IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

SALES FORECASTING USING ARIMA MODEL

[1]Sandhya C, [2] Dr. N. Radha

[1]Student,
[2]Associate Professor,
[1]Department of Data Analytics,
[1]PSGR Krishnammal College for Women, Coimbatore, Tamilnadu, India.

ABSTRACT

Forecasting is one of the techniques that incorporate the e-commerce industry for a sweet manufacturer by analyzing the historical data and key elements that would aid in the company's profitability and growth. By consolidating the data from the business time, the key factors of various attributes can be found. In this article, the Box-Jenkins of Autoregressive Integrated Moving Average (ARIMA) approach was used to apply and forecast the company's sales by taking (2,78,559) daily sales observations over a period of six months. Several suitable time series models have been created, and several performance criteria have been utilized to compare models. The analysis revealed that the ARI (0,1,1) model is sufficient for forecasting sales.

Index Terms **- ARIMA, time-series, forecasting, sales.**

I. INTRODUCTION

In recent years, e-commerce has grown rapidly in the competitive world. Forecasting the company's sales has a big impact on orders, and sales so it is very important for e-commerce enterprises to predict accurately. Sales forecasting is used by almost every company, and it has an impact on operations, marketing, and planning. Sales data details of various branches of the company are gathered and consolidated into a single dataset, using the SQL Server with the help of link tables. Which contains 2,78,559 instances. The company's daily sales are summed and sorted by date. Time series forecasting is a method of using a model to predict future values based on previously observed values. It represents the linear trend of increase or decrease in a specific period of time, but due to numerous potential uncertainties, certain phases may exhibit nonlinear fluctuation characteristics. Various methods are developed, and the best-performing method for each case study is determined. The AUTOREG(Autoregressive), ARMA(Autoregressive Moving Average),(ARIMA) Autoregressive Integrated Moving Average), Seasonal ARIMA, Simple exponential smoothing forecast, Holt's exponential smoothing method, and Holt-Winters models are all taken into consideration. Among them, Seasonal ARIMA approaches, have the lowest value of the error measurements.

II. LITERATURE REVIEW

Theresa Hoang Diem Ngo in his work used the time series which is a collection of observations for a certain variable that appear in a positive pattern over time. The most typical patterns are rising or declining trends, cycles, seasonality, and uneven oscillations. Analysts notice the outline and assume that it will repeat itself in the future to recreate a time sequence event as an occupation of its historical qualities. Using the Box-Jenkins technique, this research focuses on how to identify a suitable time series replica by comparing the aggregate autocorrelation purpose (ACF) and fractional autocorrelation purpose (PACF) to academic autocorrelation functions [1]. Eliete Nascimento Pereira employed wavelet decomposition to detect time series forecasts [7]. A combination of forecasting methodologies, including the Autoregressive incorporated Moving Average (ARIMA) and artificial Neural Networks, was employed to obtain superior merit time series forecasting. Wavelet decomposition, ARIMA, and a hybrid neural network Multilayer Perception model were introduced in this study. Time sequence forecasting is the outcome of these models being mutually linear [2]. Wahyudi, using the ARIMA model sought to prognosticate the stock price volatility of equities listed on the Indonesia Stock Exchange. The stylish ARIMA model was set up using the Indonesia Composite Stock Price Index, according to the empirical disquisition (0, 0, 1). The AIC criterion was used to determine this model [3]. In a food company, the time series technique is used to predict and forecast demand. The study demonstrates how historical demand data may be utilized to forecast future demand and how these estimates affect the supply chain. Historical demand data was used to develop several autoregressive integrated moving average (ARIMA) models using the Box–Jenkins time series procedure, and the best model was chosen based on four performance criteria: the Akaike criterion, the Schwarz Bayesian criterion, maximum likelihood, and standard error [4]. These related helps to build the forecasting using the ARIMA model to predict the future sales.

