iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** V **Month of publication:** May 2024

DOI: https://doi.org/10.22214/ijraset.2024.61858

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue V May 2024- Available at www.ijraset.com

DelayDetect: Pre-takeoff Delay Anticipation

Aswin Magesh[1], Gauri Parvati[2], George Thomas[3], Sooraj Sajeev[4], Basil Joy[5]
Department of Computer Science and Engineering, Mar Athanasius College of Engineering, Kothamangalam, Kerala

Abstract: ***Flight delays pose significant challenges to both passengers and airlines, leading to inconvenience, financial losses, and operational disruptions. A comprehensive system is introduced to address the issue of airline flight delays through the application of machine learning models. Leveraging comprehensive datasets encompassing historical flight records, meteorological data, airport congestion patterns and other variables, this model aims to construct predictive models capable of accurately forecasting the probability of flight delays. Various machine learning methodologies including random forests and regression models, are systematically examined in this predictive task. Performance evaluation is conducted employing established metrics such as accuracy and precision. In conclusion, flight delay anticipation is a valuable tool that can revolutionize the aviation industry by minimizing disruptions, reducing costs, and increasing passenger satisfaction.***

Index Terms: machine learning

I. INTRODUCTION

Flight delay anticipation involves predicting potential disruptions in flight schedules using data and statistical models. In aviation, it's vital for enhancing passenger experience by informing travelers in advance, optimizing resource allocation like crew schedules, and improving operational efficiency through preventive measures to minimize disruptions. By anticipating delays, airlines can proactively manage and mitigate the effects of schedule disruptions, ultimately leading to improved service quality and operational performance within the industry. A comprehensive system is introduced through the application of machine learning models, leveraging comprehensive datasets encompassing historical flight records, meteorological data. The system's front end will be implemented using the HTML while the back end will be built using Python. This methodology involves proposing a comprehensive machine learning based algorithm to predict any delay in the flight arrival encompassing historical data and historical data The predictive model is developed using HTML, Python, aiming to favour not only passengers but also the aviation industry. Using all the diverse datasets, the model feeds it into the classifier and whether the flight is delayed or not is predicted. Final prediction is displayed on the website as output. The system architecture comprises a front-end developed using HTML, providing user interface functionalities. The back end is built using Python, responsible for data preprocessing, data reduction and data transformation using data manipulation algorithms. The data after pre-processing is fed into the random forest classifier which predicts whether the flight is delayed or not. If the prediction is delay, the data is again preprocessed and sent into a linear regressor and XGBoost model parallelly. The final prediction of the estimated delay time by the regression models is displayed as the output on the website. In summary, this project successfully developed a predictive tool utilizing a combination of Random Forest, linear regression, and XGBoost models to forecast flight delays prior to booking. By integrating key attributes such as source and destination airports, carrier information, scheduled arrival and departure times, and weather conditions, the models demonstrated a remarkable accuracy rate of 96 percent. This achievement underscores the potential of machine learning algorithms in enhancing travel planning experiences by empowering users with valuable insights into the likelihood of flight delays. Moreover, the project contributes to the growing body of research aimed at leveraging data-driven approaches to optimize decision-making processes in the aviation industry.

II. RELATED WORKS

A. Flight Delay Prediction Using Machine Learning: A Comparative Study of Ensemble Techniques

Machine learning is a promising tool for predicting flight delays. Accurately predicting flight delays in aviation enhances operational efficiency and passenger contentment. Accurate predictions are critical to improving operational efficiency and passenger satisfaction. The study aims to develop a robust predictive model for domestic flights and identify key variables affecting delays. This investigation transcends the confines of traditional prediction methodologies by embracing the potency of ensemble techniques, thereby imbuing the model with the capacity to capture intricate patterns and dependencies within the dataset holistically. By adopting a comparative approach, this study systematically evaluates a spectrum of ensemble methods, unravelling their strengths and weaknesses in the context of flight delay prediction. The study's results highlight the strong predictive performance of stacking methods(92.4 percent) and random forest (91.2 percent), which effectively capture patterns while cautioning about the sensitivity of AdaBoostClassifier (51.6 percent) to noisy data. This research has the potential to augment the precision and applicability of flight delay prediction, fostering operational enhancements within the aviation industry while increasing passenger satisfaction

