International Journal of Frontier Technology and Business
ISSN Online: 2583-5939

OPTIMIZATION OF DEMAND FORECASTING & INVENTORY MANAGEMENT USING SARIMAX

Mr. Manan Doshi[1]

(Student, Vivekanand Business School, Mumbai)

Ms. Priya Pandey[2]

(Student, Vivekanand Business School, Mumbai)

Prof. Nikita Ramrakhiani[3]

(Asst. Professor, Vivekanand Business School, Mumbai)

Abstract

Nowadays, supply chain management has become a key element in every business. Each & every business is focusing on how to manage the logistics of a warehouse. The purpose of this project is to learn about inventory management & how to manage a warehouse portfolio. The SARIMAX Model helps in managing the inventory efficiently. The demand for a product is predicted using the Time Series Analysis Model.

Keywords: *SARIMAX, Time Series Analysis, Forecasting, Inventory Management*

Introduction:

In the year 1990, E-Commerce became a trend in the later period of the year 1990s. Information Technologies like WWW (World Wide Web), the Internet, etc. showed rapid change & created a positive impact on the social & economic environment. A lot of time was taken by various analysts, scholars & journalists to reflect upon the transformation & breakthrough so that it could be further situated in a wide & historical context. Machine learning is a field of computer science that uses algorithms to learn patterns in data and make predictions or decisions without explicit programming. It involves training a model on a dataset, allowing it to make predictions or decisions, and then refining the model based on the results. It is the method through which the computer learns the behaviour from past data and also analyses the data which further helps them to visualize future predictions.

2. Research Gap:

1. Time series uses 2 concepts: Statistics & Machine Learning. Knowledge is immensely required regarding it and only then we can achieve time series analysis.

2. In the context of time series prediction, the presence of white noise can impact the ability to establish a correlation between past data and present data. White noise is defined as a series of random numbers with a mean of 0 and a standard deviation of 1, making it difficult to predict. To overcome this challenge and improve the accuracy of time series predictions, a comprehensive understanding of statistical concepts and the differencing method in time series modeling is necessary to remove the white noise from the data.

3. When correlation is established in time series data, it is subject to limitations and assumptions. For example, in the case of sales data utilized in a project, the analysis may not account for external events such as economic recessions, natural disasters, or stock market trends, which can significantly impact sales. This highlights the importance of considering all relevant factors and potential sources of confounding when utilizing time series analysis for decision-making.

4. Data being stationary is another problem in the Time Series Analysis Model. To check whether the data is stationary or not, we have to perform the adifuller test. AdiFuller test is basically hypothesis testing where we have to find the p-value. If the p-value is greater than 0.05 then, the data is not stationary & if it is less than 0.05 then, the data is stationary. In the case of non-stationary data, first, we need to perform the moving average model concept. This model works on the pre-defined value.

5. Most time series models do not take into account the PESTLE factors *(Political, Economic, Social, Technological, Legal, and Environmental)* during their development.

6. Moving Average takes values from pre-defined values but moving average does not know what the pre-defined values must be. It does not know whether the first 2 values or the first 3 values are to be taken. The values that are under differencing do not allow to eliminate the white noise.

3. Literature Review:

This study's goal is: -

- To analyse time series forecasting techniques and provide a brief explanation of how they operate.
- We go through time series, methodologies for time series forecasting, benefits, and drawbacks of time series forecasting.
- We also go through the methodologies and uses of various time series forecasting techniques.
- The goal is to educate and disseminate information regarding time series forecasting and its associated methods.

The objectives of the research paper are:

1. To get an in-depth knowledge of Machine Learning & Statistics.
2. To achieve the sales forecast for different warehouses.
3. Inventory Management for different products in different warehouses.
4. To perform data stationarity.
5. To learn about SARIMAX Model.
6. To explore the concept of itertools.
7. To predict Inventory after 3 months.

4. Inventory Management & Sales Forecasting System:

In the sales data exploratory analysis, Power BI was employed as the data analysis software.

The sales data was pre-processed to eliminate negative values.

Data visualization through Power BI was conducted to uncover patterns and anomalies in the data.

The visualizations showed the presence of zero values in the data for the years 2011 and 2017, which could impact the analysis results.

Thus, a python code was executed to exclude this data from the analysis, ensuring the validity and reliability of the results in the exploratory analysis.

