International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 12 | December - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

FLIGHT DELAY PREDICTION BASED ON AVIATION BIG DATA AND MACHINE LEARNING

[1]Mr. S . Prudhvi raj , [2]P.Abhishek, [3]P.Pavan Sai, [4]P.Akshitha, [5]P.Nandini

[1]AssProf, [2345]Sreyas Institute of Engineering and Technology

1.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 over fitting 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 over fitting problem.

KEYWORDS: Automatic Dependent Surveillance-Broadcast (ADS-B), Air Traffic Flow Management (ATFM) , Random Forest (RF), K-nearest Neighbors (K-NN)

2.INTRODUCTION:

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. 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 multi alteration (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 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) to reduce the delay level.

Classification and regression methods are two main ways for modelling the prediction model. Among the classification models, many recent studies applied machine learning methods and obtained promising results. For instance, L. Hao et al. 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, a public dataset named VRA was used to compare the performance of several machine learning models including k-nearest neighbors (K-NN), support vector machines (SVM), 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. 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. 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, heterogeneous network traffic control, autonomous driving, unmanned aerial vehicle, wireless communications, and cognitive radio.

The above successes motivate us to apply machine learning in the field of air traffic data analytic applications. 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. However, our work covers all routes and airports which are within our ADSB platform.

2.1 RESEARCH PROBLEM:

The problem statement involves developing a flight delay prediction system by harnessing aviation big data and employing machine learning techniques. The aim is to create a model that can accurately forecast flight delays, improving passenger experience and operational efficiency for airlines. The system should process vast amounts of historical flight data, weather information, air traffic conditions, and other relevant variables.

By training on this data, the machine learning model should learn patterns and relationships that contribute to flight delays. The ultimate goal is to provide real-time or near-real-time predictions for flight delays, allowing airlines and passengers to make informed decisions. Addressing this challenge requires data preprocessing, feature engineering, model selection, and rigorous evaluation to ensure the accuracy and reliability of the predictive system.

2.2 OBJECTIVES AND GOALS:

The main objective of "Flight Delay Forecasting Based on Aviation Big Data and Machine Learning" is to improve airline operational efficiency and passenger experience by developing robust predictive models for flight delays can be predicted so By doing so the goal of this project is to enable airlines, airports to work more efficiently with passengers to make rational decisions, reduce operational disruptions, and improve overall flight schedules, and ultimately provide a more reliable and simple flight ecosystem has emerged.

2.3 BACKGROUND:

Flight delay prediction and machine learning using ADS-B data is an innovative approach in aviation. ADS-B (Automatic Dependent Surveillance-Broadcast) technology provides real-time flight information, including position, speed and altitude. By analyzing large volumes of this data, machine learning models can identify mechanisms and causes of flight delays. These models consider variables such as weather, flight congestion, and historical data to make accurate forecasts. As a result, airlines and airports can proactively execute policies, improve passenger experience and increase overall operational efficiency, ultimately reducing the impact of delays on airline operations.

3 LITERATURE SURVEY:

The section discusses the significant costs associated with flight delays and the impact on the aviation industry and customers. It also proposes a two-stage predictive model using supervised machine learning algorithms to predict flight on-time performance.

[1] In H. Khaksar study, various flight delay prediction (FDP) methods were evaluated, with a hybrid classification approach outperforming others. Fleet age and aircraft type significantly influenced Iranian flight delays, while weather conditions were a major factor for US delays. The hybrid method achieved accuracy levels of 71.39% and 76.44% for delay occurrence and 70.16% and 75.93% for delay magnitude in the US and Iranian networks. These findings are valuable for airlines, particularly in developing countries like Iran. Future research could explore alternative data mining methods and investigate combining the hybrid method with robust flight scheduling.

[2] Meng Li research introduces a machine learning model for air traffic delay prediction, combining random forest classification and approximated delay propagation. It utilizes three critical databases, optimizes feature selection, and identifies departure delay and late arriving aircraft delay as key predictors. The model can predict delays along an aircraft's itinerary and has applications in both individual airline and macro air traffic control, offering improved accuracy for daily air traffic operations.

[3] Maryam Farshchian Yazdi research presents an optimized flight delay prediction model based on deep learning with LM algorithm. It explores the impact of denoising autoencoders and dataset balancing techniques. The SDA-LM model outperforms others, achieving 92.1% accuracy on imbalanced data and 96.2% on balanced data, surpassing RNN models by 4.1% and 8.2%, respectively. Further application on different datasets is suggested for future research.

[4] Devansh Shah study highlights the effectiveness of machine learning and deep learning in predicting flight delays, with the Random Forest model achieving 92.023% accuracy. This analysis benefits airports, airlines, and passengers, focusing on scientific parameters critical to the aviation industry.

