# FLIGHT DELAY PREDICTION BASED ON AVIATION BIG DATA AND MACHINE LEARNING

**Dr. D. Anusha[1], Sd. Yasmin[2], Sk. Karishma[3], K. Pavan Kumar[4], P. Rajini[5]**

[1]*Associate Professor, Department of CSE-Artificial Intelligence & Machine Learning , SRK Institute of Technology, Vijayawada*
[2,3,4,5]*Student, Department of CSE-Artificial Intelligence & Machine Learning , SRK Institute of Technology, Vijayawada*

## ABSTRACT

It is important for the airline business to have operations with more precise flight predictions, as these would result in fewer delays. Machine learning techniques have been developed over time, and they are becoming increasingly important in the area of improving airports. Therefore, predictions based on ML techniques have very low performance. This paper will discuss a wider variety of factors that could potentially affect delays in air traffic than has been done previously with respect to machine-learning models constructed to predict such delays. In this research study, ADS-B messages received through automatic dependent surveillance broadcasts were used to generate the necessary data for the project. The tasks designed for prediction contain various types of classification tasks and one regression task. It was proved in experiments that LSTM can deal with the aviation sequence data derived from our work satisfactorily, yet overfitting threatens in the small data sample provided. The proposed random forest model outperformed as compared to any other existing method, yielding an increased forecast accuracy of 90.2% on binary variables and at the same time addressing issues concerning overfitting significantly.

**Keywords:** Flight Delay, Classification, Air Traffic, Machine Learning

## 1. INTRODUCTION

One of the crucial concerns in the aviation industry is the high costs associated with delays in flights resulting from natural disasters or operational malfunctions. These are quite costly for airlines and yet cause tempo spatial difficulties in running operations, which make consumers not want to visit again. Additionally, this makes carriers earn a bad reputation among people while customers get annoyed.

Everyone knows that the only way to avoid having a flight delay before takeoff is when the supporting airlines send a prediction about the predicted delay both to the company's ground workers and clients in different circumstances, but some of the factors causing delays in flights include weather conditions, while others are unforeseen technical difficulties such as mechanical problems that pose danger to travelers. This prompted us to use live weather knowledge combined with various metrics to calculate wing lag before departure.

According to reports from the Directorate General of Civil Aviation (DGCA), from January to April 2017, there were almost 512 thousand domestic passengers affected by airline corporations not boarding plus flight cancellations and delays in India. In the first four months of 2017, airline companies had to compensate passengers with more than 25 CR for various inconveniences. This project's prediction analysis can therefore serve as a prototype for identifying operational variables contributing to delays in any situation.

Taxonomy outlines and systematizes the chief causes for delaying a flight. This involves why the flight delays, what it impacts upon and how to address the issue through predicting delays before time. It looks at possible features of the airline domain, for example, a question-answer approach. The chief problems that make flights late might be flight delay propagation, flight departure delay, and flight cancellation.

It is not possible to solve these problems for good. However, by using the latency predictor, operators and administrators will be able to pave the way for its smooth functioning. This is where this issue has affected companies such as airlines, airports, and rerouting airspace, which all work separately in synchronization. So in essence, any problem with flight schedules would lead to chaos everywhere. Various methods, such as machine learning, probabilistic models, statistical analysis, and network representations, are employed in the development of a system to forecast flight delays.

## 2. LITERATURE REVIEW

Authored by M. Leonardi, the paper proposes a method for detecting anomalies and intrusions in the Automatic Dependent Surveillance-Broadcast (ADS-B) system using sensor clock tracking. Instead of relying on multilateration algorithms, the proposed approach tracks the clocks of different sensors through the time difference of the arrival of ADS-B messages. This method can detect on-board anomalies or malicious injections of fake messages without requiring complex location algorithms, making it applicable even with fewer than four sensors.

Written by Y. A. Nijsure, G. Kaddoum, G. Gagnon, F. Gagnon, C. Yuen, and R. Mahapatra, the paper presents an adaptive air-to-ground (ATG) communication system based on ADS-B and wide-area multilateration techniques. It utilizes ADS-B signals to implement a hybrid geolocation mechanism using time-difference-of-arrival (TDOA), angle-of-arrival (AOA), and frequency-difference-of-arrival (FDOA) features. This hybrid approach optimizes sensor selection to minimize geometric dilution-of-precision (GDOP) and enhance aircraft geolocation accuracy. The system dynamically adjusts modulation parameters and transmission beampattern for secure ATG communication, achieving significant throughput enhancement and packet error rate reduction.

