Time Series Analysis: Forecasting Product Demand

Hutchinson, Megan   
Member *(of Group 8)*  
Humber College *(Machine Learning II)*Toronto, ON

Ismail, Anas  
Member *(of Group 8)*  
Humber College *(Machine Learning II)*Toronto, ON  
Kadam, Rutuja  
Member *(of Group 8)*  
Humber College *(Machine Learning II)*Toronto, ON

***Abstract*— The following paper focuses on accurately forecasting the demand of a series of products aimed to assist a manufacturing company obtain actionable insights. The Product demand dataset was cleaned and later used as a data source in tableau to visualize the data in a descriptive fashion. The four central warehouses that were analyzed were in Brampton, Oshawa, St john’s, Surrey. Over the course of a five year stretch between 2014 to 2019, the Brampton warehouse consistently displayed the highest order demand among the four central warehouses. In addition, a machine learning model was used by implementing a centered moving average (CMA) approach to handle the time series data. Ultimately, the forecast results displayed a downward trend in product demand for all four central warehouses in the year 2020.**

***Keywords—Forecast, Time Series Analysis, Product Demand, Visualizations, Centered moving Average (****key words****)***

# Introduction

The demand for products in various industries that rely on sales is an essential component of one’s business and can single-handedly dictate how much profit a business can make in a single fiscal year. Analyzing this information on a weekly, monthly, or even on a yearly basis can prove increasingly beneficial when thinking of the capacity to make strategic decisions to push a business forward.

However, product demand is far from linear. Businesses owners tend to notice that there are tons of ebbs and flows regarding their respective market and must prepare for anything to occur. A company may find the demand for its product to be higher depending on the time of year or if it gains traction through marketing or word of mouth over time. Therefore, being able to effectively predict and forecast the demand for a company’s products over time will be a highly sought after skill seeing that it may provide it with the capability to allocate resources better, cut costs on spending, or even increase monthly funding towards advertising.

In the case of this analysis, a manufacturing company has tons of historical data pertaining to the demand of a plethora of their products over the course of a six-year period between the years of 2014 to 2019. The product demand data has been gathered from four central warehouses that are in Brampton, Oshawa, St john’s, and Surrey. The raw Product Demand dataset being utilized throughout this analysis consists of five unique attributes along with over a million individual records. This data set will be cleaned to be in the best physical state to effectively conduct a descriptive analysis using visualizations in Tableau. Subsequently, a predictive analysis will be done through the implementation of a machine learning algorithm. Ultimately, the task at hand is to accurately forecast the product demand for the manufacturing company using the Python programming language in Jupyter notebook.

# Literature review

The analysis of time series data is quite common across many industries along with the prediction of future values using machine learning algorithms. In the hopes of getting a better understanding of real-life examples, scholarly articles will be reviewed to examine their approach as it pertains to predictive analytics techniques when working with time series data.

## Data Reduction and Cleaning

Due to the size of many datasets along with the likelihood of human error occurring while imputing values, data preprocessing and cleaning is a critical component in any data driven analysis. In a study conducted by Rubi et al (2022), they had data pertaining to economic market behaviour between the years of 2001 to 2018. Data Preprocessing was done to handle missing values by using the median value, along with other variables being normalized using a zero to one scale [1]. Similarly, data cleansing was also done in another study that was looking to forecast demand for various products for a time series dataset [2]. Multiple columns were reduced as their focus was to concentrate solely on demand and sales over the 21 weeks of data gathered for all the products they were interested in [2].

In an analysis on demand forecasting throughout e-commerce platforms, data preprocessing was done to eliminate variables that had no relationship with demand [3]. The dataset was then reduced to only relevant variables and all rows containing missing data were removed [3]. In a separate analysis that also aims to forecast demand, data preparation was executed to remove records that would fall outside their target dates for their time series analysis [4]. In this case, it provided them with unnecessary data that may cause the model to perform poorly when trying to predict demand over an upcoming six-week period [4].

## Machine Learning Algorithms

In the study conducted by Rubi et al (2022), their focus was to uncover valuable information to predict stock exchange prices over monthly time series data. Two main machine learning algorithms were used in their study to derive the desired results and they were the autoregressive integrated moving average as well as multi layer perceptron artificial neural networks approach [1].

