**Forecasting Performance of ARIMA and Simple ANN Models in the Restaurant Industry**

Carl J. Loucius

Colorado State University Global

MIS581

Dr. Osama Morad

07/28/24

**Abstract**

This study investigates the forecasting performance of a traditional time series statistical model and a neural network trained on restaurant sales data. It aims to address a gap in the academic literature on constructing sales forecasting models in the restaurant industry, as little information exists on this subject. Addressing this gap allows restaurant and business owners to predict busier and slower times, allowing them to ensure adequate staffing levels and inventory management practices are in place to reflect the projected sales. Restaurant brands that can manage and control their costs in line with their revenue are more likely to remain competitive in this landscape.

The starting point of this study focused on determining whether the temporal data in this study exhibit stationary behavior since traditional time series statistical models require that the data exhibit this behavior. The findings from the analysis of this question determined that the data exhibit stationary behavior. Then, this study turned to determine the optimal ARIMA and Simple ANN models for the data using the Grid Search optimization search procedure to find the optimal values of chosen hyperparameters to be tuned. After determining the optimal models, this study evaluated the forecast performance for December 2015 by comparing the RMSE test set score. According to the results, the Simple ANN model produced a more accurate model by reducing the RMSE test set score by 24.40% from the ARIMA’s model RMSE score. However, the time series plots of the predictions vs actuals for December 2015 show a different picture. The ARIMA model appears to perform better since it can model sales noise better than the Simple ANN model. Despite these mixed results, the findings indicate that the nature of sales forecasting needs to be studied and addressed in academic literature more than ever since there is a true gap in our understanding of sales behavior and how to model it in the restaurant industry.

Table of Contents

[Forecasting Performance of ARIMA and Simple ANN Models in the Restaurant Industry 4](#_Toc173693946)

[Objectives 5](#_Toc173693947)

[Overview of Study 6](#_Toc173693948)

[Research Questions and Hypotheses 6](#_Toc173693949)

[Literature Review 7](#_Toc173693950)

[Research Methodology 9](#_Toc173693951)

[Data 9](#_Toc173693952)

[Methods 11](#_Toc173693953)

[Limitations 14](#_Toc173693954)

[Ethical Considerations 15](#_Toc173693955)

[Results and Discussion 15](#_Toc173693956)

[Descriptive Statistics and Time Series Plots: Analysis 16](#_Toc173693957)

[RQ1: Stationary Results 21](#_Toc173693958)

[RQ2: Forecasting Performance of ARIMA and Simple ANN Results 21](#_Toc173693959)

[Conclusion and Recommendations for Further Study 26](#_Toc173693960)

[References 29](#_Toc173693961)

# **Forecasting Performance of ARIMA and Simple ANN Models in the Restaurant Industry**

Industries rely on forecasting as a pivotal tool for decision-making. Forecasting allows “efficient resource allocation, cost reduction, and demand planning” for sectors such as the restaurant industry (Chaudhuri, 2023). According to Chaudhuri (2023), industries rely on these forecasts for various decisions, such as optimizing inventory and enhancing supply chain efficiency. By utilizing forecasting in decision-making, organizations can become empowered to adapt to changing market conditions and remain competitive in their respective industry (Hall, 2020).

However, the art of forecasting is complex due to the numerous methods available to business practitioners. Business practitioners can use regression, ARIMA, exponential smoothing, deep learning methods, and more to produce forecasts. The number of available methods makes it hard for the business practitioner to determine which model is the most appropriate for their business problem. To complicate matters further, business practitioners must consider the model’s performance since ineffective models can damage a company as much as effective models can empower organizations to remain competitive (“Exploring the impact of inaccurate sales forecasts”, n.d.), illustrating the complexity of the forecasting process for the business practitioner. To help alleviate this issue, research must be conducted to determine the most effective models for forecasting purposes, but this relies on understanding the industry in which the forecasts are being used.

Effective statistical machine-learning models for sales and demand forecasting could significantly benefit the restaurant industry. Business practitioners in this industry can benefit from knowing the most effective models to save time and resources for an organization without doing heavy research to determine these models. This project seeks to determine the most effective models based on the forecasts' accuracy for a sample sales dataset. The choice to focus on the accuracy of the estimates rests in the fact that sales are tied to inventory, which is a significant cost item for this industry. Furthermore, sales are tied to the business's staffing, another significant cost item in this industry. If the forecasts are off, both areas will be impacted, which can affect the company in other important ways (e.g., understaffing issues can lead to poor financial performance and influence how customers view the brand). Hence, an organization will benefit from a model that minimizes the forecasting error the most.