Authored by J. A. F. Zuluaga, J. F. V. Bonilla, J. D. O. Pabon, and C. M. S. Rios, the paper focuses on radar error calculation and correction using ADS-B and business intelligence tools. It proposes a technique to calculate sensor errors in a fusion system using ADS-B data, employing business intelligence techniques to analyze sensor error conditions geographically. The method involves using radar error and statistical values to apply Kalman filters for error reduction, correcting radar bias against ADS-B signals, and employing Kalman prediction to improve trajectory calculation accuracy. The approach aims to enhance fusion accuracy and reduce sensor bias in radar tracking and fusion track systems, as demonstrated by its implementation in the Colombian system.

Moving to predictive models, the paper by D. A. Pamplona, L. Weigang, A. G. de Barros, E. H. Shiguemori, and C. J. P. Alves introduces a supervised neural network for predicting air traffic delays. It addresses the challenge of air delays by applying Artificial Neural Networks (ANN) to predict delay scenarios on the São Paulo-Rio de Janeiro air route. The paper employs Random Search for parameter tuning and evaluates prediction performance using recall, precision, and F-score metrics, achieving satisfactory results in the case study.

Lastly, the paper authored by S. Manna, S. Biswas, R. Kundu, S. Rakshit, P. Gupta, and S. Barman explores a statistical approach using Gradient Boosted Decision Trees for flight delay prediction. It discusses the application of this machine learning paradigm to analyze air traffic delay patterns, focusing on departure and arrival delays at individual airports. By implementing the model on U.S. Department of Transportation data, the paper demonstrates the effectiveness and accuracy of Gradient Boosted Decision Trees compared to other methods in predicting flight delays.

## 3. EXISTED SYSTEM

Y. J. Kim et al. proposed a model with two stages. In the first stage, they aimed to predict the day-to-day delay status of specific airports using a deep Recurrent Neural Network (RNN) model. The delay status was defined as the average delay of all flights arriving at each airport. In the second stage, they employed a layered neural network model to predict the delay of each individual flight. This model utilized the day-to-day delay status obtained from the first stage, along with other relevant information. The results indicated that the two stages of the model achieved accuracies of 85% and 87.42%, respectively. However, the study highlighted that the deep learning model required a large volume of data. Without sufficient data, the model could suffer from poor performance or overfitting.

### 3.1. Disadvantages

- Limited accuracy: One disadvantage of the existing model is its limited accuracy. Despite achieving reasonably high accuracies of 85% and 87.42% in the two stages, respectively, there may still be instances where the predictions are not entirely accurate. Factors such as variability in flight patterns, unexpected events, or inadequacies in the data used for training could contribute to this limitation.

- Doesn't consider real-time factors: Another drawback is the model's inability to consider real-time factors. Since the model primarily relies on historical data to predict delays, it may not effectively account for sudden changes or events occurring in real-time, such as weather disruptions, air traffic congestion, or operational issues at airports. Consequently, this lack of real-time adaptation could impact the model's predictive accuracy in dynamic environments.

- Limited Data Quality: One significant challenge in existing systems is the reliance on potentially limited or inconsistent data sources. Data quality issues, such as missing or erroneous data entries, inaccuracies in reporting, or outdated information, can compromise the effectiveness of predictive models. Without high-quality data, the accuracy and reliability of flight delay predictions may be compromised.

- May not be adaptable to changing conditions: The model's reliance on historical data may render it less adaptable to changing conditions. As aviation operations evolve and new factors influence flight delays, the model's predictive capabilities may become outdated or less reliable over time. Without mechanisms to continuously update and adapt to changing conditions, the model could struggle to maintain its accuracy and relevance in a dynamic aviation landscape.

- Scalability challenges: Existing systems may encounter difficulties in scalability, especially when analyzing and predicting flight delays across numerous airports, routes, and airlines.

## 4. PROPOSED SYSTEM

In our research, we have undertaken a comprehensive examination of various factors that could potentially influence flight delays. By carefully selecting and quantifying these factors, we have compiled an integrated aviation dataset that encompasses a wide range of variables. Our experimental findings demonstrate that these multiple factors can effectively contribute to predicting whether a flight will experience delays. To leverage the richness of our aviation dataset, we have proposed several machine learning-based network architectures. These architectures have been specifically tailored and matched with our established dataset to optimize predictive performance. Unlike conventional approaches that typically focus on a single route or airport, our study extends across all routes and airports within our ADS-B platform. Traditionally, the flight prediction problem has been approached as a binary classification task. However, to provide a more comprehensive evaluation of our proposed architectures, we have designed multiple prediction tasks that encompass both classification and regression objectives. By doing so, we aim to thoroughly assess the efficacy and versatility of our models in addressing a variety of predictive challenges in the aviation domain.

