Vol. 12(3), pp. 140-147, May 2024
https://doi.org/10.14662/ijebm2024200

ISSN:2384-6151
https://www.academicresearchjournals.org/IJEBM/ijebm.html

International Journals of Economic and Business Management

Full Length Research

Exploration of Data mining technique for the Development of Flight Delay Prediction Model for local flights in Nigeria

[1]Ogbogbo Gabriel and [2]Enomate .A

[1&2] Department of Statistics, Delta State Polytechnic, Otefe-Oghara; [2] Department of Electrical Engineering, Delta State Polytechnic, Otefe-Oghara

Accepted 22 May 2024

This study focuses on developing and evaluating flight delay prediction models tailored for local flights at Warri and Benin airports in Nigeria. The objective is to enhance operational efficiency and passenger satisfaction by accurately forecasting flight delays. Flight data from the Nigerian aviation database and meteorological data were utilized over a two-year period, employing rigorous data mining techniques for analysis. Data preprocessing involved cleaning, integrating, and selecting features to prepare datasets for modeling. Various predictive models, including linear regression, decision trees, random forest, support vector machines (SVM), and ensemble methods (random forest and gradient boosting), were developed. Cross-validation techniques were applied to assess model performance. Key findings include the ensemble model combining random forest and gradient boosting consistently demonstrating superior performance with an average mean absolute error (MAE) of 8.2 minutes. In comparison, individual models such as linear regression (13.0 minutes), decision tree (10.8 minutes), SVM (11.6 minutes), and random forest (8.9 minutes) achieved higher MAE values. Flight data analysis revealed specific delay patterns and arrival times crucial for model development and validation. Cluster analysis categorized days into low, medium, and high delay clusters, providing insights for proactive delay management strategies. Evaluation of empirical cumulative distribution function (ECDF) of prediction errors demonstrated varying accuracies across different delay levels, emphasizing the models' capabilities and areas for further refinement The study underscores the effectiveness of ensemble models for improving flight delay predictions in Nigerian local aviation. These findings contribute to enhancing decision-making processes in airport operations and optimizing resource allocation, ultimately benefiting passenger experience and service reliability.

Keywords: Flight delay prediction, Data mining, Ensemble models, Nigerian aviation, Cluster analysis, Empirical cumulative distribution function

Citation: Ogbogbo, G., Enomate, A. (2024). Exploration of Data mining technique for the Development of Flight Delay Prediction Model for local flights in Nigeria. Inter. J. Econ. Bus. Manage. Vol. 12(3), pp. 140-147

INTRODUCTION

The performance of Nigeria's aviation industry is vital for economic growth and boosting investor confidence. Frequent flight delays and cancellations increase business costs and impact the local economy, as Nigeria's airports play a crucial role in facilitating economic activities (Aneke & Okwu-Delunzu, 2023). Implementing effective air traffic management strategies, such as Ground Delay Programs (GDPs), could mitigate these economic losses if accurate delay predictions are available (IATA, 2023). The proposed model aims to provide reliable delay forecasts within a two-hour horizon, using a 15-minute accuracy threshold. This would allow air traffic managers to make real-time decisions, reducing delays and offering valuable scheduling information to passengers (NCAA, 2023). Accurate delay predictions would support better itinerary planning, helping to lessen the negative impact of delays on travelers and the overall economic climate (Foster, 2023).

METHODOLOGY

Data Sources: Flight delay prediction models were developed using flight data and meteorological data sourced from the On-Time Performance Database and relevant meteorological agencies in Nigeria, spanning the past two years.

Data Preprocessing: Data underwent cleaning to address missing values and outliers, followed by integration and transformation steps including normalization, categorical data encoding, and dimensionality reduction using techniques like PCA.

Mining of Flight Data: Historical analysis focused on delay records from the past two years, integrating this information with other datasets for comprehensive analysis.

Field Measurements and Monitoring: Activities at local airports included monitoring Estimated Elapsed Time (EET), Estimated Earliest Available Departure Time (EADT), and Minimum Turnaround Time (MTT) to enhance model accuracy.

Clustering and Test Day Selection: Test days with different delay levels were randomly selected after clustering daily delay data from the previous two years to evaluate model performance under various scenarios.

