ISSN: 2455-2631

Process of Sales Prediction using ARIMA and SARIMAX to Forecast Future Sales

[1]V. Rupika Priyatham,[2] Prof. R. J. Ramasree

[1] P.G Student M.Sc. (CS),
[2] Professor, Dept. Of Computer Science,
National Sanskrit University
Tirupati, A.P, India

Abstract: Sales forecasting is an essential component of business planning and decision-making. Time series analysis is a powerful statistical method for identifying patterns in historical data and forecasting future values. The use of Autoregressive Integrated Moving Averages (ARIMA) and Seasonal Autoregressive Integrated Moving Averages with exogenous variables) (SARIMAX) models for sales prediction is the focus of this paper. The paper describes and provides a comprehensive guide to the general methodology for developing ARIMA and SARIMAX models for sales forecasting. Data visualization, making the data stationary, plotting correlation and autocorrelation charts, building the model, and making predictions are all part of the process. This paper presents a step-by-step methodology for developing ARIMA and SARIMAX models for sales forecasting, as well as the steps involved in using a machine learning model, specifically ARIMA, to forecast sales, The findings of this research paper can assist businesses in making more accurate sales forecasts and improving their decision-making processes.

Keywords: Sales, Prediction, ARIMA, SARIMAX, Forecasting, Time Series, Analysis, Accuracy

1. Introduction

Sales forecasting is an important aspect of business planning and decision-making. Time series analysis is a powerful statistical method for identifying patterns in historical data and making predictions about future values. Time series data is a type of data that is collected at regular intervals over time. It is an important and widely used data type in various fields such as economics, finance, business, engineering, and environmental science. Time series data can provide valuable insights into past trends, present conditions, and future projections. Some of the key important uses of time series data are Trend analysis, Forecasting, Monitoring and control, Decision-making, Anomaly detection etc. Overall, time series data can provide valuable insights that can help in decision-making, forecasting, and problem-solving across a wide range of fields.

1.1 ARIMA (Autoregressive Integrated Moving Averages):

ARIMA, which stands for Autoregressive Integrated Moving Averages a statistical model that, can be used for forecasting time series data. ARIMA is a univariate model that is generally used to analyze and forecast stationary time series data. ARIMA models make predictions about future values based on past values and errors. The model consists of three parts:

- The “AR” autoregressive refers to the use of past values to predict future values of the time series.
- The “MA” moving average part represents the past errors to predict future values of the forecast.
- The “I” integrated part represents the number of differences needed to make the time series stationary.

Importance and use of time series data with ARIMA

1.1.1 Importance of ARIMA

Time series data refers to a sequence of observations collected over time. Time series analysis is used to identify patterns and trends in the data to make predictions about future values. ARIMA (Autoregressive Integrated Moving Average) is a popular technique used for modeling and forecasting time series data.The importance of time series data with ARIMA lies in its ability to make accurate predictions based on historical patterns and trends. ARIMA models can capture both the short-term and long-term trends in the data, which can be used to forecast future values with a high degree of accuracy.ARIMA models are used in a variety of industries, including finance, economics, and marketing, to forecast future values of stock prices, GDP growth, and consumer demand,among others. For example, a finance company may use ARIMA models to predict future stock prices based on historical data. Similarly, a marketing company may use ARIMA models to forecast future sales of a product based on past sales data.ARIMA models can also be used for anomaly detection and outlier detection, which is useful in identifying unexpected trends or events that can impact the data. For example, a sudden increase or decrease in sales may indicate a change in consumer behavior, which can be analyzed using ARIMA models to identify the cause and make appropriate adjustments. Time series data with ARIMA is important because it allows for accurate forecasting, anomaly detection, and outlier detection, which are crucial in many industries for making informed decisions.

1.1.2 ARIMA Model Algorithm

The ARIMA (Auto Regressive Integrated Moving Average) model is a popular time series forecasting method that involves the following algorithmic steps:

➢ **Data Preparation:** The first step is to prepare the time series data for analysis. This involves checking for missing values, outliers, and any other anomalies that may affect the accuracy of the model.