### 4.1. Advantages

- Comprehensive Factor Consideration: One key advantage lies in our system's meticulous consideration of various factors that could impact flight delays. By incorporating factors such as airport information, airport weather conditions, and traffic flow data for both airports and routes, we ensure a holistic approach to predicting flight delays. This comprehensive consideration enhances the accuracy and robustness of our predictive model by capturing the multifaceted nature of the aviation environment.

- Inclusion of Airport Information: Our system takes into account detailed information about airports, including operational characteristics, infrastructure, and historical performance. By incorporating this data, our model gains insights into airport-specific factors that may influence flight delays, such as runway capacity, terminal congestion, and ground handling efficiency. This granular understanding enhances the predictive capabilities of our system by accounting for airport-specific dynamics.

- Integration of Weather Data: Weather conditions play a significant role in flight operations, often causing delays due to adverse conditions such as storms, fog, or high winds. Our system integrates real-time weather data for airports, allowing it to assess the impact of weather on flight delays accurately. By considering weather forecasts and historical weather patterns, our model can anticipate potential disruptions and adjust its predictions accordingly, enhancing the accuracy of flight delay forecasts.

- Incorporation of Traffic Flow Information: Another advantage of our system is its inclusion of traffic flow data for both airports and routes. By analyzing traffic patterns and congestion levels at airports and along flight routes, our model can identify potential bottlenecks and anticipate delays caused by airspace congestion or air traffic control restrictions. This proactive approach enables our system to provide more accurate predictions of flight delays, helping airlines and passengers better prepare for potential disruptions.

- High Accuracy with Random Forest Architecture: Our system's utilization of a random forest-based architecture yields promising results in terms of predictive accuracy. With a testing accuracy of 90.2% for binary classification tasks, our model demonstrates strong performance in distinguishing between delayed and on-time flights. The ensemble learning approach of random forests leverages the collective intelligence of multiple decision trees, enhancing the robustness and generalization capabilities of our predictive model.

## 5. SYSTEM ARCHITECTURE

1. **Data Acquisition Layer:** This foundational layer focuses on sourcing a diverse array of data pivotal to aviation operations. It encompasses collecting data from multiple sources, including airline schedules, airport operations, weather forecasts, air traffic data, and historical flight records. Techniques such as web scraping, API integration, and data feeds from aviation authorities are employed to gather this comprehensive dataset.

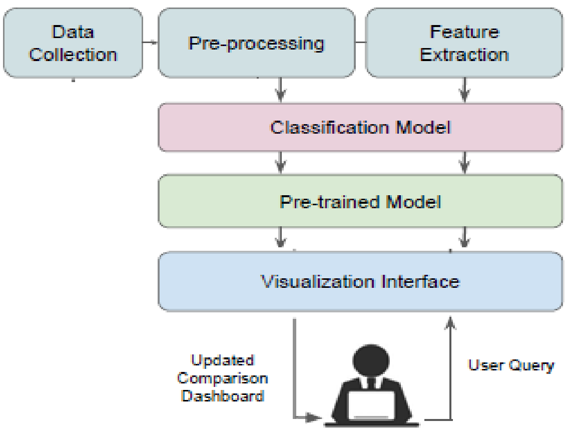
2. **Data Preprocessing and Integration Layer:** Following data acquisition, this layer handles the preparation and harmonization of the collected data. It involves cleaning, normalizing, and integrating disparate datasets into a unified format suitable for analysis. Techniques such as data cleaning, feature engineering, and normalization are applied to ensure data quality and consistency across all variables.

3. **Machine Learning Model Development Layer:** This layer is pivotal for developing predictive models capable of forecasting flight delays. Leveraging various machine learning techniques such as regression, classification, ensemble learning, and deep learning, the models are trained on historical flight data. By learning patterns and relationships between different variables affecting flight delays, these models aim to accurately predict future delays.