©IJRASET: All Rights are Reserved

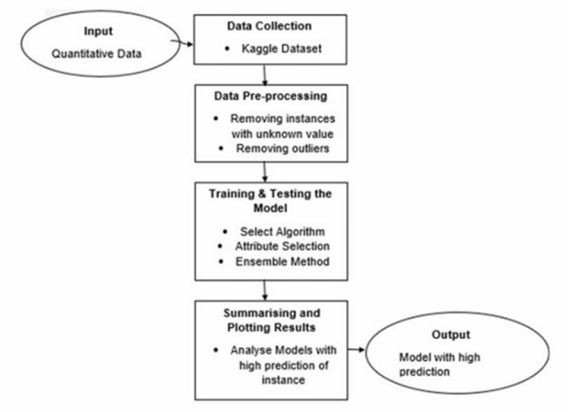


Fig. 1. Existing Design

III. PROPOSED MODEL

Data Input: The process starts with data stored in a database. Data Preprocessing: This stage involves cleaning the data. This can include filling in missing entries, removing outliers, and encoding textual data into numerical data. Data Cleaning: Imputation: This step fills in missing values in the data. For numerical data, the mean value might be used to replace missing entries. Removing Tuples with Missing Values: In this case, if a data point has missing information, the entire entry might be excluded from the dataset. Encoding String Data: This step converts textual data into numerical data. For instance, weather conditions might be encoded numerically to simplify machine learning tasks. Data Splitting: Here the data is divided into two sets: a training set and a testing set. Training Set: This set is used to train, or build, the model. The model learns from patterns in this data. Testing Set: This data is used to evaluate the performance of the trained model. The model's predictions are compared to the actual values in the testing set to determine how accurate it is. Model Building: Here a machine learning model, a random forest classifier in this case, is used to analyze the data and extract patterns. Output: Finally, the result of this process is a data model that can be used to make predictions about future data. Overall, the diagram depicts a typical data mining process where data is collected, cleaned, and prepared for analysis. Then a model is trained on a portion of the data, and the model's performance is evaluated using the remaining data.

A. Data Collection

Historical flight records: These can be obtained from airlines, flight tracking websites, or aviation authorities. They contain information such as flight numbers, departure

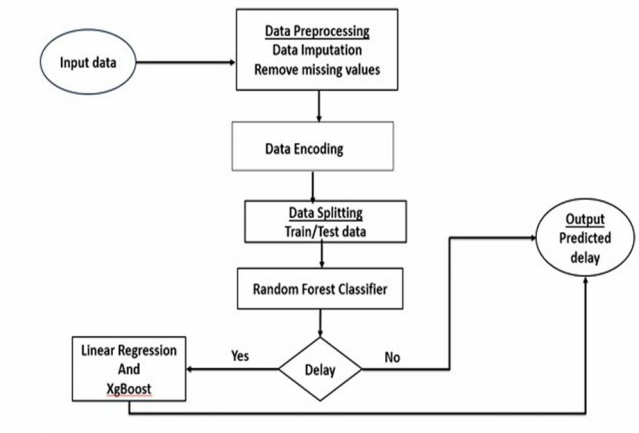


Fig. 2. Data Flow diagram of Proposed Model

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

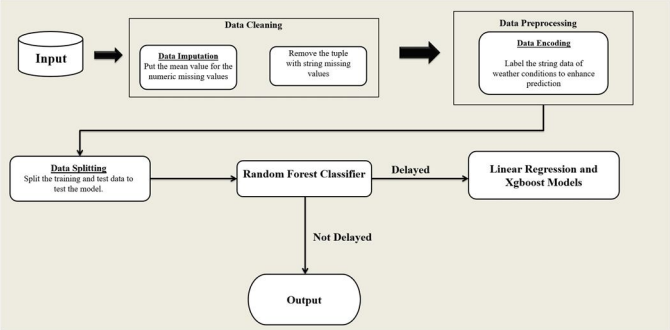


Fig. 3. Proposed Model

and arrival times, delays, cancellations, and flight routes. Meteorological data: Weather conditions significantly affect flight operations. Data sources such as NOAA (National Oceanic and Atmospheric Administration) provide historical weather data including temperature, wind speed, precipitation, and visibility. Airport congestion patterns: Airport authorities or aviation organizations may have data on airport traffic, runway utilization, and congestion levels. Other relevant variables: This could include data on air traffic control (ATC) interventions, aircraft types, airline schedules, and any other factors influencing flight operations.

