Monthly Forecasting of the Number of Delayed Flights

Fatima Ayyub

Abstract

Flight delays are one of the biggest challenges facing the aviation sector, negatively impacting passengers and airlines.The aim of this research is to develop a predictive model that can predict the number of delayed flights for the next month based on a number of weather and operational factors,using real-world data from a number of different airports in the United States. The data has undergone pre-processing, including the treatment of missing values,outliers, standardization,dimensionality reduction and a number of time series analysis techniques were used before using traditional time series models Autoregressive integrated moving average (ARIMA), AutoRegressive Integrated Moving Average with Exogenous variables(ARIMAX) and the LightGBM (Light Gradient Boosting Machine) predictive model was used,which achieved the best result from the performance measures used Mean Squared Error(MSE),Root Mean Squared Error(RMSE),Mean Absolute Error(MAE), indicating that machine learning models can help reduce the challenges of the aviation sector through advance planning.

Introduction

With the increasing number of travelers,the number of flights has increased significantly, as it has been observed that the global air passenger transport market doubles every 15 years[1].

With the increasing demand for air travel, the requirements for ensuring the efficiency of air travel and developing airport infrastructure increase. This includes expanding airport facilities, updating airline feets,and implementing effective air schedule management. Addressing these issues is crucial to provide a seamless and reliable travel experience for passengers. However, a significant challenge in delivering satisfactory services is the frequent occurrence of unexpected fight delays and cancellations[2].

In May 2025,according to the German news agency on NBC, US Transportation Secretary Sean Duffy announced a reduction in the number of flights to and from Newark Liberty International Airport,one of the busiest airports in New Jersey.This was due to a severe shortage of air traffic controllers and repeated malfunctions in radar systems, which caused severe delays.The Secretary explained in his statements, “We want to make sure that if you book a flight, the plane actually flies.That’s the priority—so you don’t get to the airport,wait four hours, and then get delayed,” said Secretary Duffy in an interview on NBC’s “Meet the Press.”

This study aims to develop a model capable of predicting the number of delayed flights per month based on a number of operational and time factors that affect flights

Related Work

Several research papers have investigated solutions to air travel challenges using data driven methods.

In [3],This research aims to optimize a HAPS mission profile for accurate aircraft flight delayprediction by analyzing the correlations between high-altitude atmospheric data from HAPS missions, varying mission specifications,and flight delays.

Comparison in research [3] which aims to improve the collection of air data using HAPS systems to facilitate the prediction of flight delays.In this research, the focus was on building a direct predictive model using historical data that includes multiple factors that affect flights.

In [4],This investigated the impact of climate change on the number of flight delays and corresponding flight delay costs by examining departure delays from John F. Kennedy Airport in New York City, USA, from 2013 to 2022 and deriving a model for future delays from 2023 to 2030

The study [4] focused on weather phenomena as a major source of aircraft flight delays. In this research, the focus was on a number of factors, including weather factors in addition to operational factors that affect air flights.

In this research, the focus was on predicting the number of delayed flights for a longer period of time than [5],as the research focused on predicting air delays during the next month, which allows the possibility of strategic planning for airlines and airports for future periods.

Methodology

Data source

The data used in this research was obtained from the UCI Machine Learning Repository.The dataset contains 171,666 instances collected from 419 airports in the United States and covers the period from 2013 to 2023.

Features	**Description**
year	The year in which the data was recorded.
month	The month of the year when the data was recorded.
carrier	Airline carrier code.
airport	Airport code related to the data.
arr flights	Total number of arriving flights.
arr delay 15	Number of flights delayed by 15 minutes or more.
carrier count	Count of delays caused by the airline.
weather count	Count of delays caused by weather conditions.
nas count	Count of delays due to the National Aviation System (NAS).
security count	Count of delays caused by security issues.
late air craft count	Count of delays caused by late-arriving aircraft from a previous flight.
arr cancelled	Number of cancelled arriving flights.
arr diverted	Number of arriving flights diverted to another airport.
arr delay	Total arrival delay time in minutes.
carrier delay	Delay time in minutes due to the airline.
weather delay	Delay time in minutes due to weather.
nas delay	Delay time in minutes due to NAS.
late$_{a}ircraft_{d}elay$	Delay time in minutes due to a late arriving aircraft.

Table 1: description of the flight delay dataset features

The features in Table 1 are suitable for the research topic as they include multiple factors to study their impact on flight delays for each month.

Data Preprocessing

Data preprocessing is a vital initial step during knowledge discovery because it determines the success of data mining projects.A dataset's quality and representation stand as the primary element because any presence of redundant, irrelevant, too noisy,or unreliable information will severely disrupt the knowledge discovery process[6].