The below diagram shows the output of our steps,

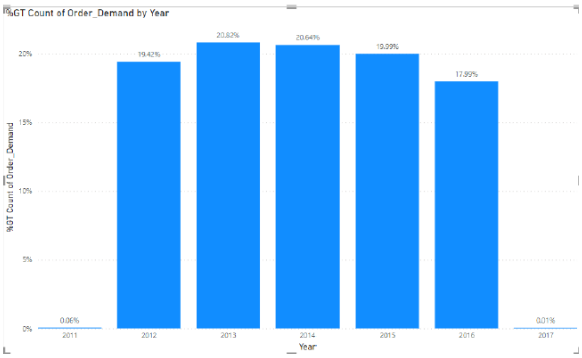

Fig: *Order_Demand by year*

The sales data was analysed to determine the distribution among warehouses using Power BI. This analysis yielded the following insights: -

i. Warehouse J has the highest sales over the 4 years followed by A, S & C. Warehouse J had the highest product category count of 3,69,078 and was 2,062.52% higher than Warehouse C having the lowest product category count of 17,067.

Further, after some analysis on Warehouse J and we found out that 5 product categories had the highest sales throughout the year. They are as follows: -

Product Category 019 had contributed 85% in Sales followed by Product Categories 005, 006, and 007.

Since Product Category 019 was the highest, so we have done a trend analysis over the years & following results were derived: -

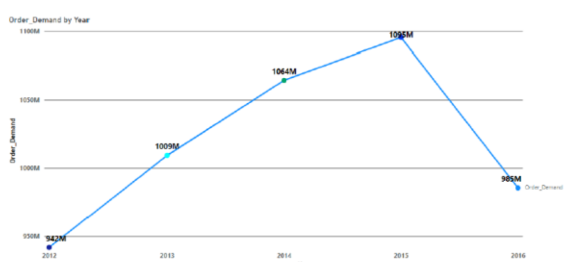

Fig: *Trend Analysis per year*

Order_Demand trended up, resulting in a 4.63% increase between 2012 and 2016.

Order_Demand started trending up on 2012, rising by 4.63% (43625222) in 4 years. Order_Demand jumped from 941742907 to 985368129 during its steepest incline between 2012 and 2016.

Outliers and missing values were addressed in the data preprocessing stage. Missing values were filled using the backward fill (bfill()) function and upper limit values (23500) were substituted for any upward outliers as the starting range of the data was from 1000 and no downward outliers were detected. This resulted in a skewed distribution as evidenced in the accompanying graphical representation,

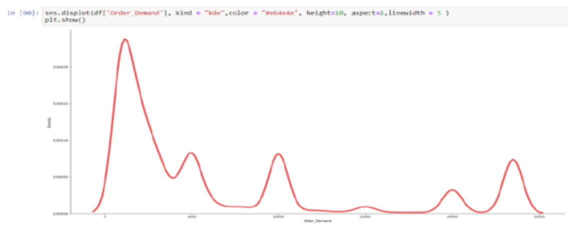

Fig: *Distplot using Python*

The data were resampled to provide a date-wise representation of the order demand for Category_019. The data was aggregated on a monthly basis, and a time series analysis was performed. As a result of the monthly resampling, 60 entries were obtained in the data frame df19.

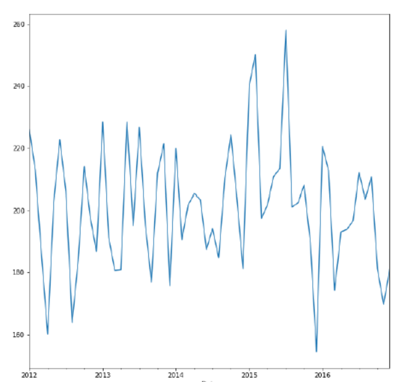

Fig: *bfill plot*

The Time Series data was decomposed. Decomposition of the time series signal leads to 3 things: Trend, Seasonal Pattern & White Noise. If the seasonal pattern is present, we are sure that we need to use the SARIMAX Model.

ACF: Auto Correlation Factor.

ACF stands for Autocorrelation Function. It is a statistical tool used to evaluate the presence of serial dependence in time series data. ACF measures the correlation between a time series and its lagged values

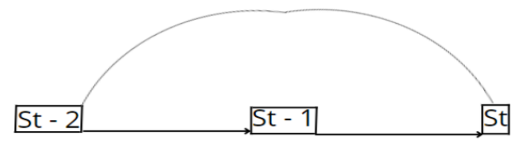

Fig: *Pictorial Representation of ACF*

In Time Series Modelling, ACF is used to determine the order of differencing needed to make the time series stationary, which is a key requirement for building a stable and robust model. A time series is considered stationary when its mean, variance, and auto covariance are constant over time.
For example, in a time series data of monthly sales of a retail store, the ACF plot can be used to identify the presence of any seasonality patterns. If the ACF plot shows a sharp drop-off at lag 12, it can be inferred that the time series data is influenced by a yearly seasonality pattern. The knowledge of this seasonality pattern can then be used to build an appropriate Time Series Model, such as SARIMA, to capture the underlying trends and patterns in the data.

PACF: Partial Correlation Factor.