B. Data Preprocessing

Cleaning: Raw data often contains errors, missing values, or inconsistencies that need to be addressed. This may involve techniques such as imputation, removing duplicates, and correcting errors. Integration: Data from different sources need to be combined into a unified dataset. This could involve matching flights based on common identifiers such as flight numbers or timestamps. Feature engineering: Creating new features from existing data to improve the predictive power of the model. For example, deriving features such as flight duration, time of day, or distance between airports from raw data. Normalization and scaling: Ensuring that all features are on a similar scale to prevent certain features from dominating others during model training. Handling categorical variables: Converting categorical variables into a numerical format suitable for machine learning algorithms, through techniques such as one-hot encoding or label encoding.

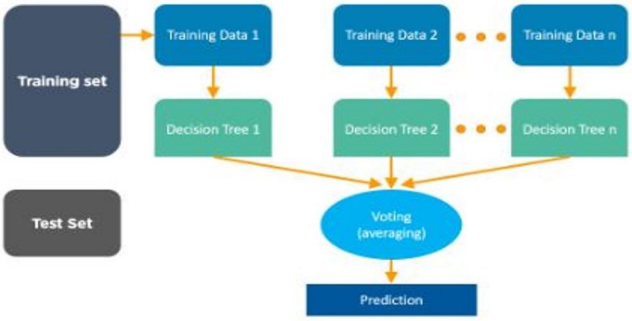


Fig. 4. Random Forest Architecture

C. Experimental Evaluation

The experimental evaluation of machine learning models revealed notable performance differences. The Random Forest Classifier exhibited flawless accuracy at 1.0, indicating robust predictive capabilities. Meanwhile, the XGBoost Regressor showcased strong performance with a mean squared error (MSE) of 48.55 and an R-squared of 0.968, suggesting highly accurate predictions. In contrast, the Linear Regression Model lagged behind with a higher MSE of 142.38 and an R-squared of 0.906, indicating relatively inferior predictive accuracy compared to the XGBoost Regressor. These results highlight the efficacy of ensemble methods like Random Forests

D. Random Forest Architecture

Random Forest is a popular ensemble learning algorithm used for classification and regression tasks. In the context of flight delay anticipation, Random Forest can be employed to predict the likelihood and duration of flight delays based on various input features such as historical flight data, weather conditions, airport congestion patterns, and other relevant variables.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

IV. RESULTS

The implementation focuses on integrating the front end, built with HTML, with the back end built using Python. The back-end manages data pre-processing, data reduction, and data transformation generation using the data manipulation algorithms, and facilitates the prediction model. The DelayDetect software reads data from an excel file, preprocesses it, and feeds the data into the classifier which tells us whether the flight gets delayed or not. The data is again preprocessed and sent into a linear regressor and XGBoost mode for delayed prediction. The displayed output on the website will be the final prediction.