Missing values

Feature	**Missing Values**
arr flights	240
arr del15	443
carrier ct	240
weather ct	240
nas ct	240
late aircraft ct	240
arr cancelled	240
arr diverted	240
arr delay	240
carrier delay	240
weather delay	240
nas delay	240
security_delay	240
late aircraft delay	240

Table 2: Number of Missing Values for Each Feature

The missing values in Table 2 were dealt with by estimating the missing value using the K Nearest Neighbors Imputer (KNN Imputer) algorithm,where the missing value was estimated based on the 5 nearest neighbors for each instant.

K Nearest Neighbor (KNN) involves identifying k similar samples by calculating the distance between the complete information of the sample with the missing data and the information of the other samples. The missing data is then estimated using the data from these k samples[7]

Smooth outliers

count	171426	171223	171426	171426	171426	171426	171426	171426	171426	171426	171426	171426	171426
mean	362.53	66.43	20.80	2.25	19.38	23.77	7.53	0.86	4239.49	1437.19	222.56	920.65	1651.70
std	992.89	179.54	50.32	7.31	61.68	72.39	43.65	3.77	12618.57	4215.68	821.09	3423.51	5221.88
min	1	0	0	0	0	0	0	0	0	0	0	0	0
25%	50.00	6.00	2.16	0.00	1.00	1.23	0.00	0.00	335.00	110.00	0.00	34.00	65.00
50%	100.00	17.00	6.40	0.40	3.91	5.00	1.00	0.00	1018.00	375.00	18.00	146.00	320.00
75%	250.00	47.00	17.26	1.86	11.71	15.26	4.00	1.00	2884.00	1109.00	146.00	477.00	1070.00
max	21977.00	4176.00	1293.91	266.42	1884.42	2069.07	4951.00	197.00	438783.00	196944.00	31960.00	112018.00	227959.00

Table 3: Descriptive statistics for numeric features

Table 3 shows the statistical table for all dimensions before dealing with outliers. It is noted that there are very high values in some dimensions, which indicate the presence of outliers that affect the distribution of the data.

count	171666	171666	171666	171666	171666	171666	171666	171666	171666	171666
mean	122.36	21.78	8.14	0.72	5.09	6.61	1307.14	505.00	47.91	197.99
std	111.44	22.22	8.38	1.03	5.73	7.51	1382.97	555.12	78.97	233.30
min	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
25%	50.00	6.00	2.16	0.00	1.00	1.23	335.00	110.00	0.00	34.00
50%	101.00	17.00	6.41	0.40	3.92	5.00	1018.25	375.00	18.00	147.00
75%	151.00	27.00	10.53	1.00	6.53	8.49	1666.00	664.00	58.00	250.00
max	550.00	108.00	39.95	4.65	27.80	36.33	6715.00	2607.00	365.00	1144.00

Table 4: Descriptive statistics after outlier treatment

Table 4 shows the distribution of the data after dealing with the outliers by replacing the outliers in the median, which helped reduce the dispersion and the effect of the outliers on the normal distribution of the data.

Standardization

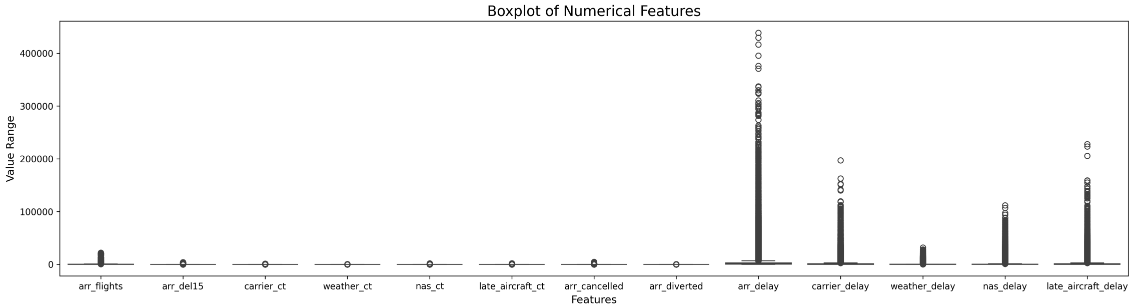


Figure 1:

Figure 1 shows how values within the IQR (Interquartile Range) appear smaller due to standardization, resulting in a compressed box.

$$\mathrm{IQR} = Q_3 - Q_1 \tag{1}$$

Equation (1) represents the range between the third quartile (Q_3) and the first quartile (Q_1), capturing the middle 50% of the data.