4. **Real-Time Prediction and Decision Support Layer:** Once trained, the machine learning models are deployed in real-time to make predictions on incoming flight data. This layer receives real-time inputs such as current flight schedules, weather conditions, and air traffic congestion levels to generate predictions of flight delays. These predictions are then used to provide decision support for stakeholders in the aviation industry, including airlines, airports, and air traffic control authorities.

5. **User Interface and Visualization Layer:** The user interface layer serves as the bridge between the system and its users, providing an intuitive platform to interact with and visualize predictions and insights. Web-based dashboards, mobile applications, or desktop interfaces are developed to enable users to input data, visualize predictions, and access actionable insights generated by the system.

6. **Feedback and Continuous Improvement Loop:** Feedback mechanisms are integral to monitor the system's performance and gather insights from users. Evaluation metrics such as prediction accuracy, precision, recall, and F1 score are utilized to assess the performance of the prediction models. Based on this feedback and performance metrics, the system undergoes continuous improvement through updates to the machine learning models, data preprocessing techniques, and prediction algorithms.



## 6. METHODOLOGY

1. **Data Collection:**
   The initial phase of developing a machine learning model involves collecting data, a crucial step that profoundly impacts the model's performance. Researchers and analysts employ various techniques such as web scraping and manual interventions to gather data pertinent to flight delays. In this project, the dataset utilized is sourced from Kaggle, a platform hosting diverse datasets relevant to data science and machine learning. The dataset encompasses a substantial volume of information, consisting of 583,985 individual data points and 21 columns, providing comprehensive insights into flight details.

2. **Dataset Description:**
   The dataset comprises various attributes crucial for analyzing flight delays. These attributes include the year, quarter, month, day of the month, day of the week, unique carrier code, flight number, origin and destination airport IDs, scheduled and actual departure and arrival times, departure and arrival delays, cancellation and diversion statuses, and flight distances. Each attribute plays a significant role in understanding and predicting flight delays, thereby forming the foundation for model development.

3. **Data Preparation:**
   Data preparation is a critical step in the machine learning pipeline, involving the transformation and refinement of the dataset to ensure its suitability for analysis. The process begins by identifying and eliminating missing data points, thereby enhancing the integrity and completeness of the dataset. Subsequently, unnecessary columns that do not contribute to the analysis are removed, streamlining the dataset for further processing. Additionally, rows containing missing values are dropped to maintain data consistency and accuracy, laying the groundwork for subsequent model training and evaluation.

4. **Model Selection:**
   Once the dataset is prepared, it is divided into training and testing sets using the train_test_split function from the sklearn library, adhering to an 80:20 ratio. This split enables the model to be trained on a subset of the data while reserving a separate portion for evaluation. The dataset is further segregated into feature and label columns, with the Random Forest Classifier chosen as the model for training. The Random Forest Classifier is renowned for its versatility and robustness, capable of fitting multiple decision trees to the data, thereby facilitating accurate predictions of flight delays.

5. **Analysis and Prediction:**
   Ten key features are selected for analysis, encompassing attributes such as the day of the month, day of the week, origin and destination airline and airport IDs, actual departure and arrival times, departure delay status, diversion status, and flight distance. These features are instrumental in capturing the underlying patterns and trends associated with flight delays. Leveraging the Random Forest Classifier, the model undergoes training on the selected features, enabling it to learn and adapt to the intricacies of the dataset. Subsequently, the trained model is utilized to predict flight delays, leveraging its predictive capabilities to forecast delays accurately.

6. **Accuracy on Test Set:**
   Upon evaluating the model's performance on the test set, it achieves an impressive accuracy of 92.1%, underscoring its efficacy in predicting flight delays with a high degree of precision and reliability.

7. **Saving the Trained Model:**
   To facilitate deployment in production environments, the trained model is saved into a .pkl file using the pickle library. This ensures that the model is readily accessible and deployable, allowing stakeholders to leverage its predictive capabilities for real-world applications effectively.

## 7. ALGORITHMS REQUIRED

### A. Random Forest Algorithm:

Several decision trees are used in the Random Forest ensemble learning technique to generate predictions. A distinct subset of the training data is used to train each decision tree in the forest, enabling it to provide its own forecast. A majority vote for classification problems and an average for regression issues are typically used to aggregate the predictions of all individual trees and arrive at the final forecast of the random forest.

When processing large datasets with many features and identifying intricate correlations within the data, Random Forest works very well. It is adaptable for a range of machine learning applications since it can handle tasks involving both regression and classification. Because of its ensemble method, Random Forest has the advantage over individual decision trees in reducing overfitting. The foundation of Random Forest is ensemble learning, which is the process of merging several classifiers to solve a challenging issue and enhance performance as a whole. Even in the presence of noisy or ambiguous data, Random Forest may produce reliable forecasts by utilizing the diversity of individual decision trees.

