

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Volume 4, Issue 1, December 2024

# **Flight Delay Detection**

Mrs. B. Ranjitha<sup>1</sup>, Ms. B. Keerthana<sup>2</sup>, Ms. B. Lohitha<sup>3</sup>, Ms. B. Tanisha Singh<sup>4</sup>

<sup>1</sup>Assistant Professor, Department of CSE, Guru Nanak Institute of Technology, Hyderabad, Telangana <sup>2,3,4</sup>Students, Department of CSE, Guru Nanak Institute of Technology, Hyderabad, Telangana

Abstract: Accurate flight delay prediction is fundamental to establish the more efficient airline business. Recent studies have been focused on applying machine learning methods to predict the flight delay. Most of the previous prediction methods are conducted in a single route or airport. This paper explores a broader scope of factors which may potentially influence the flight delay, and compares several machine learning-based models in designed generalized flight delay prediction tasks. To build a dataset for the proposed scheme, automatic dependent surveillance broadcast (ADS-B) messages are received, pre-processed, and integrated with other information such as weather condition, flight schedule, and airport information. 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.

Keywords: flight delay prediction

#### I. INTRODUCTION

IR 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. 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. The received ADS-B message along with other collected data from the Internet can constitute a huge 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 . For example, malicious or fake messages can be detected with the use of multilateration (MLAT), allowing open, free, and secure visibility to all the aircrafts within airspace. 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 arrive 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.

#### **II. LITERATURE SURVEY**

**M. Leonardi(2018)** Automatic dependent surveillance-broadcast (ADS-B) is an air traffic control system in which aircraft transmit their own information (identity, position, velocity etc.) to ground sensors for surveillance scope. The tracking of the different sensors' clocks by the use of time difference of arrival of ADS-B messages is proposed to check the veracity of the position information contained in the ADS-B messages. The method afforms detecting possible

Copyright to IJARSCT www.ijarsct.co.in DOI: 10.48175/IJARSCT-22625



172



International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

#### Volume 4, Issue 1, December 2024

on-board anomalies or the malicious injection of fake messages (intrusion) without the use of the multilateration (or any other) location algorithm. It follows that it does not need the inversion of the location problem (usually strong nonlinear and ill-posed), and, contrary to the multilateration, it works also with less than four sensors.

Y. A. Nijsure, G. Kaddoum, (2015) A novel air-to-ground (ATG) communication system, which is based on adaptive modulation and beamforming enabled by automatic dependent surveillance-broadcast (ADS-B) and multilateration techniques, is presented in this paper. From an aircraft geolocation perspective, the proposed multilateration technique uses the time-difference-of-arrival (TDOA), angle-of-arrival (AOA), and frequency-difference-of-arrival (FDOA) features within the ADS-B signal to implement the hybrid geolocation mechanism. Moreover, this hybrid mechanism aims for the optimal selection of multilateration sensors to provide a precise aircraft geolocation estimate by minimizing the geometric dilution-of-precision (GDOP) metric and imparts significant resilience to the current ADS-Bbased geolocation framework to withstand any form of attack involving aircraft impersonation and ADS-B message infringement. From an ATG communication perspective, the ground base stations can use this hybrid aircraft geolocation estimate to dynamically adapt their modulation parameters and transmission beampattern in an effort to provide a high-data-rate secure ATG communication link. Additionally, we develop a hardware prototype that is highly accurate in estimating AOA data and facilitating TDOA and FDOA extraction associated with the received ADS-B signal. This hardware setup for the ADS-B-based ATG system is analytically established and validated with commercially available universal software-defined radio peripheral units. This hardware setup displays 1.5° AOA estimation accuracy, whereas the simulated geolocation accuracy is approximately 30 m over 100 nautical miles for a typical aircraft trajectory. The adaptive modulation and beamforming approach assisted by the proposed GDOPminimization-based multilateration strategy achieves significant enhancement in throughput and reduction in packet error rate.