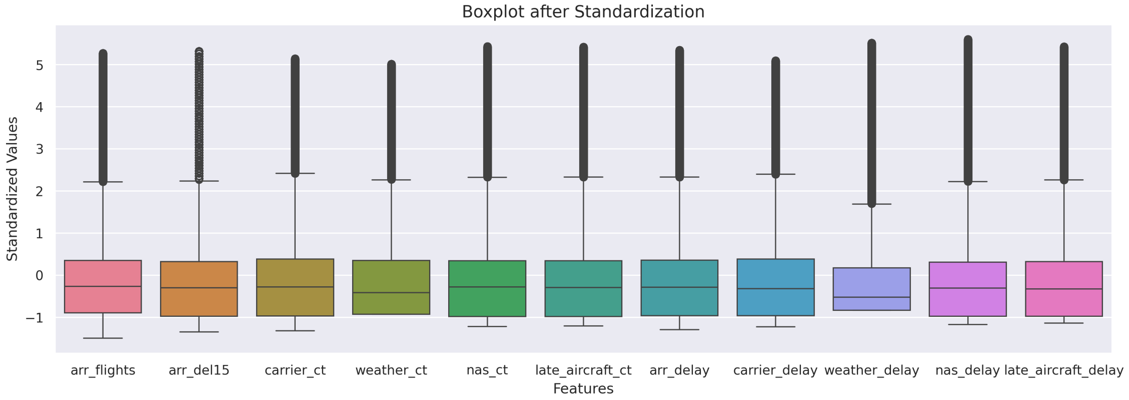


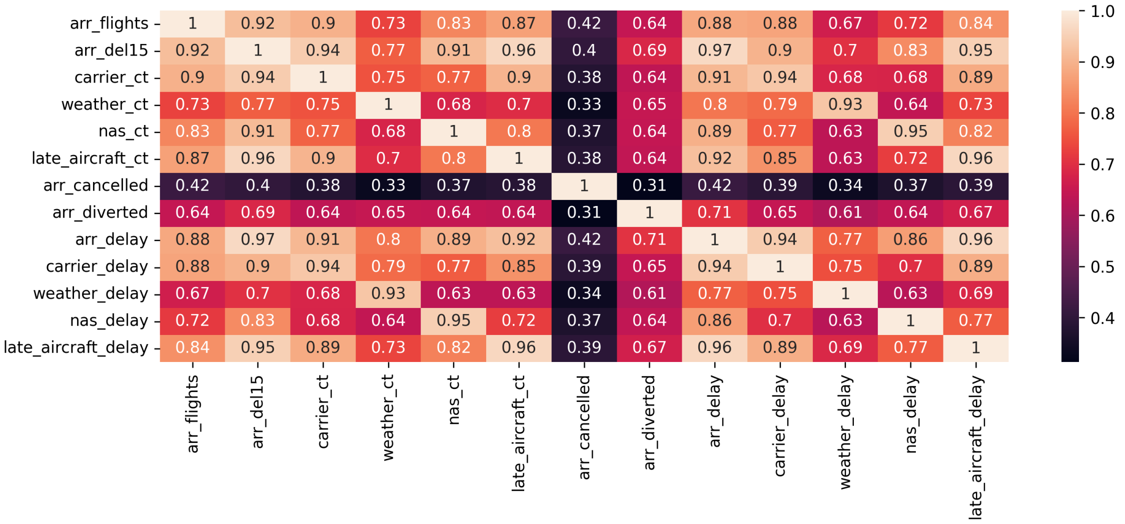
Figure 2: Standardized features.

Figure 2 shows the distribution of the data after standardization using Z-Score based on Mean Absolute Deviation(MAD)

$$mz_i = \frac{0.6745 \cdot (x_i - \tilde{x})}{MAD}$$

(2)

Reduce Dimensions



	arr_flights	arr_del15	carrier_ct	weather_ct	nas_ct	late_aircraft_ct	arr_cancelled	arr_diverted	arr_delay	carrier_delay	weather_delay	nas_delay	late_aircraft_delay
arr_flights	1	0.92	0.9	0.73	0.83	0.87	0.42	0.64	0.88	0.88	0.67	0.72	0.84
arr_del15	0.92	1	0.94	0.77	0.91	0.96	0.4	0.69	0.97	0.9	0.7	0.83	0.95
carrier_ct	0.9	0.94	1	0.75	0.77	0.9	0.38	0.64	0.91	0.94	0.68	0.68	0.89
weather_ct	0.73	0.77	0.75	1	0.68	0.7	0.33	0.65	0.8	0.79	0.93	0.64	0.73
nas_ct	0.83	0.91	0.77	0.68	1	0.8	0.37	0.64	0.89	0.77	0.63	0.95	0.82
late_aircraft_ct	0.87	0.96	0.9	0.7	0.8	1	0.38	0.64	0.92	0.85	0.63	0.72	0.96
arr_cancelled	0.42	0.4	0.38	0.33	0.37	0.38	1	0.31	0.42	0.39	0.34	0.37	0.39
arr_diverted	0.64	0.69	0.64	0.65	0.64	0.64	0.31	1	0.71	0.65	0.61	0.64	0.67
arr_delay	0.88	0.97	0.91	0.8	0.89	0.92	0.42	0.71	1	0.94	0.77	0.86	0.96
carrier_delay	0.88	0.9	0.94	0.79	0.77	0.85	0.39	0.65	0.94	1	0.75	0.7	0.89
weather_delay	0.67	0.7	0.68	0.93	0.63	0.63	0.34	0.61	0.77	0.75	1	0.63	0.69
nas_delay	0.72	0.83	0.68	0.64	0.95	0.72	0.37	0.64	0.86	0.7	0.63	1	0.77
late_aircraft_delay	0.84	0.95	0.89	0.73	0.82	0.96	0.39	0.67	0.96	0.89	0.69	0.77	1