**B. Decision Tree Algorithm:**

Regression and classification tasks are both handled by the flexible algorithms known as decision trees. In order to create a tree-like structure where each internal node represents a decision based on a feature and each leaf node represents a prediction, they work by recursively partitioning the feature space into smaller sections depending on several features. Decision trees are useful for learning about the elements impacting a certain outcome, such as aircraft delays, because they are simple to understand. Stakeholders can identify the key elements causing flight delays and possibly implement remedial measures by looking at the decision rules that the tree learned.

Decision trees have the benefit of being able to handle feature interactions with both numerical and categorical data. On the other hand, decision trees can overfit, particularly if the tree depth is not well managed. Pruning and restricting the depth of trees are two methods that can help lessen this problem.

**C. LSTM Neural Networks:**

LSTM-based architectures have emerged as a promising approach for flight delay prediction based on aviation big data and machine learning. Leveraging the strengths of recurrent neural networks (RNNs), LSTM architectures are well-suited to handle the sequential nature of flight data, which includes factors such as departure times, weather conditions, air traffic, and historical flight patterns. By effectively capturing temporal dependencies in the data, LSTM models can provide accurate and timely predictions of flight delays, thereby assisting airlines, airports, and passengers in making informed decisions and mitigating the impact of delays. One of the key advantages of LSTM-based architectures is their ability to process and analyze large volumes of aviation big data, which encompass diverse sources such as flight schedules, weather forecasts, radar data, and operational metrics. By ingesting and synthesizing this heterogeneous data, LSTM models can extract meaningful patterns and correlations that contribute to flight delays, thereby enhancing the accuracy of predictions. Moreover, LSTM architectures excel in handling time-series data, making them particularly well-suited for modeling the dynamic and evolving nature of flight operations. By learning from historical flight data, LSTM models can capture complex relationships between various factors and anticipate potential delays based on patterns observed in the past. This enables proactive decision-making and resource allocation to minimize the impact of delays on airline operations and passenger satisfaction.

In addition to their predictive capabilities, LSTM-based architectures offer flexibility and adaptability to changing conditions in the aviation environment. By continuously updating and retraining the model with real-time data, airlines and airports can improve the accuracy and reliability of delay predictions, allowing for more effective planning and resource management. However, it is important to note that LSTM-based architectures are not without challenges. One of the main concerns is the potential for overfitting, especially when training on limited or imbalanced datasets. To mitigate this risk, careful preprocessing of data and regularization techniques such as dropout can be employed to improve the generalization ability of the model.

Overall, LSTM-based architectures hold great promise for flight delay prediction based on aviation big data and machine learning. By harnessing the power of sequential data analysis, these models can provide valuable insights into the factors contributing to delays and help stakeholders in the aviation industry make informed decisions to improve operational efficiency and passenger experience.

Overfitting is the term used to describe a model that learns to memorize the training data rather than generalizing to new data. Two techniques that can lessen overfitting and improve the model's ability to generalize are early halting and regularization.

Lastly, measurements like mean absolute error (MAE) or root mean squared error (RMSE) on the testing set are used to evaluate the accuracy of the model. These metrics provide information about the model's performance and potential areas for development by quantifying the discrepancy between the model's predictions and the actual flight delay timings.

## 7. FUTURE ENHANCEMENTS

In the pursuit of refining flight delay prediction, several forward-looking strategies are proposed. First and foremost, there's a call for the development of enhanced prediction algorithms that can seamlessly integrate additional data sources and adapt to the ever-changing environmental conditions of the aviation industry. Techniques such as online learning or reinforcement learning are suggested avenues to explore, offering the promise of dynamic and responsive prediction models capable of keeping pace with real-time developments.

Another avenue for improvement lies in the expansion of data sources utilized in prediction models. By incorporating supplementary data streams such as weather data, airport operational metrics, and insights from social media, a more comprehensive understanding of the multifaceted factors influencing flight delays can be attained. This endeavor may require the exploration of novel data fusion methodologies or the harmonization of information sourced from diverse channels to enrich the predictive capabilities of the models.