- **Stationary Testing:** ARIMA requires the time series to be stationary, which means that the statistical properties of the series remain constant over time. This can be checked using statistical tests such as the Augmented Dickey-Fuller (ADF) test.
- **Differencing:** If the time series is not stationary, differencing can be applied to transform the series into a stationary one. This involves taking the first difference of the series (subtracting each value from its preceding value) or applying higher-order differences as needed.
- **Parameters Identification:** Once the series is stationary, the next step is to identify the parameters of the ARIMA model. This involves identifying the order of autoregressive (p), integrated (d), and moving average (q) terms in the model.
- **Model Fitting:** With the parameters identified, the ARIMA model can be fitted to the data using maximum likelihood estimation or another optimization technique.
- **Model Diagnostics:** After fitting the model, it is important to check its validity and accuracy. This involves checking the residuals for autocorrelation and randomness, and comparing the model's forecasted values to actual values.
- **Forecasting:** Finally, the ARIMA model can be used to forecast future values of the time series based on the identified parameters and the historical data.

1.2 SARIMAX (Seasonal Auto Regressive Integrated Moving Average with eXogenous regressors)

SARIMAX which stands for Seasonal Auto Regressive Integrated Moving Average with exogenous regressors is a statistical model used for time series forecasting. It is an extension of the ARIMA model, which is used to model non-seasonal time series data. SARIMAX adds the ability to model seasonal patterns and also incorporates exogenous (external) variables that can influence the time series being forecasted.

The SARIMAX model has several components that are used to capture the different aspects of the time series data:

- The "SAR" term refers to seasonal autoregressive respectively.
- The "SMA" term represents seasonal moving average patterns in the data.
- The "I" term represents the number of times the data needs to be differenced in order to make it stationary.
- The "S" term represents the seasonal period
- The "X" term represents the exogenous variables.

1.2.1 Importance of SARIMAX

Time series data is important because it is used to model and analyze data that is ordered chronologically, such as stock prices, weather patterns, or website traffic. Time series data can reveal patterns and trends over time, as well as help to forecast future values.SARIMAX, or Seasonal Autoregressive Integrated Moving Average with eXogenous variables, is a statistical modeling technique used to analyze and forecast time series data. It is a variation of the more general ARIMA (Autoregressive Integrated Moving Average) model that can handle seasonal data and external regressors.SARIMAX models are useful in many applications such as finance, economics, and meteorology. For example, they can be used to predict stock prices, analyze economic indicators, and forecast weather patterns. In addition, SARIMAX models can be used for anomaly detection, where they can identify unusual events in the time series data.The use of SARIMAX models requires knowledge of statistical techniques and programming skills. However, once a model is built, it can be used to make predictions and analyze the data automatically. This can help businesses and organizations make better decisions and improve their performance by providing insights into trends and patterns in the data.Seasonal Autoregressive Integrated Moving Averages (SARIMAX) is a extension of ARIMA that can handle seasonal time series data and includes exogenous these two models are the popular time series models for sales forecasting.

1.2.2 SARIMAX Model Algorithm

The SARIMAX (Seasonal Autoregressive Integrated Moving Average with Exogenous Variables) model is a popular time series forecasting model that can be used to analyze and forecast time series data with seasonal patterns and other exogenous variables. Here are the basic steps involved in building a SARIMAX model:

- **Data Preparation:** First, need to prepare time series data for analysis. This involves checking for missing values, transforming non-stationary data into stationary data, and identifying any seasonal patterns in the data.
- **Model Identification:** The next step is to identify the appropriate SARIMA (Seasonal Autoregressive Integrated Moving Average) model for data. This involves selecting the order of autoregressive (AR), integrated (I), and moving average (MA) components of the model, as well as the order of seasonal AR, seasonal MA, and seasonal differences.
- **Parameter Estimation:** Once the appropriate SARIMA model is identified, estimate the model parameters using maximum likelihood estimation. This involves using an algorithm to find the values of the model parameters that maximize the likelihood of observing the observed data.
- **Model Selection:** After estimating the parameters of the model, evaluate the goodness of fit of the model to the data. This involves calculating various diagnostic statistics, such as the Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC), to determine whether the model fits the data well.
- **Forecasting:** Once the appropriate SARIMA model and estimated the parameters are selected, use the model to make forecasts for future values of the time series. This involves using the model to predict the values of the dependent variable at each point in time, as well as estimating the confidence intervals around the forecast values.
- **Model Evaluation:** Finally, evaluate the accuracy of the SARIMAX model by comparing the forecasted values to the actual values of the time series. This involves using various statistical measures, such as the mean absolute error (MAE), mean squared error (MSE), and root mean squared error (RMSE), to assess the accuracy of the forecasts.