This study aims to identify the performance differences between utilizing a time series statistical model and a simple artificial neural network to predict restaurant sales. By identifying if there is a performance difference in the models’ forecast predictions, this project aims to provide insights for businesses on how to create more accurate forecasts that can impact their overall operations, allowing them to develop strategies and plans to maximize their potential. The goal is to illustrate to business users and others how to evaluate forecasting models and determine the most effective model by using the dataset chosen for this project, ultimately providing business users in the restaurant industry with the knowledge to do this in their own workplaces.

## **Objectives**

This study aims to determine whether forecasts using an ARIMA and Simple ANN model perform differently. By analyzing historical data on a restaurant's sales, the models aim to create a forecast that can accurately model time series noise and predict the sales for a designated time period. While this study will develop a methodology that will be utilized for this data, the methods can be tailored to meet the needs of others since the approach will be general enough that it can be applied to any restaurant sales data. Hopefully, this approach will allow restaurant organizations to develop a more accurate sales forecast model that can be utilized to maximize the potential of data-driven decision-making.

## **Overview of Study**

This study is organized as follows. The next section will identify this study's research questions and hypotheses before providing an overview of the current landscape of academic literature on creating forecasting models in the following section. The goal is not to provide a comprehensive overview of the field during this discussion but to provide enough background to understand where this study stands within this evolving landscape. Afterward, this study will delve into the research methodology utilized for this project. It will discuss the overall flow of the project, the tools and methods used for the analysis, and any limitations and ethical considerations that had to be considered. Lastly, this study will cover the results of the study’s methodology. Then, a final section will contain the conclusion and further recommendations for future studies.

# **Research Questions and Hypotheses**

First, this research paper will address one of the most critical aspects of dealing with temporal data: Is the data stationary? Time series data typically have underlying patterns such as trend or seasonality, which can cause the models to make unreliable and misleading predictions when deployed as a tool for forecasting (Prof. Frenzel, 2023; Hyndman & Athanasopoulos, 2018; Nielsen, 2020). For instance, a model seeking to estimate the mean of a time series with a nonstationary mean and variance will be questionable since the bias and error in the model will vary over time, making it problematic for decision-making purposes. Thus, if these underlying patterns are present in the dataset, it will be necessary to transform the data using some method before developing a model. In fact, many methods available to analyze time series data require that the data be stationary, which illustrates why this is the first question to answer about the dataset. Hence, the first set of hypotheses can be formulated as:

The time series dataset is nonstationary.

The time series dataset is stationary.

The second part of this project will focus on determining whether an ARIMA statistical time series model or a Simple Artificial Neural Network (Simple ANN) deep learning model produces more accurate sales forecasts using the available dataset, or stated as a question: Does an ARIMA statistical time series model or a Simple ANN deep learning model produce a more accurate forecast? As mentioned previously, the project's goal is to compare the performance of these models to determine which model will be more effective for the organization to use for forecasting purposes. These models will be evaluated based on their forecast accuracy by comparing the performance of the test models through the root mean squared error (RMSE) evaluation metric. Hence, the second set of hypotheses can be formulated as:

The forecast accuracy of the ARIMA time series model and the Simple ANN model is not significantly different.

The forecast accuracy of the ARIMA time series model and the Simple ANN model is significantly different.

# **Literature Review**

With the development of new technologies and increased data availability, businesses have gained a new way to become and remain competitive. Businesses have started to employ data analytics methods to support business decisions. These methods support the decision-making process by providing business users and leaders with actionable intelligence in a timely manner. However, the reliance and usage of these methods vary by industry. Similarly, the coverage of how these methods is used and applied in the industries reflects how much the industry relies on the methods.

According to Roy et al. (2020), their searches using the Web of Science database yielded 1610 hits (out of 11,958 publications) in the business and management categories, which is the most prominent category for articles written on the analytic applications of data, when they used the search term “restaurant.” Also, Roy et al. (2020) used the search term “restaurant\* analytic\*.” and it yielded 212 publications. This low number of publications is concerning and shows a real need for research on data analytics applications in the restaurant industry. Also, it highlights a lack of readily available articles on data analytics applications in the industry, so it will be necessary to expand the search horizon to focus on general applications of the methods.

Luckily, many articles have been written on developing forecasting models, applying time series methods, and utilizing machine learning models in other industries. These topics have been discussed extensively concerning the automated and retail industries. One of the important overall findings from these articles is that models do not perform similarly because of the type of data used. For example, Liu et al. (2021) found that sales forecasting performance can be significantly improved in the retail industry by utilizing the ConvLSTM machine learning method and a framework that incorporates the spatial information of sales data compared to some traditional time series methods such as ARIMA. However, E. et al. (2022) found that a dynamic model selection approach for demand forecasting of retail items is more effective than traditional classical forecasting methods, which contrasts with creating forecasts based on sales.