III. METHODOLOGY

To select the best model from several adequate models we should use suitable criteria that deal with measures of accuracy and also with measures of goodness of fit of a model. In this paper, we depend on some criteria such as (Ayalew et. al., 2012, Makridakis et. al., 1998, Polhemus, 2011):

Root Mean Square Error (RMSE)

$$\text{RMSE} = \sqrt{\frac{\sum_{t=1}^{n} a_t^2}{n-c}} \qquad [3.1]$$

Where t is the time period, n is the total number of observations and c is the number of parameters in the model. The RMSE has the advantage of being easier to handle mathematically.

Mean Absolute Percentage Error (MAPE)

$$\text{MAPE} = \frac{\sum \frac{|A-F|}{A} \times 100}{N} \qquad [3.2]$$

The selected model is the one with the smallest Mean Absolute Percentage Error. The MAPE has the advantage of being more interpretable and is easier to explain to non-specialists.

Method	RMSE	MAPE
Simple average method	2301528.6	25.54
Simple moving average forecast	2387101.4	26.52
Simple exponential smoothing forecast	2444217.5	26.96
Holt's exponential smoothing method	3995797.8	43.04
Autoregressive (AR) method	2098363.5	22.3
Moving Average (MA) method	2411245.9	25.57
Autoregressive integrated moving average (ARIMA)	2692928.6	28.36
Seasonal Autoregressive Integrated moving average (SARIMA)	2098363.5	22.3

Finding the method with the least RMSE and MAPE value

IV. PROCESS FLOW

DATA COLLECTION

Sales data details of various branches of the company are gathered and consolidated into a single dataset, using the SQL Server with the help of link tables. Which contains 2,78,559 instances and is saved as an Excel file. Data is imported into a data frame for further processing in Python.

DATA PRE-PROCESSING

Steps involved in data pre-processing, (i) Dropping unwanted columns. (ii) Checking for null values and dropping them. (iii) Converting the data type and sorting the data based on date. (iv) Grouping and summarizing the values based on the date. As a result, the model can be built with clean data.

EXPLORATORY DATA ANALYSIS

The data is visualized in order to identify the major points and elements influencing the company's growth. The company's top and the least sold products and monthly sales are visualized.