Figure 3: Correlation Matrix

Figure 3 indicates the Pearson Correlation Coefficient(r)between all dimensions, where dimensions with correlation coefficients less than $r = 0.7$were removed.

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} \tag{3}$$

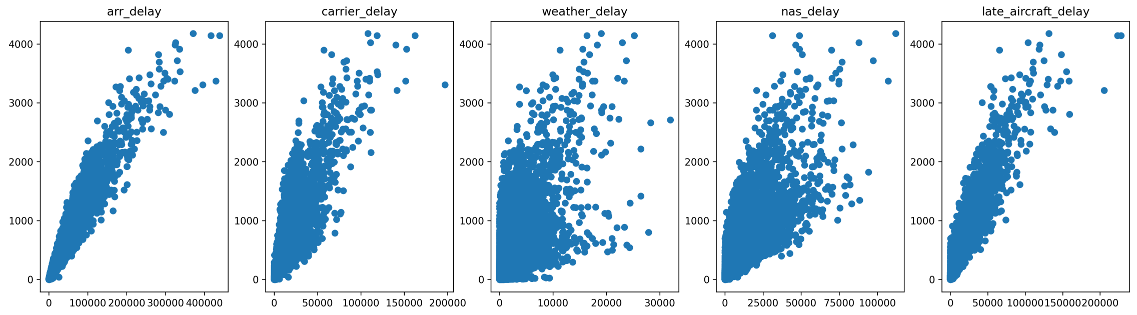


Figure 4: Scatter plot showing correlations.

Figure 4 shows that the dimensions with high correlation (r) between them also have a linear relationship with the target dimension. Therefore, the 4 highly correlated dimensions in Figure 4 were reduced to two dimensions by applying Principal Component Analysis(PCA).

PCA learns the most effective principal components to successfully reduce the dimensionality of the data while retaining most of the trends and patterns. This relies on the assumption that the given observations lie in a lower-dimensional linear subspace. Under this assumption, PCA seeks the best low-rank representation of the given data [8]

Statistic	**Value**
Chi-square value	317,649.856
p-value	0.0
Degrees of freedom	7,880

Table 5: Chi-square test results

To find out if there is a statistically significant relationship between the two dimensions, airport name and vehicle name, a chi-square test was conducted. Table 5 shows the test results.

In statistical analysis, the chi-square -test for consistency, also known as the chi-square test, is a commonly used method to determine if there is a significant association between two categorical variables in a 2×2 contingency table. This test allows researchers to assess whether the observed frequencies in the table deviate significantly from what would be expected under the assumption of independence between the variable [9]

$$\chi^2 = \sum \frac{(O-E)^2}{E} \tag{4}$$

Based on Table 5, the P-value (Probability value) is less than the usual significance level of 0.05. Therefore, the null hypothesis (H0) was rejected and it was concluded that there is a statistically significant relationship between the two dimensions: airport name and vehicle name.

Based on the results of the Chi-square test, the airport name and vehicle name dimensions can be reduced and combined into one dimension, where the dimensions have been reduced from two categorical dimensions to one categorical dimension.

Encoding

It is known that artificial intelligence algorithms are based on calculations performed using various mathematical operations.In order for these calculation processes to be carried out correctly,some types of data cannot be fed directly into the algorithms.In other words,numerical data should be input to these algorithms,but not all data in datasets collected for artificial intelligence algorithms are always numerical.These data may not be quantitative but may be important for the study under consideration.That is, these data cannot be thrown away.In such a case,it is necessary to transfer categorical data to numeric type[10].

The categorical dimension was converted to a numerical dimension by applying Frequency Encoding.

