SI 670 Final Project Report

Comparative Analysis of ARIMA and Neural Network Models for forecasting International Flight Departures

Akshat Porwal (akshatp), Anisha Vrinda Dsouza (anishavd), Prathamesh Joshi (prathuj), Saideepti Seelam Veera Naga (svnsdeep)
Project ID: 17

Abstract

This project delves into comparing ARIMA and Neural Network models for forecasting international flight departures. It aims to scrutinize their predictive accuracy, efficiency, and adaptability to flight departure patterns. Through comprehensive evaluation metrics and analysis, the project intends to elucidate the strengths and weaknesses of each model, providing valuable insights for aviation forecasting strategies and operational decisions.

In addition to the comparative analysis, this project involves tuning hyperparameters for ARIMA and exploring various configurations of neural networks. This exhaustive approach aims to fine-tune both models, optimizing their predictive capabilities for accurately forecasting international flight departures.

1. Data Description

The dataset utilized for our project constitutes the U.S. International Air Traffic data spanning the period from 1990 to 2020, sourced from the U.S. International Air Passenger and Freight Statistics Report. The Departures dataset, offers information on flights that originate at U.S. and end at non-U.S. gateways. It includes particular airline information as well as scheduled, charter, and overall flight counts.

2. EDA

2.1 Data Preprocessing

We convert the dataset into a time series format, we utilize the timestamp of the flights, whether it's based on date, month, or a combined representation. Handling missing values and outliers becomes crucial for our analysis, as these elements can significantly influence the accuracy of our time series evaluation.

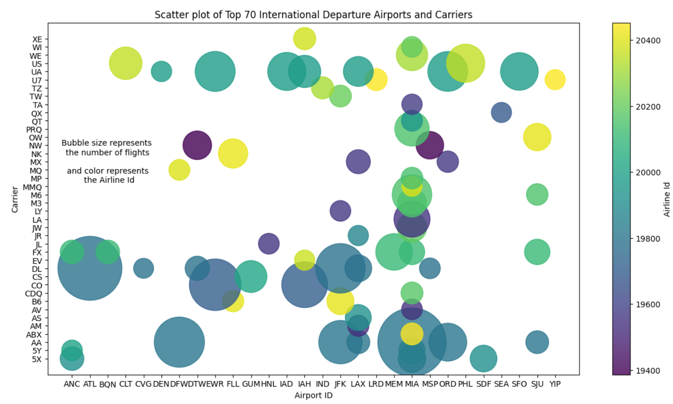


Figure 1: Scatter plot of Top 70 International Departure Airports and Carriers

2.2 Visualize Time Series

To comprehensively examine the time series data, we plan to employ line plots to gain valuable insight into the trends, patterns, and seasonality of the data moving forward. During the course of the study, we can observe the fluctuations and behaviors exhibited by the total number of flights over time through a visual inspection of these plots.

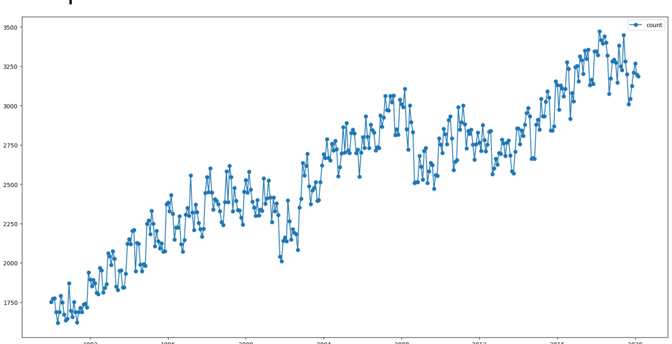


Figure 2: Time Series plot of Departure Data

The analysis of the time series data reveals a consistent upward trend in the total number of international flight departures, expanding from approximately 100 million in 1990 to exceeding 400 million in 2020. Notably, a distinctive seasonal pattern is evident, characterized by a surge in flight numbers during the summer months and a decline during the winter months. Moreover, notable short-term fluctuations are also observable, exemplified by a decrease in flights during the 2008 financial crisis.

The time series plot of passenger data exhibits a pronounced upward trend, punctuated by seasonal fluctuations. A notable spike in passenger numbers occurs during the summer months (June-August), while a corresponding decline is observed during the winter months (December-February). Additionally, the data reveals outliers, such as the precipitous drop in passenger volume following the 9/11 attacks in 2001, likely attributable to extraordinary events that disrupt travel patterns.