PACF stands for Partial Autocorrelation Function. It measures the correlation between a series and its lags after controlling for the effects of intermediate lags. In other words, PACF shows the direct relationship between a series and its lag, excluding the influence of other intermediate lags.

$$\left[S_t = \Phi_{21} S_{t-1} + \Phi_{22} S_{t-2} + E_t \right]$$

Fig: *Mathematical formula of PACF*

PACF helps in identifying the appropriate order of AR (Auto-Regressive) terms for a Time Series Model. AR terms in a Time Series Model are lags of the series. A high PACF value for a lag suggests that the series is auto correlated with that lag, and including that lag in the AR model would help improve the model's accuracy.

For example, in a Time Series Model of sales data, a high PACF value for lag 12 would suggest that the sales in the current month are directly influenced by the sales from 12 months ago, and including that lag in the AR, model would help capture the seasonal patterns in the sales data.

Using itertools, we have found values for p, d, and q and have used SARIMAX Model for forecasting.

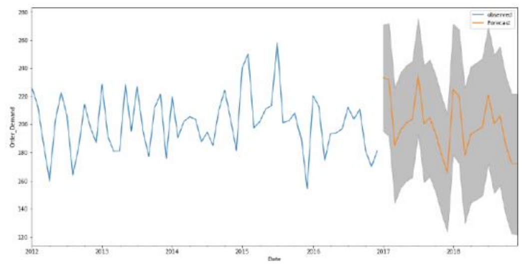

Fig: *Order_demand forecast*

	order_demand	refill_order	refill_list	order	balanced
01-01-2017	233.151914	279.782297	279.7823	233.1519	46.63038
01-02-2017	231.754245	231.474712	278.10509	231.7542	46.35085
01-03-2017	184.949491	175.58854	221.93939	184.9495	36.9899
01-04-2017	196.004916	198.216001	235.2059	196.0049	39.20098

Fig: *Inventory Management Forecast*

The above figure depicts the prediction of the product category forecast with an error rate represented by the Mean Absolute Error (MAE), which is **22%.** Thus, the accuracy of the model can be estimated at **78%.**

Further logic implementation in inventory management involves setting a safety stock value equal to 20% of the predicted stock value.

To ensure optimal inventory management, a safety stock value equal to 20% of the predicted stock level was established.

In instances where the current stock level was equal to 20% of the predicted stock, the difference between 20% and the predicted stock was calculated and a replenishment order was placed for the entire 20% predicted stock. The model was then deployed using the Streamlit framework.

Conclusion:

In the project, a 6-month inventory forecasting was performed using the SARIMAX (Seasonal Auto-Regressive Integrated Moving Average with Exogenous Factors) model.

The model provides a basis for inventory management by offering insight into expected demand patterns among different warehouse categories.

For example, if Warehouse J Category is predicted to have higher demand, and similar demand patterns are observed in other warehouses, the model can inform future inventory production decisions, indicating which warehouse category is expected to produce more inventory.

The model facilitates in mitigating overstocking of inventory in the storage facilities by accurately predicting inventory demand.

A predictive model for inventory management utilizes data analysis to identify the quantity of a product that should be kept in stock and determine which products have high demand, requiring a higher level of inventory. The model uses statistical techniques to forecast future inventory levels.

References:

1.https://www.researchgate.net/publication/335012451_An_integrated_feature_learning_approach_using_deep_learning_for_travel_time_prediction

2.https://www.sciencedirect.com/science/article/abs/pii/S0169207000000662

3.https://robjhyndman.com/papers/forecompijf.pdf

4.https://arxiv.org/abs/1710.03222

5.https://asu.pure.elsevier.com/en/publications/a-bag-of-features-framework-to-classify-time-series

6.https://www.wessa.net/download/stl.pdf

7.https://repository.upenn.edu/cgi/viewcontent.cgi?article=1074&context=marketing_papers

8.https://royalsocietypublishing.org/doi/10.1098/rsif.2013.0048

9. https://towardsdatascience.com/time-series-forecast-in-python-using-sarimax-and-prophet-c970e6056b5b

10.https://analyticsindiamag.com/complete-guide-to-sarimax-in-python-for-time-series-modeling/

11.https://medium.datadriveninvestor.com/time-series-prediction-using-sarimax-a6604f258c56

12. https://medium.com/mlearning-ai/time-series-forecasting-using-sarimax-ca98dd7238a0

13. https://neptune.ai/blog/arima-sarima-real-world-time-series-forecasting-guide#:~:text=Since%20SARIMA%20brings%20in%20seasonality,complex%20data%20spaces%20containing%20cycles.

14.https://www.fiverr.com/hamzaasif24/do-time-series-analysis-forecasting-using-sarimax-lstm-gru-in-python

15.https://www.jetir.org/papers/JETIREJ06034.pdf