On the other hand, Fortsch et al. (2022) found that automotive organizations can benefit from utilizing nonlinear forecasting models that incorporate competitors’ sales data over traditional linear models when predicting domestic sales. The difference in findings illustrates that this phenomenon will more than likely occur in the restaurant industry, too, highlighting the need for researchers to consider studying the forecasting performance of multiple models to determine the most effective method for the data. This will provide a starting point for future research to consider other aspects of forecasting in the restaurant industry.

# **Research Methodology**

This methodology section is divided into four parts. It begins by describing the dataset to be used for the project, its original intended purpose, and providing a data dictionary to define the attributes that are part of the dataset. Then, the paper will focus on describing the methods used on the dataset. This section will expand upon how the project will proceed by discussing the various methods that will be used to analyze the dataset, including the technologies that will be utilized to analyze the dataset, the statistical tests that will be utilized to evaluate the hypotheses for each research question, how the ARIMA and Simple ANN will be constructed, and the evaluation metrics that will be analyzed. Afterward, this section will briefly discuss any limitations and ethical considerations of the dataset and methods.

## **Data**

This project utilizes a dataset that Zhuang (2022) shared on the Kaggle platform. Initially, the dataset was shared on the Maven Analytics (Ruiz, 2022) platform as four separate tables. The goal of sharing the dataset was to provide a challenge where users are supposed to build a single-page dashboard to answer various business questions that can improve the restaurant’s operations, effectively making the focus on descriptive statistics and reporting. However, Zhuang modified the original dataset for the Kaggle platform to incorporate all the information from the four tables into a single table, providing these users with a more accessible dataset for the same purpose of building a single-page dashboard.

Zhuang’s (2022) dataset contains 11 features and approximately 50,000 observations of transactional data records on pizza sales throughout 2015. The variables in the dataset contain information relevant to an order line from a particular order, such as what pizza was ordered, the size, how many, and so forth. For instance, one observation might show that one small veggie pizza was ordered on July 3, 2015, at 11:30:15 am, and the total was $16.95. This observation line would also contain information about the ingredients of the veggie pizza. Table 1 provides a detailed breakdown of the various variables found in the dataset, the variable type, and a description of what the variable represents in the dataset.

**Table 1**

*Pizza Sales Dataset – Variables Description*

|  |  |  |
| --- | --- | --- |
| Variable | Variable Type | Variable Description |
| order\_details\_id | Categorical | Indicates the order line from a particular order. |
| order\_id | Categorical | The order number for a particular order. Each number corresponds to a unique order in chronological order. |
| pizza\_id | Categorical | Indicates what pizza was ordered using an integer-based system. |
| quantity | Numerical | Indicates the quantity of a specific pizza order configuration for a particular order line. |
| order\_date | Numerical | Indicates the date on which the order was placed. |
| order\_time | Numerical | Indicates the time for which the order was processed. |
| unit\_price | Numerical | Indicates the price for a single unit of the specific pizza order configuration. |
| total\_price | Numerical | Lists the total price for the order line. |
| pizza\_size | Categorical | Represents the size of the pizza that was ordered (S = small, M = medium, L = large) |
| pizza\_category | Categorical | Lists the pizza category that the pizza falls under. |
| pizza\_ingredients | Categorical | Lists the ingredients found on the pizza configuration. |

Although the dataset was provided for descriptive statistics and reporting purposes, this project will utilize it to create various predictive models for forecasting purposes. The goal is to determine the most accurate model for forecasting sales utilizing this dataset. Most of the information from the dataset will not be used since it is not pertinent in answering the business questions. In fact, none of the provided attributes will be used as variables in the model. Instead, this project will create the user-generated variables “Total Sales”, “Total Visitors”, and “Total Pizzas Sold” to represent these quantities on a specific date and use these variables for the purpose of analysis. In particular, the user-generated variable “Total Sales” will be used along with its time component to generate the forecasting models to be evaluated. In contrast, the other two variables were analyzed only for descriptive purposes since a strong correlation exists between the user-generated variables.

