ISA Journal of Engineering and Technology (ISAJET)

Homepage: https://isapublisher.com/isajet/

Volume 3, Issue 2, March-April, 2026 ISSN: 3049-1843

Flight Delay Analysis and Prediction Using Hybrid Shallow Machine Learning Algorithms

Ishaya, N. J.[1], Gurumdimma, N. Y.[2] & Deme, A. C.[3]

[1]ICT Directorate, University of Jos
[2,3]Department of Computer Science, University of Jos

Received: *01.10.2025* | ***Accepted:*** *15.10.2025* | ***Published:*** *01.03.2026*

***Corresponding Author:** Ishaya, N. J.

DOI: *10.5281/zenodo.18822983*

Abstract **Original Research Article**

The airline industry has grown tremendously in the last two decades. This growth is accompanied with challenges, one of which is flight delay. Flight delay is a serious setback that results in loss of billions of dollars each year. The aviation industry is highly dependent on punctual flight operations, and delays can incur substantial costs, inconvenience passengers, and affect airline reputation. To mitigate the impact of flight delays, this research presents an innovative approach to analyse and predict flight arrival delays using a hybrid machine learning technique. This approach combines the strengths of two machine learning algorithms, in two stages, while stage one classifies the delay, stage two estimates the delay time in minutes enhancing prediction accuracy and robustness. The model was evaluated with classification and regression metrics. After a comprehensive data preprocessing and modelling, a Recall score of 94%, R-Square score of 85.07%, MAE of 20.72 and RMSE of 15.92 minutes was obtained. The resulting model offers accurate predictions, real-time adaptability, and valuable insights into delay patterns. This paper contributes to improved operational efficiency, enhanced passenger experiences, and cost savings for airlines and airports operators.

Keywords: Flight delay, hybrid shallow architecture, machine learning, aviation, air transportation, Logistic regression, xgboost.

Copyright © 2026 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

Introduction

Artificial Intelligence (AI) is getting more popularity in the advancement of technology in this digital age, it is one of the areas of computer science that deals with the ability of system to behave and think in a similar way like human brain (Yazdi *et al.*, 2020), AI can be categorized into different subdivisions of which Machine learning (ML) is one of them (Anguila, et al., 2023; Khalil, et al., 2022). ML can be defined as a process of training a system to solve a problem or a task and to understand its environment by learning its historical data. This concept has spread steadily in the domain of computer science and has become a key tool in the design of AI (Zeng *et al.*, 2021). Areas which Machine learning can be applied include; Web search engines, email spam filters, recommender systems (Sirisati, et al., 2021) for online stores, targeted advertisement placement, credit card scoring, fraud detection, stock market trading, drugs design, medical image processing and many other applications. (Chakrabarty et al., 2019) reported that machine learning also called data mining or predictive analytics will be the driving force behind most technological innovations.

The airline industry has grown tremendously in the last two decades, with approximately 5% annually in the last 10 years International Air Transport Association (IATA) (Aviation trend, 2018). The demand for air travel grew progressively with a global air passenger growth of about 6.5% in 2018, far above the 10-year average annual growth of 5.5%. Unfortunately, the high demand of air transportation which also result in improvement of Gross Domestic Product (GDP) of US economy comes with disadvantages, one of it is flight delays (Yazdi, 2020; Ranga & Anga, 2021). The high volume of congestion of the air system caused by the disproportionate growth between flights, airport capacity and other factors causes flights to be delayed, diverted or even cancelled (Ayo-Agunbiade & Stephens, 2019).

In the preliminary review of this study, it was discovered that numerous reports from earlier researches show an increasing incidences of flight delays in airline's operations; this is generally because some internal and or external factors are not being considered during analysis. In the worst case, a flight can be cancelled instantly mostly due to extreme weather conditions (Carvallo, et al., 2020). Looking at passenger dissatisfaction and loyalty to airline operators, a survey was conducted to determine the level of satisfaction of passengers based on the services of various airline operated in the United State, the result shows that flight delay is the major factor for passenger bad experience.

There are different factors that cause flight delay, these include the followings; mechanical problems, atmospheric conditions, delay propagation and air traffic control issues among others (Wang, et al., 2021; Chunzheng, et al., 2021; Seyedmirsajad, M., & Armin, M., 2022). Furthermore, factors such as the scarcity of labour in the sectors can cause delays and the changes in traffic management using initiatives such as Ground Delay Programs (GDP) stimulate delays. GDP is seen as initiatives that represent traffic management procedures where airplanes are delayed at their departure airport to manage demand and capacity at their arrival airport (Gou, et al., 2021). Flight delay is defined as when a flight arrives or departs an airport fifteen minutes (15) later than the scheduled time.

Flight delay affect stakeholders in the aviation industry such as passengers, airlines operators, airline workers, airports, airport ground operation workers, airline service suppliers, air traffic management units and many others (Gholami & Khashe, 2022; Pinto, et al., 2023). On the other hand, all of these stakeholders also contribute to specific factors such as passenger boarding, flight and crew planning, loading fuel, queuing planes to the runway, when combined may result in a flight delay. It's always frustrating to show up to the airport only to see your flight been delayed, sometimes by hours, which may eventually affect your arrival.

Empirical studies reveal that travellers have high preferences to avoid the adverse effects of delays (Ranga & Anga, 2021). Industries, which rely on air transportation, carry large costs of delays due to a decrease in productivity. According to the U.S. Bureau of Transportation Statistics, in 2018 more than 720 million passengers boarded a domestic flight within the United States, around 19.5% of scheduled flight experienced some kind of delay, and hostile weather conditions taking about 60%. According to (Ismail, et al., 2019), passengers value the avoidance of a flight delay with around United State Dollar (USD) 1.50 per minute.

The analysis and prediction of air transport delays becomes necessary since a better knowledge of their existence, and corresponding triggers, can improve the performance of airlines and, consequently airports, in their operations by the possibility of anticipation (Hee, 2019). This thesis considers the analysis and prediction of the delays with focus on the arrivals since these are more related with the passenger's satisfaction and because one arrival delay may trigger a delay in a departure (Id, et al., 2021).

Literature Review

Machine Learning (ML) is a method of data analysis focused on the development of computer algorithms to transform data into useful information and actions (Kumar & Sreekanth., 2023; Imran, et al., 2023). Machine Learning has been widely used in various industries, examples include forecasts of weather behaviour and long-term climate changes, reduction of fraudulent credit card trans- actions, prediction of popular election outcomes, discovery of genetic sequences linked to diseases, and image recognition (Maglogiannis, et al., 2021; Reddy, et al., 2023). ML is the computational aspect of AI that employs processes such as knowledge discovery, pattern recognition, and data mining to learn from data and develop predictive work flows automatically. The standard supervised ML method involves feeding input data to a trained model. The training procedure seeks to identify a set of model parameters that maximizes the relationship between the input features and the target variables through an iterative procedure.

