1 PROPOSED SYSTEM:

->The proposed work benefits from considering as many factors as possible that may potentially influence the flight delay. For instance, airports information, weather of airports, traffic flow of airports, traffic flow of routes. The contributions of this paper can be summarized as follows:

->The system explores a broader scope of factors which may potentially influence the flight delay and quantize those selected factors. Thus we obtain an integrated aviation dataset. Our experimental results indicate that the multiple factors can be effectively used to predict whether a flight will delay.

->Several machine learning based-network architectures are proposed and are matched with the established aviation dataset. Traditional flight prediction problem is a binary classification task. To comprehensively evaluate the performance of the architectures, several prediction tasks covering classification and regression are designed.Conventional schemes mostly focused on a single route or a single airport [4], [6], [12]. However, our work covers all routes and airports which are within our ADSB platform.

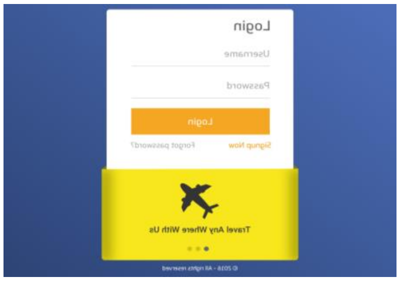
DISADVANTAGES:

->In the existing system, the system is not using Data Transformation and Balancing.

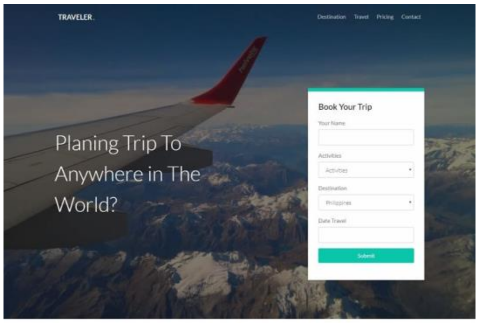
->This system is less performance due to lack of Data Cleaning and Data Integration.

4. OUTPUTSCREENS

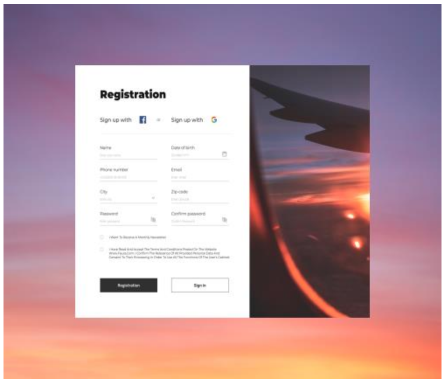
- **Login Page:**



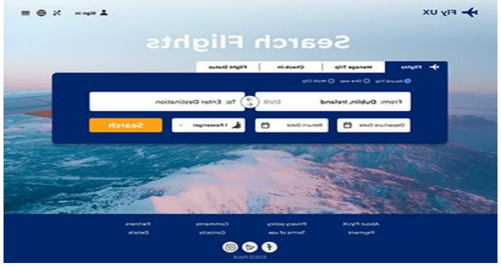
- **Book your trip:**



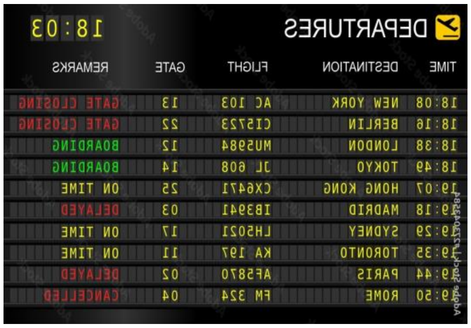
Registration Screen:



- **Search Flights:**



- **Departure Date and Timings:**



5. CONCLUSION

Machine learning algorithms were applied progressively and successively to predict flight arrival & delay.We built five models out of this. We saw for each evaluation metric considered thevalues of the models and compared them. We found out that: -In Departure Delay, Random Forest Regressor was observed as the best model with Mean Squared Error 2261.8 and Mean Absolute Error 24.1, which are the minimum value found in these respective metrics. In Arrival Delay, Random Forest Regressor was the best model observed with Mean Squared Error 3019.3 and Mean Absolute Error 30.8, which are the minimum value found in these respective metrics. In the rest of the metrics, the value of the error of Random Forest Regress or although is not minimum but still gives a low value comparatively. In maximum metrics, we found out that Random Forest Regress or gives us the best value and thus should be the model selected.

The future scope of this paper can include the application of more advanced, modern andinnovative pre-processing techniques, automated hybrid learning and sampling algorithms, anddeep learning models adjusted to achieve better performance. To evolve a predictive model, additional variables can be introduced. e.g., a model where meteorological statistics are utilized in developing error-free models for flight delays. In this paper we used data from the US only, therefore in future, the model can be trained with data from other countries as well. With the use ofmodels that are complex and hybrid of many other models provided with appropriate processing power and with the use of larger detailed datasets, more accurate predictive models can be developed. Additionally, the model can be configured for other airports to predict their flight delays as well and for that data from these airports would be required to incorporate into this research.

6. REFERENCE

[1] T. Weise, S. Bouaziz, H. Li and M. Pauly, "Realtime Performance-based Facial Animation", ACM Transactions on Graphics, Vol.30, No. 4, pp. 71-77, 2011.

[2] Q. Cai, D. Gallup, C. Zhang and Z. Zhang, "3D Deformable Face Tracking with a Commodity Depth Camera", Proceedings of 11th International Conference on European Conference on Computer Vision, pp. 229-242, 2010.

[3] G. Fanelli, M. Dantone, and L.V. Gool, "Real Time 3D Face Alignment with Random Forests-based Active Appearance Models", Proceedings of 10 th International Conference and Workshops on Automatic Face and Gesture Recognition, pp. 1-8, 2013.

[4] Z.Z. Zhang, W. Zhang, J.Z. Liu and X.O. Tang, "Multiview Facial Landmark Localization in RGB-D Images via Hierarchical Regression With Binary Patterns", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 24, No. 9, pp. 1475-1485, 2014.

[5] T. Cootes, C.J. Taylor, D.H. Cooper and J. Graham, "Active Shape Models-Their Training and Application", Computer Vision and Image Understanding, Vol. 61, No. 1, pp. 38-59, 1995.