**D. A. Pamplona et. al(2018)** Air delay is a problem in most airports around the world, resulting in increased costs for airlines and discomfort for passengers. Air Traffic Flow Management (ATFM) programs were implemented with the main objective to reduce the delay levels in the whole air transportation sector. The question is to find a suitable way to predict possible delay scenarios to better apply ATFM measures. The present work seeks to enrich the academic literature on the subject and aims to present the application of Artificial Neural Networks (ANN) to a prediction model of delays in the air route between São Paulo (Congonhas) - Rio de Janeiro (Santos Dumont). The configuration of ANN exerts a great influence on its predictive power. To better adjust the parameters of the proposed ANN and for the hyper parameterization of the network to occur, the Random Search technique is used. By using the recall, precision and Fscore metrics in the performance measurement, the prediction results show the satisfactory in the case study.

**S. Manna et. al(2017)** Supervised machine learning algorithms have been used extensively in different domains of machine learning like pattern recognition, data mining and machine translation. Similarly, there has been several attempts to apply the various supervised or unsupervised machine learning algorithms to the analysis of air traffic data. However, no attempts have been made to apply Gradient Boosted Decision Tree, one of the famous machine learning tools to analyse those air traffic data. This paper investigates the effectiveness of this successful paradigm in the air traffic delay prediction tasks. By combining this regression model based on the machine learning paradigm, an accurate and sturdy prediction model has been built which enables an elaborated analysis of the patterns in air traffic delays. Gradient Boosted Decision Tree has shown a great accuracy in modeling sequential data. With the help of this model, day-to-day sequences of the departure and arrival flight delays of an individual airport can be predicted efficiently. In this paper, the model has been implemented on the Passenger Flight on-time Performance data taken from U.S. Department of Transportation to predict the arrival and departure delays in flights

#### **Existing System**

- Y. J. Kim et al. 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.

Copyright to IJARSCT www.ijarsct.co.in DOI: 10.48175/IJARSCT-22625



173



International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

#### Volume 4, Issue 1, December 2024

• This study suggested that the deep learning model requires a great volume of data. Otherwise, the model is likely to end up with poor performance or overfitting.

#### Existing System Disadvantages

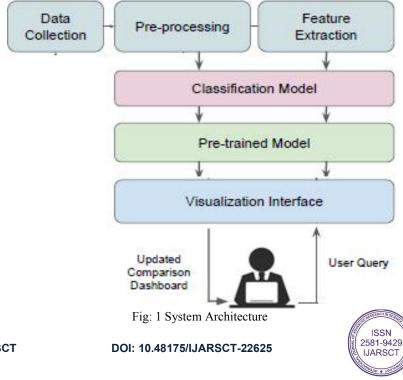
- Several reasons are restricting the existing methods from improving the accuracy of the flight delay prediction. The reasons are summarized as follows: the diversity of causes affecting the flight delay, the complexity of the causes, the relevancy between causes, and the insufficiency of available flight data.
- 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.

#### **Proposed System**

- 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. However, our work covers all routes and airports which are within our ADS B platform.

#### Proposed System Advantages

- Our work benefits from considering as many factors as possible that may potentially influence the flight delay. For instance, airport information, weather of airports, traffic flow of airports, traffic flow of routes.
- The random forest-based architecture obtained a testing accuracy of 90.2% for the binary classification, which is considered a promising result and demonstrates the strong ability of the ensemble learning.



#### **III. SYSTEM ARCHITECTURE**



International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

#### Volume 4, Issue 1, December 2024

#### Explanation

The system architecture represents how data moves and processed in flight delay detection system. The process begins with Data Collection, where relevant flight data, such as schedules, weather conditions, and historical delays, is gathered. The data then undergoes Pre-processing to clean and structure it, ensuring its suitability for analysis. Following this, Feature Extraction is performed to identify and isolate key attributes (e.g., departure time, weather, airline) that influence flight delays. The refined data is fed into a Classification Model, which predicts whether a flight will be delayed based on extracted features. This model can either be trained from scratch or enhanced using a Pre-trained Model that leverages prior knowledge, accelerating development and improving accuracy. The results are then integrated into a Visualization Interface, a dashboard that allows users to input queries, such as specific flights or routes, and view predictions and comparative analyses of delays. The dashboard is updated dynamically, enabling real-time insights and interactions. This architecture ensures an efficient pipeline for predicting flight delays while providing actionable information for decision-making.