## **Methods**

The project aims to determine whether an ARIMA or a Simple ANN model produces a more accurate daily forecast model by comparing the model's test performance's root mean squared error results. Python will be used as the primary technological tool for analysis. Python is a perfect tool since it is open-source and well-supported, with various well-developed analytic libraries (Terra, 2024). The Pandas, Matplotlib, NumPy, Scikit-sklearn, Keras, Statsmodels, and TensorFlow libraries will be incorporated into the project since they provide advanced data handling and manipulation capabilities, advanced plotting capabilities, the ability to perform statistical tests and produce the necessary results, and the ability to create various statistical and machine learning models for predictive purposes.

Furthermore, the dataset will need to undergo preprocessing before the initial exploration of the data. The data in the dataset is organized to recognize the transactional actions on a receipt, meaning that observation in the dataset records the transaction of a particular type of pizza along with additional information for a particular order (or it can be said it records transactions using seconds as the time interval). The dataset will be transformed to introduce the previously mentioned user-generated variables to represent new information tied to a different time interval component. These new variables will record the total sales, number of visitors, and number of pizzas sold on any given day. The decision to use an accumulated snapshot of each activity (sales, visitors, number of pizzas) stems from the common decision to analyze temporal data utilizing a grain of a day or month. Smaller time intervals can introduce more noise or error into the models than larger time intervals. Also, business decision leaders in the restaurant industry will be more concerned with models that forecast accurately daily since they provide information to the leaders quickly, allowing them to make decisions to impact the restaurant operations and financials quicker.

This new dataset with the accumulated variables will be the focus of the project's analysis. It will be analyzed using descriptive statistics to understand the distribution of the user-generated variables. This provides basic information on whether the variables exhibit stationary or nonstationary behavior. This analysis will be expanded upon by analyzing the time series plots to see whether a visualization displays whether the variables exhibit stationary or nonstationary behavior. The purpose of examining whether the data exhibits stationarity is to determine whether the dataset will undergo a transformation process since an ARIMA model requires that the data exhibit this behavior. Although analyzing the descriptive statistics and inspecting time series plots can reveal information about the stationarity behavior of the data, they are not meant to be a de facto method to determine whether the data exhibits this method. Instead, the Augmented Dickey-Fuller test is more appropriate to assess whether the variables exhibit stationary or nonstationary behavior (Nielsen, 2020). This test determines whether a unit root is present within the time series. If it is present, it is nonstationary and requires the data to be transformed using a method such as the logarithmic or differencing methods. This procedure is typically completed until the resulting data is stationary. Otherwise, the data is stationary, implying it does not need a transformation. By performing this test, the first research question focusing on the stationarity behavior of the data can be answered.

Furthermore, the correlations between the three user-generated variables will be calculated. These values will be examined to determine how strongly correlated the “Total Visitors” and “Total Pizzas Sold” variables are with the “Total Sales” variable. If the variables are too strongly correlated, including them with the total sales variable might not be necessary since strong correlations can negatively impact models.

After determining whether the data needs any transformations and what variables to utilize in the models, the project will construct ARIMA and Simple ANN models to compare their sales forecasting accuracy. Both models will utilize the accumulated dataset, which includes the total sales for each day in 2015. This project will utilize a test-train split on the dataset. The first eleven months of data will be used as the training data for the models, while the last month will be used as the test data. Furthermore, this project will utilize the Grid Search method to find the optimal parameters for each model. Specifically, the Grid Search procedure will determine the optimal order and seasonal order parameters for the model by using a combination of 0, 1, 2, 3, or 4 for the choice of p, d, and q in the tuple (p, d, q) for order, and choosing the combination that produces the lowest AIC score. This choice of values for order will become the selection choice used to construct the ARIMA model for the forecasting model to be compared with. On the other hand, the Grid Search procedure on the Simple ANN model will determine the optimal hyperparameter values for the batch size, number of epochs, and optimizer by selecting the combination of values that produces the model with the smallest mean absolute error value. The model with the smallest mean absolute error value will be used to create the Simple ANN forecasting model.

At this point, the optimal SARIMA and Simple ANN models can be created and compared. These models will be evaluated by comparing the root mean square error value of each model’s performance on the test data. The choice to use the root mean square error as the evaluation metric for comparison rests in the fact that the root mean square error (RMSE) is one of the more popular error measures in usage in various domains (Shcherbakov et al., 2013). The only downside is that this error measurement can be problematic when scale dependency issues exist within the dataset or a high influence of outliers (Shcherbakov et al., 2013). Thankfully, this is not an issue, which will be discussed with the descriptive statistics analysis in the paper's results section. The decision criteria to determine the effectiveness of the forecasting model will be to choose the model with the smaller test RMSE score, which is calculated by the following formula:

(1)

where is the actual sales data and is the predicted value. Hence, the optimal model will be determined by the smaller RMSE score.