ML consists of the processes as seen in figure 2.1 of discovering new, valid, useful and perceptible patterns in data as cited by (Kalyani, et al., 2020). It covers phases such as;

(i) The collection of datasets.

(ii) The preprocessing of data

(iii) The Modelling.

(iv) Performance Evaluation

(v) Implementation of the models.

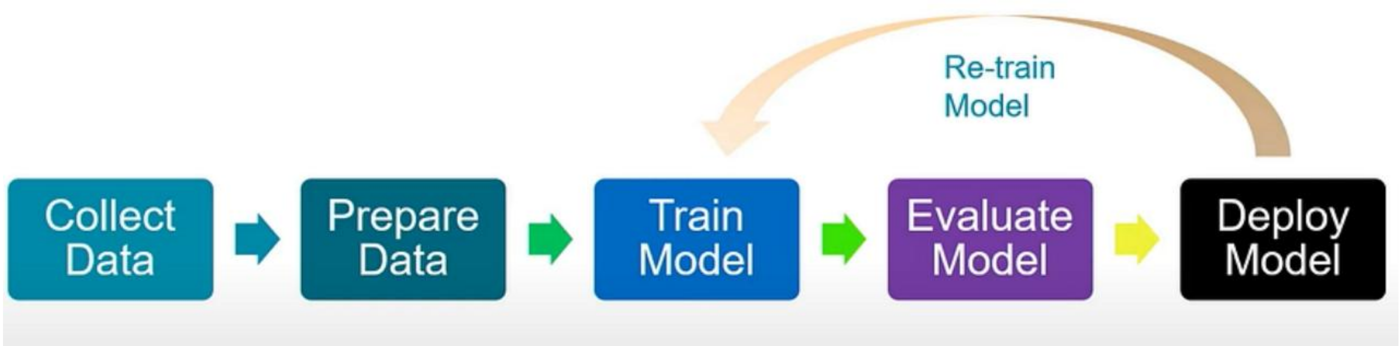


Figure 2. 1: Basic Machine Learning Processes

Source: Medium.com

Machine Learning is considered one of the steps of knowledge discovery as presented in (Jiang, et al., 2020) Figure 2.1. It represents the application of algorithms to extract patterns from the data, translating it into information (Chakrabarty, 2019; Anguita, et al., 2023). The prevalent flight delay in airline operations has led the research field to focus mainly towards developing approaches that address flight arrival delays. Several studies have been conducted to predict flight delays (Ali, et al., 2023).

Paramita, et al., (2022) demonstrated a cluster computing in a virtual environment to simulate the prediction of flight delay, the setup speed up the prediction result by 35.8% with random forest model. They suggested that other machine learning algorithms be tested on the virtual environment. Likewise, (Rik et al, 2021; Chakrabarty, 2020) considered airport planner preferences and imbalanced dataset to predict flight delay and cancellation, random forest and multilayer perceptron were used in the experiment, multilayer perceptron performs better with 78% accuracy for arrival delay and 68% accuracy for departure delay.

Ayaydin & Ali, (2022) generated and predicted flight delay in an airport using random forest, the result shows that random forest classification achieved 96.48% and 93.92% for R-Squared value. The research suggested using domestic and international dataset with weather features in future work. In another research, (Qu et al, 2020; Mas-pujol, et al., (2022) used dual channel CNN and DCNN SE-DenseNet (Squeeze and Exitation-densely connected CNN with weather data to predict flight delay. DCNN achieved a 92.1% accuracy and SE-DenseNet accuracy was 93.1%.

Bisandu et al., (2022) presented a model that uses social ski driver conditional autoregression (SSDCA)-based deep learning to classify flight delays, DBN, SVM, GBC, Deep LSTM and DRNN-SSDCA were trained and tested, the novel DRNN-SSDCA outperforms other models with 93.6% accuracy and 0.1114 RMSE. In the same way, a deep feedforward neural network and SVM to predict flight delay was used by (Bisandu et al, 2021; Alla, et at., 2021; Maged, et al., 2023), they were able achieved an accuracy of 90%.

Ali, et al., (2023) presented a classification model in which the main objective was to exclusively predict flight delays of individual flights caused by climatic conditions. They used data collected from the BTS airline On-time Performance dataset for the years of 2005 to 2015 using features like flight schedules.

Yushan & Liu (2020) used on-time performance and quality controlled local climatological data for flight delay classification. In this work, different models were tested, the result shows that MLP was the best model having achieved 89.32%. an optimization of feature extraction methods was suggested in subsequent work. A combination of feedforward ANN and XGBoost was used to make prediction, this hybrid model analysed structured air traffic data on departure delay severity prediction on ten major airports in the US which have very high ground and air congestion (Zhou, et al., 2022; Allharbi & Prince, 2020). The model achieved 74.38% accuracy. Parameter tuning and more data was suggested in future work.

The prediction of flight using spacio-temporal trajectory convolutional neural network and airport situational awareness map data with Logistic regression, multilayer perceptron (MLP) support vector machine (SVM), TrajCNN with weather features. The models were trained and tested and TrajCNN got a better result of 37.25 RMSE, the study suggested that more dataset should be use in subsequent work (Pixeto, 2020; Shi, et al., 2021)). Wang et al, 2022 predicted flight delay at Beijing Capital International airport based on stacking ensemble models, they included arrival/departure cruise pressure. The result shows an MAE of 12.58.

Anderson, et al., (2019) developed a process using the Support Vector Machine (SVM) with this model, a climate forecast for a particular area could be used as input in estimating an airport's delay. It also has the ability to predict the impact of weather on future flights, that is, how long could we expect a flight to be delayed due to the weather condition? This model, according to the authors, achieved 83% accuracy of the time (Ball, et al., 2019; Meel, et al., 2020).

According to (Esmaeilzadeh & Mokhtarimousayi, 2020; Balamurgan, et al., 2022) predicted delays in two ways; through classification where they classified a departure delay as being more or less than a predefined threshold, through regression, where they estimated a future delay in departure on a specific route. Results of the study showed an average precision of 81% in the performance of Random Forests algorithm selected for the classification, and a mean prediction error of the regression of 21 minutes. It should be noted that, for both means of forecasting, the error increased as the forecast advance also increased.

Kumar (2019); Khan, et al., (2021) predicted the delay using random forest with flight data and with a particular emphasis on departure delay features, he got a RMSE of 20 minutes, which is a good result, however he recommended adding weather feature in future work.

Zeng, et al., (2021) used multiple machine learning algorithms and a Deep Neural Network to predict the occurrence of a delay in a specific flight (through a classification problem) and to predict its numerical value. Gradient Boosting Classifier and Extra Trees Regressor performed the best from the set of supervised learning.