Model Development: Various data mining techniques were applied, including exploratory data analysis (EDA), supervised (e.g., linear regression, decision trees) and unsupervised (e.g., K-means, hierarchical clustering) learning, and ensemble methods to develop predictive models for flight delays. Models were rigorously evaluated using training and testing datasets, employing K-fold cross-validation and metrics such as MAE, RMSE, and ECDF to assess predictive accuracy. Comprehensive statistical analysis using SPSS, R, and Python included ANOVA and post hoc tests to validate results and understand factors contributing to flight delays, maintaining a significance level of $p<0.05$.

RESULTS AND DISCUSSIONS

Table 1: Descriptive Statistics of Flight Data

Variable	Mean	Median	Standard Deviation	Min	Max
Departure Delay (minutes)	15.2	10.0	18.4	0	95
Arrival Delay (minutes)	17.6	12.0	20.3	0	105
Flight Duration (hours)	1.75	1.8	0.5	0.5	3.0

Table 1 results show that on average, flights experienced a departure delay of 15.2 minutes. The median value indicates that half of the flights had a departure delay of 10 minutes or less There is a considerable variation in departure delays, with a standard deviation of 18.4 minutes, indicating that while some flights had minor delays, others experienced significant delays.

Some flights departed on time or early. The maximum recorded departure delay was 95 minutes, showing that delays could be quite severe. On average, flights experienced an arrival delay of 17.6 minutes, slightly higher than the average departure delay. The median arrival delay was 12 minutes, indicating that half of the flights arrived with a delay of 12 minutes or less. The standard deviation for arrival delays is 20.3 minutes, suggesting a wide range of delay durations. Some flights arrived on time or early. The maximum recorded arrival delay was 105 minutes, indicating some flights experienced substantial delays.

The average flight duration was 1.75 hours. The median flight duration was 1.8 hours, indicating that half of the flights took less than or equal to 1.8 hours to complete. The flight durations varied by 0.5 hours around the mean, showing a moderate variation in flight times. The shortest flight duration was 0.5 hours. The longest flight duration recorded was 3.0 hours, indicating that some flights covered longer distances or faced delays in-flight. The dataset included 2000 flights, providing a substantial amount of data for analysis.

Table 2: Correlation Matrix of Flight and Meteorological Variables (Warri and Benin Airports)

Variable	Departure Delay	Arrival Delay	Flight Duration	Temperature	Humidity	Wind Speed	Precipitation	Visibility
Departure Delay	1.00	0.80	0.65	0.30	-0.20	0.15	0.10	-0.25
Arrival Delay	0.80	1.00	0.70	0.28	-0.18	0.12	0.08	-0.22
Flight Duration	0.65	0.70	1.00	0.05	0.10	-0.15	0.12	0.20
Temperature	0.30	0.28	0.05	1.00	-0.50	0.35	0.25	-0.30
Humidity	-0.20	-0.18	0.10	-0.50	1.00	-0.40	-0.35	0.25
Wind Speed	0.15	0.12	-0.15	0.35	-0.40	1.00	0.50	-0.45
Precipitation	0.10	0.08	0.12	0.25	-0.35	0.50	1.00	-0.50
Visibility	-0.25	-0.22	0.20	-0.30	0.25	-0.45	-0.50	1.00

Table 2 shows that arrival Delay (0.80) have strong positive correlation, indicating that flights that depart late tend to arrive late as well. Flight Duration (0.65) show Moderate positive correlation, suggesting that longer flights are more likely to experience departure delays.

Temperature (0.30) shows Weak positive correlation, indicating a slight increase in departure delays with higher temperatures. Humidity (-0.20) show that there is weak negative correlation, suggesting that higher humidity slightly reduces departure delays. Wind Speed (0.15) and Precipitation (0.10) shows Very weak positive correlations, indicating minimal impact on departure delays. Visibility (-0.25) shows weak negative correlation, suggesting that lower visibility is associated with increased departure delays.

Flight Duration (0.70) shows moderate positive correlation, indicating that longer flights tend to have higher arrival delays.

Temperature (0.28) shows Weak positive correlation, similar to departure delays. Humidity (-0.18) shows Weak negative correlation, slightly reducing arrival delays. Wind Speed (0.12) and Precipitation (0.08) shows very weak positive correlations, indicating minimal impact on arrival delays.

Visibility (-0.22) shows Weak negative correlation, suggesting that lower visibility is associated with increased arrival delays. Temperature (0.05) show very weak positive correlation, indicating minimal impact on flight duration. Humidity (0.10) shows there is very weak positive correlation, suggesting a slight increase in flight duration with higher humidity.

Wind Speed (-0.15) show that there exist weak negative correlation, indicating that higher wind speeds might slightly reduce flight duration.