## **Limitations**

The chosen dataset's biggest limitation is its small sample size for the time-dependent observations. The dataset contains information about sales, the number of visitors, and the number of pizzas sold for a year, so there is insufficient training data to determine whether the temporal data is truly stationary. The analysis would benefit from additional time-dependent observations. The minimum amount of collected data should represent at least two years of observations for the training data since it will indicate whether sales are stationary. However, this will not invalidate the study results since the findings are still meaningful. Instead, this limitation can be easily fixed in another study or even by business users in the real world when they attempt to build a forecasting model.

## **Ethical Considerations**

Although the dataset chosen for this project is already anonymized, ethical considerations for this data must be considered since business users will not likely be subjected to anonymized data in a business setting. An organization trying to set up a similar data analytics project for sales forecasts will draw on its own resources, which will be impacted by how the organization currently handles its incoming data. To prepare the data for handling, the business users must ensure that the data remains confidential and anonymous by removing any personal identifier associated with sales tracking. Examples of information that should be removed from the data are personal credit card numbers, identifiers that can associate orders with customers, and so forth. These pieces of information do not add any relevance to this type of project at hand, so they should be removed. Furthermore, this information should remain anonymous and confidential since the customer was not informed of their data being used in this manner. This will eliminate the chance for the customer to experience any potential harm because of data from this model being leaked with confidential information.

# **Results and Discussion**

This study’s primary goal is to evaluate the sales forecasting performance of an ARIMA time series statistical model and a Simple ANN deep learning model. The purpose is to provide a starting point for discussing effective sales forecasting models in the restaurant industry since more accurate predictions allow these organizations to determine more realistic strategic strategies to manage the organization’s resources effectively. Suggesting the two different methods to forecast sales is important since it allows for discussing how the two types of methodological approaches, statistical and deep learning, differ. However, the starting point of this next discussion will need to focus on the results generated from descriptive statistics and the time series plots of the data since it is vital to understand whether the data exhibits stationary behavior, which will be covered in the following section. Afterward, the results of the performance difference in the sales forecasts of the optimal ARIMA and Simple ANN models will be discussed, along with some additional remarks on the forecasting performance of the models.

## **Descriptive Statistics and Time Series Plots: Analysis**

Figure 1 displays the summary statistics for the user-generated variables. According to Figure 1, each variable contains 365 observations corresponding to each day in 2015. The sales per day (average sales) are $2284.52, with a standard deviation of $402.02. The orders per day (average orders) are 60 orders, with a standard deviation of 9.95 orders. Lastly, the number of pizzas sold per day (average number of pizzas sold) is 138 pizzas, with a standard deviation of 24 pizzas. Notably, these variables' averages do not deviate far from the median (50% quantile), indicating that each variable shouldn’t have much skewness in their data distributions and can likely follow the normal curve. This is the first indicator that the variables might be stationary, but it will be better by analyzing the histograms.

**Figure 1**

*Descriptive Statistics Summary*

A table with numbers and a few black text

Description automatically generated

The histograms in Figure 2 show the data distribution for each variable. The histograms for “Total Sales” and “Total Pizzas Sold” show distributions that should have a small skewness value since the distributions show a slight skewness of the data, indicating that the data is not normally distributed. The slight skewness of the data for these two variables does not necessarily imply the data is nonstationary, but it should be checked with more visualizations. The “Total Orders” histogram shows normally distributed data, indicating that this data should be stationary. However, these results must be supplemented with the analysis of time series plots to determine the stationarity behavior of the data further.

**Figure 2**

*Histograms for “Total Sales”, “Total Orders”, and “Total Pizzas Sold”*

A graph of a number of red columns

Description automatically generated with medium confidence

Figures 3 – 5 display time series plots for each variable. These time series plots were constructed using the rolling mean and rolling standard deviation for a seven-day window, along with the time series plot for the actual observed values. Each plot does not show any trend or seasonality among the data. There are common high and low peaks that the variables should experience. The plot has no discernible patterns except that the end of November and December experienced a sharp decline in observed values. However, these sharp declines do not indicate a trend or seasonality.

**Figure 3**

*Time Series Plot for “Total Sales”*

A graph showing sales and statistics

Description automatically generated with medium confidence

**Figure 4**

*Time Series Plot for “Total Visitors”*

A graph showing a number of numbers

Description automatically generated with medium confidence

**Figure 5**

*Time Series Plot for “Total Pizzas Sold”*

A graph showing a line of pizzas

Description automatically generated

Furthermore, each variable's rolling mean and standard deviation do not display any weird behavior or pattern. They appear normal since the values appear to balance each other out. Thus, it would be safe to say that each variable displays stationary behavior, which would answer the first research question. Still, it must be confirmed with a standard statistical test to determine quantitively whether the series exhibits stationarity.