#### **IV. METHODOLOGY**

#### Modules Name:

#### **Data Collection:**

This is the first real step towards the real development of a machine learning model, collecting data. This is a critical step that will cascade in how good the model will be, the more and better data that we get, the better our model will perform.

There are several techniques to collect the data, like web scraping, manual interventions and etc. The dataset used in this Flight Delay dataset taken from Kaggle

#### Dataset:

The dataset consists of 583985 individual data. There are 21 columns in the dataset, which are described below. YEAR — Year in which flight took place

QUARTER — Quarter in which flight took place (1–4) MONTH — Month in which flight took place (1–12) DAY\_OF\_MONTH — Day of the month in which flight took place (1–31)

 $DAY_OF_WEEK - 1$  for Monday, 2 for Tuesday, etc. in which flight took place UNIQUE\_CARRIER - Airline carrier code

TAIL\_NUM — Aircraft tail number FL\_NUM — Flight number

ORIGIN\_AIRPORT\_ID — ID of origin airport

ORIGIN — Code of origin airport(ATL, DFW, SEA, etc.) DEST\_AIRPORT\_ID — ID of destination airport DEST — Code of destination airport (ATL, DFW, SEA, etc.) CRS\_DEP\_TIME — Scheduled departure time

DEP\_TIME — Actual departure time

DEP\_DELAY — Departure Delay in minutes

DEP\_DEL15 — 1 if departure is delayed by 15 minutes or more else 0 CRS\_ARR\_TIME — Scheduled arrival time ARR\_TIME — Actual arrival time

ARR\_DELAY — Arrival Delay in minutes

ARR\_DEL15 — 1 if arrived late by 15 minutes or more else 0 CANCELLED — 1 if Flight was cancelled else 0 DIVERTED — 1 if Flight was diverted else 0 CRS\_ELAPSED\_TIME — Scheduled flight time in minutes ACTUAL ELAPSED TIME — Actual flight time in minutes DISTANCE — Distance traveled in miles

#### **Data Preparation:**

We will transform the data. By getting rid of missing data and removing some columns. First we will create a list of column names that we want to keep or retain.

Next we drop or remove all columns except for the columns that we want to retain. Finally we drop or remove the rows that have missing values from the data set.





International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

#### Volume 4, Issue 1, December 2024

#### Model Selection:

While creating a machine learning model, we need two dataset, one for training and other for testing. But now we have only one. So lets split this in two with a ratio of 80:20. We will also divide the data frame into feature column and label column.

Here we imported train\_test\_split function of sklearn. Then use it to split the dataset. Also, test\_size = 0.2, it makes the split with 80% as train dataset and 20% as test dataset.

The random\_state parameter seeds random number generator that helps to split the dataset. The function returns four datasets. Labelled them as train\_x, train\_y, test\_x, test\_y. If we see shape of this datasets we can see the split of dataset. We will use Random Forest Classifier, which fits multiple decision tree to the data. Finally I train the model by passing train\_x, train\_y to the fit method.

Once the model is trained, we need to Test the model. For that we will pass test\_x to the predict method.Random Forest is one of the most powerful methods that is used in machine learning for classification problems. The random forest comes in the category of the supervised classification algorithm.

This algorithm is carried out in two different stages the first one deals with the creation of the forest of the given dataset, and the other one deals with the prediction from the classifier.

