Aero Forecast: Elevating Travel Experiences Through Flight Delay Anticipation

Karthick Balaje Gopalakrishnan Elango, Saaijeesh Sottalu Naresh, Nitharshan Coimbatore Venkatesan
Department of Data Science, University of Colorado Boulder, United States

ABSTRACT:

In the realm of air travel, where smooth journeys are desired despite the challenges of delays, predictive systems play a crucial role. The goal of this project, called Aero Forecast, is to improve travel experiences by forecasting aircraft delays using data mining. Our strategy combines very precise prediction algorithms with real-time monitoring by considering multiple parameters at both big and small scales. To provide aviation authorities, airlines, and airports with useful insights, Aero Forecast incorporates a variety of data sources, including weather, technical evaluations, and air traffic patterns. We work to lessen the financial losses and passenger angst associated with aircraft delays by doing in-depth analysis and modelling. This project complements current efforts in the airline sector to improve communication during interruptions, increase operational efficiency, and optimize the traveler experience overall. Aero Forecast offers a major step forward in building a more robust and effective air travel system that will benefit passengers and stakeholders equally, by expanding our understanding of the factors that contribute to aircraft delays.

Keywords:

Modelling; Forecasting; Data Mining

1. INTRODUCTION

Air travel, a marvel of modern engineering, has revolutionized the way we navigate our interconnected world, bridging vast distances and connecting people and cultures. However, amidst the wonders of aviation, the spectre of travel delays looms as a common challenge that impacts the seamless experience we envision when taking to the skies. The significance of air travel in our daily lives is undeniable, as it facilitates swift journeys for business, leisure, and urgent matters. Whether it's a crucial business meeting, a long-awaited family reunion, or a dream vacation, air travel offers unparalleled speed and efficiency. Yet, the reality of travel delays, caused by various factors such as weather conditions, technical issues, or air traffic congestion, introduces a dimension of unpredictability.

The inconvenience of travel delays extends beyond the personal sphere, affecting the global network of transportation, commerce, and cultural exchange. While the marvel of air travel has accelerated the pace of our lives, delays can disrupt schedules, causing a ripple effect on interconnected flight routes and affecting the overall efficiency of the aviation system. Despite these challenges, the continuous evolution of air travel strives to address and mitigate delays. As we navigate the skies, the interconnected dance between the marvels of air travel and the challenges of delays shapes our modern relationship with the boundless opportunities and occasional setbacks that come with traversing the world by air.

2. DATA COLLECTION

2.1 Flight Status Dataset from Bureau of Transportation Statistics (BTS)

The Bureau of Transportation Statistics (BTS) flight status dataset is a vital resource for comprehending the complexities of airline operations, such as delays and cancellations. This dataset, which covers the period from January 2018 to the present, provides a wide range of parameters, including date, airline, origin city, destination city, departure time, and arrival time. Thorough pretreatment steps are necessary to fill in missing values and standardize data formats before analysis.

	flightdate	airline	origin	dest	cancelled	diverted	crsdeptime	deptime
0	2018-01-23	Endeavor Air Inc.	ABY	ATL	False	False	1202	1157.0
1	2018-01-24	Endeavor Air Inc.	ABY	ATL	False	False	1202	1157.0
2	2018-01-25	Endeavor Air Inc.	ABY	ATL	False	False	1202	1153.0
3	2018-01-26	Endeavor Air Inc.	ABY	ATL	False	False	1202	1150.0
4	2018-01-27	Endeavor Air Inc.	ABY	ATL	False	False	1400	1355.0

Figure 1. Flight status Dataset

2.2 NASA Gov API for Climatic Data

The NASA Gov API presents an invaluable resource for accessing climatic data pertaining to diverse cities. This API provides detailed information on important meteorological variables, such as temperature, humidity, wind speed, precipitation, and dew throughout different time periods. To fully utilize this API, customized queries that include start and end dates, latitude, longitude, and community must be formulated. In the data retrieval process involves carefully crafting queries to obtain relevant meteorological information for pre-identified locations and periods.

	origin	date	temp	dew	min_temp	max_temp
0	DEN	2018-01-01	9.28	-11.18	-11.24	-1.95
1	DEN	2018-01-02	16.18	-13.67	-12.00	4.18
2	DEN	2018-01-03	15.35	-15.13	-5.77	9.58
3	DEN	2018-01-04	10.71	-10.07	-2.56	8.15
4	DEN	2018-01-05	12.71	-8.99	-1.90	10.81

Figure 2. Weather Dataset

3. DATA CLEANING

The precision and dependability of data are critical for well-informed decision-making and efficient operations in the aviation sector. This report describes our meticulous methodology for gathering and preparing data for flight status datasets from the Bureau of Transportation Statistics (BTS) for the years 2018 to 2020, as well as meteorological data from the NASA Gov API. Our approach places a high priority on careful data management to guarantee accuracy and dependability, providing a strong framework for further analysis and decision-making.

During the data gathering stage, we meticulously obtained flight status datasets from the Bureau of Transportation Statistics (BTS) for the years 2018 to 2020. These datasets function as an all-inclusive database for flight operations data, which includes vital details like arrival and departure timings, delays, and cancellations. In parallel, we obtained meteorological information from the NASA Gov API, which provides essential meteorological details including wind, precipitation, and temperature. This meteorological data is essential for strengthening our analytical skills, increasing the robustness of our analyses, and comprehending the significant effects of weather occurrences on flight operations.