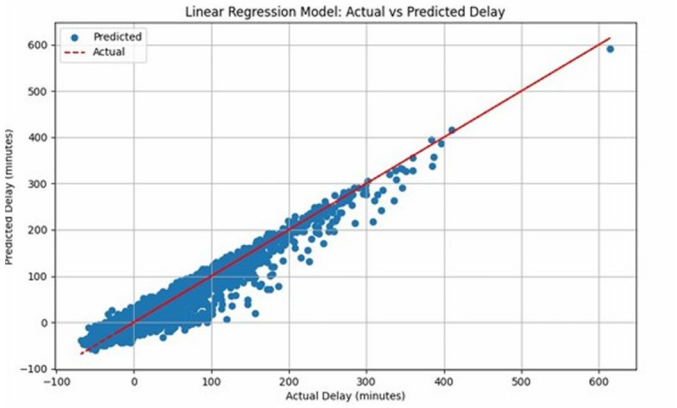


Fig. 5. Linear Regression Model Performance

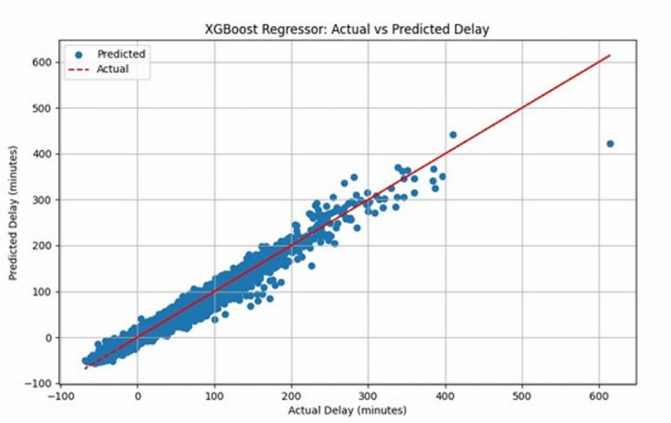


Fig. 6. XGBoost regressor performance

V. FUTURE SCOPE

The project ”DelayDetect: Pre-takeoff Delay Anticipation” presents a rich landscape for further exploration and enhancement in predictive analytics. Firstly, refining and fine-tuning existing models through advanced algorithmic

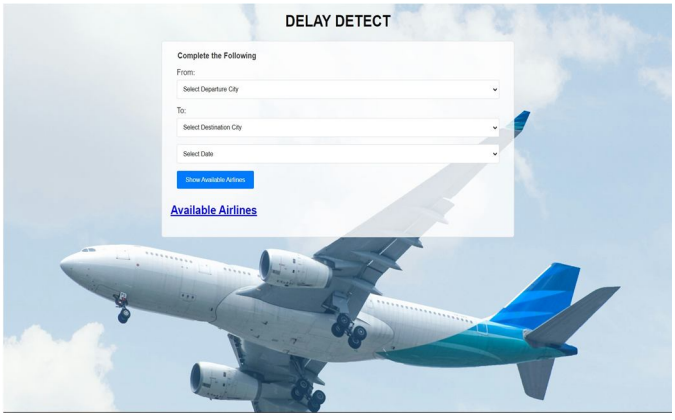


Fig. 7. Implementation

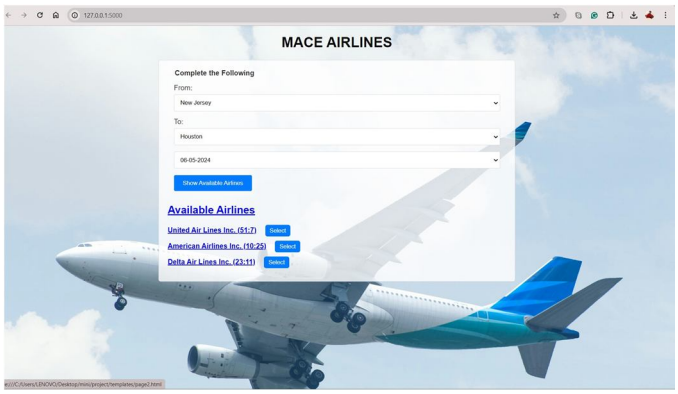


Fig. 8. Implementation

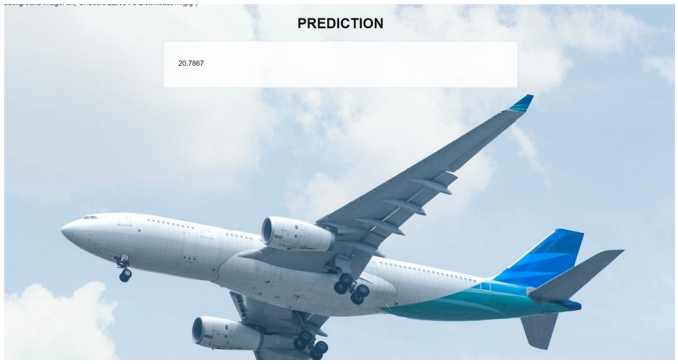


Fig. 9. Implementation

Techniques, such as ensemble learning and deep learning architectures, could lead to even greater prediction accuracy and reliability.