#### **Analyze and Prediction:**

In the actual dataset, we chose only 10 features: DAY\_OF\_MONTH — Day of the month in which flight took place (1–31) DAY\_OF\_WEEK — 1 for Monday, 2 for Tuesday, etc. in which flight took place OP\_CARRIER\_AIRLINE\_ID — ID of origin airline ORIGIN\_AIRPORT\_ID — ID of origin airport DEST\_AIRPORT\_ID — ID of destination airport DEP\_TIME — Actual departure time ARR\_TIME — Actual arrival time DEP\_DEL15 — 1 if departure is delayed by 15 minutes or more else 0 DIVERTED — 1 if Flight was diverted else 0 DISTANCE — Distance traveled in miles ARR\_DEL15 — 1 if arrived late by 15 minutes or more else 0

#### Accuracy on test set:

We got an accuracy of 92.1% on test set.

#### Saving the Trained Model:

Once you're confident enough to take your trained and tested model into the production-ready environment, the first step is to save it into a .h5 or .pkl file using a library like pickle .

Make sure you have pickle installed in your environment.

Next, let's import the module and dump the model into .pkl file.

#### Implementation

The existing system and the proposed system differ significantly in their scope, input data, and algorithmic approaches. The existing system primarily focuses on single airports and uses a two-stage deep learning model comprising a Deep Recurrent Neural Network (RNN) for predicting daily airport delay statuses and a layered neural network for predicting individual flight delays. In contrast, the proposed system broadens the scope by incorporating data from all routes and airports within an ADS B platform, allowing for a comprehensive analysis. While the existing system relies on limited input data, the proposed system integrates a richer dataset, including weather conditions, air traffic, geographic factors, and operational constraints.

The algorithms also vary considerably between the two systems. The existing model uses an **RNN** for temporal analysis and a fully connected neural network for flight-level predictions. On the other hand, the proposed system employs a

Copyright to IJARSCT www.ijarsct.co.in





International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

#### Volume 4, Issue 1, December 2024

wider array of algorithms tailored to classification and regression tasks. These include traditional models like logistic regression and support vector machines (SVMs), ensemble methods such as Random Forest and Gradient Boosting Machines (GBMs), and advanced deep learning architectures like Deep Neural Networks (DNNs), Convolutional Neural Networks (CNNs), and Transformer models.

The proposed system emphasizes handling multi-source data and avoiding overfitting by leveraging advanced techniques and regularization. Additionally, the proposed model expands the prediction tasks to include both binary classification and continuous regression, providing a more comprehensive framework compared to the existing system's narrower focus. These enhancements make the proposed system more robust and better suited for predicting flight delays across diverse scenarios.

#### V. ALGORITHMS USED

#### **Existing Algorithms**

#### Deep Recurrent Neural Network (RNN):

RNN is utilized to predict the day-to-day delay status of specific airports. RNNs are well-suited for sequential data, as they have the ability to maintain a memory of previous inputs through their hidden states, enabling them to learn temporal dependencies in time-series data. This makes them an effective choice for analyzing daily patterns of delays influenced by factors such as weather or traffic trends. However, traditional RNNs often suffer from vanishing gradient problems when handling long sequences, which may limit their ability to capture long-term dependencies unless variants like Long Short-Term Memory (LSTM) or Gated Recurrent Units (GRU) are implemented.

#### **Proposed Algorithm**

#### **Random Forest:**

Random forest is an ensemble learning method that combines the outputs of multiple decision trees to improve accuracy and robustness. Each decision tree in the forest is trained on a random subset of the data, and the final prediction is based on the majority vote (classification) or average prediction (regression) of all trees. Random Forest handles overfitting well due to its aggregation mechanism and can automatically estimate feature importance, making it particularly useful for high-dimensional datasets like aviation data. However, it may struggle with time-series data unless temporal features are explicitly engineered.

#### **Linear Regression:**

Linear regression is a basic algorithm for regression tasks, predicting continuous outcomes such as the duration of a flight delay. It models the relationship between the dependent variable (e.g., delay time) and independent variables (e.g., weather conditions, air traffic) as a linear equation. While linear regression is straightforward to implement and interpret, it assumes a linear relationship and is sensitive to multicollinearity among input features. For flight delays influenced by multiple non-linear factors, this algorithm may serve as a benchmark or be enhanced through feature engineering.





