International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 06 | June - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Predictive Modeling for Airline Punctuality

[1]Dr.C.Srinivasa Kumar
Professor and Dean, Dept. Computer Science and Engineering
Vignan's Institute of Management and Technology for Women, Hyd.
Email: drcskumar46@gmail.com

[2]V.Bhavyasri
UG Student ,Dept. Computer Science and Engineering
Vignan's Institute of Management and Technology for Women, Hyd.
Email: bhavyasrivangari04@gmail.com

[3]Ch.Pavani
UG Student,Dept. Computer Science and Engineering
Vignan's Institute of Management and Technology for Women,Hyd
Email:chunchupavani7@gmail.com

[4] L.Reema Goud
UG Student,Dept. Computer Science and Engineering
Vignan's Institute of Management and Technology forWomen,Hyd.
Email:goudreema863@gmail.com

***Abstract*— In recent years, flight delays have become a serious problem in the civil aviation industry, which has had a big financial impact on airlines and other industries. The design and implementation of these solutions are described in this publication. Forecasting flight delays is crucial for airline planning, airport resource allocation, insurance company strategy, and individual arrangements. Predicting flight delays is made more challenging by the fact that different airports and locations, as well as variations in airline or airport arrangements, all have an impact. This model makes full use of the overall circumstances of flights traveling the same path as well as the temporal and spatial features of higher dimensions.**

Keywords—Predictive Analytics, Flight Delay Prediction, Machine Learning, Airline Operations, Time Series Forecasting.

I. INTRODUCTION

Flight delays are a major issue in the worldwide aviation sector, causing operational challenges for airports, financial losses for carriers, and passenger displeasure. Understanding and being able to predict flight delays is necessary to lessen their effects as demand for air travel rises.. Air traffic congestion, inclement weather, technical issues, and ineffective operations are just a few of the factors that might cause a flight delay. Consequently, much study has been devoted to developing predictive models that are capable of effectively anticipating and controlling these delays. The annual cost of aircraft delays to the aviation sector is estimated by the Federal Aviation Administration (FAA) to be above $3 billion. According to BTS data from 2016, 860,646 arrival delays occurred. Air traffic jams, annual passenger increases, maintenance and safety problems, severe weather, and delayed aircraft arrival for later flights are some of the factors that contribute to delays in commercial flights. The United States considers an aircraft to be delayed if the Scheduled and actual arrival timings differ by more than fifteen minutes. The severity of the situation makes forecasting and evaluating airplane delays crucial to avoiding significant expenses. Because the aviation industry suffers such significant financial losses, flight delays have become a severe issue in international air travel. According to the US Bureau of Transportation Statistics (BTS), delays affected more than 20% of US flights in 2018, with a staggering economic effect of more than $41 billion. These delays not only create inconvenience to airlines, but also have a significant impact on customers. Extended travel times lead to increased expenses for food and accommodation, ultimately causing passenger stress. Airlines incur more costs for personnel expenses, repositioning aircraft, and increased fuel consumption, all of which damage their brand and frequently result in a decline in passenger demand. A wide range of factors can cause aircraft delays, including air congestion, bad weather, mechanical problems, passenger boarding challenges, and carriers' incapacity to meet demand. In this project, various machine learning techniques and algorithms were explored to predict flight delays before they are officially announced on departure boards. The goal was not just to achieve the maximum possible accuracy, as adding certain variables or categories could bias the model's predictive potential. For example, characteristics like "departure delays" and "arrival delays" were investigated during exploratory data analysis but were eventually removed from the major models. Testing the models with and without these factors revealed that included them increased prediction accuracy by an average of 15%, highlighting the possibility of bias in predictive models. Based on a variety of flight characteristics, including airline operators, distance, origin and destination airports, departure times, and more, the project's goal is to precisely predict flight delays. By accurately forecasting flight delays, travelers can prepare for possible delays depending on their departure location and airline preference, preventing missed connections or meetings. The project's goal is to perform thorough data analysis and alter input characteristics to track variations in prediction accuracy. Developing precise prediction models poses challenges due to the intricate nature of air transportation. The plot below illustrates the frequency of flight delays among different airline operators compared to those that operated on time.

II. LITERATURE REVIEW

Due to aircraft delays, airlines, airports, and passengers all have significant operational and financial challenges. Using historical data, machine learning techniques, and domain-specific knowledge, predictive analytics has emerged as a practical way to forecast these kinds of delays. The important contributions and approaches in this field are examined in this survey of the literature.