3. Literature Review

In the aviation industry, departure forecasting is essential for improving overall efficiency, resource management, and operational planning.

AutoRegressive Integrated Moving Average (ARIMA) models have been widely used in the aviation sector historically, due to groundbreaking studies like Box and Jenkins (1970). In a variety of fields, these conventional models have shown to be successful in capturing linear trends and seasonality. Among the noteworthy applications are those made by Tuckwell and Uys (2002), who used ARIMA models to anticipate demand for air travel and set the stage for later research.

According to recent research, hybrid models that combine conventional methods and neural networks are beneficial for departure prediction. An ARIMA-Long Short-Term Memory (LSTM) hybrid model was proposed by Wang et al. (2016), utilizing both pattern recognition and historical context. Hong et al.'s (2020) demonstration of ensemble approaches shows how integrating different forecasting models can improve departure prediction accuracy.

Despite progress, many issues remain, such as the interpretability of neural network models and the requirement for real-time flexibility. In a thorough analysis, Chen et al. (2019) emphasized the difficulties with interpretability in NNs and offered potential solutions. As suggested by Zheng et al. (2015), including outside variables—like meteorological data—represents a new line of inquiry to strengthen departure forecasting models' resilience.

A dynamic research landscape that highlights the particular complexities and difficulties involved with predicting departure trends within the aviation industry is presented by the synthesis of approaches and continuous efforts to overcome obstacles.

4. Time Series Analysis

Time series modeling involves analyzing and forecasting data points collected and ordered over time. Its primary objective is to capture and understand the underlying patterns, trends, and dependencies within the sequential data. This modeling technique is fundamental across various domains, aiding in predicting future values based on historical observations.

4.1 Time Series Decomposition

Time series decomposition is a statistical technique that divides a time series into its basic components, which include trend, seasonality, and residual. The goal of this technique is to better understand the underlying structure and behavior of time series, as well as to aid in forecasting and anomaly identification.

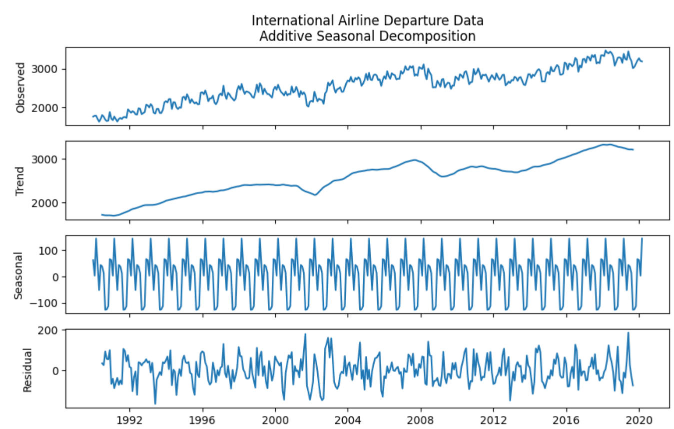


Figure 3: Time Series Decomposition into Trend, Seasonality and Residuals based on Additive Model

4.1 Stationarity of Time Series

We evaluate the stationarity of the time series using three approaches: original, differencing, and log transformation. The original series is examined first to identify trends and patterns. Next, differencing the series and applying a log transformation helps mitigate trends and achieve stability.

By calculating the rolling mean and rolling standard deviation for each transformed series, we visually assess their fluctuations and variances over time. A stationary series would display a relatively constant mean and variance across different time periods, providing a foundation for more reliable time series modeling and forecasting techniques.

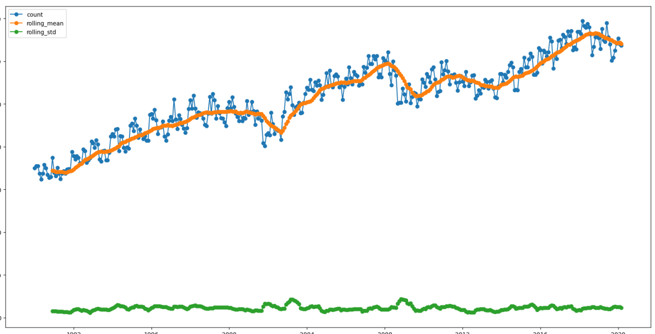


Figure 4: Time series and rolling Mean and rolling Standard deviation for Original Series

The above plot displays the time series of the flight departures, along with the rolling mean and rolling standard deviation of the data. The rolling mean, which is a moving average of the data over a specified window size, shows that the number of international flight departures has increased steadily over time, with some seasonal fluctuations. On the other hand, the rolling standard deviation, which is a moving standard deviation of the data over a specified window size, indicates that the variability in the number of flight departures has remained relatively constant over time, suggesting that the underlying distribution of flight departures is stable.