However, before discussing the statistical test results to determine stationarity, it would be important to discuss any correlations among the variables. Figure 6 displays a scatter plot matrix. According to Figure 6, the “Total Orders” and “Total Pizzas Sold” variables are correlated with the “Total Sales” variable. “Total Pizzas Sold” is very strongly correlated with “Total Sales” while “Total Orders” has more of a moderate correlation with “Total Sales”. This should not be shocking since these variables directly relate to “Total Sales” (e.g., sales increase with the number of pizzas sold or orders received in a day). Hence, “Total Orders” and “Total Pizzas Sold” will not be used during the rest of this analysis. Instead, the rest of this paper will analyze “Total Sales” to determine whether the data is stationary and whether an ARIMA or a Simple ANN model forecasts better.

**Figure 6**

*Scatter Plot Matrix*

A collage of blue and white graphs

Description automatically generated with medium confidence

## **RQ1: Stationary Results**

Now that the descriptive statistics and time series plots of the dataset have been discussed, it is time to turn the attention to whether the data is nonstationary or stationary. According to the analysis in the previous section, it is reasonable to assume the data is stationary. However, as mentioned in the Methods section, the Augmented Dickey-Fuller Test is a common statistical test applied to time series data to determine whether the data is nonstationary (the null hypothesis assumes the data is nonstationary). The results shown in Figure 7 indicate the time series data is stationary since the ADF statistic is less than the critical value at the 5% significance level, meaning that the null hypothesis can be rejected. Hence, the time series data is stationary, implying no need for transformation methods such as differencing to be applied to the dataset before generating a time series statistical model for forecasting purposes.

**Figure 7**

*ADF Test Results*

A close-up of a number

Description automatically generated

## **RQ2: Forecasting Performance of ARIMA and Simple ANN Results**

This last section analyzes the results from evaluating the forecasting performance of the optimal ARIMA and Simple ANN models. As mentioned previously in the Methodology section, finding the optimal ARIMA and simple ANN models is the first step to answering the research question. First, the dataset was divided into a training and test set. The training set contained all the information from January 2015 to November 2015, while the testing set contained only the observations from December 2015. Afterward, the Grid Search optimization procedure was used to find the optimal models by searching through various hyperparameters and selecting the optimal values for the hyperparameters based on a loss function. Then, each model was constructed with the optimal hyperparameter values. These models were used to construct the sales forecasts and the RMSE for each model's training and testing partitions. The RMSE for the testing partitions was used as the evaluation criteria to determine which model was more effective.

Only one hyperparameter was optimized for the ARIMA model. As mentioned previously, the order hyperparameter, a tuple of three values, was optimized. Each value of the tuple could take on the value of 0, 1, 2, 3, or 4. According to the Grid Search procedure, the optimal value for the hyperparameter order is (4, 0, 4) since it produced the model with the lowest AIC value of 4804.633. Thus, the ARIMA model was created using (4, 0, 4) as the order.

Figure 8 shows the summary statistics for the ARIMA statistical model. These results show the performance of the statistical model created using the training data. According to Figure 8, each statistical term (const, ar.L1, ar.L2, ar.L3, ar.L4, ma.L1, ma.L2, ma.L3, ma.L4, and sigma2) is statistically significant since the p-value, 0.000 for each term, does not exceed the significance level, 0.05. Furthermore, the Ljung-Box p-value statistic (0.83) is greater than the significance level, implying that the model is a good fit since the residuals from the model are white noise (Kumar, 2022). However, the other two statistics shown in the bottom section of the summary results in Figure 8 indicate that the model might be problematic for forecasting purposes. The heteroskedasticity test indicates variance within the residuals since its p-value is significant. This is problematic when using the model for forecasting since it can produce inaccurate forecasts. (Cerqueira, 2022) Also, the p-value for the Jarque-Bera test is significant, indicating that the residuals do not pass the normality assumption, implying that this model’s forecasts can be inaccurate. Regardless, this model will be used as the optimal model for the ARIMA model.