Khan, et al., (2022) worked on delay prediction for international flights, he used GB regression technique with flight information and weather data to predict the arrival delay, The study achieved a good result with RMSE value of 23.5 and MAE of 14.3. He recommended a comprehensive outlier clean up on features used as input to the model.

Divya & Diyyasindhu, (2023) developed models using logistic regression to predict the existence of arrival delays and their values in advance. The regression approach of this work got a root mean square error of 30 minutes and R-squared value of 80%, depending on the forecast horizon.

Based on the reviewed literature, it was discovered that majority of the research focused on classification with only few on regression techniques. Hence, the need to use hybrid machine learning technique which combine the strength of both classification and regression models to know how long a flight will be delayed in minutes. Feature selection parameter tuning technique was a critical challenge as seen in most of the reviewed works. It was also discovered that there were issues of bias-variance trade off and outliers which most researchers do not put into consideration.

3.0 Research Methodology

We will use this concept to build a model that will make flight delay predictions, starting with data collection and description, data pre-processing which includes data cleaning, exploratory data analysis, feature engineering and feature selection method used for the modelling and the evaluation of the model performance.

The phases for the development of the model are presented in figure 3.1. The phases consist of several process or steps that are followed to develop the machine learning model.

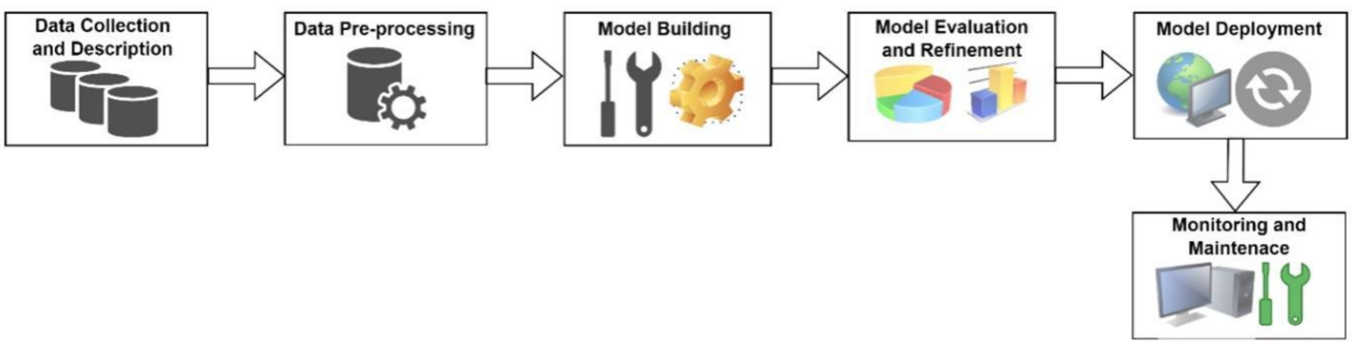


Figure 3. 1: Phases of ML Model Development

The first data was provided by the US Department of Transportation's Bureau of Transportation Statistics (BTS) and contained a combined domestic flight records of over 7.2 million collected between January and December 2018. The BTS On-Time Performance database provides detailed logs of individual flights reported by U.S. certified air carriers with originally 109 features. This dataset did not include weather data which had to be extracted from a different source.

The second data contained hourly weather data condition and weather events for 2018 and was provided by Automated Surface Observation System (ASOS)", sometimes called "Automated Weather Observation System (AWOS)". This type of system is located at airports for aviation support and weather prediction, providing essential observations for the National Weather Service (NWS), and the Federal Aviation Administration (FAA).

Missing values are known as data that are "missing" for some features in the dataset, they are critical to the management of the data because they can represent an issue in data quality and compromise the interpretation of data (Smishad & Jason, 2020).

The weather has a great impact on flight operations and it is a major predictor of the various flight events. Therefore, weather data was introduced to complement the already existing flight historical data. The weather data was imported from a weather observation website and therefore need to be clean and reformatted to the standard acceptable for data analysis (Schltz, et al., 2021).

In order to acquire additional data to complement the initial ones for the creation of further features for this study, we collected holiday data from Office of Personnel Management (OPM). Holiday data is cleaned and ready for further processing, this is simply because it has just few entries compare to the flight and weather datasets. After cleaning and saving the datasets as data frame, they are being merged into a single data frame for further processing.

Exploratory Data Analysis

Data visualization gives insight about the data: discover patterns, spot anomalies or test assumptions. The purpose is to expose the underlying features of the data and acquire knowledge to serve as a foundation for the modelling task. Below is the analysis of the data.

In figure 3.2, the months of June, January and February have higher percentage on the arrival delay. The seasons that have more impact on the delay are Winter and Summer in accordance to the months as also concluded by (Li, Jing & Dong, 2023). This probably happens because of the weather in winter and due to the summer vacations when most people travel. The distribution per month denotes a slight increase in flights on the holiday season.

Figure 3. 2: Impact of year, Month, week feature on Arrival Delay

The same can be seen in the distance between origin and destination of a flight. As illustrated in Figure 3.3. The distance is categorized into distance groups from 1 to 12, representing shorter to long distance. Most of the delays occur in the short distances, long distances have lower number of arrival delays.

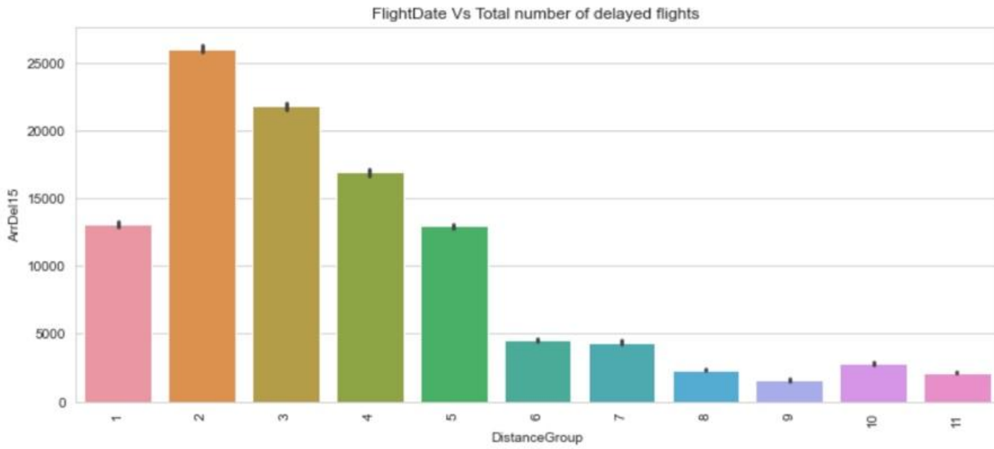


Figure 3. 3: Impact of Distance feature on Arrival delay

When looking at the schedule and real departure time when grouping the hour intervals in phases of the day, there is a part that has more delays in arrival in comparison to the others. The three phases of the day were represented for a better understanding of the behaviour of the arrival delay by the course of the day; morning, afternoon and evening. The highest percentage of arrival delays occur in the evening time compare to the afternoon and morning as seen in figure 3.4 Hence, it is best to schedule a flight in the early morning, because the average delays are low.