[5] SP Lakshmi Narayanan paper analysis offers valuable insights for all stakeholders in the aviation industry, addressing financial losses, negative airline reputation, and sustainability concerns like increased fuel consumption and emissions. It not only predicts delays but also provides statistical descriptions, airline rankings, and peak delay hours, serving as a prototype for aviation authorities, including the Indian scenario.

[6] Micha Zoutendijk presents the application of Mixture Density Networks and Random Forest regression for probabilistic flight delay prediction, achieving estimations with a CRPS of 11 min days in advance. These predictions improve decision-making for airport coordinators. Section 3 employs probabilistic predictions in linear programming for flight-to-gate assignments, reducing daily conflicts by up to 74%. Future work aims to extend the approach to larger airports, considering various factors, and integrate it into other airport operation models such as sequencing and electric taxiing planning.

[7] Mithun Mallick analysis of various metrics, including RMSE, MAE, and R-squared, reveals that Linear Regression consistently outperforms Random Forests and Gradient-Boosted Trees in predicting flight delays. Linear Regression demonstrates lower RMSE and MAE values and higher R-squared values, indicating its superior predictive performance.

[8] Rahul Garg Machine and deep learning algorithms, including Support Vector Machine, Random Forest, and KNN, are used to predict flight delays and analyze influencing factors. Naïve Bayes, considering independence among predictions, is preferred for real-time accuracy and scalability, benefiting airlines, passengers, and ground personnel.

[9] BALAMURUGAN.R study applied machine learning to predict flight arrival and departure delays, finding Random Forest Regressor as the best model in various metrics. Future research can explore advanced pre-processing techniques, hybrid learning, deep learning, and expand the model to include data from different countries and airports for more accurate predictions.

[10] Mohammed Ayaz Hussain Khan, Flight delay prediction is crucial, and research has focused on enhancing models for precision. Combining multidimensional data with diverse techniques for feature selection and regression offers promising tools for cancer domain inference. MLP algorithm stands out, achieving an impressive 82% accuracy, making it a state-of-the-art choice for structured data tasks.

4 METHODOLOGY:

4.1 ALGORITHMS:

In our project we are using two machine learning algorithms they are:

1.Random Forest

2.LSTM

Random forest :

Random forest is a supervised learning algorithm that works by generating multiple decision trees during training and then combining their predictions to produce a reliable and accurate prediction Each decision tree is built on random training data done on (hence the term "random") . This randomness in the data and incoming features helps prevent overfitting and increases the generalizability of the model.

IMPLEMENTATION:

Using random forest algorithms to predict flight delays is a common and effective method in machine learning, especially when dealing with large flight datasets A step-by-step guide to how you can use Random Forest here is the role for this purpose:

In aviation, where punctuality and efficiency are paramount, the use of machine learning to predict flight delays has become increasingly important and one such powerful system a used in this case is the Random Forest algorithm. Forecasting flight delays based on large flight data is a challenging task due to the multidimensional variables consumed including weather, air traffic, and airport activity Random Forest algorithm, the leading ensemble learning method these challenges are addressed by integrating predictions from multiple decision trees. It provides exceptional accuracy and robustness, making it ideally suited to dealing with uncertainty and variability in flight data. By training on historical flight data, the random forest algorithm can recognize patterns and relationships among various factors affecting flight delays, enabling it to make accurate predictions in real time in This predictive capability not only helps airlines optimize scheduling and baggage allocation But it also benefits passengers by providing valuable information to better plan their journeys In big flight data and machine learning meanwhile, unplanned forest systems play an important role in increasing the reliability and efficiency of flight operations, ultimately contributing to a smoother travel experience for independents fill in all the boxes.