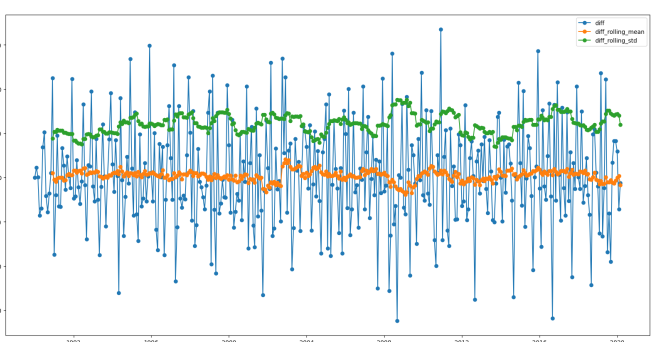


Figure 5: Time series and rolling Mean and rolling Standard deviation for Differenced Series, where difference is First Order

The plot in Figure 5 reveals that the differencing process has effectively eliminated the non-stationary behavior observed in the original time series. The rolling mean and rolling standard deviation lines of the differenced series are relatively constant over time, indicating that the mean and variance of the series are no longer changing over time. This suggests that the differenced series is more amenable to time series analysis methods that assume stationarity.

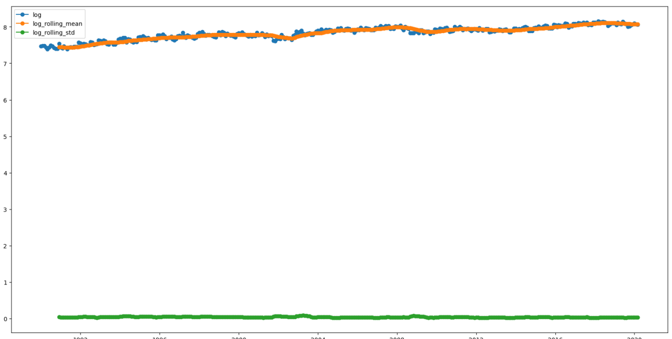


Figure 6: Time series and rolling Mean and rolling Standard deviation for Log Transformed Series

The log time series of international flight departures exhibits a clearer upward trend compared to the original series, indicating exponential growth in air travel demand. The log-transformed series also demonstrates greater stationarity, with relatively constant rolling mean and rolling standard deviation lines. This suggests its suitability for time series analysis methods that assume stationarity. Additionally, the log-transformed series highlights the accelerated growth in air travel demand in recent years.

4.2 Augmented Dickey–Fuller (ADF) test

In the Augmented Dickey-Fuller (ADF) test, which is commonly used to test for stationarity in a time series dataset, the p-value serves as a measure to determine the significance of the test results.

For a significance level (alpha) of 5%, the threshold for the p-value in the ADF test is typically less than 0.05. The hypothesis conditions are:

- Null Hypothesis (H0): The null hypothesis assumes that the time series data is non-stationary, implying it possesses a unit root (i.e., it's not stationary).
- Alternative Hypothesis (H1): The alternative hypothesis contradicts the null hypothesis. It

suggests that the time series data is stationary (i.e., no unit root).

Interpreting the p-value in the ADF test:
- If the p-value is less than the chosen significance level (alpha), such as 0.05, it provides evidence to reject the null hypothesis (H0). In this case, it suggests that the time series data is stationary.
- Conversely, if the p-value is greater than the significance level (alpha), there's insufficient evidence to reject the null hypothesis. It indicates that the time series data is non-stationary and likely possesses a unit root.

For our data, the ADF test outputs gave:

```
Original Time Series
    ADF Statistic: -1.829860
    p-value: 0.365741
Differenced Time series
    ADF Statistic: -4.895917
    p-value: 0.000036
Log Time series
    ADF Statistic: -2.313881
    p-value: 0.167530
```

Thus we conclude that graphically as well as by ADF test that the First Ordered Differenced Time Series is a stationary time series.

4.3 ACF and PACF plots

Next we proceeded with ACF (Autocorrelation Function) and PACF (Partial Autocorrelation Function) plots which are graphical tools crucial in time series analysis. ACF represents the correlation between a time series and its lagged values at different time intervals, showcasing the relationship between points separated by various time spans. PACF, on the other hand, portrays the correlation between observations at specific lags, excluding the influence of intermediate time points.