**Figure 8**

*Summary Statistics for ARIMA Model of Order (4,0,4)*

A screenshot of a document

Description automatically generated

Unlike the ARIMA model, the Simple ANN had four optimized hyperparameters: optimizer, activation, epochs, and batch size. Each hyperparameter was given three values to be tested, creating 81 potential models. The optimizer hyperparameter had the options of “adam”, “sgd”, and “adamax”. The choices for the activation hyperparameter were “relu,” “sigmoid,” and “tanh.” The choices for the number of epochs were 25, 50, and 100, while the choices were 1, 5, and 10 for the batch size. The Grid Search procedure determined that the optimal value set is {adam, relu, 100, 1} since it had the smallest mean squared error value of $392.29. Thus, the Simple ANN model was constructed using the optimal value set {adam, relu, 100, 1}.

Since creating a Simple ANN model does not produce a statistical summary for the model's performance like the ARIMA model, the rest of this section will now focus on analyzing the forecast performance of the optimal models. Table 2 shows each model's resulting training and testing RMSE scores and the percent difference in the forecasting performance. According to the table, the ARIMA model has the smaller training RMSE score, indicating this model worked better at predicting the correct values used in the training set. However, the model’s test RMSE score is $524.48, a significant increase. This indicates that the ARIMA model has overfitting issues. The Simple ANN model’s training RMSE score is $396.75. Although it is bigger than the ARIMA’s training RMSE score, it is not bad. On the other hand, the test RMSE score ($421.60) for the Simple ANN model is significantly better since it is 24.4% less than the ARIMA’s model score. Hence, the Simple ANN model produced a more accurate sales forecast for December 2015, so the null hypothesis that the forecast accuracies of the model are not significantly different can be rejected.

**Table 2**

*Training and Testing Set RMSE Scores – ARIMA and Simple ANN Model*

|  |  |  |
| --- | --- | --- |
| Model | Training RMSE Score | Test RMSE Score |
| ARIMA | $337.39 | $524.48 |
| Simple ANN | $396.75 | $421.60 |
| Percent Difference | +14.96% | -24.40% |

However, it is important to note that even though the Simple ANN model performed better, the time series plots reveal some interesting things about the forecasts. Figure 10 displays a time series plot of the actual sales (blue line), ARIMA predictions (red line), and the Simple ANN predictions (green line) for each day in December. It is apparent from the visualization that the ARIMA model is better at modeling since its plot has peaks and valleys in sales like the plot for the actuals. The Simple ANN model produces a plot resembling a straight line, indicating slight variation in the predicted values, which would be problematic for restaurant operators.

**Figure 10**

*Time Series Plot of Predictions vs Actual Sales*

A graph showing a line of sales

Description automatically generated with medium confidence

Restaurant operators use sales forecasts daily to manage their operations. They make daily decisions based on these forecasts to manage costs such as inventory and labor. According to Figure 10, the Simple ANN model would be problematic for restaurant operators since they would make roughly the same daily sales prediction for each day of December. They would produce schedules and order products based on those forecasts, leading them to make decisions that likely understaff the restaurant and/or have the restaurant running out of product. They would still have a problem if they utilized the ARIMA predictions according to Figure 10, but they would fare better at times with these forecasts because more of the predictions are closer to the actual sales. Regardless, the visualizations reveal information about how the models’ predictions compared with the actuals, and this information reveals some possible issues with the Simple ANN model that would alarm restaurant operators. As Table 3 shows, the Simple ANN model performs better when we examine the forecast performance for the whole month since its percent error +5.77% with respect to the actual sales, while the ARIMA model had a percent error of +7.39%, but the time series visualization of the forecast raises questions for future researchers.

**Table 3**

*Forecasting Performance for December 2015 – ARIMA and Simple ANN Model*

|  |  |  |
| --- | --- | --- |
| Model | Total Sales | Percent Error (w.r.t. Actual) |
| Actual | $66,985.68 | - |
| ARIMA | $71,936.26 | +7.39% |
| Simple ANN | $70,853.52 | +5.77% |

# **Conclusion and Recommendations for Further Study**

In conclusion, although sales forecasting will hardly ever be entirely accurate, forecasts with relatively low error can benefit the restaurant industry since the profit margin can be relatively small for various reasons. This study has shown that the dataset did not exhibit any seasonality or trend in the variables since it rejected the null hypothesis about the data being nonstationary for the first research question. Hence, the data did not have to undergo any procedures such as a logarithmic transformation or differencing to transform the data to become stationary. Furthermore, this study has shown that the Simple ANN produced a more accurate sales forecast for December 2015 than the ARIMA model utilizing this dataset, indicating that the null hypothesis for the second research question since the RMSE for the testing partition of the Simple ANN was significantly smaller than the ARIMA’s model score.