On the other hand, a plethora of machine learning algorithms were used in another study that aimed to forecast the demand while working with Walmart’s retail data across a number of different departments [2]. Some of the algorithms used were simple moving average, Naïve Bayes, Prophet, and exponential smoothing [2]. Later on, they decided to attempt an ensemble learning method by incorporating linear regression, random forest, and weighted average to their analysis. In the study conducted by Jain et al (2020), seasonal autoregressive integrated moving average and long short-term memory network to predict product demand. The two models were tested separately, and the results were then compared to one another to examine which of the two was more accurate.

Furthermore, Rožanec et al (2021) actually considered twenty-one different machine learning algorithms and approaches in attempts to forecast the demand of products across European manufacturers within the global automotive industry. Ultimately, the decided to go with the mean squared error approach, support machine, and random forest out of the options they were considering [4]

## Key Findings

In the study conducted by Rubi et al (2022), they noticed that their forecast was more accurate when using the artificial neural network as opposed to the autoregressive integrated moving average. They also tested the mean absolute percentage error and it appeared that the estimated error of the artificial neural network was substantially lower than the autoregressive method [1]. In the future, Rubi et al (2022) mentioned that they would prefer to explore other algorithms to assist them in predicting demand in the long run.

In the time series analysis done by Zhang et al (2022), it appears that the ensemble model using linear regression fits the best in terms of predicting demand. However, the accuracy results displayed only differed by 3% in comparison to random forest and weighted moving average [2]. They believe that by combining predictive algorithms, one can better equip themselves in attaining more accurate predictions [2].

On the flip side, Jain et al (2020) the root mean squared error was used to compare the performance of the two machine learning algorithms they chose to predict the demand for e-commerce platforms. Their findings display that the autoregressive approach had an RMSE of 1.24 whereas their long short-term model had an RMSE of 1.55 [3]. Therefore, they concluded that the autoregressive approach would be better for them for forecasting demand down the line.

In the study conducted by Rožanec et al (2021), the most effective models for them were the support vector machine and the random forest with both of which obtaining near-human performance for the ratio of forecasts with an error below 30%. After the analysis, their goal is to develop effective error bounding strategies for demand forecasts and provide consistent explanations of what the data is explaining. All in all, each of the four scholarly articles reviewed implemented different machine learning algorithms and experienced different results as far as which appeared to be the most accurate.

# Methods

As mentioned earlier, the product demand dataset contained five variables along with over a million individual records. Prior to being able to derive any valuable information regarding the predicting the status of deliveries, the data set had to be analyzed, reduced, and cleaned. The goal was to ensure that the product demand data set contained no null values, outliers, and duplicate data prior to implementing the use of a machine learning algorithm or approach. A deep dive in product demand forecasting was done to build business acumen prior to conducting any data analysis techniques.

## Data Reduction and Cleaning

The time series data being used contained a mammoth 1,048,575 individual records. Null values were searched for to either remove or replace with the median value for the given record. There were a total of 11,239 missing values found in the date column. In addition, the date column also contains dates that are not on the calendar and can potentially be attributed to human error. For example, years that are not considered to be “leap” years contain dates of February 29th in both 2015 and 2019. There were a total of 1,750 records that contained incorrect dates. All these dates were removed, and the data set was reduced to 1,035,586.

## Data Visualization in Tableau

The preprocessed data set was then added as a source in Tableau to derive some descriptive statistics and display them in the form of visualizations. A sheet was created to assess the sum of order demand for each central warehouse in the form of a bar graph. Another visualization was created in the form of a line chart to display the data across different years.

Similarly, another visualization was created in the form of a bar chart comparing the top five products over the course of the six years. In addition, another bar chart was created to distinguish the top 5 products in each year throughout the entire data set.

## Forecasting Approach

After extensive research, the centered moving average approach was implemented to forecast product demand. The objective was to determine the total number of orders for each year and then use the trailing moving average to calculate what the total for 2020 would be. For this analysis, only the date and order demand variables were used in the Jupyter notebook. In addition, product demand was also forecasted for each warehouse. The forecast was done using a window of three and only the order demand, warehouse, and date columns were used. For both sets of forecasts, a table and line chart were created to visualize these results.

# Results

## Descriptive Analysis in Tableau

As mentioned in the methods, order demand was assessed from both a total sum perspective and individual yearly approach to get a better idea as to which warehouse was performing the best. Figure 1.0 highlights the large discrepancy between Brampton and the rest of the other warehouses as far as their order demand. Brampton has the highest order demand with over 3 million orders over the six year stretch whereas St john’s is the lowest with under 100 thousand orders over the same period. Similarly, when looking at the graphical representation in figure 1.1, Brampton outperforms all the other warehouses on a consistent basis. External factors such as population size within the respective cities were out of the scope of this analysis and could potentially be valid reasons as to why it is evident that Brampton is outperforming the other warehouses by a margin as large as there is in both figures 1.0 and 1.1.