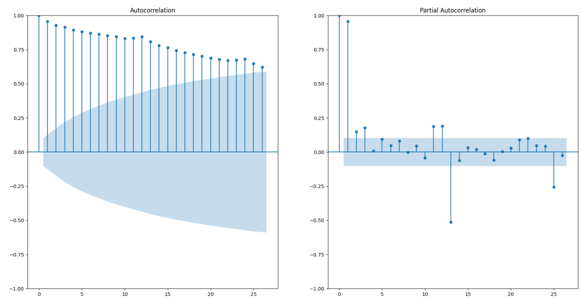


Figure 7: ACF and PACF components of Original time series

ACF and PACF components of Original time series show that the seasonality occurs at the intervals of 12 months, and the PACF plot confirms the correlation between t_k and t_{k-1} (its first order difference) for original time series.

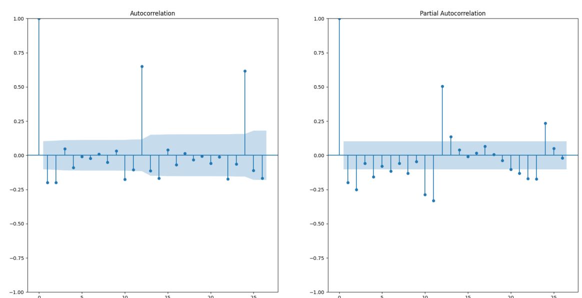


Figure 8: ACF and PACF components of Differenced time series

ACF and PACF components of Differenced time series show that the seasonality occurs at the intervals of 12 months, and the ACF plot does not show a correlation between t_k and t_{k-1} for differentiated time series.

5. ARIMA Models

ARIMA (AutoRegressive Integrated Moving Average) models are a popular approach in time series forecasting. They amalgamate autoregression (AR), differencing (I), and moving averages (MA) to predict future values based on historical data. These models are characterized by their parameters: p (order of AR), d (degree of differencing), and q (order of MA). ARIMA models are versatile and capable of handling diverse patterns within time series data, enabling predictions and trend identification.

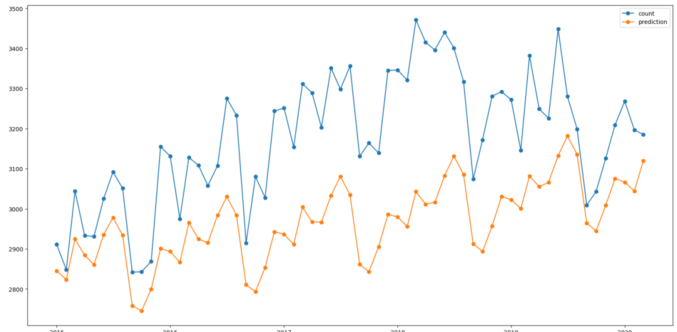


Figure 9: Test output for Original Time Series for ARIMA (12, 2, 12)

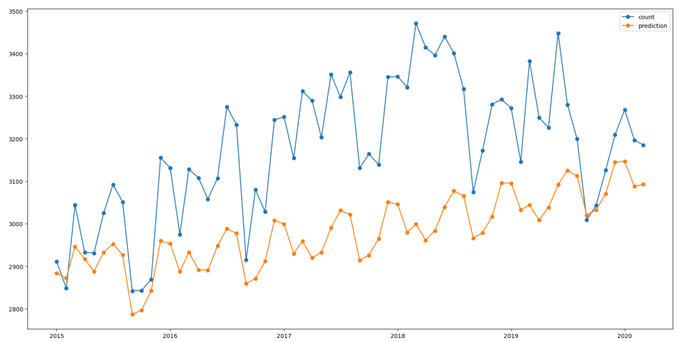


Figure 10: Test output for Original Time Series for ARIMA (12, 3, 12)

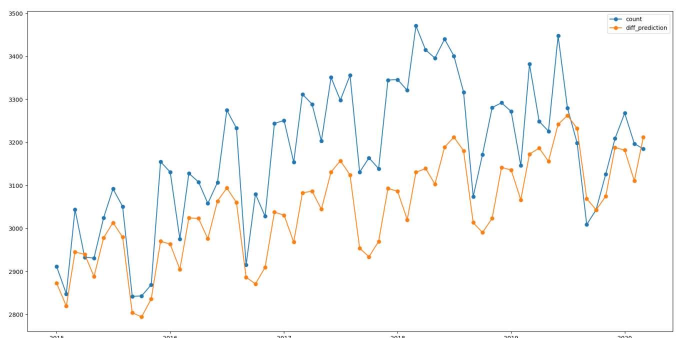


Figure 11: Test output for Differenced Time Series for ARIMA (12, 0, 12)