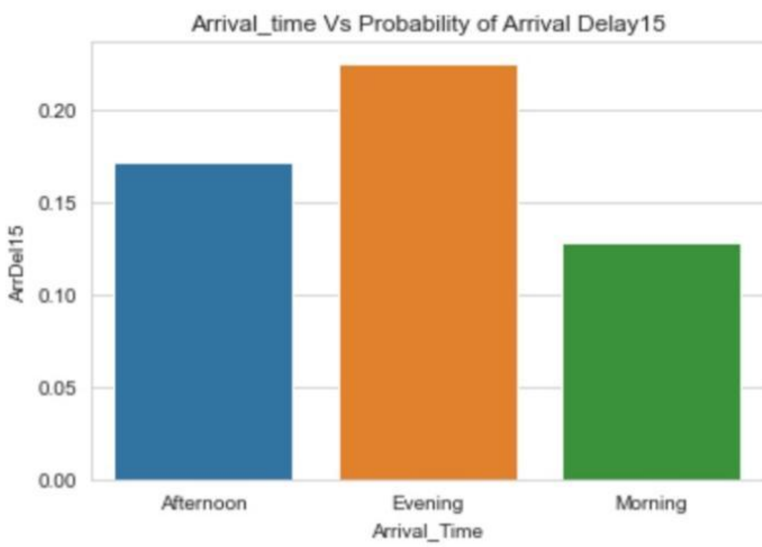


Figure 3. 4: Impact of CRS arrival time in parts of the day

This higher percentage of delayed flights in the evening interval is due to the departure time of some flights. In the case of the scheduled arrival time, it is noticed that around 42% of total delays at arrival occurs in flights scheduled to arrive in the afternoon interval, and 35% occur in flights scheduled to arrive in the evening interval. Arrival delay could be affected by other elements besides just the flight duration such as baggage, maintenance, cabin crew, and others.

About the airline company operating the flight is known that over 60% of the total delays are composed of flights operated by Southwest airline, Delta AirLines, SkyWest airline and United airline as seen in figure 3.5.

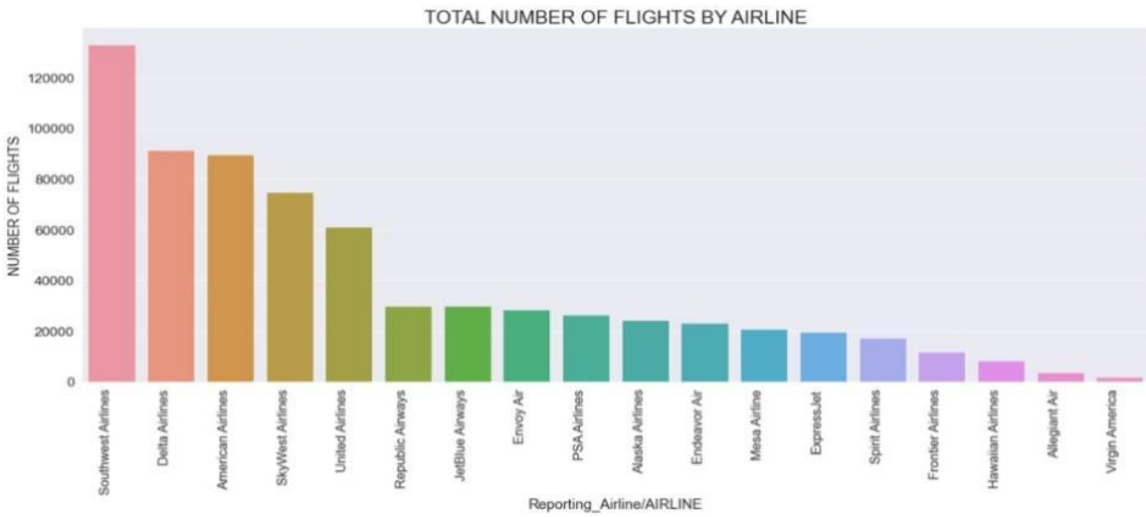


Figure 3. 5: Impact of Airline company feature on Arrival delay

In this paper, weather effect indicates a general weather trend in different seasons, which does not refer to every day's corresponding weather condition. To describe weather in a location, usually we need several indicators to illustrate the condition. These indicators include; wind, visibility, and sky conditions.

Visibility is the ability to see an object in the atmosphere. Visibility is the greatest horizontal distance, at which selected objects can be seen, identified, and/or measured with instrumentation. For a safe flying, the pilot needs a minimum amount of visibility for landing at the airport. Clear clean air has a better visibility than air polluted with dust or other particles. This depends on a number of factors which are all weather related. There is a correlation between flight delay and sky conditions, which means different sky conditions will have a different influence on flying.

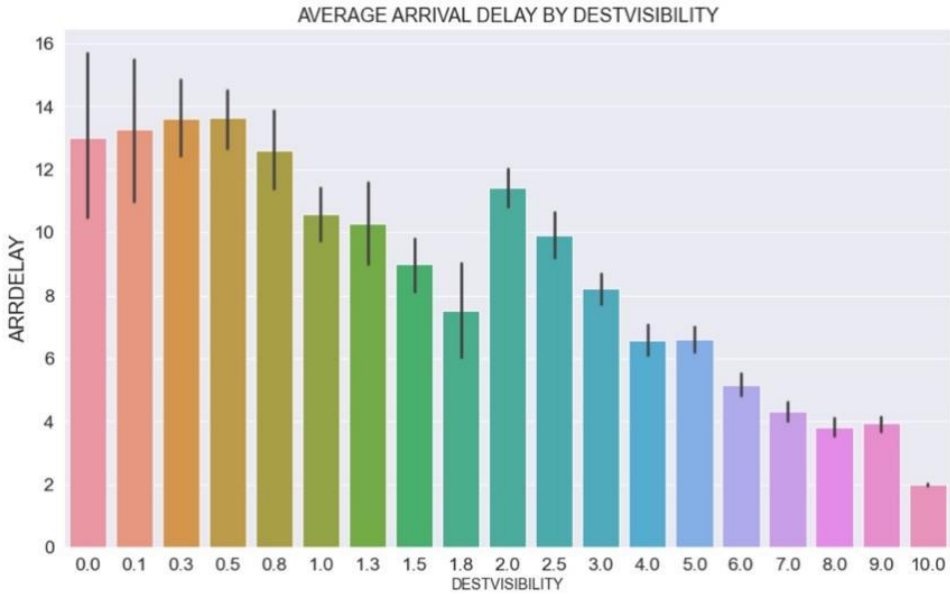


Figure 3. 6: Impact of Visibility on flight arrival

The visibility features show that the flights that contribute the most to the delay are the ones with visibility between 0.0 to 0.8 and it starts to drop and rise at 2.0 and continue to fall. As the visibility decreases, the percentage of delayed flights increases as seen in figure 3.6.

