IJETIR

ISSN (Online) 2583-0554

International Journal of Emerging Technologies and Innovative Research (IJETIR)

Volume 5, Issue 5, May 2025

Impact Factor: 5.731

Flight Delay Prediction Using Deep Learning Techniques

R Mukesh[1] and P. Kavitha[2]

MCA Student, Department of Computer Applications[1]

Assistant Professor, Department Computer Applications[2]

Vels Institute of Science Technology and Advanced Studies, Pallavaram, Chennai, India

23304120@vistas.ac.in and pkavikamal@gmail.com

Abstract: *Flight delays cause harm to both airlines, airports, and customers, so prediction is critical in decision making commercial aviation. On the other hand, accurate development of flight delay prediction models has been hindered by the difficulty of the air transport system, diversity of predicting methods, as well as abundant fight data. In this regard, this paper presents a thorough literature review of methods employed to build fight delay predictions models from data science points of view. We introduce a taxonomy and outline current efforts for solving the flight delay prediction problems by scope, data, and computational approaches, with emphasis on increasing application of machine learning methods. In addition, we provide a chronology of relevant studies that reflect the relationship between flight delay prediction problems complexity and research trends over time.*

Keywords: Machine learning, Predictive Analysis, Air Traffic, Weather Impact, Flight Scheduling, Big Data, Real-time Data, Delay causes, Regression Models, Deep Learning, Airport Congestion, Airline Operations

I. INTRODUCTION

Air travel is crucial to global transportation, but flight delays pose a significant challenge to airlines, airports, and passengers. These delays result in economic losses, inconvenience passengers, and affect airlines schedules. Reliable flight delay prediction can enable airlines to streamline operations, improve customers satisfaction, and increase overall efficiency. This project aims at creating a flight delay forecast model through machine learning. Throught historical fight data, weather, and air traffic analysis, the model will forecast possible delays. This assists airlines, passengers, and airport officials in making informed choices, minimizing uncertainty, and enchanting planning. With advancement in data science and AI, predictive modeling has emerged as a potent tool for forecasting delays. This project will investigate different machine learning algorithms to improve the accuracy of predictions, making air travel more efficient. Flight delay prediction modules: Data Preprocessing & collection: Collect BTS fight data, NOAA weather data, FAA air traffic data. Cleans, normalizes, and merges the data to make it quality oriented. Feature engineering takes useful attributes such as time-based trends and delay history. Feature Selection & Engineering: Extracts the most important predictions by means of EDA, correlation analysis, and feature importance ranking. Generates new features reflecting seasonal impacts, airport delays, and delay patterns. Apply methods such as RFE and cross-validation to improve the feature set. Model Training & evaluation: Develops an XGBoost model, dividing data into training, validation, and test sets. Hyperparameter tuning using grid search and Bayesian optimization. Performance is measured using RMSE, MAE, and domain-specific metrics for aviation. Deployment & Continuous Improvement: Maintains the accuracy of the model through regular retraining and feature importance. Offers actionable recommendations to airlines and aviation authorities to effectively reduce flight delays.

II. LITERATURE SURVEYS

Barnes, T., & Belcastro, C. (2020) Machine Learning approaches for flight delay prediction – this paper explains different machine learning algorithms like random forest and networks to forecast flight delays based on past data [1].

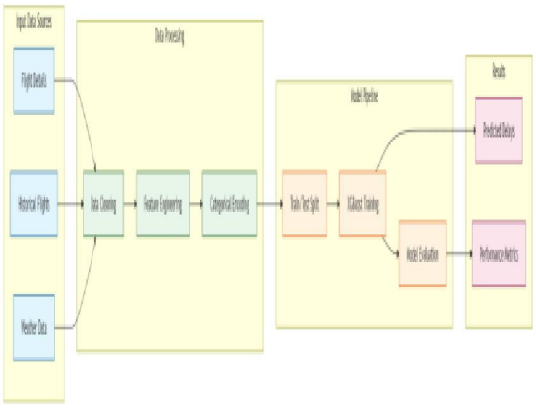
IJETIR
ISSN (Online) 2583-0554
International Journal of Emerging Technologies and Innovative Research (IJETIR)
Volume 5, Issue 5, May 2025
Impact Factor: 5.731