As a result, we adopted a methodical approach in our data preparation efforts to guarantee the accuracy and consistency of the datasets. First, we carefully selected only active flights from the flight status records, carefully excluding any cancelled or diverted flights. The goal of this curation procedure was to provide datasets that faithfully represented the operational environment of flight operations. In addition, we diligently resolved any missing values in important columns including "DepTime," "DepDelayMinutes," "DepDelay," "AirTime," "ArrTime," "ArrDelayMinutes," "ArrDelay," and "ActualElapsedTime." By utilizing a range of methods, including imputation and comparison, we were able to guarantee thorough data coverage while upholding standard procedures. Additionally, to improve dataset usability and clarity, unnecessary columns were carefully eliminated. Concurrently, extensive cleaning processes were performed on the climate data, which strengthened its quality and coherence for further analytical efforts. These processes included handling missing values, renaming columns, and formatting date columns.

This report's emphasis on the rigorous data preparation and gathering processes demonstrates our commitment to maintaining data integrity and dependability in aviation studies. We strengthen a strong basis for drawing meaningful conclusions and enabling well-informed decision-making in the aviation sector by implementing strict data management procedures. We guarantee our studies' accuracy and coherence by carefully screening datasets, filling in the blanks, and simplifying data clarity. Our dedication to these procedures demonstrates our quest for excellence in providing insights that can be used to improve operational effectiveness and safety in aviation operations.

4. EXPLORATORY DATA ANALYSIS

4.1 Busiest Airports

Los Angeles International Airport (LAX) is the busiest hub for arrivals and departures out of the top five airports, according to the comparative graphic. On the other hand, the geography map shows the locations of significant inland and coastal airports, including those in Boston, Chicago, San Francisco, and Seattle, which correspond to their status as key centers for international travel and commerce. The Northeast region, which includes Washington, D.C., has a high concentration of airports, indicating a high concentration of population and commercial activity. Airports in Dallas/Fort Worth and Las Vegas show a lot of business and tourist traffic, while inland hubs like Denver and Chicago are well-positioned for domestic travel. The complex interactions between air traffic and urban-industrial settings are highlighted by this distribution.

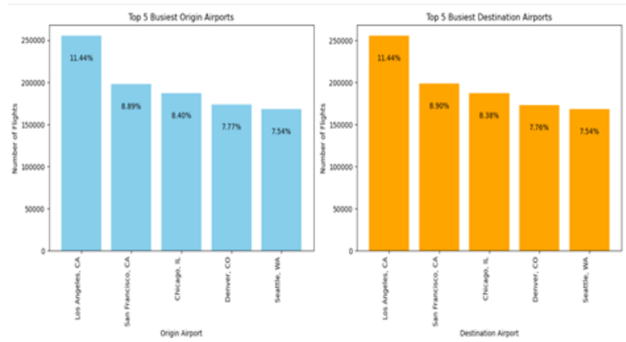

Figure 3. Busiest Origin and Destination airports

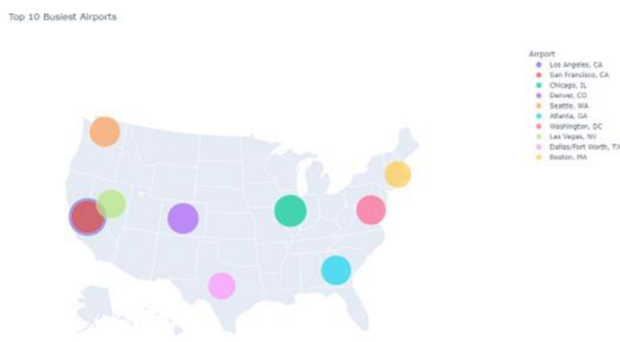

Figure 4. Top 10 busiest airports

4.2 Frequency of flights by Time Block and their delay patterns

The bar chart depicting flight counts across different time blocks reveals intriguing insights into aviation trends. According to the data, lunchtime is the busiest time of day, with the afternoon, evening, and morning blocks following suit. This shows that demand for air travel varies throughout the day, peaking in the midday and early evening hours. Furthermore, a consistent pattern can be observed in the graph that shows the average arrival and departure delays throughout different time intervals, with delays increasing from early morning to late afternoon. In the aviation industry, a thorough understanding of these temporal trends is essential for improving customer happiness and operational efficiency. It also helps airlines and travelers with resource management and strategic planning.

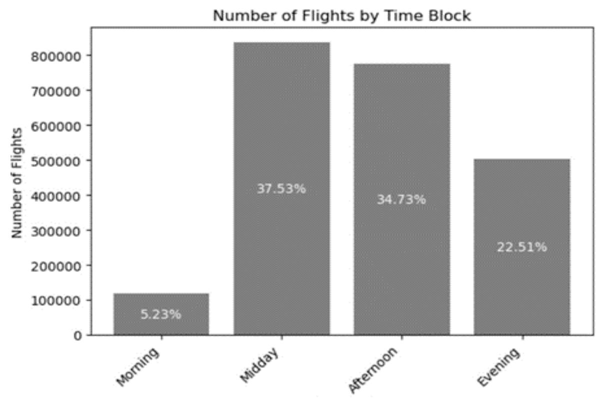

Figure 5. Number of flights by each time block

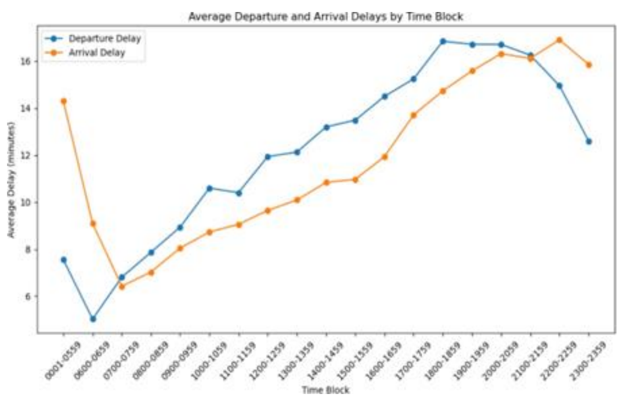

Figure 6. Average departure and arrival delays by time block

4.3 Monthly flight activity and average delay by season

There are clear seasonal variations in the line graph that shows the total monthly flight activity from 2018 to 2020; it peaks in August and October and decreases in May. Within the aviation industry, this visualization helps with resource allocation and trend analysis. The graph that shows the average arrival and departure delays by season also demonstrates seasonal differences, with summer seeing the greatest delays because of increased air traffic and weather-related disruptions. Compared to fall and spring, winter shows more delays. This provides information about scheduling and controlling customer expectations. Strategic decision-making is informed by this data, which highlights the effects of seasonal variations on airline operations and customer experience.