Holiday occurrences can cause a more significant influx of passengers causing a possible delay. It makes intuitive sense that airlines are probably going to be under stress near and during holidays, so this feature could probably be a decent predictor of a flight delays. The distribution per month denotes a slight increase in flights on the holiday season. The number of flights during 2018 are plotted daily to understand if there are any density increase during the national holidays. It seems that during the national holidays the flight distribution is redefined, but there is no visible increase in the number of flights.

Algorith Selection

Two algorithms are selected for building the hybrid model, these models are Logistic Regression (LR) for binary classification and eXtreme Gradient Boosting (XGBoost) for regression analysis. These algorithms were selected based on their track record for high performance on general predictive classification and regression modelling respectively. These algorithms are also believed to perform prediction with high accuracy in flight delay analysis as seen in some articles.

This approach focuses on combing the result of flight classifier and regressor in a two-stage model by simply combining the results from the classifier into the regressor through a pipeline.

In the model, a Logistic Regression will be used to represents the existence of delay or not. If the actual time of arrival is greater than 15 minutes from the scheduled time then the feature assumes the value 1, otherwise it assumes the value of 0. XGBoost, will be utilize as the regression algorithm to determine how long the delayed flight will be in minutes as depicted in the flowchart in figure 3.8

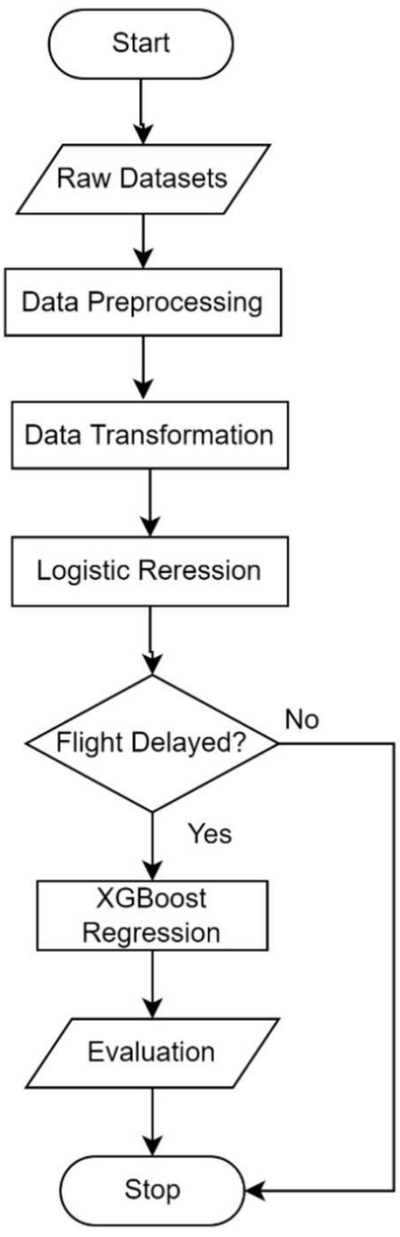


Figure 3. 8: Model Flowchart

Performance Metrics

The different metrics used for classification problems can best be explained through a confusion matrix as depicted in table 3.2. Where the definitions of True Positive (TP), False Positive (FP), False Negative (FN) and True Negative (TN) are given, which corresponds to what the model predicted and what the actual class was (True or False).

Table 3.12: Confusion Matrix for a general Binary Classification

		Actual Class	
		Positive (1)	**Negative (0)**
Predicted Class	**Positive (1)**	**TP**	**FP**
	Negative (0)	**FN**	**TN**

Accuracy, Precision, F1 Score, and Recall performance measures and confusion matrices are generated using table 3.1. The True Positive (TP), also known as sensitivity, is used when both the observed and anticipated values are positive. False Negative (FN), also known as Specificity, is when both the observed and predicted values are negative. False Positive (FP), also known as Specificity, is when some observations are negative but the expected value is positive. True Negative (TN), also known as Specificity, is when both the observed and forecasted values are negative.

i. **Accuracy:** Is the number of correct predictions divided by the total number of made predictions in terms of elements of the confusion matrix.

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN}$$

ii. **Precision:** Precision gives a measure of all the positive classes which are predicted, how many are actually positive. It can be taught of as the ability of the model to identify only relevant instances within the dataset.

$$Precision = \frac{TP}{TP+FP}$$

iii. **Recall:** Recall gives a measure of all the actual positive classes; how much were predicted correctly. It can be taught of as the ability of a model to find all the relevant instances within the dataset.

$$Recall = \frac{TP}{TP+FN}$$

iv. **F1-score:** In order to compare models on a single metric, the F1-score is introduced, which is harmonic mean of precision and recall.

$$\mathbf{F1-Score} = \frac{2 * \frac{TP}{TP+FP} * \frac{TP}{TP+FN}}{\frac{TP}{TP+FP} + \frac{TP}{TP+FN}}$$

v. **Mean Absolute Error (MAE):** The MAE is the average of the absolute difference between the actual values and the predicted values, also called the residuals.

$$\boldsymbol{MAE = \frac{1}{n}\sum_{i=1}^{n} \hat{y}_i - y_i}$$

vi. **Root Mean Squared Error (MSE):** The MSE is simply defined as the average of squared differences between the predicted output and the true output.

$$RMSE = Sqrt \sum_{i=1}^{n} (\hat{y}_i - y_i)^2$$

vii. **R-Squared:** The R2 coefficient represents the proportion of variance of outcomes that a model can predict based on its properties.

$$R - Squared = 1 - \frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}$$

y_i: Actual Value

$\hat{y}$: Predicted Value

y: Mean of all instances of y_i

n: The total number of instances of y_i

Results and Discussions

The result achieved a good precision, recall, and f1-score of 92%, 98%, and 95% respectively in correctly classifying the training instances. However, it was observed that the overall performance of precision, recall, and F1 classifiers declined with the test dataset. Sometimes accuracy alone can be misleading, especially in the case of imbalanced datasets where the focus is on predicting specific classes. This research will focus on "Recall Score" as the key metric for model deployment.