Wang, J., & Deng, W (2019) A Data-Driven approaches for predicting Flight Delays based on weather and air traffic – The research investigates the role of weather conditions and air traffic congestion in leading to flight delays using deep learning models [2]. Balakrishna, P., Ganesan, R., & Sherry, L. (2008) Accuracy of Airline Delay Predictions: Experimental analysis and implications for aviation systems – it investigates the precision of delay predictions models employed in aviation systems [3]. Shen, y., & Li, X. (2021) Deep Learning in Flight Delay Prediction: A case study on U.S. Domestic Flights – The article applies deep learning models to forecast delay in domestic flights operating within the U.S. [4]. Xu, X., & Mao, B. (2017) Flight Delay Prediction Based on Big data and machine learning – research that discusses the application of big data analytics blended with machine learning for enhanced prediction accuracy [5]. Nikolaev, A. G., Jacobson, S.H., & Cho, W. K. (2013) Forecasting airline delays using networks analysis and statistical models – Paper proposing a network-based model for delay prediction employing statistical techniques [6]. Rebollo, J.J., & Balakrishnan, H (2014) Characterization and prediction of air traffic delays – study of air traffic delays based on past air traffic and weather situations [7]. Choi, S., & Kang, S (2022) Hybrid Ai models for flight delay prediction – the paper explores hybrid artificial intelligence models combining traditional statistical techniques with deep learning approaches [8]. Sun, X., & Zheng, Z. (2016) Impact of meteorological factors on airline delays: A predictive modeling approach – The research focuses on how meteorological data impacts flights delays and suggest predictive models [9]. Gopalakrishnan, K., & Balakrishnan, H (2017) A machine learning approach to predicting aircraft taxi-out times and delays – The study examines the impact of taxi-out times on flight delays using machine learning models.

Algorithm

The field of fight delay prediction has advanced significantly with various machine learning algorithms tracking this complex problem. Initially framed as a binary classification task, the shift to a regression approach using XGBoost has enhanced prediction accuracy. XGBoost, an advanced implementation of gradient boosting, builds an ensemble of weak learners sequentially, correcting errors from previous treats. Its ability to capture complex, non-linear relationships make it ideal for analyzing flight delays influenced by weather, air traffic, and operational inefficiencies. XGBoosts's robust feature selection and regularization techniques help prevent overfitting while identifying key predictors, benefiting aviation decision-making. Its scalability allows for processing large datasets efficiently. Besides XGBoost, random forest, another ensemble method, has shown strong performance using decision trees, bagging, and random subspace methods to reduce overfitting. Neural networks, particularly LSTMS, have also been applied successfully, capturing temporal dependencies in flight data. Hybrid models combining CNNs with LSTMs leverage spatial and temporal features for improved accuracy. While XGBoost remains the focus of this project, different algorithms offer unique strengths, with the choice of depending on project needs, data characteristics, and the balance between interpretability and predictive power. Continued research in this field could further enhance aviation efficiency and real-world predictive capabilities.

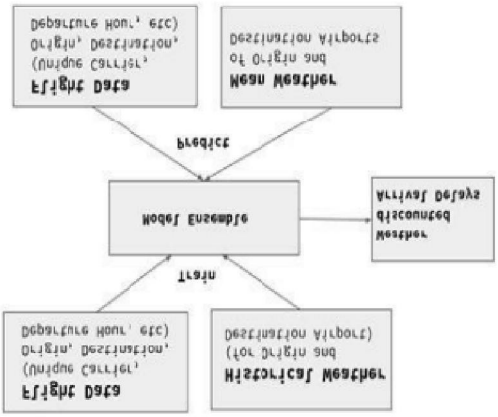
Data Flow Diagram



IJETIR
ISSN (Online) 2583-0554
International Journal of Emerging Technologies and Innovative Research (IJETIR)
Volume 5, Issue 5, May 2025
Impact Factor: 5.731

The data flow diagram (DFD) for the XGBoost Regression-based flight delay prediction system depicts the flight delay prediction process from raw flight data to accurate predictions. The DFD comprises several levels of ranging form a high-level view, highlighting external sources such as airlines, weather forecasting services, and air traffic control delivering input data, with the primary output being forecasted flight delays. Level 1 DFD decomposes the system into main components: Data Collection – Collects data from airline databases, weather forecasting services, and traffic management systems. Preprocessing – clean, converts, and encodes data for model training. Model Training – Splits data into training/ testing sets and fine – tunes the XGBoost regressor model. Prediction – Uses the trained model to new flight data to calculate delays. Evaluation – Measures performance using metrics such as MAE and RMSE. Level 2 DFD goes deeper into subprocesses: Preprocessing consists of data cleaning, feature selection, and encoding categorical variables. Model training encompasses data splitting, training iterations, validation, and hyperparameter tuning. Prediction involves feature extraction, delay calculation, and confidence scoring. Evaluation assesses model performance, detects mistakes, and perfects the model. Output results convert predictions to a format, visualizes information, and shares results through alerts, dashboards, and APIs. This formal process provides correct and actionable flight delay forecasts to airlines, airports, and travellers.