The frequency encoding method is almost the same as the count encoding method, only,in this method, the textual data are exchanged not by the number of repetitions,but by the frequency of repetitions[11]

$$\mathrm{TEnc}\left(b_i\right) = \frac{\sum_{j=1} j}{\sum_{i=1} i} \tag{5}$$

Models

Three models were used in this research ARIMA (Autoregressive Integrated Moving Average),ARIMAX (AutoRegressive Integrated Moving Average with Exogenous variables),LightGBM(Light Gradient Boosting Machine),These models were chosen to achieve the research objective of forecasting time series data.

ARIMA model

The ARIMA model is the most broadly applied approach to working with time series and its analysis. It was first introduced by Box and Jenkins in 1970 (Box and Jenkins,1970).It is useful in such a way that it may characterize various time series data like pure autoregressive, pure moving average, and combined approach.Thus,ARIMA (p,d,q) is the general model,where p,d,and q are autoregressive parameter, number of differencing operators, and moving average parameter,respectively [12] where p represents the number of lag observations in the model (autoregressive part),d is the number of times that the raw observations are differenced (integrated part) to achieve stationarity,q denotes the size of the moving average window (moving average part)[13]

An AR(AutoRegressive) is employed to forecast a time series where AR(1) denotes the first-order autoregressive and Yt is regressed on Yt−1.The autoregressive model of pth order is represented by AR(p).In multiple regression models,the variable of interest is predicted using a linear combination of a set of predictors.A linear combination of a set of past values of the variable is used to build the autoregression model.The term autoregression implies that it is a regression of a variable versus itself. Hence,any autoregressive model of pth order can be mathematically represented as [12]

$$y_t = c + \emptyset_1 y_{t-1} + \emptyset_2 y_{t-2} + \emptyset_3 y_{t-3} + \ldots + \emptyset_p y_{t-p} + \varepsilon_t \tag{6}$$

MA(Moving Average) The dependent variable in moving average process is normally estimated considering both a constant and a moving average of error terms, that is,it is also a regression which is based on current and lagged error terms that behave like a first-order moving average process denoted by MA(1).Instead of using predecessor values of the forecast variable in the regression process,past forecast errors are used in a moving average model in a regression-like model.Additionally,q number of error terms included in the model typically follows the qth order moving average process,denoted by MA(q) which is written as[12]

$$y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \theta_3 \varepsilon_{t-3} + \ldots + \theta_q \varepsilon_{t-q} \tag{7}$$

In time series analysis,most economic variables exhibit non-stationarity, meaning their statistical properties such as mean and variance change over time. Non-stationary data often contains trends, seasonality, or other systematic patterns that need to be removed to make the data suitable for modelling and forecasting. The process of differencing is commonly used to transform non-stationary data into a stationary form. For stationary time series data, the Autoregressive Moving Average (ARMA) model,denoted as ARMA(p, q),can be applied.However,if the data is non-stationary,the Autoregressive Integrated Moving Average (ARIMA) model, denoted as ARIMA(p,d,q),is used.In ARIMA, the 'Integrated'component(d)represents the number of differences required to make the time series stationary.[13] Stationarity is a necessary condition in building a ARIMA model and differencing is often used to stabilize the time series data.The main methods to check the stationarity of time series include the sequence trend diagram,autocorrelation function (ACF), partial autocorrelation function (PACF),augmented dickey-fuller (ADF) unit root test, phillips and perron(PP)test, nonparametric test and so on.[14] In this study, the ACF,PACF plots, and ADF test were used to identify the stationarity of time series

The ARIMA model predicts only one time series. In this research, the prediction depends on a number of external factors, so the ARIMAX model was used.

ARIMAX Model

ARIMAX is an extension of the ARIMA model that incorporates independent variables into the forecasting process.The ARIMAX approach decomposes the time series output into autoregressive(AR),moving average (MA), and integrated (I) components, along with external factors(X).These external factors integrate current values with past values of the independent variables into the model, allowing it to capture more complex data patterns [14]

LightGBM Model

The LightGBM model has many hyperparameters,some of which can be adjusted to improve model performance, while others can shorten training time due to the large volume of data. In short,a hyperparameter is a parameter whose value controls the learning process and determines the value of a model parameter which is eventually learned by the learning algorithm. Therefore,the hyperparameters one of the most critical variables and can determine the conclusion in machine learning [15]