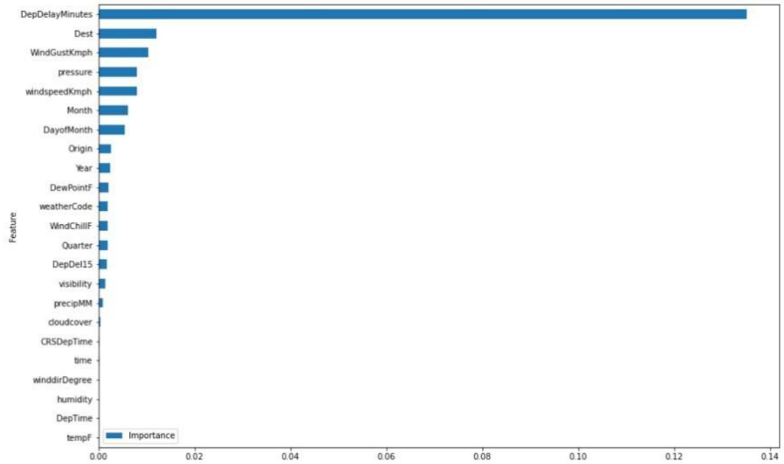
Table 4. 1: Baseline Logistic Regression Result for Imbalance Dataset

Data	**accuracy**	**precision**	**Recall**	**F1-score**
Train		**0.92**	**0.98**	**0.95**
Test	**0.91**	**0.89**	**0.68**	**077**

4.5 Feature Importance

In this section, the results of the most important features for the classification model are presented. The most important features include DepDelayMinutes, Dest, WindGustKmph, Presure, windSpeedkmph, month, Daysofweek, origin, year, dewpoint, and more as presented in figure 4.55 below:

CC BY NC



Classification Result

The first stage of machine learning, known as classification, seeks to forecast whether or not a scheduled aircraft will be delayed. Delay flights are those with arrival delays of more than 15 minutes. The target variable "ArrDel15" is set to 1 for flights that are delayed and to 0 for planes that arrive on time. Based on the classification performance measures provided in the preceding chapter, the model's performance is examined. The result of table 4.2 shows that Recall score improve from 64% of baseline to 96% after handling data imbalance.

Table 4. 2: Logistic Regression Result for Balanced Dataset

Class	**accuracy**	**precision**	**Recall**	**F1-score**
0		**0.98**	**0.96**	**0.97**
1	**0.96**	**0.87**	**0.94**	**0.90**

4.6.2 Regression Result

The second stage of the machine learning model is regression. If the classifier labels the flight as "Delayed," it forecasts the arrival delay in minutes. The regression model is trained using the flights that had an "ArrDelayMinutes" value greater than 15. Table 7 presents a summary of the models' performance.

R2 is a metric used to assess how well a regression model can forecast the variances in a collection of data. The performance of the model was evaluated using the mean absolute error (MAE), root mean square error (RMSE), and R2 of the arrival delays. The R2, MAE, and RMSE of the anticipated delays are displayed in Table 4.3.

Table 4. 3: Regression Result for XGBoost

Metrics	Testing
MAE	**10.5267**
RMSE	**15.4520**
R-Squared	**0.9360**

4.6.3 Pipeline Analysis

A pipeline was built to depict the data flow after the Logistic and XGBoost Regression models had been trained and tested. The pipeline was implementation the result is presented in table 4.3 below. The MAE is 12.3954, RMSE is 17.5131 and R^2 is 0.9449 which is equivalent to 95%. The regressor is found to function better with a pipeline. This may be due to the correctly classified data points.

Table 4. 4: Regression Result for pipeline Analysis

Metrics	Testing
MAE	**12.3954**
RMSE	**17.5131**
R-Squared	**0.94489**

4.7 Distribution of Delay

The dataset is divided into ranges of arrival delay minutes in this section, and each range's performance of the pipelined XGBoost Regressor is examined. There is a 0 to 1210 minutes range in the flight arrival delay. According to the arrival delay's frequency distribution plot figure 4.6, the majority of data points are found in the 0 to 200 range. The majority of the data points had 'ArrDelayMinutes' values between 0 and 100 minutes, as shown in Figure 4.7. The number of data points reduces as the range widens, suggesting that there are fewer flights with extremely high flight delays. The values of R2 scores rise as the number of data points decreases for each range, with the exception of the 101–200 range, as shown in Table 4.7. This is explained by the dataset's lower proportion of data points with significant delays.

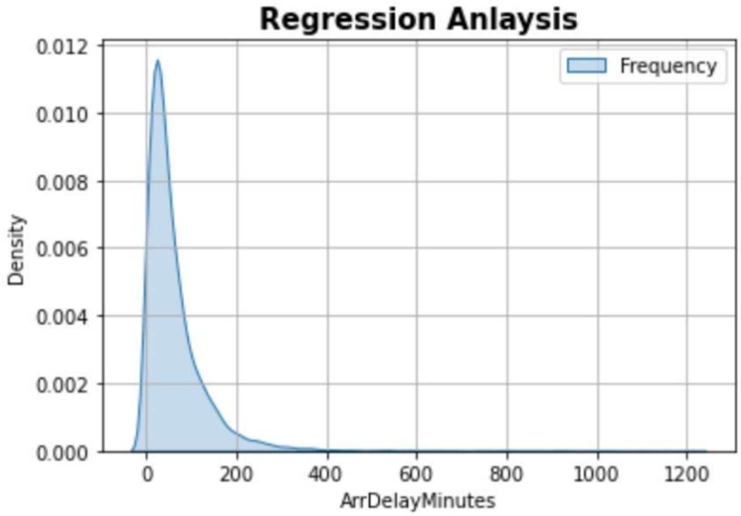


Figure 4.6: Arrival delay class frequency distribution

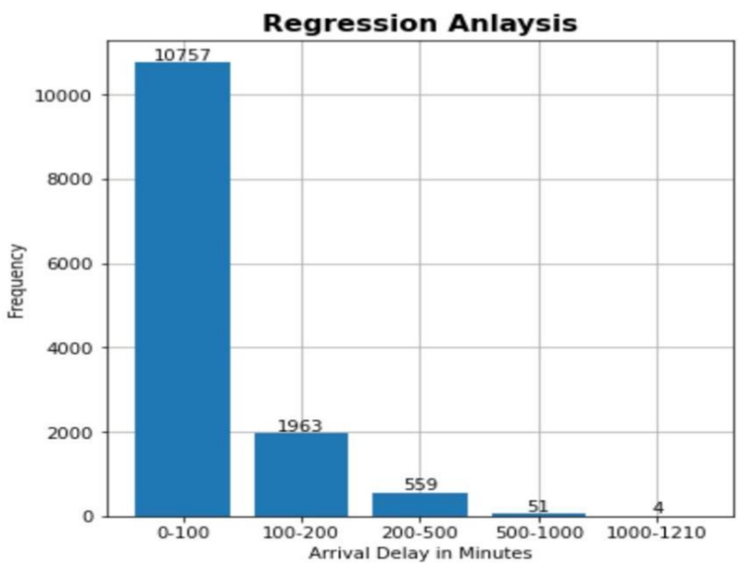


Figure 4. 7: Distribution of delay in minutes

Regression testing is providing blocks of data to the regressor that fall within a predetermined range for the arrival delay of the flight in minutes, such as 15–100 minutes or 100–200 minutes. The distribution of the initial test dataset in relation to the fight's arrival delay is displayed in table 4.5. These blocks are supplied to the regressor in order to verify its performance, and the MAE, RMSE and R^2 from each block are then calculated.