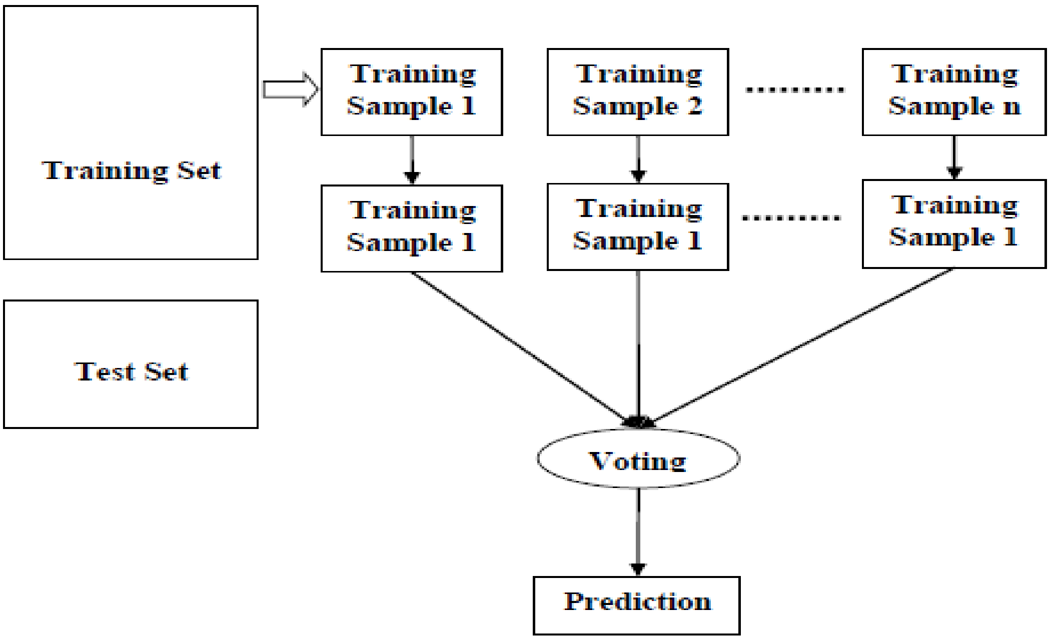


Figure 1

LSTM (LONG-SHORT TERM MEMORY):

Long-term short-term memory (LSTM) is a type of recurrent neural network (RNN) architecture designed to solve the missing flow problem of traditional RNNs LSTM can recognize and recall long sequences of data by memory cells an internal controller, which can store and update information over time They have three gates – an input gate, a forget gate, and an output gate – that control the flow of information into and out of memory, allowing capture remote reliance on sequential data LSTMs are widely used in various internal machine learning applications such as natural language processing, speech recognition and time series prediction.

IMPLEMENTATION:

In flight delay forecasting, long-term and short-term memory (LSTM) is a powerful tool for use with big flight data and machine learning. When applied to big flight data, LSTM models can help airlines, airports, and travelers predict and mitigate the impact of flight delays. The main advantage of LSTM is its ability to capture time dependence and patterns in its data, making it particularly suitable for forecasting flight delays, generally revealing time-related features and possible dependencies. Aviation big data includes a wide variety of information, such as flight information, historical delay records, weather data, and even lifestyle factors that can affect airline operations Machine learning methods are used and with good reason of this data, a It can be messy and complex. By incorporating LSTM into the analysis, machine learning models can effectively use this wealth of information to predict flight delays .In summary, working on flight delay forecasting based on big flight data and machine learning, LSTM offers a sophisticated and robust approach to an important challenge in the aviation industry enabling stakeholders to make informed decisions, improve efficiency, reduce the impact of flight delays and enhance overall travel experience for passengers Empowers This technology represents a powerful tool for aviation the industry, contributing to efficiency and customer satisfaction

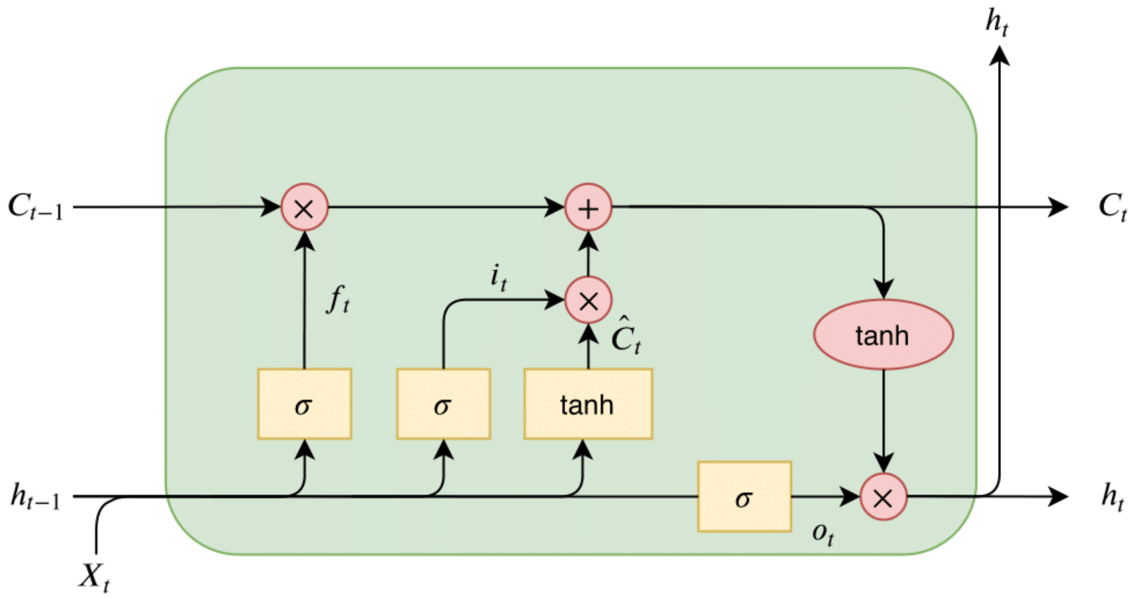


Figure 2

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 12 | December - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