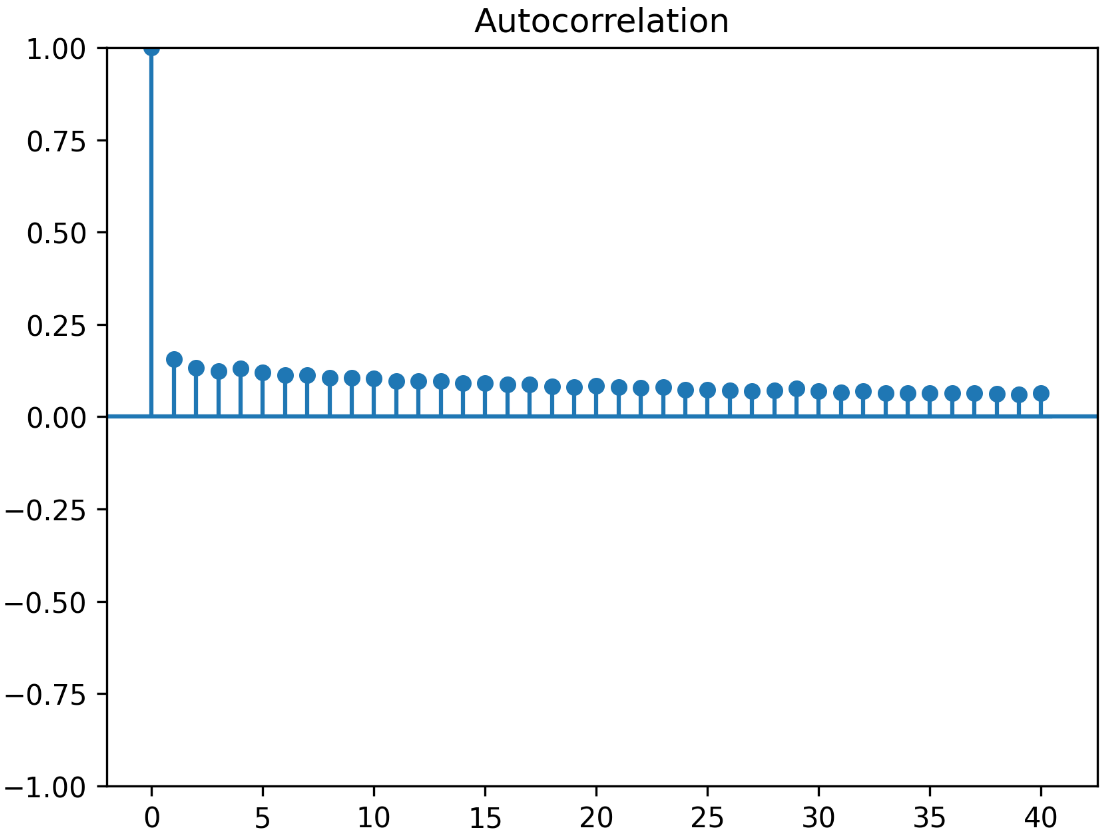


Figure 5: ACF

Figure 5 shows that there is very little positive autocorrelation in the target column and accordingly the value of q=0 was set.