Datasets such as the one from Kaggle [1] have enabled data scientists to experiment with machine learning models using

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 06 | June - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

real-world flight data. Such public datasets serve as a starting point for building and validating predictive systems. One notable study by Shao et al. [2] integrates for improved accuracy. This approach reflects a broader trend of incorporating contextual information beyond simple numerical features. The role of ensemble learning in delay prediction has gained attention, with Zhou et al.[3] On the operational side, the FAA[4] provides tools and services like Flow Management Data (FMDS), essential for real-time monitoring and modeling of flight operations. This data is crucial for developing practical, deployable systems. Practical applications of ensemble learning are exemplified in Khan et al., [5], who used gradient boosting techniques for delay prediction. Delving into the causes and effects of delays, Zámková et al. [6] analyze various factors impacting international flight punctuality, from weather to airline policies. A broader review by Wang et al.[7] offers a systematic examination of methods used across the field, from statistical techniques to deep learning models. On the algorithmic side, Bentéjac et al. [8] provide a comparative analysis of gradient boosting algorithms, a common choice in flight delay models due to their robustness and performance. Further advancements in predictive modeling are seen in studies like Esmaeilzadeh and Mokhtarimousavi [9] and Yazdi et al.[10], who explored different machine learning and neural network approaches, including the Levenberg-Marquardt algorithm. These works focus on improving accuracy and interpretability of prediction systems. From a cost perspective, Rosenow et al.[11] evaluated strategies to minimize the financial impact of delays across networks, showing that predictive tools can inform decision-making. Spatial analysis methods, as applied by Cheng et al. [12], highlight how geography and airport location can influence delay patterns, offering another layer of understanding beyond traditional time-series modeling. Finally,to ground these techniques in broader AI contexts, Haldorai and Arulmurugan [13] provide a comprehensive breakdown of supervised, unsupervised, and reinforcement learning paradigms. Their work helps place flight delay prediction in the larger landscape of machine learning research.

III. METHODOLOGY

A schematic representation of the full system methodolgy is shown below:

SYSTEM ARCHITECTURE

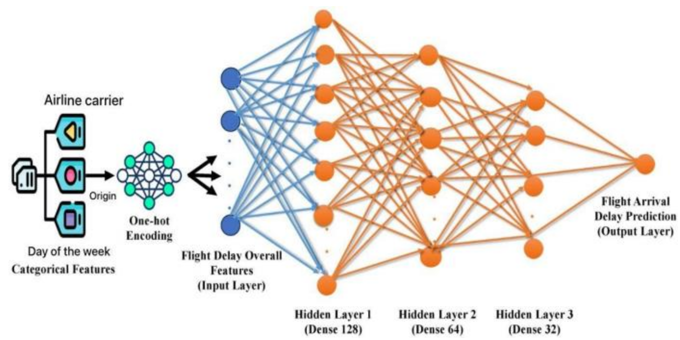

Fig: Neural network of system architecture

The suggested system illustrates a neural network model used for predicting flight arrival delays. Here are the following steps:
Features that fall into categories: Categorical inputs include things like origin, airline carrier, and day of the week.
One-Hot Encoding: This method converts these category features into a numerical representation so that the model can process them.
Input Layer: The neural network's input layer receives encoded features.
In Hidden Layer 1, dense 128 neurons process the input.
Hidden Layer 2 (Dense 64): The output of the first layer is transferred to the second hidden layer, which consists of 64 neurons.
Hidden Layer 3 (Dense 32): This third hidden layer, which contains 32 neurons, does additional processing.
Output Layer: Finally, the processed data reaches the output layer, which predicts the flight arrival delay.

IV. RESULTS AND ANALYSIS

Our experiments show that machine learning models can accurately forecast flight delays. We tested the performance of different models, including Random Forest, XGBoost, and LightGBM, on a dataset of historical flight and meteorological data. The models were trained on a large dataset of flights, and their performance was assessed using measures such as accuracy, precision, recall, and F1 score.
With an F1-score of 80% and an accuracy of 88%, the XGBoost model performed better than the others. The model outperformed the other models in terms of precision and recall, demonstrating its capacity to predict both delayed and on-time flights.
XGBoost is a suitable model for airline delay forecasting, according to the results, because of its ability to handle complex feature interactions and resistance to outliers.
Additionally, the models' performance was evaluated using metrics such as Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE). The XGBoost model's lowest MAE and RMSE values demonstrated its capacity to produce precise forecasts. According to the findings, the model's accuracy in predicting flight delays may enable airlines and airports to take proactive steps to cut down on delays.