Fig 2: Home page

Copyright to IJARSCT www.ijarsct.co.in





International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

#### Volume 4, Issue 1, December 2024



#### Fig 3: Login page

A login page for users who are trying to predict flight delay. To access the resources or system it usually requires username and password.

# ABSTRACT

Accurate flight delay prediction is fundamental to establish the more efficient airline business. Recent studies have been focused on applying machine learning methods to predict the flight delay. Most of the previous prediction methods are conducted in a single route or airport. This paper explores a broader scope of factors which may potentially influence the flight delay, and compares several machine learning-based models in designed generalized flight delay prediction tasks. To build a dataset for the proposed scheme, automatic dependent surveillancebroadcast (ADS-B) messages are received, pre-processed, and integrated with other information such as weather condition, flight schedule, and airport information. 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.



#### Fig 4: Abstract Page

After entering the login details, we see an abstract page popping up where it describes about the project, the purpose of the project and also which algorithm has been used to build this project.



#### Fig 5: Upload dataset page

After we click on next in the abstract page, we see a page for uploading dataset where we need to provide the dataset to train the model. Here we uploaded the dataset and clicked on upload button





International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

#### Volume 4, Issue 1, December 2024

IJARSCT

| FL   | IGHT DELA                                    | <b>NY</b>   |  |   |   |   |   |  |
|------|--|---|--|---|---|---|---|--|
| 2019 | 5  | 6   | WN   | 19393   | WN  | N7885A  | 2536  | 10792  |
| 2019 | 5  | 6   | WN   | 19393   | WN  | N7725A  | 2613  | 10792  |
| 2019 | 5  | 6   | WN   | 19393   | WN  | N//14B  | 4360  | 10/92  |
| 2019 | 5  | 6   | WN   | 19393   | WN  | N437WN  | 591   | 10792  |
| 2019 | 5  | 6   | WN   | 19393   | WN  | N8306H  | 3533  | 10792  |
| 2019 | 5  | 6   | WN   | 19393   | WN  | N8522P  | 2599  | 10792  |
|      | 2019<br>2019<br>2019<br>2019<br>2019<br>2019 | 2019 5   2019 5   2019 5   2019 5   2019 5   2019 5 | 2019 5 6   2014 5 6   2019 5 6   2019 5 6   2019 5 6 | 2019 5 6 WN   2019 5 6 WN | 2019 5 6 WN 19393   2019 5 6 WN 19393 | 2019 5 6 WN 19393 WN   2019 5 6 WN 19393 WN | 2019 5 6 WN 19393 WN N7005A   2019 5 6 WN 19393 WN N7725A   2019 5 6 WN 19393 WN N7725A   2019 5 6 WN 19393 WN N/714B   2019 5 6 WN 19393 WN N437WN   2019 5 6 WN 19393 WN N437WN   2019 5 6 WN 19393 WN N437WN | 2019 5 6 WN 19393 WN N7005A 2536   2019 5 6 WN 19393 WN N7725A 2613   2019 5 6 WN 19393 WN N7725A 2613   2019 5 6 WN 19393 WN N/148 4360   2019 5 6 WN 19393 WN N437WN 591   2019 5 6 WN 19393 WN N830611 5533 |

|      | FL   | IGHT DELA | ١Y |    |       |    |        |     |       |
|------|------|-----------|----|----|-------|----|--------|-----|-------|
| 1300 | 2019 | 31        | 4  | UA | 19977 | UA | N39726 | 104 | 12016 |

#### Click to Train | Test

Fig 6:Preview Page

After previewing the page and checking whether the dataset has been uploaded properly or not then click the train test button to train the model.

| Flight Delay Prediction     |
|-----------------------------|
| DAY_OF_MONTH                |
| 1                           |
| DAY_OF_WEEK :               |
| 3                           |
| OP CARRIER AIRLINE ID-      |
| 20366                       |
| ORIGIN_AIRPORT_ID :         |
| 13930                       |
| DEST AIRPORT ID             |
| 11977                       |
| DEP_TIME .                  |
| 1003                        |
| ARR TIME :                  |
| 1117                        |
| DEP_DEL15                   |
| 0                           |
| DIVERTED :                  |
| 0                           |
| DISTANCE ·                  |
| 1/4                         |
| Predict Departure Delay     |
| Assival Delay Prediction is |
| Analysis                    |