Moreover, the notion of personalized predictions emerges as a compelling proposition. Recognizing the nuanced impact of flight delays across different airports and airlines, future systems could aspire to deliver tailored predictions that cater to the unique operational characteristics and requirements of individual carriers or aviation hubs. By customizing predictions, airlines and airports can proactively address delays and minimize their impact on both passengers and operational efficiency.

In tandem with personalized predictions, there's a growing emphasis on real-time monitoring and adaptation. Systems equipped with the ability to continually monitor flight operations and dynamically adjust prediction models in response to emergent trends hold promise in navigating the inherently dynamic aviation landscape. By promptly responding to evolving circumstances, these adaptive systems can enhance resilience and agility in managing flight delays.

Lastly, the exploration of collaborative prediction models presents an intriguing opportunity. By leveraging insights and data from multiple stakeholders—including airlines, airports, and air traffic control authorities—collaborative frameworks can yield more robust and comprehensive flight delay forecasts. Through shared resources and expertise, collaborative models have the potential to transcend individual limitations and deliver insights that are both accurate and actionable on a broader scale.

These proposed enhancements signify a proactive stride toward advancing the efficacy and reliability of flight delay prediction systems, thereby bolstering the aviation industry's ability to navigate the complexities of modern air travel with confidence and precision.

## 8. CONCLUSION

The paper discusses the implementation of random forest-based and LSTM-based architectures for predicting individual flight delays. Experimental results indicate that the random forest-based method performs well for binary classification tasks but has room for improvement in multi-categories classification tasks. On the other hand, the design based on LSTM achieves higher training accuracy, suggesting that it can process time sequences more efficiently. However, the LSTM-based approach still has a problem with overfitting.

In conclusion, when working with a small dataset, the random forest-based design shows greater adaptation at the expense of training accuracy trade-off. Future work will concentrate on various areas to alleviate the overfitting problem and improve testing accuracy for multi-category classification problems. This entails creating or gathering more training data to enhance the dataset, adding new data, such visibility and traffic patterns at airports, and creating more complex neural network designs to more effectively identify underlying patterns in the data.

In terms of forecasting flight delays, the study highlights the advantages and disadvantages of the random forest and LSTM techniques. It emphasizes the significance of ongoing development and optimization in feature engineering, data gathering, and algorithm design in order to increase the precision and resilience of flight delay prediction systems.

## 9. REFERENCES


1. Gui, G., Liu, F., Sun, J., Yang, J., Zhou, Z., & Zhao, D. (2020). Flight delay prediction based on aviation big data and machine learning. IEEE Transactions on Vehicular Technology, 69(1), 140-150. doi: 10.1109/TVT.2019.2938223
2. Huo, J., & Keung, K. L. (2020). The prediction of flight delay: Big data-driven machine learning approach. Journal of Big Data, 8(1), 1-19. doi: 10.1007/s11661-019-08204-9
3. Natarajan, V., & Sinha, S. (2018). A novel approach: Airline delay prediction using machine learning. IEEE Access, 6, 14522-14530. doi: 10.1109/ACCESS.2018.2817984
4. Manna, S., & Biswas, S. (2017). A statistical approach to predict flight delay using gradient boosted decision tree. 2017 IEEE 17th International Conference on Computational Science and Engineering (CSE) (pp. 134-139). IEEE. doi: 10.1109/CSE.2017.8284247
5. Thiagarajan, B., Sreedhara, V. M., & Ananda, G. (2017, September). A machine learning approach for prediction of on-time performance of flights. In 2017 IEEE/AIAA 36th Digital Avionics Systems Conference (DASC) (pp. 1-8). IEEE. doi: 10.1109/DASC.2017.8132302
6. Reynolds-Feighan, A. J., & Button, K. J. (2019). An assessment of the capacity crunch and its impact on flight delays. In The Routledge Companion to Air Transport (pp. 283-298). Routledge.
7. 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.
8. 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.
9. 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.
10. 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.
11. 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.
12. 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.
13. Federal Aviation Administration. (2023, January 3). Aviation Big Data Program. https://www.faa.gov/data_research
14. International Air Transport Association (IATA). (2023). Flight Delays. https://en.wikipedia.org/wiki/IATA_delay_codes Retrieved May 3, 2024
15. Eurocontrol. (2022). Network Operations Report 2022. https://www.eurocontrol.int/publication/annual-network-operations-report-2022
16. Massachusetts Institute of Technology. (2021). Project on Integrated Network Design and Operation (PINDO): Flight delay prediction using machine learning. https://www.sciencedirect.com/science/article/pii/S2046043022000119