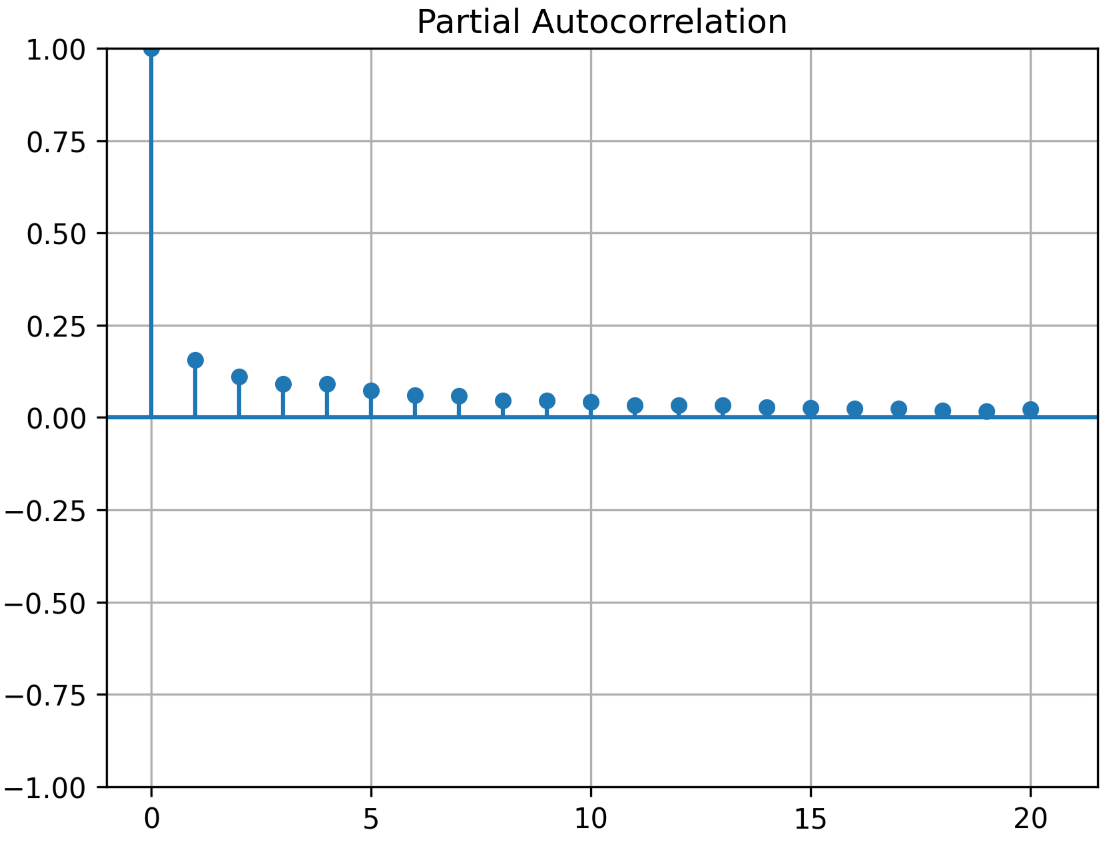


Figure 6: PAC

Figure 6 shows that the partial autocorrelation in the target column is very small and therefore the value of P=0 was set.

Test	**Value**
ADF Statistic	-28.869
p-value	0.000

Table 6: ADF Test Results for Stationarity

Table 6 shows the results of the ADF test for the target series to evaluate the stability of the time series and determine the value of d. The results of the ADF test indicate that there is strong evidence against the non-stationarity hypothesis. Therefore, the series is considered stationary and there was no need to conduct a differentiation. Therefore, the value of d=0 was set.

Experiment

Split Data

The data was arranged chronologically from oldest to newest before starting the segmentation and training processes.

Split the data into a training set and a test set, with 80% of the data divided into the training set and 20% into the test set.

Three forcasting models were implement ARIMA,ARIMAX,LightGBM ,and to evaluate the models used Mean Squared Error(MSE),Mean Absolute Error (MAE),Root Mean Squared Error(RMSE)

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 \tag{8}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|x_i - x| \tag{9}$$

$$\mathrm{RMSE} = \sqrt{\frac{\sum (y_i - y_p)^2}{n}} \tag{10}$$

All experiments in this study were conducted using the Google Colaboratory (Colab) Python development environment.

Result

Table 7: Model Evaluation Metrics

Model	**MAE**	**MSE**	**RMSE**
ARIMA	0.9793	1.8828	1.3722
ARIMAX	0.2902	0.4580	0.6767
LightGBM	0.1892	0.2882	0.5369

Discussion

Table 7 shows that the LightGBM model achieved the lowest error metrics, which indicates its ability to deal with a large set of features, in addition to its support for the Gradient Boosting mechanism, which works to improve the results cumulatively. This is consistent with what was found in the study [16], where the LightGBM model outperformed the other algorithms used in the study when using the same data set with all models. This enhances the reliability of the model in analyzing time series.

The ARIMA model achieved less effectiveness, while the ARIMAX model, which accepts external variables, achieved better performance than ARIMA.

The ARIMA model achieved less effectiveness,while the ARIMAX model, which accepts external variables, achieved better performance than ARIMA. This indicates that incorporating external variables improves the accuracy of prediction,and that models that include external factors are more suitable for the task of predicting airline delays.This is consistent with a study [17] on predicting Hand,Foot and Mouth Disease(HFMD),where the performance of the LSTM (Long Short-Term Memory) and ARIMA models was compared.The results showed that adding external variables improved the performance of LSTM and gave much better results than ARIMA.

Conclosion

In this paper,the performance of the predictive models ARIMA,ARIMAX,LightGBM was compared to determine their performance in predicting the number of monthly delayed flights.The LightGBM model demonstrated its ability to capture non-linear relationships in time series data by achieving the lowest error measures (MAE,MSE,RMSE).The paper also included pre-processing, standardization,and dimensionality reduction to improve the performance of the models.

There are several limitations,including that the data used in this research paper was collected from different airports in the same country,which may reduce the generalizability of the results across different countries.

Future research could benefit from incorporating additional features,such as the months with the highest demand for flights. Deep learning models could also be used to improve interpretability and prediction.

References

[1] Economics-IATA,Air Passenger Market Analysis 2014,International Air Transport Association,2014

[2] M.Efthymiou,E.T.Njoya,P.L.Lo,A.Papatheodorou, and D.Randall,"The impact of delays on customers' satisfaction:an empirical analysis of the British Airways on-time performance at Heathrow airport," Journal of Aerospace Technology and Management,vol.11,2018,Art.no.e0219.