4.2 DATASET:

In our project we use dataset flight dataset:

	departure_	departure_	flight_nam	flight_num	Date	departure_	pre_flight	expectatic	historical_	arrival_co	arrival_air	airport_de	airport_route_delay_reason
2	USA	ORD	FrontierAir	FN456	########	22:44	Safe	32	25	India	AMD	Leave port	airtraffic control problem
3	USA	ORD	HawaiianA	FN123	########	13:53	Notsafe	16	115	India	AMD	Weather	airtraffic control problem
4	USA	DFW	JetBlue	FN456	########	18:13	Safe	18	40	India	GOI	Weather	airtraffic control problem
5	USA	LAX	FrontierAir	BD123	########	04:33	Safe	18	58	India	BLR	Weather	Holiday
6	USA	ORD	SpiceJet	FN789	########	14:04	Notsafe	36	62	India	CCA	Leave port	Holiday
7	USA	DEN	AllegiantAi	FN123	########	13:46	Safe	21	77	India	CCA	Leave port	Holiday
8	USA	ALT	FrontierAir	AD123	########	20:48	Notsafe	59	55	India	AMD	Aiport traf	Holiday
9	USA	MIA	FrontierAir	AE123	########	21:35	Safe	34	120	India	CCA	Aiport traf	airtraffic control problem
10	USA	MCO	HawaiianA	AD123	########	13:04	Safe	50	20	India	CCU	Weather	Holiday
11	USA	ALT	JetBlue	AB123	########	07:40	Safe	27	103	India	COK	Leave port	airtraffic control problem
12	USA	ORD	HawaiianA	AD123	########	22:16	Notsafe	20	25	India	GOI	Leave port	Holiday
13	USA	SFO	SpiceJet	FN456	########	14:02	Safe	23	48	India	MAA	Weather	airtraffic control problem
14	USA	SEA	HawaiianA	CA432	########	08:17	Safe	10	56	India	PNQ	Aiport traf	airtraffic control problem
15	USA	MCO	SpiceJet	FN456	########	06:55	Safe	52	105	India	AMD	Weather	Holiday
16	USA	LAX	AllegiantAi	FN456	########	03:57	Safe	51	34	India	AMD	Aiport traf	Holiday
17	USA	MCO	JetBlue	AB123	########	07:15	Notsafe	47	27	India	AMD	Weather	unexpected problem
18	USA	SEA	AllegiantAi	AS123	########	19:16	Safe	53	77	India	GOI	Leave port	airtraffic control problem
19	USA	MIA	JetBlue	AE123	########	04:56	Safe	12	85	India	BLR	Weather	airtraffic control problem
20	USA	LAX	AllegiantAi	CA432	########	20:32	Notsafe	45	53	India	MAA	Aiport traf	Holiday
21	USA	LAX	SpiceJet	BD123	########	11:10	Notsafe	22	119	India	AMD	Leave port	unexpected problem
22	USA	DFW	AllegiantAi	BD123	########	00:34	Safe	39	56	India	PNQ	Leave port	unexpected problem
23	USA	ALT	AllegiantAi	AS123	########	12:55	Notsafe	40	37	India	PNQ	Leave port	Holiday
24	USA	DFW	JetBlue	AE123	########	00:41	Safe	30	68	India	COK	Aiport traf	unexpected problem
25	USA	DFW	AllegiantAi	FN789	########	07:46	Safe	16	76	India	PNQ	Weather	Holiday
26	USA	DEN	AllegiantAi	FN123	########	11:58	Notsafe	30	77	India	PNQ	Leave port	Holiday
27	USA	SEA	JetBlue	CA432	########	04:16	Notsafe	59	105	India	BLR	Leave port	airtraffic control problem
28	USA	ALT	FrontierAir	AE123	########	12:01	Safe	14	52	India	GOI	Leave port	airtraffic control problem
29	USA	MCO	JetBlue	FN123	########	22:03	Safe	29	113	India	GOI	Weather	airtraffic control problem