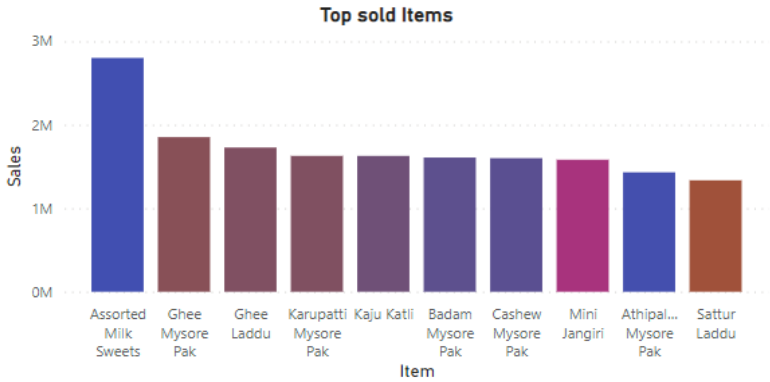


Figure 4.1Top Most Sold Products

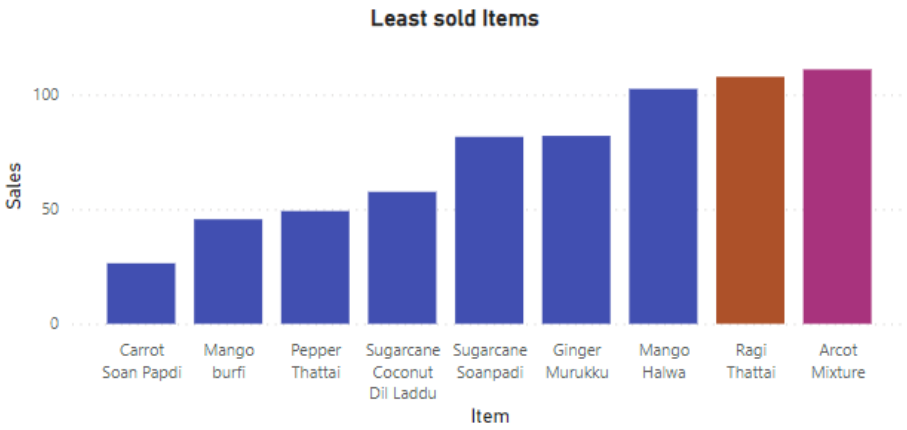


Figure 4.2 Least Sold Products

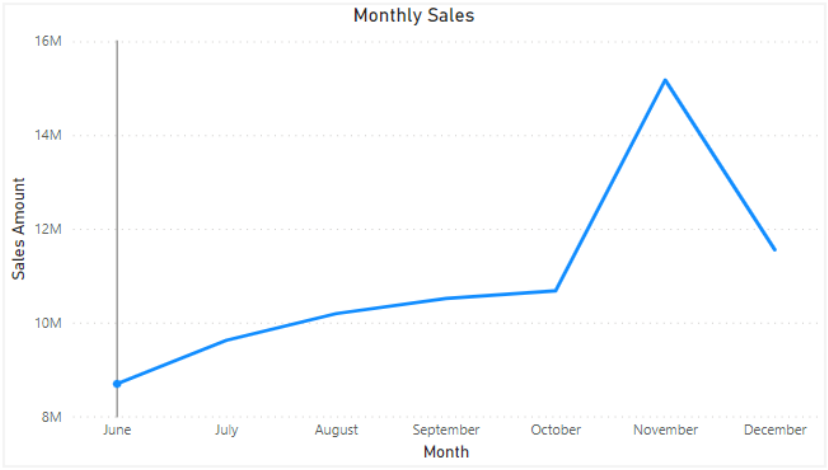


Figure 4.3 Monthly sales of the company

ARIMA MODEL

ARIMA (Auto-Regressive Integrated Moving Average) is a class of models that 'explains' a time series based on its own previous values, that is, its own lags and lagged prediction errors, so that equation can be used to anticipate future values. Any 'non-seasonal' time series that exhibits patterns and is not a random white noise can be modeled with ARIMA models.

An ARIMA model is characterized by 3 terms: p, d, q

where,

p is the order of the AR term

q is the order of the MA term

d is the number of differences required to make the time series stationary

If a time series has seasonal trends, seasonal terms must be added, and the time series becomes SARIMA, short for 'Seasonal ARIMA.'

An Auto-Regressive (AR) model is one where Y_t depends only on its own lags. That is, Y_t is a function of the 'lags of Y_t'.

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \ldots + \beta_p Y_{t-p} + \epsilon_1 \quad [4.1]$$

The moving average (MA) model is one where Y_t depends only on the lagged forecast errors.

$$Y_t = \alpha + \epsilon_t + \emptyset \epsilon_{t-1} + \emptyset_2 \epsilon_{t-2} + \ldots + \emptyset_q \epsilon_{t-q} \quad [4.2]$$

where the error terms are the errors of the autoregressive models of the respective lags. The errors Et and E(t-1) are the errors from the following equations,