[3] K.Cho,T.Berberian,M.E.Poretti,M.G. Balchanos,A.P.Payan,and D.Mavris,"High Altitude Platform Systems(HAPS) mission planning for the prediction of potential flight delays for commercial aviation," Advanced Engineering Informatics, published online Jan.3,2025.[Online]. Available: https://doi.org/10.2514/6.2025-1789.

[4] A.C.Wimmer,Forecasting Flight Delays with Climate Data and Implications for the Airline Industry,M.S.thesis,Universidade Catolica Portuguesa,Portugal,2024.

[5] S.Kim and E.Park,"Prediction of flight departure delays caused by weather conditions adopting data-driven approaches,"Journal of Big Data, vol. 11,no.11,2024.[Online].Available: https://doi.org/10.1186/s40537-023-00867-5.

[6] B.Konda,"The impact of data preprocessing on data mining outcomes,"World Journal of Advanced Research and Reviews,vol.15,no.3,pp.540–544,2022.[Online].Available: https://doi.org/10.30574/wjarr.2022.15.3.0931

[7] J.Li,S.Guo,R.Ma,J.He,X.Zhang,D.Rui, Y.Ding,Y.Li,L.Jian,J.Cheng,and H.Guo,"Comparison of the effects of imputation methods for missing data in predictive modelling of cohort study datasets," BMC Medical Research Methodology,vol.24,no.41,2024.[Online].Available: https://doi.org/10.1186/s12874-024-02173-x

[8] F.Tonin,Q.Tao,P.Patrinos,and J.A.K.Suykens,"Deep Kernel Principal Component Analysis for Multi-level Feature Learning,"arXiv preprint arXiv:2302.11220,Feb.2023.[Online]. Available: https://arxiv.org/abs/2302.11220

[9] M.Aslam and F.Smarandache,"Chi-square test for imprecise data in consistency table," Frontiers in Applied Mathematics and Statistics,vol.9,2023.[Online].Available: https://www.frontiersin.org/articles/10.3389/fams.2023.1279638/full

[10] A.Iustin,"Effective Methods of Categorical Data Encoding for Artificial Intelligence Algorithms,"Mathematics, vol.12,no.16,p.2553,2024. [Online].Available: https://www.mdpi.com/2227-7390/12/16/2553

[11] M.A.Prakash,K.I.Gandhi,R.Sriram, and Amaysingh,"An Effective Comparative Analysis of Data Preprocessing Techniques,"in Smart Intelligent Computing and Communication Technology, Clifton,VA,USA:IOS Press,2023,pp.14–19. [Online].Available: https://ebooks.iospress.nl/doi/10.3233/APC210005

[12] H.R.Alsamamra,S.Salah,andJ.H. Shoqeir,"Performance analysis of ARIMA model for wind speed forecasting in Jerusalem, Palestine," Energy ExplorationExploitation,vol.42,no.5, pp. 1727–1746,Sep.2024. [Online]. Available: https://doi.org/10.1177/01445987241248201

[13] C.Spulbar and C.C.Ene,"Predictive Analytics in Finance Using the ARIMA Model:Application for Bucharest Stock Exchange Financial Companies Closing Prices,"Studies in Business and Economics,vol.19,no.3,pp.142–153,2024. DIO:10.2478/sbe-2024-0042.

[14] I.G.I.Sudipa,K.M.Aman,I.M.S. Sandhiyasa,K.J.Atmaja,and I.G. Sudiantara,"Predictive time-series modelling of rice price fluctuations in East Nusa Tenggara using ARIMAX:A data driven case study,"PowerTech Journal, vol.48,no.4,pp.111–120,Nov.2024. [Online].Available: https://powertechjournal.com

[15] Using the Light Gradient Boosting Machine (LightGBM) Model," International Journal on Informatics Visualization, Jun.2025[Online].Available: https://doi.org/10.13140/RG.2.2.37740.49622

[16]A.D Hartanto,Y.N.Kholik and Y. Pristyanto,"Stock Price Time Series Data Forecasting Using the Light Gradient Boosting Machine(LightGBM) Model," INTERNATIONAL JOURNAL ON INFORMATICS VISUALIZATION,vol.7,no.1,pp.73-78,Mar. 2023.[Online].Available: https://www.researchgate.net/publication/377404962

[17] Q.Zhang,S.Sun,H.Zhang,and Y.Liu, "Application of ARIMA and LSTM models in predicting the incidence of hand,foot and mouth disease in Ningbo,China," BMC Infectious Diseases,vol.20,no.1,pp.1–10, Jan.2020.[Online].Available:https://doi.org/10.1186/s12879-020-05245-4