Figure 7. Monthly flight activity

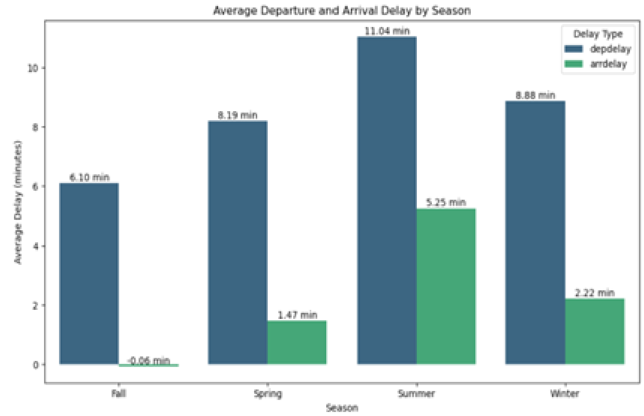

Figure 8. Average departure and arrival delay by season

4.4 On-Time Departure and Arrival performance with distribution of departure delays

The chart that displays the on-time performance of the top ten airlines shows a downward trend in punctuality, with certain airlines performing better in terms of arrivals or departures. Most airlines continue to operate effectively despite small differences. The stacked bar chart sheds lighter on delay trends by showing that major airlines give priority

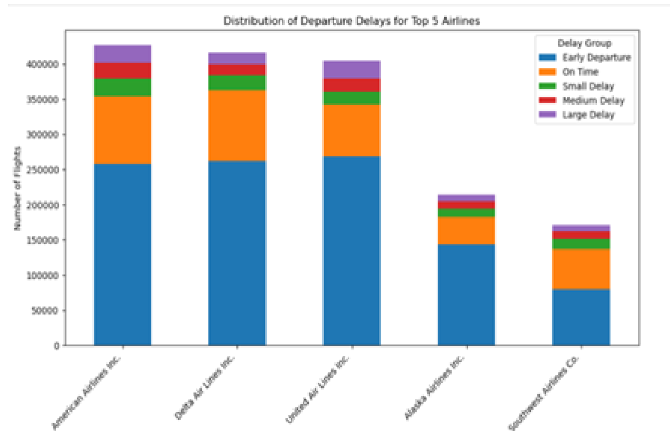

Figure 9. Distribution of departure delays for top 5 airlines

to departing early, whereas smaller airlines are more likely to encounter minor to moderate delays. Trans States carriers are remarkable for having the longest median departure delay, indicating that a considerable portion of flights have delays. Notably, carriers exhibit persistent outliers in delay durations, highlighting the necessity of having strong backup plans if the industry experiences occasional, protracted delays.

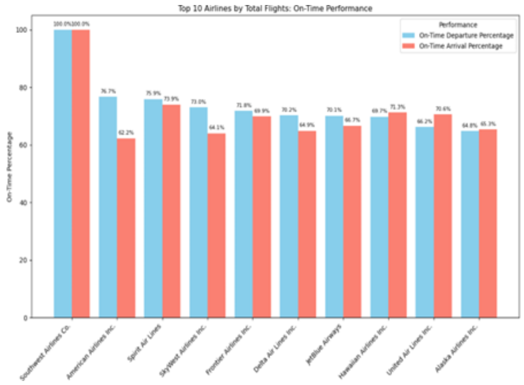

Figure 10. Top 10 airlines by total time: On Time performance

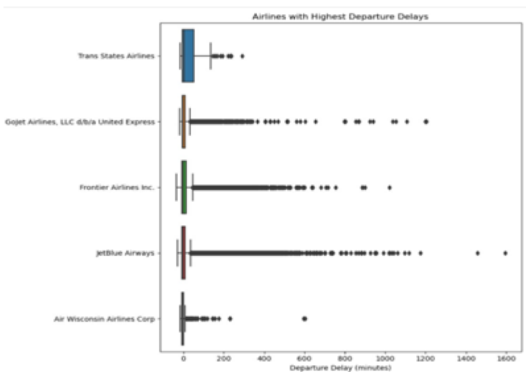

Figure 11. Airlines with highest departure delays

4.4 Common flight routes

The top 5 flight routes are shown in the pie chart, with a focus on West Coast cities and routes connecting East and West Coast cities. The majority, or 23.7%, consists of flights between the major Californian cities of Los Angeles and San Francisco, suggesting that these routes are frequently travelled. Another sizable fraction (18.1%) consists of flights from Las Vegas to Los Angeles, which are probably mostly driven by business and leisure travel. There is a significant demand for air travel between these coastal population centers, as seen by the 16.4% share of transcontinental flights between New York City and Los Angeles. The substantial amount of air traffic between Southern California and Las Vegas is highlighted by flights operating between Los Angeles and Las Vegas. Overall, the data underscores Los Angeles' importance as a major aviation hub and the frequency of travel within California and between West and East Coast gateways.