2. Literature Survey

Time series Analysis, SARIMAX, stationarity concept, univariate analysis along machine learning will help the industry to study the data time to time and help them to make data driven decisions [1].During retail stage of food supply chain (FSC), food waste and stock-outs occur mainly due to inaccurate sales forecasting which leads to inappropriate ordering of products. The daily demand for a fresh food product is affected by external factors, such as seasonality, price reductions and holidays. In order to overcome this complexity and inaccuracy, the sales forecasting should try to consider all the possible demand influencing factors, to over come this issue, SARIMAX model is very Suitable[2].Seasonal Autoregressive Integrated Moving Average with external variables (SARIMAX) model which tries to account all the effects due to the demand influencing factors, to forecast the daily sales of perishable foods in a retail store. It is found that with respect to performance measures, the proposed SARIMAX model improves the traditional Seasonal Autoregressive Integrated Moving Average (SARIMA) model. A large number of real-world applications conducted by various individuals were also studied, and it was discovered that ARIMA is a real-world toll for time series prediction, forecasting, and analysis with accuracy [3].In addition to the autoregressive, differencing, and average terms for each season, a seasonal model must account for the number of occurrences in each season. ARIMA models can be created and implemented using a variety of software tools, such as Python and others[4].ARIMA performs better for the next few years, there may be deviation from the expected results but at present ARIMA model is best for sales forecasting and static time series data[5].

3. Proposed Methodology

The process of using ARIMA and SARIMAX models for sales prediction involves the following steps:

Data Visualization: Visualizing sales data is the first step in developing an ARIMA or SARIMA model. This entails plotting the time series data to better understand its behaviour over time. The goal is to search the data for trends, patterns, or cycles, as well as to identify any outliers or extreme values that may affect the analysis.

Stationary Analysis: The time series data must then be made stationary in the second step. A stationary time series has a constant mean and variance over time and does not change in behavior. To determine if the time series is stationary, statistical tests such as the Augmented Dickey-Fuller (ADF) test or the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test are used. Non-stationary time series can be made stationary using differencing, logarithmic transformation, or seasonal differencing. Using differencing or other methods, make the time series data stationary by removing the trend and seasonality.

Correlation and Autocorrelation Analysis: The correlation and autocorrelation charts are plotted in the third step. The autocorrelation function (ACF) and partial autocorrelation function (PACF) plots can aid in determining the ARIMA model's order. ACF depicts the correlation of a time series with its lagged values, whereas PACF depicts the correlation of a time series with its lagged values after the effect of intermediate lags is removed.

Model Construction: The fourth step is to build the ARIMA or Seasonal ARIMA from the data. Determine the order of the ARIMA model based on the results of the ACF and PACF plots, which includes the autoregressive order (p), integration order (d), and moving average order (q). Seasonal ARIMA is used if the time series has a seasonal component and includes an additional set of seasonal parameters (P, D, Q) and the seasonal period (m). Maximum likelihood estimation (MLE) or the method of moments (MOM) should be used to estimate the model parameters. Check the residuals for normality, stationarity, and independence to validate the model. Building the model in accordance with the identified order.

Prediction: The model is then used to make predictions as the final step. Once validated, the model can be used to make predictions for future time periods. The model parameters are used to calculate forecasted values and confidence intervals. The model's performance is monitored, and it is updated on a regular basis based on new data. Based on the model parameters, compute the forecasted values and confidence intervals. Monitor the model's performance and update it on a regular basis based on new data. Finally, by using the proposed model mentioned below used to forecast future values.