However, the visualizations of the predictions vs actual values suggest something else. Although the Simple ANN model produced a more accurate forecast for sales for December 2015, the ARIMA model produces a time series plot indicating that this model follows the true noise associated with sales better than the Simple ANN model. The time series plot for the Simple ANN model produced a plot that was almost a horizontal line, indicating little variation in the predictions. Unlike the ARIMA model, the Simple ANN model failed to capture the true noise associated with daily sales. Regardless, this does not invalidate the results of this study. Instead, it generates more questions for further research on this topic.

Based on the findings of this project, several recommendations can be made to further the research on sales forecasting in the restaurant industry. First, the dataset can be expanded upon by adding additional observations to extend the dataset to include data beyond 2015. Adding additional observations can allow the models to improve their forecast accuracy since they will better understand how seasonality and/or trends impact sales, ultimately creating a better model that can produce a more accurate forecast that models daily sales behavior. This type of model will be more effective for restaurant operators who are concerned with daily operations. Secondly, this study can be expanded upon by testing the performance of additional models. It is clear that not all models perform equally, so any additional research on the performance of other models will add to the discussion. Lastly, this study can be expanded upon by introducing new data sources such as demographics, weather patterns, competitors’ sales data, socioeconomic factors, social media, and data collected from newer technologies such as home delivery platforms. Hence, these additional data sources can add richness to the models and possibly help produce more accurate forecasts by adding variables that can be significant in the models, allowing restaurants to remain competitive by giving them a tool to help accurately manage incoming revenue streams along with costs.

# **References**

Cerqueira, V. (2022, December 19). *How to detect heteroskedasticity in time series.* Medium. https://towardsdatascience.com/how-to-detect-heteroskedasticity-in-time-series-3413a8aa8da9

Chaudhuri, S. (2023, October 25). *Assessment of accuracy metrics for time series forecasting.* Medium. https://medium.com/analytics-vidhya/assessment-of-accuracy-metrics-for-time-series-forecasting-bc115b655705

E., E., Yu, M., Tian, X., & Tao, Y. (2022). Dynamic model selection based on demand pattern classification in retail sales forecasting. *Mathematics*, *10*(17), https://doi.org/10.3390/math10173179

*Exploring the impact of inaccurate sales forecasts.* (n.d.). Sands Partners. Retrieved on June 16, 2024 from https://sandspartners.com/the-impact-of-inaccurate-sales-forecasts/

Fortsch, S. M., Choi, J. H., & Khapalova, E. A. (2022). Competition can help predict sales. *Journal of Forecasting, 41*(2), 331-344. https://doi.org/10.1002/for.2818

Hall, R. (2020, October 21). *Why forecasting is important for business success*. BAASS Business Solutions Inc. https://www.baass.com/blog/why-forecasting-is-important-for-business-success

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and practice*. OTexts.

Kumar, S. (2022, October 21). *Reading ARIMA results.* Medium. https://medium.com/@sachinnayak18/reading-arima-results-35eb1df7a565

Liu, Y., Lan, K., Huang, F., Cao, X., Feng, B., & Zhu, B. (2021). Aggregate store sales forecasting framework based on ConvLSTM. *ICCDA ’21: Proceedings of the 2021 5th International Conference on Compute and Data Analysis*, 67-72. https://doi.org/10.1145/3456529.3456540

Nielsen, A. (2020). *Practical time series analysis: Prediction with statistics & machine learning.* O’Reilly.

Professor Frenzel. (2023, September 30). *Time-Series analysis – Part 3 – Making a time-series stationary.* Medium. https://prof-frenzel.medium.com/kb-time-series-data-part-3-6e32032f7b49

Roy, D., Spiliotopoulou, E., & de Vries, J. (2022). Restaurant analytics: Emerging practice and research opportunities. *Production and Operations Management, 31*(10), 3687-3709. https://doi.org/10.1111/poms.13809

Ruiz, E. (2022, September 21). *Introducing the maven pizza challenge* [Dataset]. Maven Analytics. https://www.mavenanalytics.io/blog/maven-pizza-challenge

Shcherbakov, M. V., Brebels, A., Shcherbakova, N. L., Tyukov, A. P., Janovsky, T. A., & Kamaev, V. A. (2013). A survey of forecast error measures. *World Applied Sciences Journal, 24*(2013), 171-176. https://doi.org/10.5829/idosi.wasj.2013.24.itmies.80032

Terra, J. (2024, February 16). *Why Python is essential for data analysis and data science.* Simplilearn. https://www.simplilearn.com/why-python-is-essential-for-data-analysis-article

Zhuang, S. L. (2022, September 21). *Pizza restaurant sales* [Dataset]. Kaggle.com. https://www.kaggle.com/datasets/shilongzhuang/pizza-sales