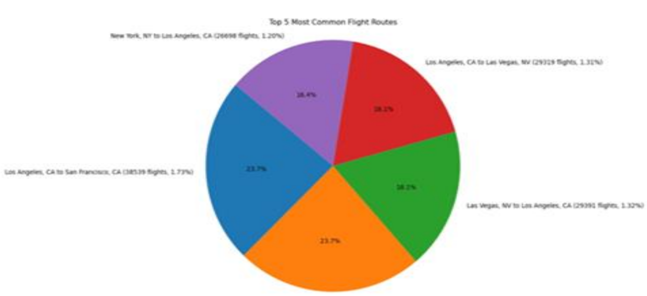

Figure 12. Top 5 most common flight routes

5. DATA MODELLING

Building on the fundamental data analysis, our team moved on to deploy advanced predictive models with the goal of improving the accuracy and reliability of flight delay predictions. This phase used a variety of machine learning techniques, each designed to capitalize on the patterns and complexities revealed in our dataset. Here's a thorough overview of the approaches and insights generated by these models.

5.1 Data Preparation

During the data preparation portion of our study, we carefully chose and processed flight information, focusing on the biggest airports to guarantee our analysis accurately reflects high-traffic scenarios. The dataset contains various predictor variables that are critical for interpreting flight patterns, such as dates, times, airline identifiers, and airport codes. We developed a tripartite classification system for the dependent variable, flight arrival status. aircraft arriving earlier than 7 minutes ahead of schedule are labelled 'Early Arrival,' those arriving between 7 minutes early and 12 minutes late are declared 'On Time,' and aircraft arriving more than 12 minutes late are labelled 'Delayed Arrival.' This classification allows us to make more detailed predictions about flight punctuality. The emphasis on the busiest airports places our analysis in the context of high-stakes, high-volume travel situations, where precise predictions are most needed and effective for operational optimization and traveler convenience. Made entirely of data from the year 2020. This temporal separation ensures that our model's performance is evaluated based on its ability to generalize to unseen, future data, which is critical for the practical deployment of our flight delay prediction system.

	year	month	dayofmonth	dot_id_marketing_airline	originairportid	destairportid
1	2018	1	1	19790	11278	12478
2	2018	1	1	19790	13930	10397
3	2018	1	6	19790	12264	12478
4	2018	1	13	19790	12264	12478
5	2018	1	14	19790	12264	12478

Figure 13. Flight status data for modelling

Prior to the model training phase, we performed thorough data preparation to assure the integrity and quality of the information used in our predictive analytics. We collected significant flight data from 2018 and 2019, including departure and arrival timings, airport IDs, temperature readings, and more. This dataset was then rigorously cleaned and preprocessed, which included addressing missing values, normalizing numerical inputs, and encoding categorical variables to make it appropriate for our machine learning techniques. The final generated dataset was divided into training and testing sets, with the training set drawn from 2018 and 2019 data and the testing set consisting solely of data from 2020. This temporal separation ensures that our model's performance is evaluated based on its ability to generalize to unseen, future data, which is critical for the practical deployment of our flight delay prediction system.

yminutes	depdelay	...	arr_min_temp	arr_max_temp	arr_sp_humid	arr_rel_humid	arr_precip
49.0	49.0	...	-11.31	-6.06	1.65	94.31	0.03
0.0	-6.0	...	-7.10	-0.31	1.53	57.00	0.00
0.0	-4.0	...	-10.60	-8.61	1.53	93.75	0.17
0.0	-3.0	...	-2.82	10.01	5.13	87.88	30.52
0.0	-5.0	...	-6.64	-2.94	1.83	74.94	0.00

Figure 14. Weather data for modelling

5.2. Decision Tree Classifier

Before Hyperparameter Tuning

We used Decision tree classifier, which was commended for its simplicity and interpretability. The model was trained on a carefully resampled dataset to avoid the impacts of class imbalance, resulting in equitable learning across all flight status categories. The decision tree gave us preliminary insights into the important elements impacting delays, with an accuracy of around 56%. To assess its overall performance, detailed metrics like as precision, recall, and F1-score were assessed, with findings indicating reasonable success in identifying various forms of delays.

After Hyperparameter Tuning

Decision trees are a non-linear predictive modelling approach that can handle qualitative predictors and deliver clear decision-making insights. We improved the Decision Tree's performance by modifying hyperparameters such as tree depth and the minimum amount of data per leaf, resulting in a 70.29% accuracy. This optimized model improves precision to 67.72% and recall to 70.29% while also achieving an F1-score of 67.59%. The confusion matrix indicated the model's improved ability to effectively classify various flight statuses, demonstrating its efficacy in anticipating flight delays.

Model	Accuracy	Precision	Recall	F-1 Score
Decision Tree (Before Hyperparameter Tuning)	0.5657	0.5918	0.5657	0.5753
Decision Tree (After Hyperparameter Tuning	0.7029	0.6777	0.7029	0.6758

Table 1. Scores

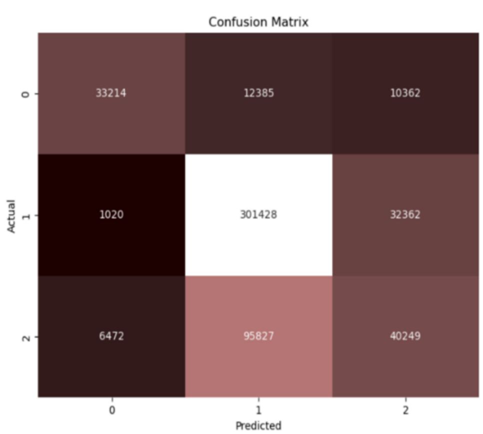

Fig15. Confusion matrix

- For the "Delay Arrival" class (class 0), the model accurately identifies 33,214 instances and incorrectly classifies 12,385 instances as "Early Arrival" and 10,362 instances as "On Time."

- For the "Early Arrival" class (class 1), the model correctly predicts 301,428 instances but misclassifies 1,020 instances as "Delay Arrival" and 32,362 instances as "On Time."