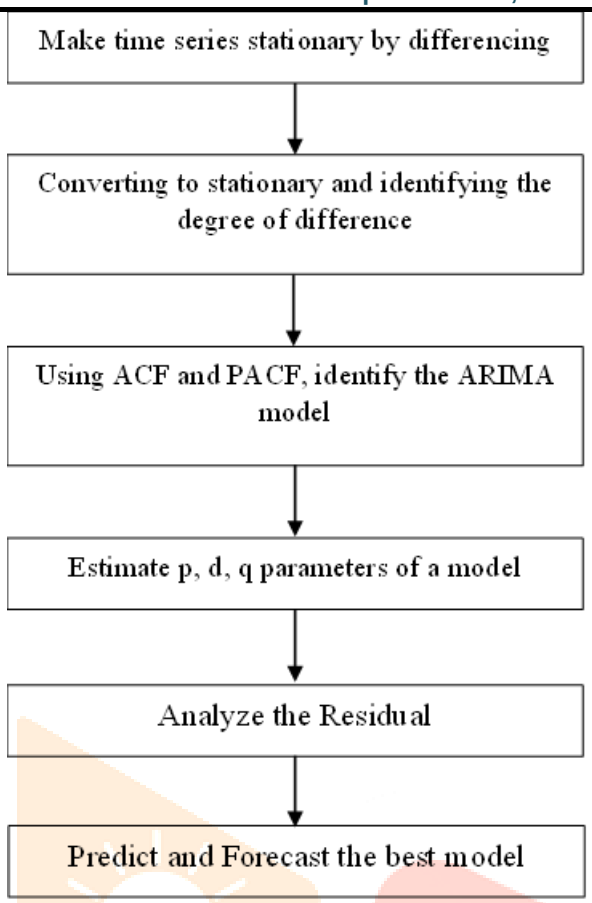
$$Y_t = \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \ldots + \beta_0 Y_0 + \epsilon_1 \quad [4.3]$$

$$Y_{t-1} = + \beta_1 Y_{t-2} + \beta_2 Y_{t-3} + \ldots + \beta_0 Y_0 + \epsilon_{t-1} \quad [4.4]$$

That is AR and MA models respectively.

An ARIMA model is one where the time series was differenced at least once to make it stationary and you combine the AR and the MA terms. So, the equation becomes:

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \ldots + \beta_p Y_{t-p} \epsilon_1 + + \emptyset \epsilon_{t-1} + \emptyset_2 \epsilon_{t-2} + \ldots + \emptyset_q \epsilon_{t-q} \quad [4.5]$$



Arima model

Predicted Y_t = Constant + Linear combination Lags of Y (upto p lags) + Linear Combination of Lagged forecast errors (upto q lags)

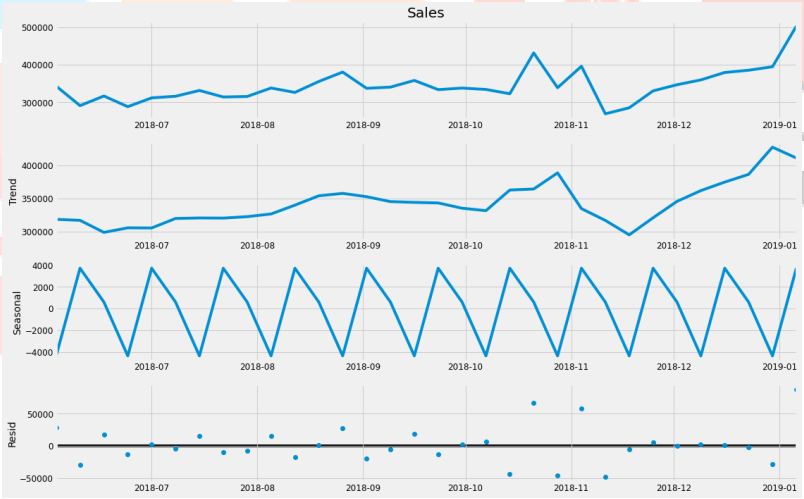


Figure 4.4 Seasonal Decomposition Plotting

V. RESULT AND ANALYSIS

Implementing ARIMA model and model that will predict the sales. 70% of the data will be used for training and 30% will be used for testing in this project. The model will be trained, and it will be used to forecast. Based on the forecast made, a graph will be plotted for future sales.