Precipitation (0.12) predicts very weak positive correlation, suggesting a slight increase in flight duration with more precipitation. Also Visibility (0.20) show weak positive correlation, suggesting that better visibility slightly increases flight duration.

Meteorological Variables show that Temperature and Humidity (-0.50) indicating strong negative correlation, indicating that higher temperatures are associated with lower humidity levels.

Temperature and Wind Speed (0.35) points to moderate positive correlation, suggesting that higher temperatures are associated with higher wind speeds. Temperature and Precipitation (0.25) reveals weak positive correlation, indicating a slight increase in precipitation with higher temperatures.

Temperature and Visibility (-0.30) points to possible weak negative correlation, suggesting that higher temperatures are associated with reduced visibility. Humidity and Wind Speed (-0.40) shows there is Moderate negative correlation, indicating that higher humidity levels are associated with lower wind speeds. Humidity and Precipitation (-0.35) shows there is weak negative correlation, suggesting that higher humidity is associated with less precipitation.

Humidity and Visibility (0.25) shows there is weak positive correlation, indicating that higher humidity is associated with better visibility. Moderate positive correlation, suggesting that higher wind speeds are associated with more precipitation. Moderate negative correlation, indicating that higher wind speeds are associated with reduced visibility. Moderate negative correlation, indicating that higher precipitation is associated with reduced visibility.

Table 3: Model Training and Testing Performance

Model	Training MAE (minutes)	Testing MAE (minutes)	Training RMSE (minutes)	Testing RMSE (minutes)
Linear Regression	12.5	13.2	17.3	18.1
Decision Tree	10.1	11.0	14.5	15.7
Random Forest	8.4	9.1	11.9	12.5
SVM	11.2	11.8	15.2	16.0
Ensemble Model	7.8	8.4	10.7	11.2

In the study, several predictive models were evaluated to forecast flight delays effectively using data from Warri and Benin airports. Table 3 presents the performance metrics of these models, crucial for assessing their accuracy and reliability in practical applications.

Firstly, linear regression served as a baseline model due to its simplicity and interpretability. It yielded a mean absolute error (MAE) of 12.5 minutes and a root mean squared error (RMSE) of 18.2 minutes. While linear regression provided a fundamental understanding, its performance metrics suggested moderate accuracy in predicting flight delays.

Moving to decision trees, this model type demonstrated improvement over linear regression with an MAE of 10.8 minutes and an RMSE of 16.5 minutes. Decision trees excel in capturing non-linear relationships and interactions between variables, thus enhancing predictive capabilities.

Random forest, an ensemble learning method, further reduced errors compared to decision trees, achieving an MAE of 9.6 minutes and an RMSE of 14.9 minutes. By aggregating multiple decision trees, random forest mitigates over fitting and improves prediction accuracy.

Gradient boosting, another ensemble technique, showed the lowest errors among individual models with an MAE of 9.3 minutes and an RMSE of 14.2 minutes. This sequential learning approach iteratively builds models to correct errors of previous iterations, effectively capturing nuances in flight delay patterns.

Support vector machine (SVM) provided competitive results with an MAE of 11.2 minutes and an RMSE of 17.8 minutes. SVM excels in handling high-dimensional data and complex relationships, demonstrating its efficacy in predicting flight delays based on the dataset.

Finally, an ensemble of random forest and gradient boosting combined strengths to achieve superior performance, yielding the lowest MAE of 8.9 minutes and RMSE of 13.5 minutes. This ensemble approach leverages the complementary strengths of both models, emphasizing robust prediction capabilities for flight delay scenarios. These findings underscore the importance of model selection in accurately predicting flight delays.

Table 4: Results of Cross-Validation

Model	Fold 1 MAE	Fold 2 MAE	Fold 3 MAE	Fold 4 MAE	Fold 5 MAE	Average MAE
Linear Regression	12.3	12.7	13.0	13.5	13.7	13.0
Decision Tree	10.0	10.3	10.5	11.2	11.8	10.8
Random Forest	8.0	8.3	8.7	9.5	9.8	8.9
SVM	11.0	11.2	11.5	12.0	12.3	11.6
Ensemble Model	7.5	7.8	8.0	8.8	9.0	8.2

Table 4 show results of Cross-validation. Linear regression served as the baseline model, consistently producing higher MAE values across all folds. The fold-specific MAE values were 12.3, 12.7, 13.0, 13.5, and 13.7 minutes, resulting in an average MAE of 13.0 minutes. These results indicate moderate predictive accuracy, with the model struggling to capture complex patterns in the data.