1) *Advanced Algorithmic Techniques:* Explore ensemble learning and deep learning architectures to refine and fine-tune existing models for improved prediction accuracy and reliability.
2) *Enhanced Feature Engineering:* Incorporate a broader range of relevant features into the dataset, including historical flight data, air traffic congestion patterns, and maintenance schedules, to provide richer contextual information for more precise predictions.
3) *Real-Time Data Integration:* Integrate real-time data streams such as weather updates, air traffic control communications, and aircraft sensor readings to enhance the models' adaptability to dynamic conditions and improve timeliness in predicting delays.
4) *Innovative Methodologies:* Investigate novel methodologies like anomaly detection and causal inference techniques to gain deeper insights into the underlying factors contributing to flight delays, enabling proactive mitigation strategies.
5) *Industry Collaboration:* Collaborate with airlines, airports, and aviation regulators to access proprietary data sources and domain expertise, fostering interdisciplinary research efforts aimed at optimizing pre-takeoff delay anticipation systems.
6) *Operational Efficiency and Passenger Satisfaction:* Ultimately, advancements in pre-takeoff flight delay anticipation have the potential to revolutionize air travel by minimizing disruptions, enhancing operational efficiency, and improving passenger satisfaction.

In summary, the outlined future scope helps us to get informed about the flight delays. By integrating with flight booking systems the delay prediction model, enhancing booking functionalities, and expanding content offerings it could be more useful, we aim to continue considering both passengers as well as the aviation industry. These advancements promise to strengthen accessibility, reliability, convenience, and operational efficiency.

VI. CONCLUSION

The project "DelayDetect: Pre-takeoff Delay Anticipation" represents a significant advancement in predictive analytics within the realm of flight delay forecasting, leveraging a sophisticated combination of Random Forest, linear regression, and XGBoost models.

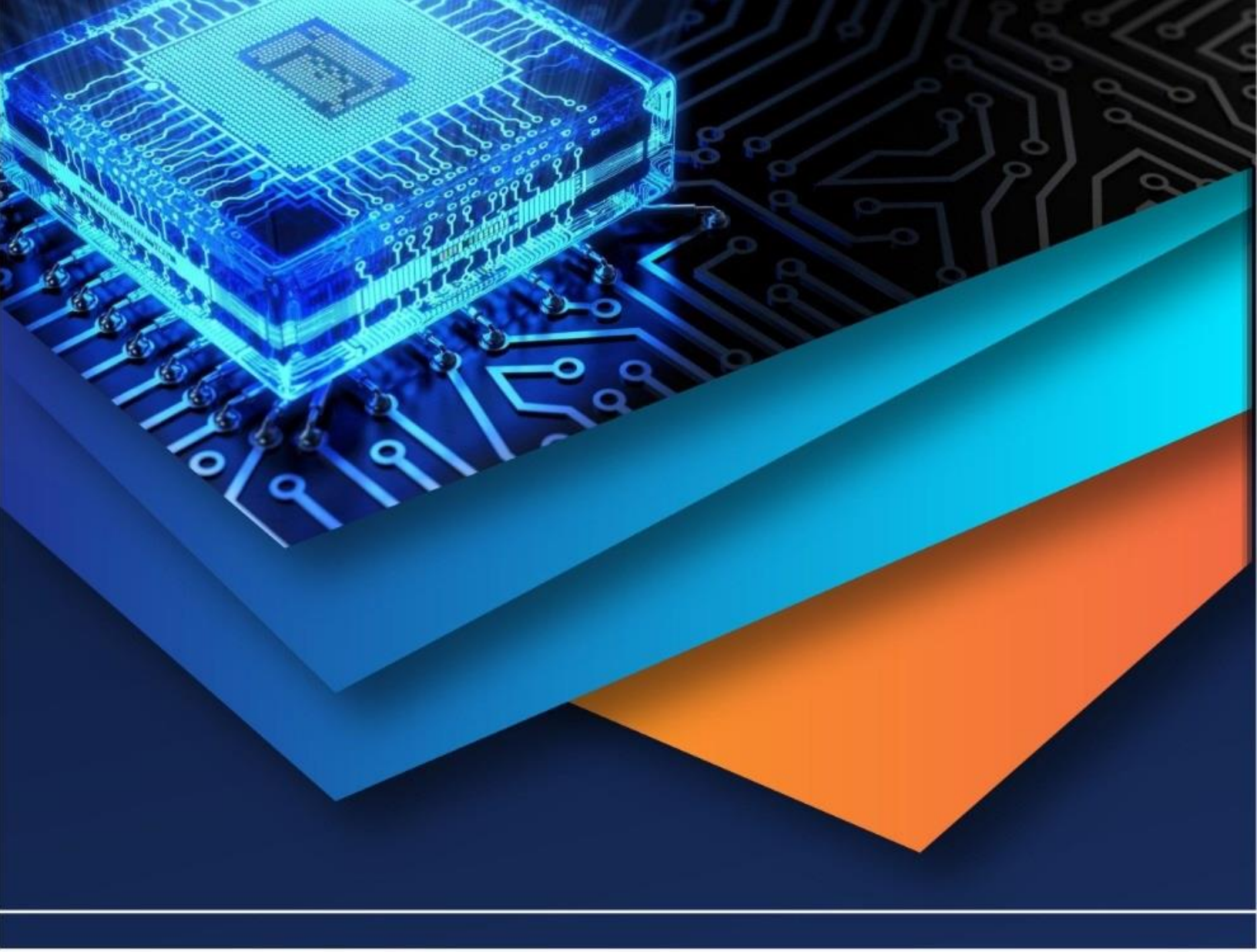
Through the comprehensive integration of crucial attributes such as source and destination airports, carrier specifics, scheduled arrival and departure times, and pertinent weather conditions, the ensemble of models has yielded an impressive accuracy rate of 96 percent. This remarkable achievement not only underscores the efficacy of machine learning algorithms in the domain of travel planning but also signifies a transformative shift in empowering users with invaluable insights into the likelihood of flight delays before making booking decisions. Moreover, the project's contribution extends beyond its immediate application, serving as a noteworthy addition to the burgeoning body of research dedicated to harnessing data-driven methodologies to optimize decision-making processes within the aviation industry.