- For the "On Time" class (class 2), the model accurately predicts 40,249 instances but mistakes 6,472 instances as "Delay Arrival" and 40,249 instances as "Early Arrival."

5.2 Naive Bayes Classifier

The Naive Bayes Classifier, which is known for its simplicity and efficiency when dealing with huge datasets, was used because of its ability to perform well under the premise of feature independence. The model attained an accuracy of 70.56%, with a precision of 67.43% and a recall of 70.56%. It obtained an F1 score of 65.66%. The confusion matrix demonstrated the model's performance across several classes, showing its usefulness in accurately predicting flight delays while being computationally less costly than more complicated models. This makes it especially useful in cases needing quick decisions based on massive amounts of data.

Model	Accuracy	Precision	Recall	F-1 Score
Naïve Bayes	0.7056	0.6743	0.7056	0.6565

Table 2. Scores

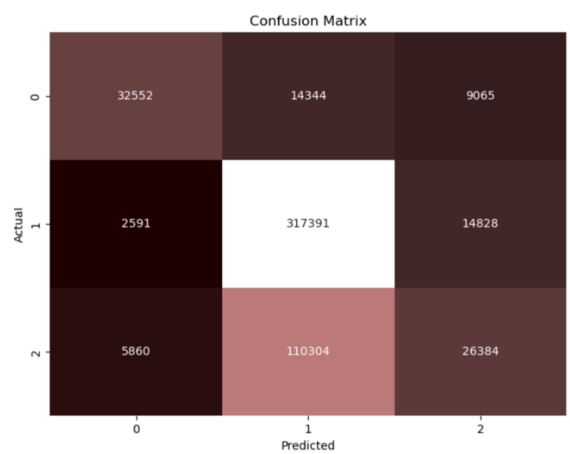

Fig16. Confusion matrix

- For the "Delay Arrival" category (class 0), the model correctly identifies 32,552 instances, but it erroneously categorizes 14,344 instances as "Early Arrival" and 9,065 instances as "On Time."

- Regarding the "Early Arrival" group (class 1), the model accurately predicts 317,391 instances, yet it mislabels 2,591 instances as "Delay Arrival" and 14,828 instances as "On Time."

- In the case of the "On Time" category (class 2), the model correctly predicts 26,384 instances, but it misclassifies 5,860 instances as "Delay Arrival" and 110,304 instances as "Early Arrival."

5.3 XGBoost Classifier with Hyper parameter Tuning

XGBoost is known for its high performance and speed, particularly with structured data. It is a gradient boosting framework that employs a gradient boosting method to optimize predictions. The model is well-known for producing excellent results when dealing with a wide range of data kinds, sizes, and distributions. The XGBoost model obtained 71.27% accuracy, 68.72% precision, and 71.27% recall by fine-tuning important hyperparameters such as learning rate, number of estimators, and objective function. The F1-score, a key metric in unbalanced classrooms like ours, was 67.90%. The confusion matrix provides a thorough perspective of the model's performance across several classes, demonstrating high prediction accuracy. This model's strength is its ability to provide fast, accurate predictions even with complicated and huge datasets, making it a vital tool in our predictive analytics arsenal for forecasting flight delays.

Model	Accuracy	Precision	Recall	F-1 Score
XGBoost Classifier (*Hyper Parameter Tuning*)	0.7056	0.6743	0.7056	0.6565

Table 3. Score

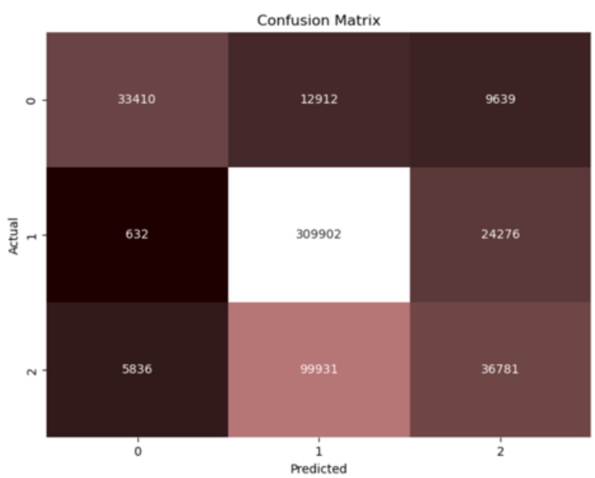

Figure 17. Confusion matrix

- In the "Delay Arrival" category (class 0), the model accurately identifies 33,410 instances but erroneously categorizes 12,912 instances as "Early Arrival" and 9,639 instances as "On Time."

- For the "Early Arrival" class (class 1), the model correctly predicts 309,902 instances, yet it mislabels 632 instances as "Delay Arrival" and 24,276 instances as "On Time."

- In the "On Time" class (class 2), the model accurately predicts 36,781 instances, but it misclassifies 5,836 instances as "Delay Arrival" and 99,931 instances as "Early Arrival."

5.4 Logistic Regression

Logistic Regression is a popular statistical method that predicts binary outcomes and is known for its simplicity and interpretability. This paradigm performs well in cases when relationships can be represented linearly. In our situation, Logistic Regression was used to forecast flight delays, with an accuracy of 65.67%. The precision was 59.72%, and the recall was 65.67%, which matched the accuracy. The F1-score, which is critical for balancing precision and recall, was measured at 60.21%. The confusion matrix described the model's performance across various flight statuses, emphasizing strengths in properly recognizing 'On-time' flights while suggesting opportunities for improvement in discriminating between 'Delayed' and 'Highly Delayed' aircraft. Although Logistic Regression is not as robust as more complex models, its simplicity and effectiveness make it a viable option for early investigation and baseline modelling.