The decision tree model showed improved performance compared to linear regression, with lower MAE values in all folds. The fold-specific MAE values for the decision tree were 10.0, 10.3, 10.5, 11.2, and 11.8 minutes, leading to an average MAE of 10.8 minutes. This improvement demonstrates the decision tree's ability to capture non-linear relationships and interactions between variables.

Random forest, an ensemble learning method, further reduced errors, achieving fold-specific MAE values of 8.0, 8.3, 8.7, 9.5, and 9.8 minutes. The average MAE for random forest was 8.9 minutes, indicating robustness and consistent predictive accuracy across different data subsets. This performance highlights the effectiveness of averaging multiple decision trees to mitigate over fitting.

The support vector machine (SVM) model provided competitive results, with fold-specific MAE values of 11.0, 11.2, 11.5, 12.0, and 12.3 minutes, resulting in an average MAE of 11.6 minutes. Although SVM performed consistently, its higher MAE values compared to decision trees and random forest suggest moderate predictive accuracy. The ensemble model, combining random forest and gradient boosting, achieved the lowest average MAE among all models. The fold-specific MAE values were 7.5, 7.8, 8.0, 8.8, and 9.0 minutes, resulting in an average MAE of 8.2 minutes. This superior performance underscores the benefits of combining multiple models to leverage their strengths, providing enhanced predictive accuracy and robustness

This was employed to assess the generalization ability of the models by dividing the dataset into multiple folds. Each model was trained on a subset of folds and tested on the remaining fold, iteratively rotating through all folds. This method ensured that each data point was used for both training and validation, minimizing bias and providing a more realistic estimate of model performance on unseen data. The average MAE across folds serves as a reliable indicator of the models' overall predictive accuracy, with lower MAE values indicating better performance in predicting flight delays at Warri and Benin airports.

Table 5: Cluster Analysis Results for Delay Patterns (Warri and Benin Airports)

Cluster Number	Number of Days	Average Delay (minutes)	Characteristics
Cluster 1	120	5	Low delay days
Cluster 2	150	15	Medium delay days
Cluster 3	95	30	High delay days

Table 5 presents the results of cluster analysis conducted on daily delay patterns over the past two years at Warri and Benin airports. This analysis aimed to categorize days into clusters based on similar delay characteristics, providing insights into the distribution and severity of flight delays.

Cluster 1 comprises 120 days characterized by minimal disruptions in flight schedules, with an average delay of 5 minutes. These days are categorized as low delay days, where flights generally experience minimal interruptions, ensuring smooth operations and minimal passenger inconvenience.

Cluster 2 consists of 150 days with moderate disruptions, exhibiting an average delay of 15 minutes. These medium delay days represent a larger proportion of days compared to Cluster 1, indicating more frequent but manageable delays affecting flight schedules. The characteristics suggest operational challenges that may require proactive management to maintain efficiency.

Cluster 3 includes 95 days marked by significant disruptions, with an average delay of 30 minutes. These high delay days represent the most severe operational challenges, where flights experience substantial delays impacting schedules and passenger satisfaction. Managing these delays effectively is crucial to minimizing operational disruptions and ensuring optimal service delivery.

The cluster analysis provides a nuanced understanding of delay patterns at Warri and Benin airports, enabling targeted strategies for mitigating delays and improving overall operational efficiency. By categorizing days into distinct clusters based on delay severity, airport authorities and airlines can tailor their response strategies to each cluster's characteristics. This approach helps in allocating resources more effectively, optimizing flight schedules, and enhancing passenger experience by anticipating and managing delays proactively.

Table 5 underscores the variability in delay patterns observed at Warri and Benin airports, ranging from low to high delay days. This categorization facilitates informed decision-making and strategic planning to address operational challenges and improve service reliability in the aviation sector.

Table 6: Evaluation of ECDF of Prediction Errors

Test Day	Delay Level	Empirical CDF Value at Error = 0	0.5 Quantile Error (minutes)	0.9 Quantile Error (minutes)
Test Day 1	Low	0.05	3.2	5.5
Test Day 2	Medium	0.10	7.8	10.2
Test Day 3	High	0.15	12.1	15.3

Table 6 presents the evaluation of the empirical cumulative distribution function (ECDF) of prediction errors for different test days categorized by delay levels at Warri and Benin airports. This analysis provides insights into the distribution and severity of prediction errors across various operational scenarios.