Chart, bar chart, waterfall chart

Description automatically generated

**Figure 1.0:** This figure highlights the difference in order demand across the four warehouses over the course of a six-year period.

Chart, diagram, line chart

Description automatically generated

**Figure 1.1:** This figure highlights the discrepancy between the order demand across the different warehouses on a yearly basis. The colours for the warehouse are set up the same as figure 1.0.

Similarly, a descriptive analysis was also done on the products category within the data set to assess which of them had the largest order demand over the course of the six year stretch and each individual year as well. Figure 1.2 displays that the product category 019 has the highest order demand over the entire history of the data set amongst the top five product categories. In addition, figure 1.3 illustrates the top five product categories across each given year.

Table

Description automatically generated with low confidence

**Figure 1.2:** This figure displays the top five products over the course of the six-year period in which the data was collected regarding order demand.

Chart, bar chart

Description automatically generated**Figure 1.3:** This figure emphasises the difference in the top five performing product categories in a given year across all warehouses.

By analyzing the data in a descriptive fashion, it offers up the foundation to being forecasting the product demand using the centered moving average approach.

## Predictive Analysis in Jupyter Notebook

Chart, line chart

Description automatically generated

**Figure 2.0:** This figure highlights the order demand along with the forecast of it as well in the form of a line chart for the upcoming year.

Chart, line chart

Description automatically generated

**Figure 2.1:** This figure emphasizes the forecast of the product demand over the upcoming year across the four warehouses of focus.

Based on the machine learning algorithm used, there will be a decline in the total number of orders that is expected for 2020 as shown in figure 2.0. Similarly, when forecasting the product demand for each individual warehouse, there will also be a decline in the number of orders that each warehouse will have placed in the year 2020. The value obtained for the order demand using the moving average approach in the year 2020 was 8.776047e+0.8 whereas the year prior was 1.061653e+09. The line chart was used to provide a simpler visual representation of the expected decline in 2020.

# Conclusion

In conclusion, the forecast of the product demand deemed that the volume is scheduled to decrease subsequent to using the central and trailing moving average approach. In addition, the descriptive analysis has provided the information that the Brampton warehouse has the largest volume of order demand on a consistent basis for the manufacturing company. It may prove wise to begin implementing techniques that are working well in the Brampton warehouse across the other sites in order for the overall order demand to increase company wide, albeit there may be additional factors outside the scope of this analysis contributing to those results. In hindsight, the predictive analysis conducted using the moving average approach may have limited the capacity to elicit highly accurate results as opposed to other machine learning algorithms. The moving average model is often used in time series forecasting, however, the problem with it is its dependence on historical data. Regunath (2023) makes the argument that the autoregressive approach is the best when trying to forecast demand in time series data. In addition, the same approach was also used throughout a couple of the scholarly articles that were examined in the literature. Therefore, although the moving average approach is widely used in time series analyses when forecasting demand, the autoregressive approach may prove to be substantially more accurate than the centered and trailing moving average methods.

##### References

1. Rubi, M. A., Chowdhury, S., Rahman, A. A. A., Meero, A., Zayed, N. M., & Islam, K. A. (2022). Fitting Multi-Layer Feed Forward Neural Network and Autoregressive Integrated Moving Average for Dhaka Stock Exchange Price Predicting. *Emerging Science Journal*, *6*(5), 1046-1061.
2. Zhang, Y., Zhu, H., Wang, Y., & Li, T. (2022). Demand Forecasting: From Machine Learning to Ensemble Learning. In *2022 IEEE Conference on Telecommunications, Optics and Computer Science (TOCS)* (pp. 461-466). IEEE.
3. Jain, A., Karthikeyan, V., Sahana, B., Shambhavi, B. R., Sindhu, K., & Balaji, S. (2020). Demand forecasting for e-commerce platforms. In *2020 IEEE International Conference for Innovation in Technology (INOCON)* (pp. 1-4). IEEE.
4. Rožanec, J. M., Kažič, B., Škrjanc, M., Fortuna, B., & Mladenić, D. (2021). Automotive OEM demand forecasting: A comparative study of forecasting algorithms and strategies. *Applied Sciences*, *11*(15), 6787.
5. Regunath, G. (2023). *10 incredibly useful time series forecasting algorithms*. Advancing Analytics.