flight_dataset

Figure 3

A1 departure_country

30	USA	JFK	HawaiianA	FN456	########	17:27	Safe	40	81	India	CCU	Aiport traf	airtraffic control problem
31	USA	SEA	AllegiantAi	AE123	########	07:03	Notsafe	49	61	India	MAA	Weather	Holiday
32	USA	LAX	FrontierAir	BD123	########	08:47	Notsafe	52	71	India	CCA	Leave port	unexpected problem
33	USA	SFO	JetBlue	FN456	########	21:12	Notsafe	34	102	India	AMD	Aiport traf	unexpected problem
34	USA	MCO	AllegiantAi	AB123	########	01:18	Safe	36	96	India	CCA	Weather	airtraffic control problem
35	USA	ORD	AllegiantAi	AB123	########	16:54	Safe	13	101	India	HYD	Leave port	airtraffic control problem
36	USA	SFO	HawaiianA	AE123	########	17:23	Notsafe	31	65	India	COK	Leave port	airtraffic control problem
37	USA	JFK	AllegiantAi	AB123	########	11:12	Safe	18	24	India	AMD	Weather	Holiday
38	USA	SFO	FrontierAir	BD123	########	12:53	Safe	18	111	India	GOI	Leave port	unexpected problem
39	USA	LAX	SpiceJet	FN456	########	01:21	Safe	43	25	India	HYD	Weather	Holiday
40	USA	DEN	SpiceJet	CA432	########	19:42	Notsafe	48	75	India	BLR	Aiport traf	airtraffic control problem
41	USA	DEN	AllegiantAi	FN123	########	14:54	Safe	52	64	India	MAA	Aiport traf	unexpected problem
42	USA	JFK	AllegiantAi	AC123	########	00:27	Notsafe	42	100	India	BLR	Weather	unexpected problem
43	USA	JFK	JetBlue	FN456	########	16:13	Safe	40	62	India	CCA	Aiport traf	airtraffic control problem
44	USA	DFW	AllegiantAi	FN789	########	18:22	Notsafe	53	39	India	HYD	Leave port	airtraffic control problem
45	USA	DEN	JetBlue	AC123	########	09:40	Notsafe	42	106	India	AMD	Leave port	airtraffic control problem
46	USA	DEN	SpiceJet	AD123	########	16:13	Safe	51	113	India	CCU	Weather	Holiday
47	USA	MCO	JetBlue	AB123	########	16:31	Notsafe	56	63	India	BLR	Leave port	Holiday
48	USA	LAX	JetBlue	AB123	########	22:26	Notsafe	32	50	India	BLR	Weather	airtraffic control problem
49	USA	MIA	HawaiianA	AB123	########	20:30	Safe	10	100	India	GOI	Aiport traf	airtraffic control problem
50	USA	DEN	SpiceJet	FN456	########	19:03	Safe	11	35	India	AMD	Weather	Holiday
51	USA	LAX	JetBlue	FN789	########	20:03	Notsafe	49	96	India	CCA	Weather	unexpected problem
52	USA	ORD	HawaiianA	FN789	########	14:30	Notsafe	19	89	India	AMD	Leave port	airtraffic control problem
53	USA	DFW	FrontierAir	FN456	########	06:03	Notsafe	10	54	India	GOI	Leave port	unexpected problem
54	USA	ORD	SpiceJet	CA432	########	00:15	Safe	35	44	India	CCA	Aiport traf	airtraffic control problem
55	USA	MCO	SpiceJet	AC123	########	21:49	Notsafe	35	67	India	COK	Leave port	unexpected problem
56	USA	JFK	JetBlue	AC123	########	13:39	Notsafe	11	107	India	HYD	Weather	unexpected problem
57	USA	SFO	HawaiianA	BD123	########	16:42	Notsafe	41	68	India	MAA	Weather	airtraffic control problem
58	USA	DEN	SpiceJet	FN456	########	20:34	Safe	60	37	India	CCA	Weather	airtraffic control problem

flight_dataset

Figure 4

The modules and features are essential in the development of a machine learning model for flight delay prediction. By assessing historical data and real-time information related to these variables, machine learning algorithms can learn patterns, correlations, and dependencies that enable accurate predictions of flight delays, ultimately helping airlines, airports, and passengers make informed decisions and mitigate the impact of delays:

1. Departure Country: This feature represents the country from which the flight is departing. It can be important in predicting delays because different countries may have varying air traffic regulations, weather conditions, and operational factors that can influence flight delays.

2. Departure Airport: This feature indicates the specific airport from which the flight is departing. Different airports may have different congestion levels, infrastructure, and operational procedures, which can impact flight delays.