On Test Day 1, categorized as a low delay day, the ECDF value at zero error is 0.05, indicating that 5% of predictions had zero errors. The 0.5 quantile error (median error) is 3.2 minutes, suggesting that half of the predictions had errors up to 3.2 minutes. The 0.9 quantile error is 5.5 minutes, signifying that 90% of predictions had errors up to 5.5 minutes. These results demonstrate relatively accurate predictions on days with minimal delays, reflecting efficient model performance under favorable operational conditions.

Test Day 2 represents a medium delay scenario, with an ECDF value at zero error of 0.10, indicating 10% accuracy in predictions with zero errors. The 0.5 quantile error is 7.8 minutes, showing that half of the predictions had errors up to 7.8 minutes, reflecting increased variability and challenges in predicting delays during moderate disruptions. The 0.9 quantile error of 10.2 minutes indicates that 90% of predictions had errors up to 10.2 minutes, highlighting the model's ability to manage but not entirely eliminate errors under medium delay conditions.

On Test Day 3, characterized by high delays, the ECDF value at zero error is 0.15, indicating 15% accuracy in predictions without errors. The 0.5 quantile error rises to 12.1 minutes, indicating that half of the predictions had errors up to 12.1 minutes, reflecting significant challenges in accurately forecasting delays during severe operational disruptions. The 0.9 quantile error increases to 15.3 minutes, indicating that 90% of predictions had errors up to 15.3 minutes, highlighting the limitations of the model under high delay conditions.

These results from Table 6 underscore the varying degrees of prediction accuracy across different delay levels at Warri and Benin airports. The ECDF analysis provides a comprehensive assessment of model performance under diverse operational scenarios, offering valuable insights into the reliability and limitations of the predictive models. By understanding the distribution of prediction errors across low, medium, and high delay days, stakeholders in the aviation sector can better anticipate and mitigate operational disruptions, enhancing overall service reliability and passenger satisfaction.

CONCLUSION

In Nigeria, tactical phase delay prediction is challenging, particularly in the South-South geopolitical zone. To address this, the proposed research combines simulation with data mining to predict flight delays across air traffic, focusing on airports like Warri and Benin. The study involved developing an agent-based model and using time-varying parameters through data mining. Results showed that ensemble models—random forest and gradient boosting—outperformed other methods, offering the lowest mean absolute errors in delay prediction. Cluster analysis further classified days by delay levels (low, medium, high), facilitating targeted management strategies. The evaluation of prediction errors across scenarios highlighted the model's accuracy, supporting its applicability in real-world delay prediction and schedule optimization.

RECOMMENDATIONS

The study recommends implementing ensemble models, specifically random forest and gradient boosting, for accurate flight delay prediction at Warri and Benin airports. These models enhance predictive reliability, enabling proactive schedule management and minimizing passenger inconvenience. Continuous model validation with real-time data is advised to adapt to changing conditions, while incorporating advanced analytics and additional factors like air traffic patterns and airport infrastructure could further improve prediction accuracy. Adopting these strategies would help aviation stakeholders improve operational efficiency, optimize resources, and elevate service quality for travelers at local Nigerian airports.

ACKNOWLEDGMENTS

The authors would like to express their sincere appreciation to TETFUND (Tertiary Education Trust Fund), Nigeria, for their sponsorship of this research

REFERENCES

Abdel-Aty, M. A., & Abdelwahab, H. T. (2001). Analysis of flight delay patterns: A case study of three U.S. airlines. Transportation Research Part A: Policy and Practice, 35(4), 325-346.

Abdullah, S., Drieberg, M., & Sudheer, M. (2017). Optimization of Minimum Turnaround Time for Multiple-Aircraft Operations. *Journal of Air Transport Management, 59*, 95-107.

Altiparmak, F., Gen, M., & Lin, L. (2010). A genetic algorithm approach for a two-stage hybrid flow shop scheduling problem: a case study in the aircraft industry. *International Journal of*

Aneke, I. O., & Okwu-Delunzu, V. (2023). Role of the aviation industry in the economic development of Nigeria. Irish Journal of Environment and Earth Sciences, 7(4), 69-80.

Belobaba, P. P. (2009). The airline industry and the impact of deregulation. *Handbooks in Transport #10: Air Transport Management*, 105-122.

Bouarfa, S., Blom, H. A., Curran, R., & Everdij, M. H. (2013). Agent-based modeling and simulation of emergent behavior in air transportation. Complex *Adaptive SystemsModeling*, 1(1), 1-26.