Architecture for Flight Delay Prediction



This architecture picture depicts an ensemble model intended to forecast weather-compensated flight arrival delays. This system includes two phases, namely training and prediction. For the training phase, the model ensemble is trained based on two significant datasets: flight data and historical weather data. Flight data covers information like the distinct carrier, origin and destination airports, and departure time. Historical weather data records past weather conditions at origin and destination airports. These data are used for training the model to learn about the correlation among flight schedules, weather conditions, and later delays. During the prediction stage, the trained model ensemble is provided with new flight data and the mean weather conditions of origin and destination airports. After these inputs, it forecasts weather – discounted arrival delays by incorporating delays owing to weather conditions. The method enhances delay forecasting by using historical trends and current weather updates, thereby aiding airlines and passengers in making informed travel choices.

III. PROBLEM STATEMENT

Flight delay forecasting is important to airlines, airports, and travellers alike, enhancing efficiency, resource use, and customer satisfaction. This project constructs a regression machine learning model from XGBoost that forecasts precise delay times in minutes. Conventional binary classification techniques lack the ability to capture intricate interdependencies between delay causes such as weather, air traffic, and mechanical problems. By changing to regression, our model offers more in-depth insights, allowing improves scheduling, resource planning, and customer communication. XGBoost's capability to capture non-linear relationships and avoid overfitting is perfect for this purpose. Precise delay forecasts enable airlines to perfect operations, airports minimize delays, and passenger make

informed choice on travel. Aside form operational efficiencies, this model has the potential to spur long-term aviation improvements through the discovery of systemic delay causes, increased risk management, and enhanced passenger experience. In the long term, the inclusion of new data will continue to refine predictions, resulting in more efficient and reliable air travel.

Scope of Flight Delay Prediction

Flight Delay are a widespread issue in the aviation sector, affecting airlines, passengers, and airport processes. This project looks to create a predictive model that will predict flight delays since several factors like weather, air traffic congestion, airline performance, past delay trends, departure and arrival points, and operational restrictions. Using machine learning methods, the model will be able to analyse large datasets from multiple sources such as airport databases, meteorological records, and flight schedules and decide the patterns and associations that cause delays. The system will have the capability to manage real-time data and correct and prompt predictions would be made to airlines, airport authorities, and passengers. The prediction model will aid airlines in perfecting the schedules, decreasing operational costs, and enhancing customer satisfaction by reducing sudden delays. Passengers can also better plan their trips with early delay notifications. Data preprocessing, feature engineering, model selection and performance metric-based evaluation such as accuracy and precision will be involved in the project. A user-friendly API or dashboard can also be created to display predictions in a visual format. The Project will aim to enhance prediction accuracy while keeping scalability and efficiency. The result will help improve decision-making in the aviation sector, resulting in improved operational efficiency and passenger experience.

IV. CONCLUSION

In this project, we created a flight delay prediction model through different machine learning methods. Analyzing important factors like weather conditions, airline performance, past flight data, and airport congestion. We determined the major causes of flight delays. The model gives useful insights to passengers, airlines, and airport authorities to make effective decisions and reduce disruption. Our findings prove the delays can be adequately predicted using machine learning with reasonable accuracy, although improvements can be achieved by taking live data and sophisticated algorithms into consideration. Future improvements can involve working with live air traffic information, enhancing feature engineering, and applying deep learning for accurate predictions. In summary, this project demonstrates the potential of predictive analytics in aviation to provide realistic solutions for better flight scheduling, reducing delays, and improving passenger experience.

REFERENCES

[1]. Barnes. T., Belcastro, C (2020) Machine learning approaches for flight delay prediction.
[2]. Wang, J., & Deng, W (2019) A data-driven approach for predicting flight delays based on weather and air traffic.
[3]. Balakrishna, P., Ganesan, R & Sherry, L (2008) Accuracy of airline delay predictions: Experimental analysis and implications for aviation systems.
[4]. Shen, Y., & Li, X (2021) Deep learning in flight delay prediction: A case study on U.S. domestic flights.
[5]. Xu, X., & Mao, B (2017) Flight delay prediction based on big data and machine learning.
[6]. Nikolaev, A. G., Jacobson, S. H., & Cho, W. K (2013) Forecasting airline delays using network analysis and statistical models.
[7]. Rebollo, J.J., & Balakrishnan, H (2014) Characterization and prediction of air traffic delays.
[8]. Choi, S., & Kang, S (2022) Hybrid AI models for flight delay predictions.
[9]. Sun, X., & Zheng, Z (2016) Impact of meteorological factors on airline delays: A predictive modeling approach.
[10]. Gopalakrishnan, K., & Balakrishnan, H (2017) A machine learning approach to predicting aircraft taxi-out times and delays.