3. Flight Name: The flight name typically refers to the airline operating the flight. Airlines have their own scheduling and operational practices, which can influence the likelihood of delays.

4. Flight Number: The flight number is a unique identifier for a particular flight route. Historical data related to a specific flight number can help predict delays associated with that specific route.

5. Date: The date of the flight is a crucial factor. Flight delays can be influenced by factors like holidays, weekends, and seasonal weather patterns. Historical data on specific dates can provide insights into delay trends.

6. Departure Scheduled Time: This feature represents the planned departure time for the flight. Deviations from the scheduled time can lead to delays. Historical data on scheduled departure times can be used to assess the impact on delays.

7. Preflight Expectation Flight Delay: This feature may represent the airline's or airport's expectation of a delay before the flight departs. It can be based on real-time factors like weather conditions or air traffic, and it provides valuable insights into potential delays.

8. Historical Flight Delay: Historical flight delay data for the same route or flight number can be used to understand past patterns and trends. This data is crucial for training machine learning models to predict future delays.

9. Arrival Country: Similar to the departure country, the arrival country represents the destination of the flight. Different countries may have varying air traffic control and airport operations, which can influence delays.

10. Arrival Airport: This feature indicates the specific airport where the flight is scheduled to land. Like departure airports, different arrival airports have unique characteristics that can impact flight delays.

11. Airport Delay Reason: This module can provide information on the reasons for delays directly related to the airport. These reasons can include issues like runway maintenance, security checks, gate availability, or ground handling delays.

12.Airport Route Delay Reason: This feature focuses on delay reasons related to the specific route or airspace. It can include factors like air traffic congestion, weather patterns along the flight path, or routing decisions that affect the flight's efficiency.

5. PROPOSED SYSTEM:

The proposed flight delay prediction system utilizes aviation big data and advanced machine learning techniques to integrate historical flight data, real-time updates, weather information, and more. By employing machine learning models, it enhances prediction accuracy, offers insights into delay factors, and ensures scalability through cloud computing.

- The system's user-friendly interface aids decision-making, while continuous learning from new data improves adaptability, enhancing operational efficiency and passenger services. Multiple machine learning-based network architectures are matched with aviation datasets, addressing traditional binary classification tasks and incorporating comprehensive prediction tasks covering classification and regression.
- Unlike conventional approaches limited to single routes or airports, this work encompasses all routes and airports within our ADS-B platform.

Architecture:

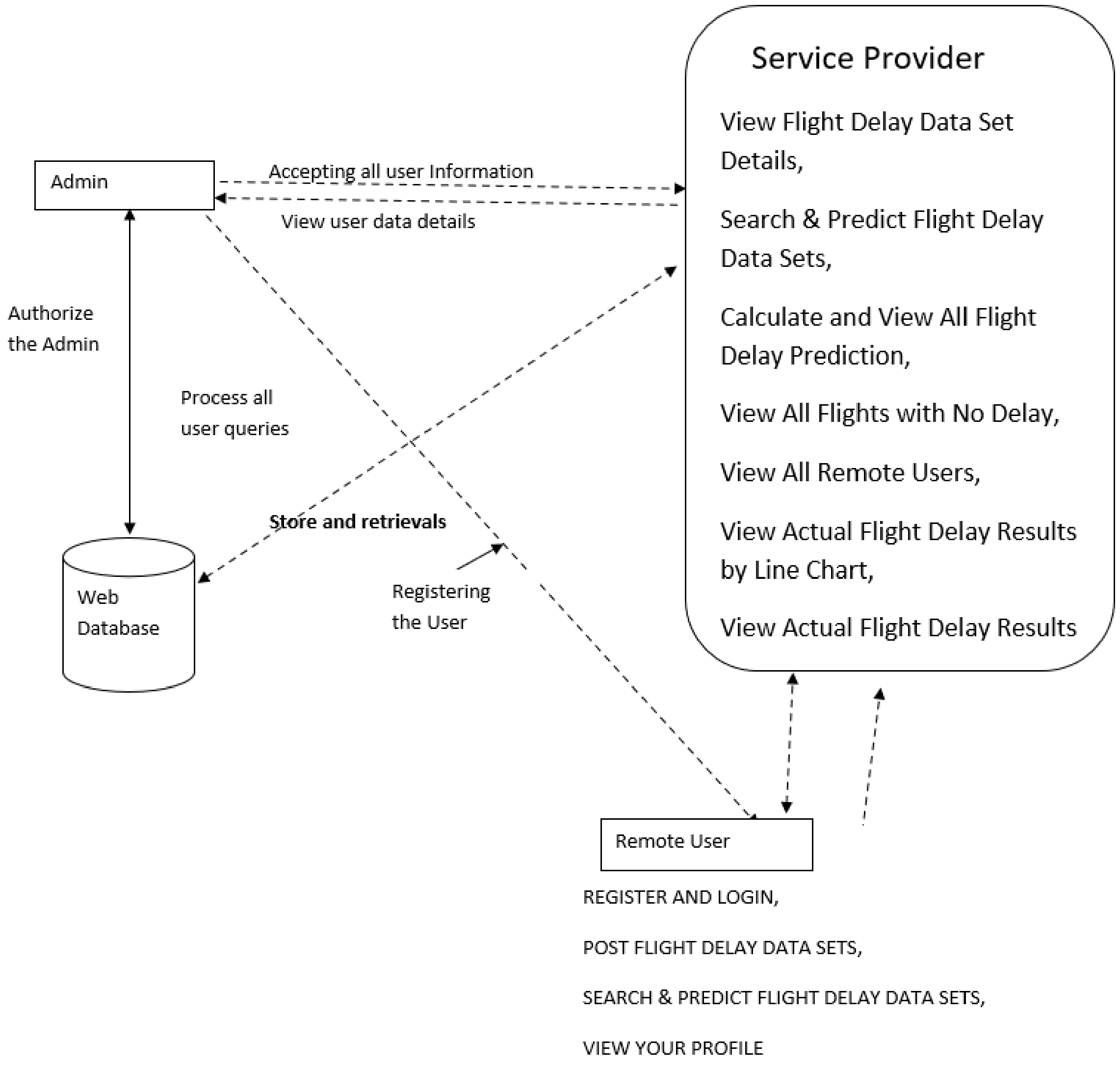


Figure 5

Service Provider

In this module, the Service Provider has to login by using valid user name and password. After login successful he can do some operations such as View Flight Delay Data Set Details, Search & Predict Flight Delay Data Sets, Calculate and View All Flight Delay Prediction, View All Flights with No Delay, View All Remote Users, View Actual Flight Delay Results by Line Chart, View Actual Flight Delay Results, View Flight Delay Prediction Results.

View and Authorize Users

In this module, the admin can view the list of users who all registered. In this, the admin can view the user's details such as, user name, email, address and admin authorize the users.

Remote User

In this module, there are n numbers of users are present. User should register before doing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user will do some operations like POST FLIGHT DELAY DATA SETS, SEARCH & PREDICT FLIGHT DELAY DATA SETS, VIEW YOUR PROFILE.

6.RESULT:

This is the search bar where a user can Enter the flight number as keyword and check if any delay for the particular flight.

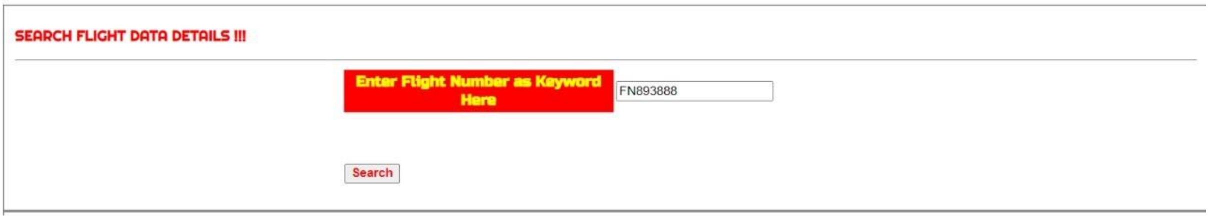
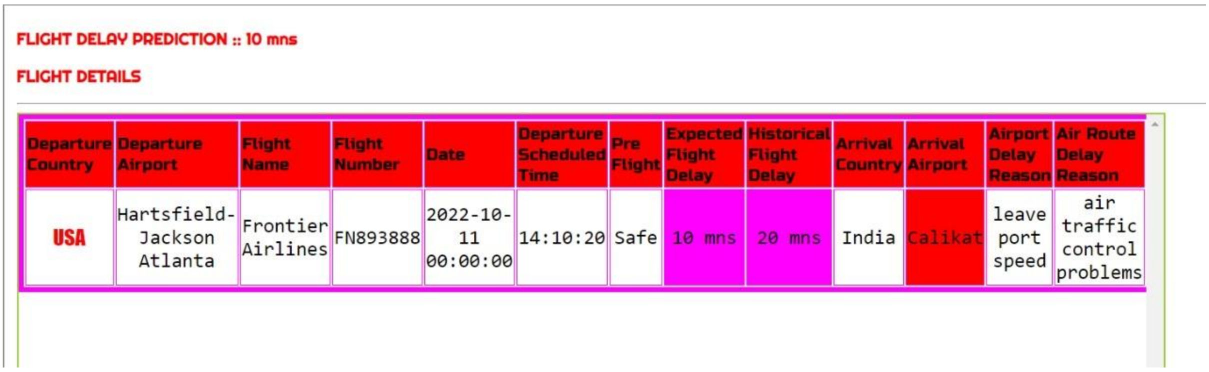


Fig 6.1: Search Flight Data Details

After entering the flight number in the search bar, the delay is predicted as shown above.