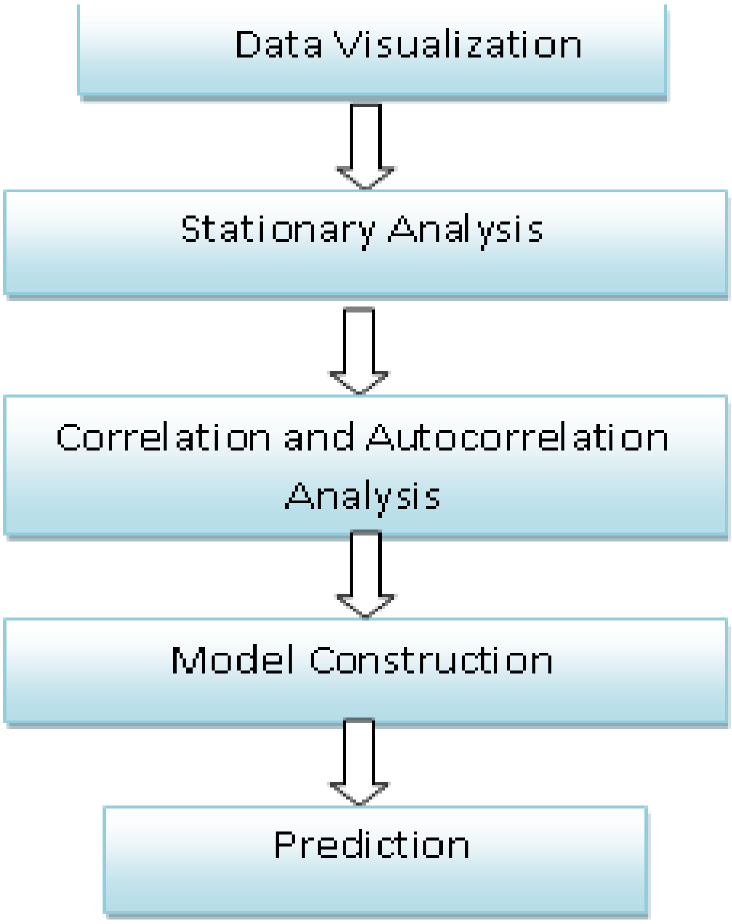

Figure 1 – Contextual methodology to predict future sales

4. Conclusion

Ultimately, the ARIMA and SARIMAX models are effective sales forecasting tools. These models can make accurate predictions about future values by detecting patterns in historical data. Visualizing the data, making it stationary, plotting the correlation and autocorrelation charts, constructing the model, and making predictions are all part of the general process for developing these models. These models can help businesses to make data-driven decisions for future sales and plan accordingly.

REFERENCES:

[1]. Malde Ratik Vimal , Shaikh Mohammad Bilal Naseem, Time Series Analysis:Forecasting With Sarimax Model And Stationarity Concept 2020 JETIR December 2020, Volume 7, Issue 12

[2]. Nari Sivanandam Arunraj, Deggendorf Institute of Technology, Germany Diane Ahrens, Deggendorf Institute of Technology, Germany Michael Fernandes, Deggendorf Institute of Technology, Germany Application of SARIMAX Model to Forecast Daily Sales in Food Retail Industry

[3]. Sandhya C, 2 Dr. N. Radha, SALES FORECASTING USING ARIMA MODEL,INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT) ISSN:2320-2882

[4]. Nadia Tarannum J1, Sri Vidya M S A Brief Introduction to Demand Forecasting using ARIMA models International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 08 Issue: 06 | June 2021 www.irjet.net p-ISSN: 2395-0072.

[5]. Sana Prasanth Shakti, Mohan Kamal Hassan, Yang Zhenning, Ronnie D. Caytiles# and Iyengar N.Ch.S.N. Annual Automobile Sales Prediction Using ARIMA Model, International Journal of Hybrid Information Technology Vol. 10, No. 6 (2017), pp.13-22, http//dx.doi.org/10.14257/ijhit.2017.10.6.02.