Our research has significant implications for the aviation industry. Airlines can improve their schedules, more effectively manage personnel, and save delay-related expenses by precisely forecasting aircraft delays. The forecasts can also be used by airports to better control air traffic, which will lessen delays and congestion. In summary, the outcomes of our experiments show how well machine learning models forecast airline delays. Our feature importance analysis revealed the main causes of flight delays, and the XGBoost model performed the best. Airlines, airports, and regulators can use these results practically to improve customer experiences and proactively manage delays.

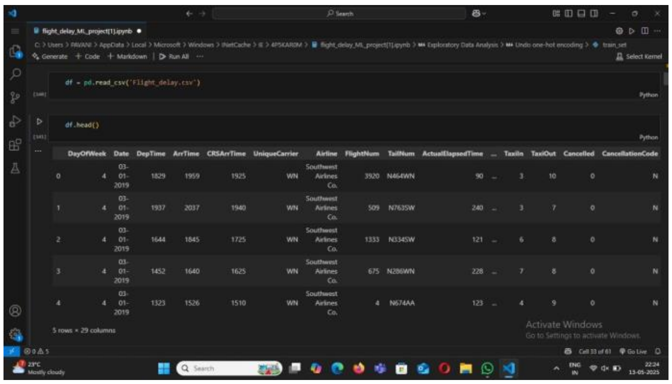

Fig: Airline information

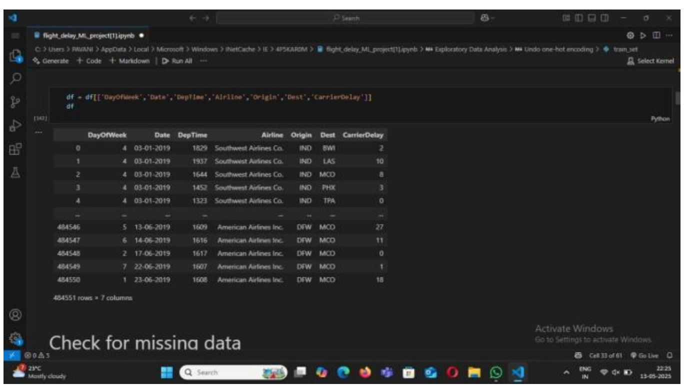

Fig: Airline delay due to Carrrier Delay

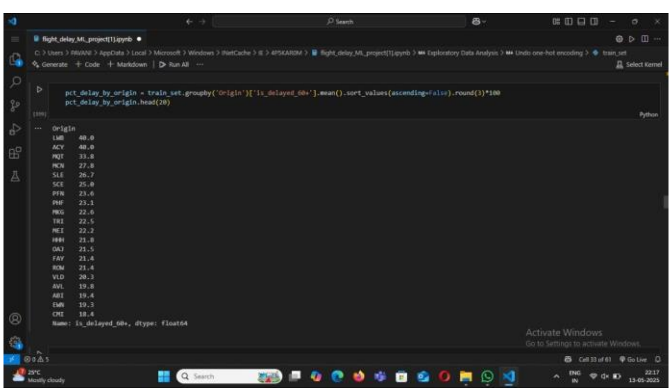

Fig: Percentage of flights that are delay for more than 1hr

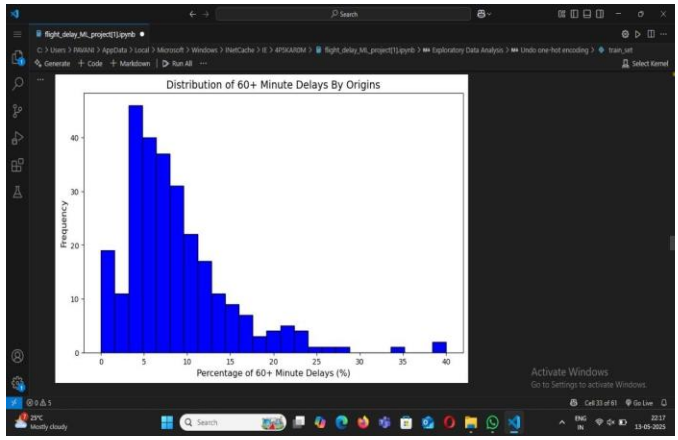

Fig: Distribution of delays by origins delay more than 1hr

V. CONCLUSION

By projecting delays based on historical and real-time data, machine learning-based flight delay prediction helps airlines and travelers better plan and manage their journeys. ML models can detect trends and reasonably forecast possible delays by utilizing information such as weather, airline schedules, air traffic, and aircraft details. This can lower expenses related to delays, increase customer happiness, and improve operational efficiency. However, feature selection, data quality, and timely updates all have a significant impact on the model's performance, highlighting the significance of reliable data gathering and ongoing optimization.