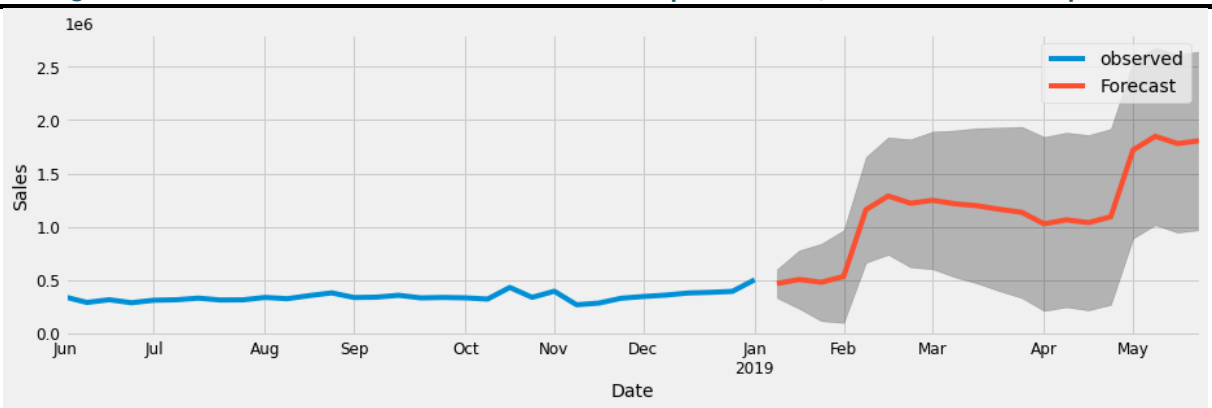


Figure 5.1 Plotting the future sales

CONCLUSION

ARIMA is a popular technique for analysing stationary univariate time series data. Model identification, model estimation, and model checking are the three main stages in building an ARIMA model, with model classification being the most important stage. As a result, the survey provides insight into the various time series prediction and forecasting models using ARIMA. 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.

REFERENCES

[1] Wahyudi, S. T. (2017). "The ARIMA Model for the Indonesia Stock Price", International Journal of Economics & Management, 11(1), 223-236.
[2] Eliete Nascimento Pereira, "Time Series Forecasting By Using A Neural Arima Model Based On Wavelet Decomposition", Independent journal of management & production (ijm&p), v. 7, n. 1 pp:252-270,2016.
[3] Jamal Fattah, Latifa Ezzine, Zineb Aman, Haj El Moussami, Abdeslam Lachhab, "Forecasting of demand using ARIMA model", International Journal of Engineering Business Management, v-10, 2018.
[4] Theresa Hoang Diem Ngo, "The Box-Jenkins Methodology for Time Series Models", SAS Global Forum 2013, pp 454-2013.
[5] Lisa Bianchi, Jefft Jarrett and R. Choudary Hanumara. "Improving forecasting for telemarketing centers by ARIMA modeling with intervention", International Journal of Forecasting 14, pp. 497-504, 1998.
[6] HAMILTON J.D. 1994. Time series analysis. Princeton. Princeton University Press. ISBN 978-0691042893 pp. 799.
[7] Douglas C. Montgomery, Cheryl L. Jennings, Murat Kulahci, "Introduction to Time Series Analysis and Forecasting", 2015, 2nd Edition, Wiley.
[4] Deschamps, E. (2001). "The Impact of Institutional Changeon Forecast Accuracy: A Case Study of Budget Forecasting in Washington State", Paper presented at the International Symposium on Forecasting.
[8] Mentzer, J. T., Moon, M. A., Smith, C. D., "Conducting a sales forecasting audit", International Journal of Forecasting 18, 19, PII no: S0169-2070(02)00032-8, 2002.
[9] P. Sobreiro, D. Martinho and A. Pratas, "Sales forecast in an it company using time series",13th Iberian Conference on Information Systems and Technologies (CISTI), pp. 1-5, 2018.
[10] Rokach, L., Ensemble methods for classifiers. Data Mining and Knowledge Discovery Handbook; Springer:Cham, Switzerland, pp. 957–980, 2005.
[11] Ayalew S., Babu M. C. and Raw L.K, "Comparison of New Approach Criteria for Estimating the Order of Autoregressive Process". Journal of Mathematics, 1, 10-20, 2012.
[12] Brockwell PJ and Davis RA, "Time series: theory and method", Berlin: Springer-Verlag, 1987.
[13] Hamilton JD, "Time series analysis", Princeton: PrincetonUniversity Press, 1994.