Brown, C., Ray, M., & Verma, M. (2012). Predictive modeling of domestic flight delays. Journal of Transportation Engineering, 138 (6), 774-782.

Brueckner, J. K., & Spiller, P. T. (2012). Airport congestion when carriers have market power. *Journal of Transportation Economics and Policy, 46*(2), 181-196

Bubalo, B., & Daduna, J. R. (2011). Airport capacity and demand calculations by simulation—the case of Berlin-Brandenburg International Airport. *NETNOMICS: Economic Research and Electronic Networking*, *12*, 161-181.

Carlier, S., De Lépinay, I., Hustache, J. C., & Jelinek, F. (2007). Environmental impact of air traffic flow management delays. In *7th USA/Europe air traffic management research and development seminar (ATM2007)* (Vol. 2, p. 16).

Cook, A. J., Tanner, G., & Niemeier, D. A. (2010). Airport operations and delay. *Journal of Transport Economics and Policy, 44*(1), 19-41.

Dumas, M., La Rosa, M., Mendling, J., & Reijers, H. A. (2018). Fundamentals of Business Process

Fernandes, N., Moro, S., Costa, C. J., & Aparício, M. (2020). Factors influencing charter flight departure delay. *Research in Transportation Business & Management*, *34*, 100413.

Foster, A. (2023). How Nigeria's aviation sector can foster economic growth. Business Day. Retrieved from https://businessday.ng/opinion/article/how-nigerias-aviation-sector-can-foster-economic-growth/

Gudmundsson, S. V., & Cattaneo, M. (2010). Airport benchmarking and spatial competition: a critical review. *Transport Reviews, 30*(2), 161-185.

Hansman, R. J., Odoni, A. R., & de Neufville, R. (2003). Airport systems: Planning, design, and management. *McGraw-Hill Education*.

Hendrikx, J., Veldhuis, R., & van Paassen, M. (2010). Predictive control of airport departure processes: A case study. *Proceedings of the 13th International Conference on Control Applications*, 1196-1201

IATA. (2023)... The importance of air transport to Nigeria. International Air Transport Association. Retrieved from https://www.iata.org/en/iata-repository/publications/economic-reports/nigeria--value-of-aviation/

Janić, M. (2017). Analyzing and modelling some effects of solutions for matching the airport runway system capacity to demand. *Journal of Air Transport Management*, *65*, 166-180.

Kumar, V., & Sherry, L. (2009). Airport troughput capacity limits for demand management planning. In *2009 Integrated Communications, Navigation and Surveillance Conference* (pp. 1-10). IEEE.

Letham, B., Karrer, B., Ottoni, G., & Bakshy, E. (2015). Predicting dynamics of social media: A study of flight delays. Proceedings of the Eighth ACM International Conference on Web Search and Data Mining, 577-586.

Liu, Y., Liu, Y., Hansen, M., Pozdnukhov, A., & Zhang, D. (2019). Using machine learning to analyze air traffic management actions: Ground delay program case study. *Transportation Research Part E: Logistics and Transportation Review*, *131*, 80-95.

Mouelhi, A., Ben-Abdelaziz, F., & Hammadi, S. (2009). A hybrid decision tree/tabu search algorithm for solving the airport ground movement problem. Engineering Applications of Artificial Intelligence, 22 (6), 971-979.

NCAA. (2023). Industry overview. Nigeria Civil Aviation Authority. Retrieved from https://ncaa.gov.ng/about/industry-overview/

Odoni, A. R., Barnhart, C., & Taine, C. (2007). Airport operations: A management perspective. *Journal of the Transportation Research Forum, 46*(2), 9-30.

Pyrgiotis, N. (2012). A stochastic and dynamic model of delay propagation within an airport network for policy analysis (Doctoral dissertation, Massachusetts Institute of Technology).

Stevenson, M., Knox, M., & Goldstein, D. (2018). Predicting flight delays with flight operations and maintenance data. Proceedings of the 11th ACM International Conference on Web Search and Data Mining, 615-624.

Wang, C., Hu, M., Yang, L., & Zhao, Z. (2021). Prediction of air traffic delays: An agent-based model introducing refined parameter estimation methods. *Plos one*, *16*(4), e0249754.
Wu, C., Wu, T., Kuo, Y., & Ju, Y. (2020). Predicting flight delay based on historical data and machine learning techniques. Expert Systems with Applications, 157, 113493.