VI. FUTURE SCOPE

Predictive analytics will increasingly rely on real-time data from weather systems, air traffic control, aircraft sensors, and airport operations.

Route planning, infrastructure development, and sustainability initiatives will all benefit from long-term delay trend analysis. Reducing delays will also reduce pollution and needless fuel consumption, which will help airlines operate more sustainably.

VII. REFERENCES

[1]. [Online] "The world of air transport in 2019," ICAO. The annual report for 2019 may be found at https://www.icao.int/Pages/the-world-of-air-transport-in-2019. aspx.In accordance with%20to%20ICAO [retrieved August 21, 2023].

1 Kaggle, "Flight delay prediction," January 2023. [Online]. Available: https://www.kaggle.com/datasets/divyansh22/flight-delay-prediction/versions/2?resource=download [Accessed 12 May 2023].

[2] "Flight delay prediction using airport situational awareness map," by W. Shao, A. Prabowo, S. Zhao, S. Tan, P. Koniusz, J. Chan, X. Hei, B. Feest, and F. D. Salim, in Proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, November 5-8, 2019Chicago,USA2019,doi:https://doi.org/10.48550/arXiv.1911.01605

[3]. X. Zhou, C. Li, and J. Li, "A flight delay prediction model usingJ. Chung and J. Teo, "Single classifier vs. ensemble machine learning approaches for mental health prediction," Brain Informatics, vol.10, no. 1, 2023, pp. 1–10.

[4] "Flow Management Data and Services (FMDS)," Federal Aviation Administration, 2022. [Online]. This link is accessible: https://www.faa.gov/atmm#:~:text=The FAA oversees the safest [Accessed 4 June 2023].

[5]. R. Khan, S. Akbar and T. A. Zahed, "Flight delay prediction based on gradient boosting ensemble techniques," in 16th Int. Conf. on Open Source Systems and Technologies (ICOSST), December 14 - 15, 2022, Lahore, Pakistan, 2022, https://easychair.org/publications/preprint/PNnP

[6]. M. Zámková, S. Rojík, M. Prokop and R. Stolín, "Factors affecting the international flight delays and their impact on airline operation and management and passenger compensations fees in air transport industry: Case study of a selected airlines in Europe," Sustainability, vol. 14, no. 14763, pp. 1-16, 2022.

[7]. T. Wang, Y. Zheng and H. Xu, "A review of flight delay prediction methods," in Proc. of Int. Conf. on Big Data Engineering and Education (BDEE), August 5-7, 2022, Chengdu, China, 2022, doi:10.1109/BDEE55929.2022.00029

[8]. C. Bentéjac, A. Csörgő and G. Martínez-Muñoz, "A comparative analysis of gradient boosting algorithms," Artificial Intelligence Review, vol. 54, pp. 1937-1967, 2021.

[9]. E. Esmaeilzadeh and S. Mokhtarimousavi, "Machine learning approach for flight departure delay prediction and analysis," Transportation Research Record, vol. 2674, no. 8, p. 145–159, 2020.

[10] M. F. Yazdi, S. R. Kamel, S. J. M. Chabok and M. Kheirabadi, "Flight delay prediction based on deep learning and Levenberg-Marquart algorithm," Journal of Big Data, vol. 7, no. 1, 2020.

[11] J. Rosenow, P. Michling, M. Schultz and J. Schönberger, "Evaluation of strategies to reduce the cost impacts of flight delays on total network costs," Aerospace, vol. 7, no. 11, p. 165, 2020.

[12]. "Study of flight departure delay and causal factor using spatial analysis," Journal of Advanced Transportation, vol. 2019, pp. 1-11, 2019, by S. Cheng, Y. Zhang, S. Hao, R. Liu, X. Luo, and Q. Luo.

[13]. "Supervised, unsupervised and reinforcement learning - a detailed perspective," vol. 11, pp. 429-433, Journal of Advanced Research in Dynamical and Control Systems, A. Haldorai and R. Arulmurugan, 2019.