Fig 7: Entry of flight details

After entering all the details we have to click on Predict Departure Delay to know whether the flight is delayed or not

Copyright to IJARSCT www.ijarsct.co.in





International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

#### Volume 4, Issue 1, December 2024

| Flight Delay Prediction          |
|----------------------------------|
| DAY OF MONTH                     |
| DAY_OF_MORTH                     |
| DAY OF WIEK                      |
| DAY_OF_WEEK                      |
| OP CARRIER AIRLINE ID            |
| OP_CARRER_AIRLINE_ID             |
| ORIGIN AIRPORT_ID                |
| ORIGIN_AIRPORT_ID                |
| DEST AIRPORT ID :                |
| DEST_AIRPORT_ID                  |
| DEP_TIME                         |
| DEP_TIME                         |
| ARR_TIME :                       |
| ARR_TIME                         |
| DEP_DEL15                        |
| DEP_DEL15                        |
| DIVERTED :                       |
| DIVERTED                         |
| DISTANCE                         |
| DESTANCE                         |
| Predict Departure Delay          |
| Arrival Delay Prediction is [1.] |
| Analysis                         |

#### Fig 8: Result page If the result i.e. If the arrival Delay prediction is 1 the it indicates that the flight is not delayed CONCLUSION AND FUTURE WORK

In this paper, random forest-based and LSTM-based architectures have been implemented to predict individual flight delay. The experimental results show that the random forestbased method can obtain good performance for the binary classification task and there are still room for improving the multi-categories classification tasks. The LSTM-based architecture can obtain relatively higher training accuracy, which suggests that the LSTM cell is an effective structure to handle time sequences. However, the overfitting problem occurred in the LSTM-based architecture still needs to be solved. In summary, the random forest-based architecture presented better adaptation at a cost of the training accuracy when handling the limited dataset. In order to overcome the overfitting problem and to improve the testing accuracy for multi-categories classification tasks, our future work will focus on collecting or generating data, integrating more information like airport traffic flow, airport visibility into our dataset, and designing more delicate networks.

#### HOME

#### Fig 9: Conclusion page

By clicking on Future in the previous page we get a conclusion and future work page where it concludes the whole project and tells about the future enhancements

#### VII. CONCLUSION

In this paper, random forest-based and LSTM-based 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. The LSTM-based architecture can obtain relatively higher training accuracy, which suggests that the LSTM cell is an effective structure to handle time sequences. However, the overfitting problem occurred in the LSTM-based architecture still needs to be solved. In summary, the random forest-based architecture presented better adaptation at a cost of the training accuracy when handling the limited dataset.

#### **VIII. FUTURE ENHANCEMENT**

In order to overcome the overfitting problem and to improve the testing accuracy for multi- categories classification tasks, our future work will focus on collecting or generating more training data, integrating more information like airport traffic flow, airport visibility into our dataset, and designing more delicate networks.

#### REFERENCES

- [1]. M. Leonardi, "Ads-b anomalies and intrusions detection by sensor clocks tracking," IEEE Trans. Aerosp. Electron. Syst., to be published, doi: 10.1109/TAES.2018.2886616.
- [2]. Y. A. Nijsure, G. Kaddoum, G. Gagnon, F. Gagnon, C. Yuen, and R. Mahapatra, "Adaptive air-to-ground secure communication system based on ads-b and wide-area multilateration," IEEE Trans. Veh. Technol., vol. 65, no. 5, pp. 3150–3165, 2015.