Model	Accuracy	Precision	Recall	F-1 Score
Logistic Regression	0.6567	0.5972	0.6567	0.6021

Table 4. Score

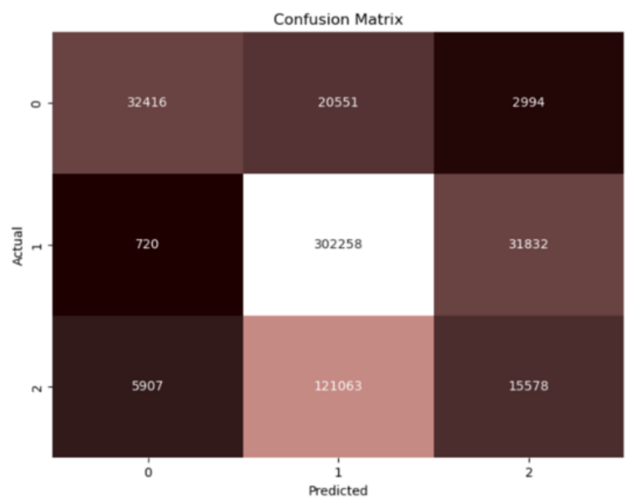

Figure 18. Confusion matrix

- In the "Delay Arrival" category (class 0), the model accurately identifies 32,416 instances. However, it incorrectly categorizes 20,551 instances as "Early Arrival" and 2,994 instances as "On Time."

- For the "Early Arrival" class (class 1), the model correctly predicts 302,258 instances. Nevertheless, it misclassifies 720 instances as "Delay Arrival" and 31,832 instances as "On Time."

- In the "On Time" class (class 2), the model accurately predicts 15,578 instances. Yet, it misclassifies 5,907 instances as "Delay Arrival" and 121,063 instances as "Early Arrival."

5.5 Random Forest Classifier

The Random Forest Classifier, an ensemble learning method that creates numerous decision trees and integrates them to get a more accurate and consistent forecast, performed admirably in our project. It attained an accuracy of 70.67%, with a precision of 68.26% and a recall of 70.67%. The F1-score, a key parameter for determining the balance of precision and recall, was 68.07%. The confusion matrix demonstrated that the model did an excellent job of discriminating between different flight statuses, particularly detecting 'On-time' flights. The Random Forest's strength is its ability to properly handle huge datasets with various factors, making it an ideal tool for complicated predictive tasks such as anticipating aircraft delays. The model's resistance to overfitting and ability to generate relevance scores for various features make it particularly useful for improving operational decision-making in the aviation industry.

Model	Accuracy	Precision	Recall	F-1 Score
Random Forest Classifier	0.6567	0.5972	0.6567	0.6021

Table 5: Score

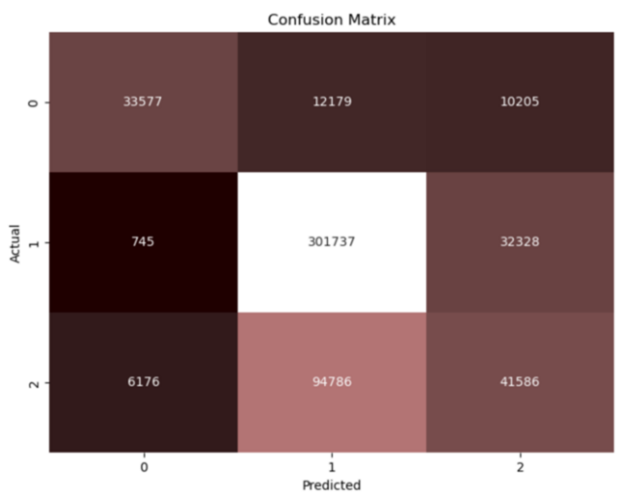

Figure 19. Confusion matrix

- In the class representing "Delay Arrival" (class 0), the model accurately identifies 33,577 instances. However, it mistakenly categorizes 12,179 instances as "Early Arrival" and 10,205 instances as "On Time."
- For the "Early Arrival" category (class 1), the model makes correct predictions for 301,737 instances. Nevertheless, it mislabels 745 instances as "Delay Arrival" and 32,328 instances as "On Time."
- In the "On Time" class (class 2), the model accurately predicts 41,586 instances. Nonetheless, it erroneously categorizes 6,176 instances as "Delay Arrival" and 94,786 instances as "Early Arrival."

5.6 Feed Forward Neural network.

Our study used a feed forward neural network with TensorFlow and Keras to improve flight delay predictions. This deep learning model had numerous layers with ReLU activation and a SoftMax output layer for multiclass classification, which was suitable for dealing with the complicated patterns in our large flight data set. The model's strength is its versatility and capacity to discern non-linear correlations in huge datasets, making it ideal for prediction tasks. Trained throughout 20 epochs with constant improvements in validation accuracy, the model exhibited its capacity to learn successfully, obtaining a test accuracy of roughly 71.11%. This strong performance demonstrates the model's applicability in real-time forecasting scenarios, where correct delay forecasts have a major impact on operational efficiency and passenger happiness. The neural network's advanced learning structure provides a scalable and adaptive solution, confirming its suitability for ongoing aviation industry difficulties.

5.7 Handling Class Imbalance

To effectively address the issue of class imbalance in our dataset, a strategic strategy involving resampling techniques was used, which improved the predicted accuracy and dependability of our models. The dataset was initially segmented using the 'delay_group' classification, with significant resampling performed to balance the distribution across three categories: 648,452 samples for 'Highly Delayed' (delay_group 2), 407,369 for 'On-time' (delay_group 0), and 900,000 for 'Delayed' (delay_group 1). After integrating these subsets, the entire dataset was rigorously randomized to ensure a random distribution, which is necessary for unbiased model training. The dataset was then refined by removing less significant variables such as various temporal and weather-related attributes, which expedited the modelling process while maintaining predictive capabilities. The dataset was then refined in subsequent rounds by removing less significant variables such as various temporal and weather-related attributes, which expedited the modelling process while maintaining predictive capabilities.