6. Neural Networks

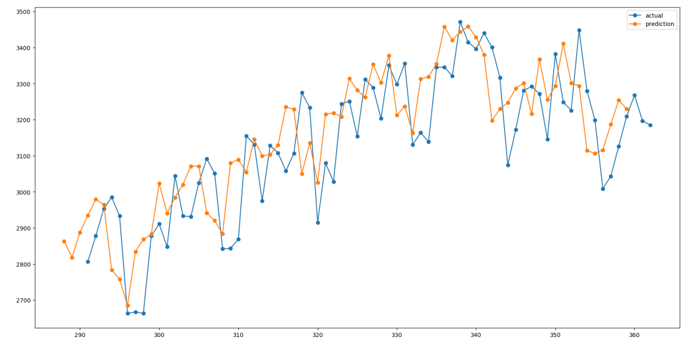
Neural networks for time series analysis entail structuring sequential data into input-output pairs, training the network to recognize patterns within these sequences, optimizing the network's parameters to minimize prediction errors, and utilizing the trained model to forecast future values in the time series.

6.1 Dense Neural Network

For the current exploration, we have considered the window size = 3 and epochs = 100 for generating the input data and the output will be the next preceding time series value.

Thus we chose optimum model with layers as:

```
Layer1: (Dense(64, activation='relu'))
Layer2: (Dense(32, activation='relu'))
Output: (Dense(1))
```



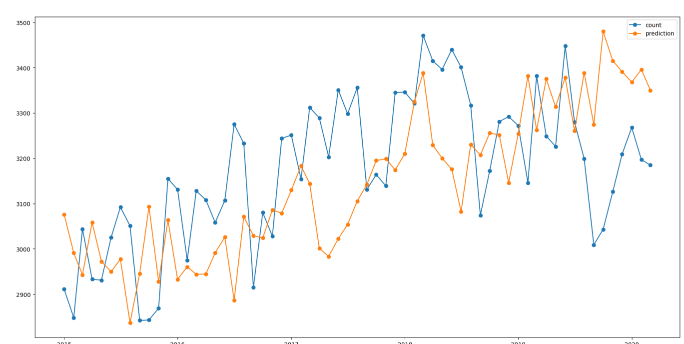
Graph 11: Test output for Original Time Series for Dense Neural Network

6.2 LSTM Neural Network

RNNs, including LSTMs, are beneficial for time series analysis as they handle sequences effectively. They learn from historical patterns and can predict future values based on the sequential nature of the data. LSTMs, in particular, excel in preserving long-range dependencies, making them suitable for longer time series forecasting tasks where capturing context over extended periods is crucial.

For the LSTM Model, we constructed a window size of 48, and 50 epochs. A model was constructed as follows:

Layer (type)	Output Shape	Param #
lstm_9 (LSTM)	(None, 64)	16896
dense_10(Dense)	(None, 1)	65



Graph 12: Test output for Original Time Series for LSTM Neural Network

7. Observations

For comparative analysis of various models on parameters like accuracy and speed, we chose two criteria, RMSE value and model execution time respectively. The observations for them are represented in Table 1.

Model	RMSE	Execution Time
ARIMA (12, 2, 12) Original TS	232.99	2.65 s
ARIMA (12, 3, 12) Original TS	227.76	2.48 s
ARIMA (12, 0, 12) Differenced TS	159.96	2.14 s
Dense Neural Nets	121.15	35 s
LSTM Neural Nets	174.27	109 s

Table 1: Comparative Analysis of Accuracy and Speed for various models.

8. Inferences and Comparative Analysis

The comparative analysis of ARIMA (AutoRegressive Integrated Moving Average) models and neural networks hinges on two pivotal criteria: accuracy, as measured by RMSE (Root Mean Square Error) values, and speed, represented by model execution times. The observations, as presented in Table 1, illustrate the performance of various models across these dimensions.