FLIGHT DELAY PREDICTION :: 10 mns

FLIGHT DETAILS

Departure Country	Departure Airport	Flight Name	Flight Number	Date	Departure Scheduled Time	Pre Flight	Expected Flight Delay	Historical Flight Delay	Arrival Country	Arrival Airport	Airport Delay Reason	Air Route Delay Reason
USA	Hartsfield-Jackson Atlanta	Frontier Airlines	FN893888	2022-10-11 00:00:00	14:10:20	Safe	10 mns	20 mns	India	Calikat	leave port speed	air traffic control problems

Fig 6.2: Flight Data Predictions

This line chart shows the actual flight delay results for four airlines (FN892737, FN890190, FN890222) throughout a day. The X-axis displays time in hours and minutes, while the Y-axis displays flight delay in minutes. Each line represents a specific airline, and data points on the lines indicate the flight delay at that time. The chart shows that flight delays fluctuate throughout the day, with some airlines experiencing more delays than others.

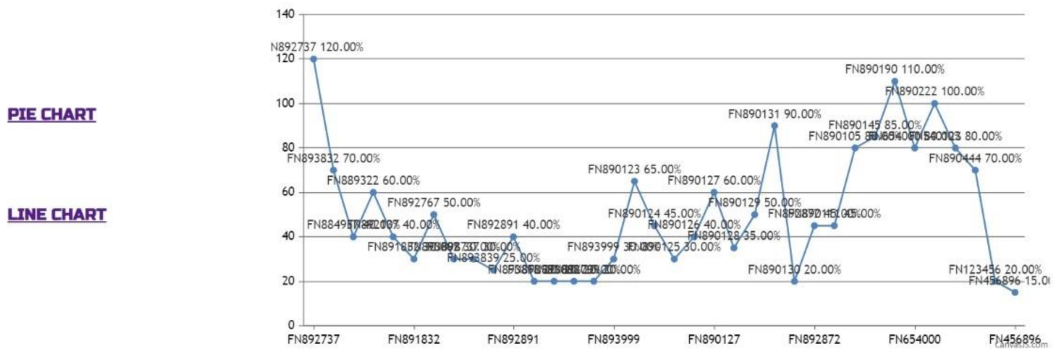


Fig 6.3: Flight Data Line Chart

This pie chart displays the percentage of flights delayed for each of 12 airlines. The data shows that FN890444 has the highest percentage of delayed flights (70%), while FN892981 has the lowest (20%). Four airlines have 60% or more of their flights delayed, while three airlines have 40% or less. This information can be useful for passengers to consider when choosing an airline, as it may indicate which airlines are more likely to experience delays.

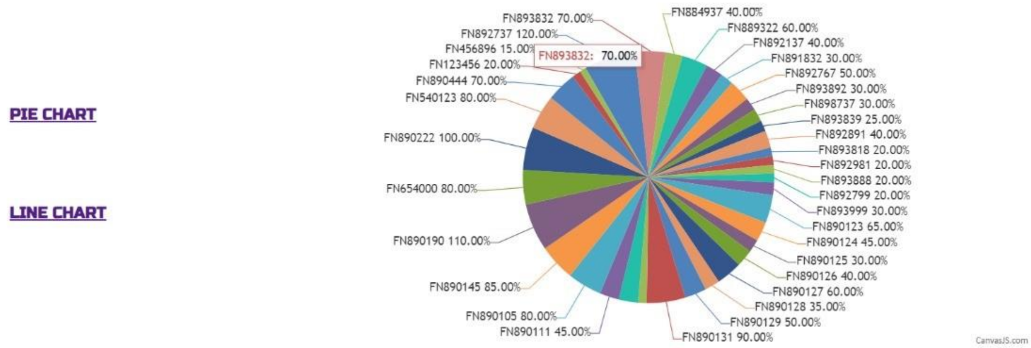


Fig 6.4: Flight Data Pie Chart

7.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 over fitting 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 more training data, integrating more information like airport traffic flow, airport visibility into our dataset, and designing more delicate networks.

8.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.

[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, “Newyork, 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.