The success of this endeavor not only highlights the efficacy of ensemble learning techniques but also underscores the potential for further innovation and refinement in predictive modeling approaches for flight delay prediction. Moving forward, future endeavors could explore avenues for enhancing the models' predictive capabilities by refining algorithmic parameters, expanding the dataset to encompass additional relevant features such as historical flight data and airport-specific metrics, and integrating real-time data sources for even more precise and up-to-date predictions. By embracing a continuous improvement mindset and harnessing the power of cutting-edge technologies, such as machine learning and big data analytics, the aviation industry stands to benefit immensely from more accurate and reliable predictive tools, ultimately enhancing operational efficiency, passenger satisfaction, and overall travel experiences.

REFERENCES

[1] 123456) khaksar2019airline, title=Airline delay prediction by machine learning algorithms, author=Khaksar, Hassan and Sheikholeslami, Abdolrreza, journal=Scientia Iranica, volume=26, number=5, pages=2689– 2702, year=2019
[2] sternberg2017review, title=A review on flight delay prediction, author=Sternberg, Alice and others, journal=arXiv preprintarXiv:1703.06118, year=2017
[3] gui2019flight, title=Flight delay prediction based on aviation big data and machine learning, author=Gui, Guan and others, journal=IEEE Transactions on Vehicular Technology, volume=69, number=1, pages=140–150, year=2019
[4] esmaeilzadeh2020machine, title=Machine learning approach for flight departure delay prediction and analysis, author=Esmaeilzadeh, Ehsan and Mokhtarimousavi, Seyedmirsajad, journal=Transportation Research Record, volume=2674, number=8, pages=145–159, year=2020
[5] jiang2020applying, title=Applying machine learning to aviation big data for flight delay prediction, author=Jiang, Yushan and others, booktitle=2020 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech), year=2020, organization=IEEE
[6] park2019flight, title=Flight delay prediction using machine learning: A case study of Korea, author=Park, Taesung and Ko, In-Young, booktitle=Proceedings of the 6th International Conference on Industrial Engineering and Applications (ICIEA 2019), organization=Atlantis Press, year=2019
[7] zhang2020flight, title=Flight delay prediction based on deep learning models, author=Zhang, Yimeng and others, journal=IEEE Access, volume=8, pages=7367–7381, year=2020
[8] bhowmik2019predicting, title=Predicting flight delays using ensemble machine learning, author=Bhowmik, Tanay and others, booktitle=Proceedings of the 2019 International Conference on Artificial Intelligence and Computer Science (AICS 2019), organization=ACM, year=2019
[9] fu2018novel, title=A novel flight delay prediction model using hybrid machine learning algorithms, author=Fu, Wenjie and others, booktitle=Proceedings of the 2018 International Conference on Computer, Information and Telecommunication Systems (CITS 2018), organization=IEEE, year=2018
[10] wang2017flight, title=Flight delay prediction based on machine learning and weather data, author=Wang, Jie and others, booktitle=2017 IEEE 2nd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), organization=IEEE, year=2017

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)