Furthermore, the training data was subjected to the Synthetic Minority Over-sampling Technique (SMOTE) to improve minority class representation. This method generates fresh examples in the minority classes synthetically, ensuring that all classes have an equal influence in the training process. This method not only eliminates model bias towards the majority class, but it also improves overall performance measures such as accuracy, precision, and recall. These rigorous preparations guaranteed that our predictive algorithms are well-suited to handling real-world data, resulting in strong and equitable flight delay projections. After adding strategies to address class imbalance in our dataset, we found varying performance across models, each customized to better forecast flight delays.

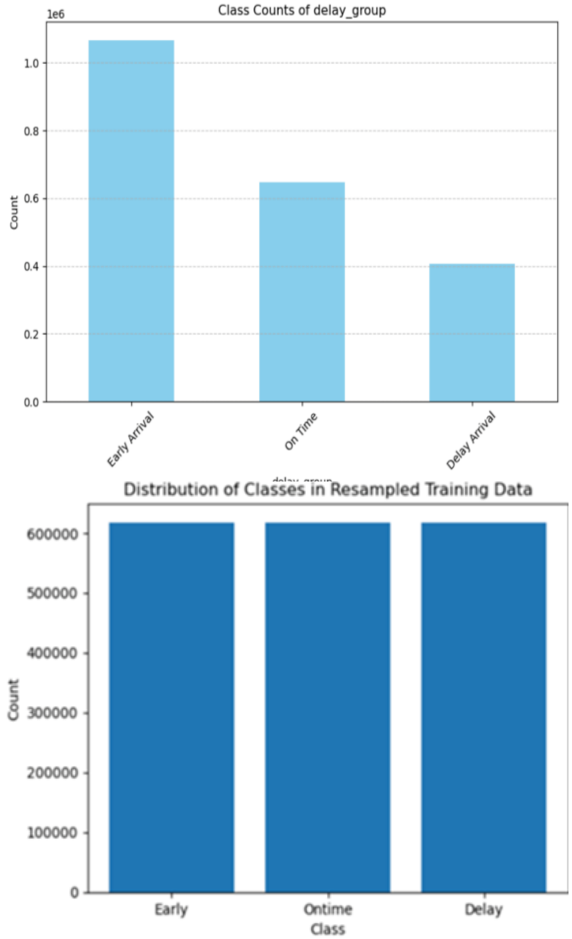

Figure 20. Before and after handling Class imbalance

5.8 Performance Comparison of Machine Learning Models After Handling Class Imbalance

Even with class imbalance corrections, the Decision Tree achieved a moderate accuracy of 55.04%, precision of 59.08%, and F1-score of 56.39%. This approach, while simpler and speedier, has shortcomings in dealing with complex class boundaries, as evidenced by the mixed results in its confusion matrix. In contrast, the XG Boost model performed significantly better, with an accuracy of 69.48%, a precision of 67.46%, and an F1-score of 67.91%. This model's capacity to manage nonlinearities and feature interactions increased its effectiveness in managing imbalanced data, as seen by its performance indicators. The Naive Bayes model likewise performed well, after handling the presence of class imbalance, with an accuracy of 70.49%, precision of 67.24%, and F1-score of 65.82%. The model's ability to handle huge datasets with numerous features proved useful in this instance. Finally, Logistic Regression demonstrated the difficulties of linear models in dealing with imbalanced data, achieving an accuracy of 51.61%, precision of 58.78%, and an F1-score of 53.72%. Despite its simple implementation, it struggled with the intricate patterns found in the data.

5.9 Conclusion of Predictive Modelling Before and After Managing Class Imbalance

Before addressing class imbalance and hyperparameter tuning

Initially, our predictive models, which included Decision Tree, XG Boost, Naive Bayes, Logistic Regression, Random Forest, and Neural Networks, were applied to the flight delay data without regard for class imbalance or hyperparameter optimization. The performance of these models varied greatly due to the dataset's intrinsic class imbalance. This imbalance often resulted in models that were biased towards the dominant class, resulting in worse predicted accuracy, particularly for minority groups. Models such as Logistic Regression and Decision Tree struggled more, highlighting the need for specialized methodologies to successfully manage dataset.

After addressing class imbalance and hyperparameter tuning

Decision Tree: Following hyperparameter tuning and class balance modifications, the Decision Tree's performance improved modestly, showcasing its simplicity and speed of deployment, albeit it remained limited in handling complicated patterns.

XG Boost: This model improved significantly, thanks in large part to hyperparameter optimization and class imbalance handling. It stood out for its high accuracy and strong handling of complicated datasets, making it an ideal candidate for complex predictive tasks.

Naive Bayes: After changes, Naive Bayes retained decent performance, notably in terms of efficiency with large datasets and speedy execution, making it appropriate for early exploratory research.

Logistic Regression: Despite corrections for class imbalance, Logistic Regression showed modest improvement, highlighting the difficulties linear models confront in complicated, imbalanced.

Random Forest: Like XG Boost, Random Forest benefited from class imbalance, demonstrating enhanced accuracy and robustness while collecting different data properties without overfitting.

Neural Networks: The Neural Network's deep learning skills allowed it to adapt well to the altered data, highlighting the benefits of flexible, layered designs in managing non-linear data relationships and complicated patterns.