Table 4. 5: XGBoost Results for each Delay block

Arrival delay(mins)	RMSE	MAE	R^2
15-100	9.834668	6.634876	0.856493
101-200	15.399585	9.194358	0.658591
201-500	18.425296	10.801726	0.914323
501-1000	20.275852	11.399412	0.982030
1001-1210	5.276057	4.457500	0.995092

Table 4.5 shows that the R2 values for the block 101–200 is relatively low when compared to other blocks.

4.8 Results Comparison

In this section, we describe some comparisons of the proposed hybrid model with other models as described in literature review section.

Table 4.6: *Performance Comparison of our hybrid model and other Models*

Author/Data	Model	RMSE	MAE	R-Squared
Alla, et at., (2021)	MLP	18.48	12.40	0.9560
Yhdego, et al., (2023)	Hybrid RNN	1.516	1.218	0.928
Syed, et al., (2023)	RF	54.948	30.80	-
Pophale, et al., (2023)	Quantum ML	3.840	4.203	0.9152
Our Hybrid model (2023)	LR-XGBoost	17.513	12.395	0.9445

Table 4.6 shows the results of various metrics used to evaluate the performance of the proposed hybrid RNN model, together with Hybrid RNN, and Random Forest (RF) regressor models. These metrics were MAE (Mean Absolute Error), RMSE (Root Mean Squared Error), and R-squared. The values of these metrics indicate the accuracy of the model in predicting flight arrival delays. Lower

values for MAE, RMSE, and higher values for R-squared indicate better performance. The LR-XGBoost model had an accuracy of 94.45%, which is comparatively higher than the Hybrid RNN model indicating a better fit to the data points. The LR-XGBoost model has a RMSE and MAE of 17.51, and 12.40 respectively, which is better than that of RF, having a RMSE of 54.95, and MAE of 30.80 respectively. However, Hybrid RNN has a better RMSE of 1.516 and MAE of 1.218. As can be observed in Table 4.6, the proposed hybrid LR-XGBoost model outperformed the other models in terms of R-squared value.

Conclusion

This study applied Logistic Regression for binary classification of flight delays and XGBoost for regression-based prediction of delay duration. While Logistic Regression provided a simple and interpretable model for delay classification, XGBoost outperformed it in predicting the actual delay time in minutes. The results suggest that airlines could benefit from using XGBoost to predict flight delay durations more accurately, which could help in operational planning and passenger communication.

In a nutshell, the combination of logistic regression and XGBoost regression in flight delay prediction contributes to more accurate, comprehensive, and data-driven decision-making in the aviation industry. It exemplifies the potential of advanced analytics to improve operations, passenger experiences, and overall efficiency in the field of aviation.

Further research could explore the integration of real-time data, such as live weather reports and air traffic control updates, to enhance the accuracy of flight delay predictions. Additionally, combining machine learning models with optimization algorithms could help airlines minimize delays and improve customer satisfaction.

References

Alharbi, B., & Prince, M. (2020). A hybrid artificial intelligence approach to predict flight delay. *Int J Eng Res Technol*, 13(4), 814-822.

Ali, S. M., Khan, A. A. A., Khan, A. W., & Anita, T. (2023). Predicting Flight Delays with Error Calculation Using Machine Learned Algorithm. *Mathematical Statistician and Engineering Applications*, 72(1), 1428-1434.

Anderson, A., Kumar, A., & Christopher, A. (2019). Analysis of flight delays in aviation system using different classification algorithms and feature selection methods. *The Aeronautical Journal*, *123*(1267), 1415-1436. doi:10.1017/aer.2019.72

Anguita, J. M., & Olariaga, O. D. (2023). Prediction of departure flight delays through the use of predictive tools based on machine learning/deep learning algorithms. *The Aeronautical Journal*, 1-23.

Aviation Trend, (2018). Civil aviation authority. Airline for America, *U.S. Airline Industry Review: Allocating Capital to Benefit Customers, Employees and Investors*, 2017.

Ayaydin, A. & M. Ali Akçayol. (2022). Deep Learning Based Forecasting of Delay on Flights. *International Journal of Informatics Technologies*, *15*(3), 239–249. https://doi.org/10.17671/gazibtd.1060646

Ayo-Agunbiade Oluwafisayo T., Stephens Mobolaji S. (2019). Causal Factors of Flight Delay in Nigerian Airport: A Case Study of Murtala Muhammed International Airport. *Bangladesh Journal of Multidisciplinary Scientific Research*; Vol. 1, No. 2 ISSN 2687-850X E-ISSN 2687-8518 July-September; 2019.

Balamurugan, R., Maria, A. V., Baranidaran, G., MaryGladence, L., & Revathy, S. (2022, April 1). Error Calculation for Prediction of Flight Delays using Machine Learning Classifiers. *In 2022 6th International Conference on Trends in Electronics and Informatics (ICOEI) (pp. 1219-1225). IEEE Xplore*. https://doi.org/10.1109/ICOEI53556.2022.9776709

Ball, M., Barnhart, C., Dresner, M., Hansen, M., Neels, K., Odoni, A., Peterson, E., Sherry, L., Trani, A., Zou, B. (2019). Total Delay Impact Study. In *NEXTOR Research Symposium, Washington DC.* 2010. http://www.nextor.org. Accessed July 14, 2019.

Bisandu, D. B. & Moulitsas, I. (2023). A Hybrid Ensemble Machine Learning Approach for Arrival Flight Delay Classification Prediction Using Voting Aggregation Technique. *CERES (Cranfield University)*. https://doi.org/10.2514/6.2023-4326

Carvalho, L., Sternberg, A., Maia Gonçalves, L., Beatriz Cruz, A., Soares, J. A., Brandão, D., Carvalho, D., & Ogasawara, E. (2020). On the relevance of data science for flight delay research: a systematic review. *Transport Reviews*, 1–30. https://doi.org/10.1080/01441647.2020.1861123

Chakrabarty, N. (2019). A Data Mining Approach to Flight Arrival Delay Prediction for American Airlines. *2019 9th Annual Information Technology, Electromechanical Engineering and Microelectronics Conference (IEMECON)*, 102-107.