Copyright to IJARSCT www.ijarsct.co.in





International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

#### Volume 4, Issue 1, December 2024

- [3]. J. A. F. Zuluaga, J. F. V. Bonilla, J. D. O. Pabon, and C. M. S. Rios, "Radar error calculation and correction system based on ads-b and business intelligent tools," in Proc. Int. Carnahan Conf. Secur. Technol., pp. 1–5, IEEE, 2018.
- [4]. D. A. Pamplona, L. Weigang, A. G. de Barros, E. H. Shiguemori, and C. J. P. Alves, "Supervised neural network with multilevel input layers for predicting of air traffic delays," in Proc. Int. Jt. Conf. Neural Networks, pp. 1–6, IEEE, 2018.
- [5]. S. Manna, S. Biswas, R. Kundu, S. Rakshit, P. Gupta, and S. Barman, "A statistical approach to predict flight delay using gradient boosted decision tree," in Proc. Int. Conf. Comput. Intell. Data Sci., pp. 1–5, IEEE, 2017.
- [6]. L. Moreira, C. Dantas, L. Oliveira, J. Soares, and E. Ogasawara, "On evaluating data preprocessing methods for machine learning models for flight delays," in Proc. Int. Jt. Conf. Neural Networks, pp. 1–8, IEEE, 2018.
- [7]. J. J. Rebollo and H. Balakrishnan, "Characterization and prediction of air traffic delays," Transp. Res. Part C Emerg. Technol., vol. 44, pp. 231–241, 2014.
- [8]. L. Hao, M. Hansen, Y. Zhang, and J. Post, "New york, new york: Two ways of estimating the delay impact of new york airports," Transp. Res. Part ELogist. Transp. Rev., vol. 70, pp. 245–260, 2014.
- [9]. ANAC, "The Brazilian National Civil Aviation Agency." anac.gov, 2017. [online] Available:http://www.anac.gov.br/.
- [10]. S. Zhang, X. Li, M. Zong, X. Zhu, and R. Wang, "Efficient knn classification with different numbers of nearest neighbors," IEEE Trans. Neural Netw. Learn. Syst., vol. 29, no. 5, pp. 1774–1785, 2017.
- [11]. J. Sun, Z. Wu, Z. Yin, and Z. Yang, "Svm-cnn-based fusion algorithm for vehicle navigation considering atypical observations," IEEE Signal Process. Lett., vol. 26, no. 2, pp. 212–216, 2018.
- [12]. Y. J. Kim, S. Choi, S. Briceno, and D. Mavris, "A deep learning approach to flight delay prediction," in Proc. Digit. Avion. Syst. Conf., pp. 1–6, IEEE, 2016.
- [13]. Y. Cong, J. Liu, B. Fan, P. Zeng, H. Yu, and J. Luo, "Online similarity learning for big data with overfitting," IEEE Trans. Big Data, vol. 4, no. 1, pp. 78–89, 2017.
- [14]. F. Tang, Z. M. Fadlullah, B. Mao, and N. Kato, "An intelligent traffic load prediction-based adaptive channel assignment algorithm in sdn-iot: A deep learning approach," IEEE Internet Things J., vol. 5, pp. 5141–5154, Dec 2018.
- [15]. N. Kato, Z. M. Fadlullah, B. Mao, F. Tang, O. Akashi, T. Inoue, and K. Mizutani, "The deep learning vision for heterogeneous network traffic control: Proposal, challenges, and future perspective," IEEE Wireless Commun., vol. 24, pp. 146–153, June 2017.
- [16]. J. Wang, J. Liu, and N. Kato, "Networking and communications in autonomous driving: A survey," IEEE Commun. Surveys Tuts., vol. 21, pp. 1243–1274, April 2019.
- [17]. Y. Kawamoto, H. Nishiyama, N. Kato, F. Ono, and R. Miura, "Toward future unmanned aerial vehicle networks: Architecture, resource allocation and field experiments," IEEE Wireless Commun., vol. 26, pp. 94– 99, February 2019.
- [18]. D. Takaishi, Y. Kawamoto, H. Nishiyama, N. Kato, F. Ono, and R. Miura, "Virtual cell-based resource allocation for efficient frequency utilization in unmanned aircraft systems," IEEE Trans. Veh. Technol., vol. 67, no. 4, pp. 3495–3504, 2018.