1. Accuracy:

- ARIMA: Demonstrates efficacy in handling linear relationships and stationary time series. However, its performance may degrade when confronted with datasets exhibiting intricate patterns.
- Neural Networks: Possesses the capability to capture complex nonlinear relationships and patterns, potentially outperforming ARIMA on data characterized by intricate structures. However, caution is warranted to prevent overfitting, necessitating appropriate regularization strategies.

2. Speed:

- ARIMA: Generally excels in speed for smaller datasets and simpler structures, owing to its linear nature and less computationally intensive operations.
- Neural Networks: May exhibit slower performance, particularly with larger datasets, deep architectures, or extensive training epochs. The computational time is a crucial consideration in assessing overall efficiency.

3. Robustness:

- ARIMA: Hinges on stationarity assumptions and may encounter challenges with highly volatile or non-stationary data.
- Neural Networks: Exhibits greater adaptability to diverse data types, showcasing resilience in handling non-stationarity, outliers, and noisy data through judicious preprocessing and regularization.

4. Transparency and Interpretability:

- ARIMA: Simpler and more transparent due to its linear nature, making it easier to understand and interpret the model coefficients and forecasts.
- Neural Networks: Often considered black-box models due to their complex structures and numerous parameters, making it challenging to interpret how predictions are made.

9. Conclusion

This project embarked on a comprehensive exploration of time series forecasting for international flight departures, employing a comparative analysis of ARIMA and Neural Network models.

The Augmented Dickey-Fuller (ADF) test guided our efforts towards achieving stationarity, with differencing proving to be a crucial step in mitigating non-stationarity in the original time series data.

A meticulous comparative analysis of time series forecasting models, namely ARIMA and Neural Networks, was conducted for predicting international flight departures. ARIMA models, particularly (12, 2, 12) and (12, 3, 12) for the original time series, showed competitive performance with RMSE values of 232.99 and 227.76, respectively, alongside fast execution times of approximately 2.5 seconds. Differentiating the time series enhanced ARIMA's performance, reducing RMSE to 159.96 with a quicker execution time of 2.14 seconds.

Dense Neural Networks demonstrated superior accuracy with an RMSE of 121.15, showcasing their effectiveness in capturing intricate patterns. However, their longer execution time (35 seconds) raises considerations for real-time applications. LSTM Neural Networks, while exhibiting an RMSE of 174.27, displayed promising results, emphasizing their ability to capture long-term dependencies, albeit with a longer execution time of 109 seconds.

The incorporation of feedback from reviews led to valuable refinements, addressing concerns about dataset scope, computational resources, interpretability, and the potential inclusion of weather data. The literature review highlighted the dynamic landscape of time series forecasting methodologies in aviation, providing context and inspiration for our analyses.

In essence, this project contributes to the evolving field of aviation forecasting by offering insights into the strengths and trade-offs between traditional and modern forecasting models. The results and methodologies presented herein serve as a foundation for further research, encouraging a continuous exploration of innovative techniques to enhance the precision and applicability of time series forecasting in the aviation industry.

10. Future Work

In furthering time series analysis, several steps can enhance the accuracy and performance of forecasting models. Initial steps involve analyzing residuals from fitted models like ARIMA or LSTM to detect any patterns or anomalies, ensuring these residuals exhibit stationary behavior. Following this, the focus could shift towards tuning the hyperparameters of LSTM networks.

Experimentation with variations in LSTM units, learning rates, and other parameters can optimize model performance. Feature engineering remains a crucial aspect, exploring additional domain-specific features or transformations, such as lagged variables or moving averages, to improve predictive power.

References:

- Chen, J., Song, L., Wainwright, M. J. (2019). Are neural networks interpretable for time series? arXiv preprint arXiv:1905.13448
- Gajdoš, A., Hančl, J., & Hančlová, J. (2016). Selecting the optimal ARIMA model in time series forecasting
- Brockwell, P. J., & Davis, R. A. (2016). Introduction to Time Series and Forecasting. Springer.
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A Simple Way to Prevent Neural Networks from Overfitting. Journal of Machine Learning Research, 15, 1929-1958.
- Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing..
- Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. Neural Computation.
- MurçaM.C.R. et al.
  Flight trajectory data analytics for characterization of air traffic flows: A comparative analysis of terminal area operations between New York, Hong Kong and Sao Paulo
  Transp. Res. C (2018)
- Weatherford, Larry & Gentry, T & Wilamowski, Bogdan. (2003). Neural network forecasting for airlines: A comparative analysis. Journal of Revenue and Pricing Management. 1. 319-331. 10.1057/palgrave.rpm.5170036.