Model	Accuracy (*Before Handling Imbalance*)	Accuracy (*After Handling Imbalance*)
XGBoost Classifier	71.26%	69.48%
Feed Forward Neural Network	75.73 %	71.26%
Random Forest Classifier	70.67%	67.68%
Naïve Bayes	70.56%	70.48%
Decision Tree	70.29%	69.21%
Logistic Regression	65.67%	51.60%

Table 6. Scores

In conclusion, correcting class imbalance and optimizing model parameters is crucial for improving the predictive power and dependability of machine learning models. Advanced techniques such as XG Boost, Random Forest, and Neural Networks gained greatly from these changes, demonstrating their promise for deploying effective, real-time prediction systems for flight delays. Future studies should concentrate on refining these models, investigating more sophisticated balancing strategies, and incorporating real-time data updates to improve operational decision-making and passenger happiness in the aviation sector.

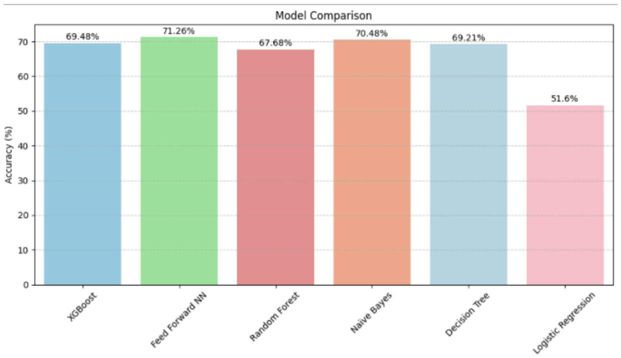

Figure 21. Model Comparison

6. CONCLUSION

Aero Forecast's thorough data analysis has produced important insights that highlight the significant influence of weather on flight delays and highlight significant inefficiencies in airline operating procedures. Using advanced models like Decision Trees, Naive Bayes, and XGBoost, we have uncovered actionable insights that enable airlines to make proactive schedule adjustments, resource allocation decisions, and improve overall timeliness.

The knowledge gained from this study is critical to cutting operational costs, minimizing passenger complaints, and improving the overall effectiveness of air travel. These developments directly support our goal of providing a flawless travel experience. Airlines may proactively manage customer expectations and resource allocation by precisely forecasting probable delays, which ultimately leads to a more dependable and resilient service.

7. FUTURE WORKS

In the next phase of our project, we intend to create Sky Status, a cutting-edge application that will offer passengers with real-time updates on flight delays. Sky Status will be developed in various stages, starting with a user-friendly front end, and progressing to a comprehensive backend system to control data flow and interact with external APIs. We will use Aero forecast and other advanced machine learning approaches to improve the forecasting accuracy of flight delays, including models like XG Boost and Neural Networks. Data management will take a dual-database approach, with user inputs kept in a flexible NoSQL database and structured data handled via SQL databases for fast real-time analytics. Security protocols will be strictly enforced to preserve data integrity and user privacy. Sky Status will be delivered on a scalable cloud platform, with regular upgrades and improvements depending on user feedback and technical breakthroughs. This project is expected to dramatically improve the passenger experience by providing timely and accurate flight status information, establishing a new benchmark in the aviation industry for real-time passenger information systems.

8. REFERENCES

1. *Kim, Y.J.; Choi, S.; Briceno, S.; Mavris, D. A deep learning approach to flight delay prediction. In Proceedings of the AIAA/IEEE Digital Avionics Systems Conference, Sacramento, CA, USA, 25–29 September 2016*

2. *Lambelho, M.; Mitici, M.; Pickup, S.; Marsden, A. Assessing strategic flight schedules at an airport using machine learning-based flight delay and cancellation predictions. J. Air Transp. Manag. 2020, 82, 101737.*

3. *Choi, S.; Kim, Y.J.; Briceno, S.; Mavris, D. Prediction of Weather-Induced Airline Delays Based on Machine Learning Algorithms. In Proceedings of the IEEE/AIAA 35th Digital Avionics Systems Conference (DASC), Sacramento, CA, USA, 25–29 September 2016.*

4. *Alonso, H.; Loureiro, A. Predicting flight departure delay at Porto Airport: A preliminary study. In Proceedings of the 2015 7th International Joint Conference on Computational Intelligence (IJCCI), Lisbon, Portugal, 12–14 November 2015; IEEE: New York, NY, USA, 2015; Volume 3, pp. 93–98.*

5. *Chen, J.; Li, M. Chained predictions of flight delay using machine learning. In Proceedings of the AIAA Scitech 2019 Forum, San Diego, CA, USA, 7–11 January 2019; pp. 1–25.*

6. *Kalliguddi, A.M.; Leboulluec, A.K. Predictive Modeling of Aircraft Flight Delay. Univers. J. Manag. 2017, 5, 485–491.*

7. *Manna, S.; Biswas, S.; Kundu, R.; Rakshit, S.; Gupta, P.; Barman, S. A statistical approach to predict flight delay using gradient boosted decision tree. In Proceedings of the ICCIDS 2017—International Conference on Computational Intelligence in Data Science, Chennai, India, 2–3 June 2018; pp. 1–5.*

8. *Yu, B.; Guo, Z.; Asian, S.; Wang, H.; Chen, G. Flight delay prediction for commercial air transport: A deep learning approach. Transp. Res. Part E Logist. Transp. Rev. 2019, 125, 203–221.*

9. *Thiagarajan, B.; Srinivasan, L.; Sharma, A.V.; Sreekanthan, D.; Vijayaraghavan, V. A machine learning approach for prediction of on-time performance of flights. In Proceedings of the AIAA/IEEE Digital Avionics Systems Conference, St. Petersburg, FL, USA, 17–21 September 2017; pp. 5–10.*

10. *Ayhan, S.; Costas, P.; Samet, H. Predicting estimated time of arrival for commercial flights. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Data Mining, London, UK, 19–23 August 2018; pp. 33–42.*