Chakrabarty, N., Kundu, T., & Dandapat, S. (2019). Flight Arrival Delay Prediction Using Gradient Boosting Classifier. *ResearchGate*, *January*. https://doi.org/10.1007/978-981-13-1498-8

Chunzheng Wang, Minghua Hu, Lei Yang , Zheng Zhao (2021). Prediction of air traffic delays: An agent-based model introducing refined parameter estimation methods. https://doi.org/10.1371/journal.pone.0249754

Divya, J., & Divyasindhu, P. (2023). Flight Delay Prediction using GANN-An Improved Artificial Neural Network Model Integrating Genetic Algorithm. *In 2023 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI)* (pp. 1-10). IEEE. https://doi.org/10.1109/accai58221.2023.10200001

Esmaeilzadeh, E., & Mokhtarimousavi, S. (2020). Machine Learning Approach for Flight Departure Delay Prediction and Analysis. *Transportation Research Record: Journal of the Transportation Research Board*, *2674*(8), 145–159. https://doi.org/10.1177/0361198120930014

Gholami, S., & Khashe, S. (2022). Alexa, Predict My Flight Delay. arXiv preprint arXiv:2208.09921.

Gholami, S., & Khashe, S. (2022). Flight Delay Prediction Using Deep Learning and Conversational Voice-Based Agents. *Am. Acad. Sci. Res. J. Eng. Technol*. Sci, 89, 60-72.

Guo, Z., Yu, B., Hao, M., Wang, W., Jiang, Y., & Zong, F. (2021). A novel hybrid method for flight departure delay prediction using Random Forest Regression and Maximal Information Coefficient. *Aerospace Science and Technology*, 116, 106822.

Hee Kyung Kim and Chang Won Lee. (2019). Development of a Cost Forecasting Model for Air Cargo Service Delay Due to Low Visibility. https://www.mdpi.com/journal/sustainability. doi:10.3390/su11164390.

Id, C. W., Hu, M., Yang, L., & Zhao, Z. (2021). Prediction of air traffic delays: An agent-based model introducing refined parameter estimation methods. *PLOS ONE*. https://doi.org/10.1371/journal.pone.0249754

Imran, S. H., Anjum, A., Ananthasiri, C. H., & Viraja, D. U. (2023). Build a Machine Learning Model For Prediction Flight Delays With Error Calculation. *International Journal of Management Research and Reviews*, 13(2), 12-17.

Ismail, B. M., Siti, M. S., & Shafaatunnur, H. (2019). A Preliminary Study on Learning Challenges in Machine Learning-based Flight Delay Prediction. *International Journal ofInnovativeComputing*.9(1)1-5.DOI: https://doi.org/10.11113/ijic.v9n1.204

Jiang, Y., Liu, Y., Liu, D., & Song, H. (2020). Applying Machine Learning to Aviation Big Data for Flight Delay Prediction. *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)*. https://doi.org/10.1109/dasc-picom-cbdcom-cyberscitech49142.2020.00114

Kalyani, N., Jeshmitha, G., BinduSriSai, U., Samanvitha, M., M., J., & Kiranmayee, B. (2020). Machine Learning Model - based Prediction of Flight Delay. *2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, 577-581.

Khalil, A., Awan, M. J., Yasin, A., Singh, V. P., & Shehzad, H. M. F. (2022). Flight web searches analytics through big data. *International Journal of*

Computer Applications in Technology, *68*(3), 260. https://doi.org/10.1504/ijcat.2022.124949

Li, Q., Guan, X., & Liu, J. (2023). A CNN-LSTM framework for flight delay prediction. *Expert Systems with Applications*, *227*, 120287. https://doi.org/10.1016/j.eswa.2023.120287

Li, Q., Jing, R., & Dong, Z. (2023). Flight Delay Prediction With Priority Information of Weather and Non-Weather Features. *IEEE Transactions on Intelligent Transportation*

Maglogiannis, I., Macintyre, J., & Iliadis, L. (Eds.). (2021). Artificial Intelligence Applications and Innovations. *AIAI 2021 IFIP WG 12.5 International Workshops. IFIP Advances in Information and Communication Technology.* doi:10.1007/978-3-030-79157-5

Peixoto, F. (2020). A Simple overview of Multilayer Perceptron(MLP). *Analytics Vidhya*. https://www.analyticsvidhya.com/blog/2020/12/mlp-multilayer-perceptron-simple-overview/

Pinto Neto, E. C., Baum, D. M., Almeida Jr, J. R. D., Camargo Jr, J. B., & Cugnasca, P. S. (2023). Deep Learning in Air Traffic Management (ATM): A Survey on Applications, Opportunities, and Open Challenges. Aerospace, 10(4), 358.

Ranga Swamy Sirisati, K. Prasanthi, & Anga Gautami Latha. (2021). An Aviation Delay Prediction and Recommendation System Using Machine Learning Techniques. *Algorithms for Intelligent Systems*, 239–253. https://doi.org/10.1007/978-981-33-6307-6_25

Reddy, R. T., Pati, P. B., Deepa, K., & Sangeetha, S. T. (2023, April). Flight Delay Prediction Using Machine Learning. *In 2023 IEEE 8th International Conference for Convergence in Technology (I2CT) (pp. 1-5). IEEE.*

Schultz, M., Reitmann, S., & Alam, S. (2021). Predictive classification and understanding of weather impact on airport performance through machine learning. *Transportation Research Part C: Emerging Technologies*, *131*, 103119. https://doi.org/10.1016/j.trc.2021.103119

Shi, T., Lai, J., Gu, R., & Wei, Z. (2021). An Improved Artificial Neural Network Model for Flights Delay Prediction. *International Journal of Pattern Recognition and Artificial Intelligence*, *35*(08), 2159027. https://doi.org/10.1142/s0218001421590278

Sirisati, R. S., Prasanthi, K. G., & Latha, A. G. (2021). An Aviation Delay Prediction and Recommendation System Using Machine Learning Techniques. *In Proceedings of Integrated Intelligence Enable Networks and Computing*: IIENC 2020 (pp. 239-253). Springer Singapore.

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. https://doi.org/10.1371/journal.pone.0249754

Wang, L., Mao, J., Li, L., Li, X., & Tu, Y. (2023). Prediction of estimated time of arrival for multi-airport systems via “Bubble” mechanism. *Transportation Research Part C: Emerging Technologies*, *149*, 104065. https://doi.org/10.1016/j.trc.2023.104065

Yazdi, M.F., Kamel, S.R., Chabok, S.J., & Kheirabadi, M. (2020). Flight delay prediction based on deep learning and Levenberg-Marquart algorithm. *Journal of Big Data*, *7*, 1-28.

Yu, Y., Chen, H., Yuan, L., & Zhang, B. (2021). Flight delay classification warning based on evolutionary under-sampling bagging ensemble learning. *In Fifth International Conference on Traffic Engineering and Transportation System (ICTETS 2021)* (Vol. 12058, pp. 476-480). SPIE.

Zeng, W., Li, J., Quan, Z., & Lu, X. (2021). A Deep Graph-Embedded LSTM Neural Network Approach for Airport Delay Prediction. *Journal of Advanced Transportation*, *2021*. https://doi.org/10.1155/2021/6638130

Zhou, H., Li, W., Jiang, Z., Cai, F., & Xue, Y. (2022). Flight Departure Time Prediction Based on Deep Learning. *Aerospace*, 9(7), 39