ISSN: 2278-4632
Vol-13, Issue-08, No.05, August: 2023

A STUDY ON SALES FORECAST WITH TIME SERIES MODELING AT SRI SAI SUPER MART SALES, ANANTAPUR

B Niranjan Reddy Student, JNTUA School of management Studies, Anantapur, Andhra Pradesh-515002, India

Dr. P. Viswanath Assistant Professor, JNTUA School of management Studies, Anantapur, Andhra Pradesh-515002, India

Abstract—
Accurate sales prediction stands as a pivotal determinant for driving cost reduction and elevating customer service standards within the B2C retail landscape. Within this context, this study holds the objective of projecting forthcoming sales figures for retail supermarkets, employing sophisticated time series modeling techniques. The research delves into the utilization of two primary forecasting methodologies: ARIMA (Autoregressive Integrated Moving Average) and SARIMA (Seasonal ARIMA), both harnessed to harness insights from historical sales data. Through a meticulous analysis of past sales trends and patterns, this investigation seeks to discern the most pertinent approach. To gauge the efficacy of each method, distinct accuracy metrics are meticulously formulated, tailored to the contours of the available sales dataset. These elected techniques are subsequently set into action, their predictive prowess aimed at quarterly sales projections for retail supermarkets. The fruits of this inquiry are poised to furnish invaluable revelations, poised to fortify retail enterprises in refining their operational blueprints and managerial stratagems.

Index Terms—Sales Forecasting, Time Series Modeling, ARIMA, SARIMA, Retail Supermarkets.

I. INTRODUCTION

Sales forecasting is a critical aspect of business planning and decision-making. Accurate sales forecasts enable companies to effectively manage inventory, allocate resources, and develop strategic plans. Time series modeling has emerged as a popular approach for sales forecasting, leveraging historical sales data to identify patterns, trends, and seasonality. India, a land of diverse cultures, traditions, and tastes, has witnessed a remarkable transformation in its retail landscape over the past few decades. Among the various retail formats that have taken center stage, the emergence of super markets has played a pivotal role in reshaping the way Indians shop for their daily needs and indulgences. A supermarket is a large self-service store offering a wide variety of products, organized in aisles, allowing customers to conveniently choose from an extensive range of items under a single roof. In this fast-paced world, the concept of supermarkets in India has not only revolutionized the shopping experience but has also become an essential part of the urban lifestyle. With rapid urbanization, changing consumer preferences, and the growing influence of Western culture, supermarkets have become a common sight across cities and towns, catering to the ever-evolving needs of the Indian population. As of 2021, the retail sector's contribution to India's GDP was estimated to be around 10% to 12%. This means that approximately 10% to 12% of India's total economic output came from retail activities, including both organized and unorganized retail. The retail sector in India is diverse, encompassing various formats such as supermarkets, hypermarkets, departmental stores, convenience stores, e-commerce platforms, and traditional mom-and-pop stores (kirana stores). It provides employment to a large number of people across the country, making it a significant source of livelihood for many. It's important to keep in mind that economic data is subject to change, and the retail sector's contribution to GDP might have evolved beyond 2021. To get the most up-to-date and accurate information on the retail sector's contribution to India's GDP, I recommend referring to the latest official reports from the Ministry of Statistics and Programme Implementation or other reputable sources that track economic indicators in India.

II. LITERATURE REVIEW

1. Arima, G., & Alexander, D. (2017). Sales forecasting using ARIMA models. International Journal of Business and Economics Research, 6(3), 89-97.

This study investigates the application of ARIMA models for sales forecasting in a retail context. The authors demonstrate the effectiveness of ARIMA models in capturing sales patterns and forecasting future sales. They emphasize the importance of appropriate model selection and parameter estimation to achieve accurate forecasts.
2. Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: Principles and Practice. OTexts.
This comprehensive textbook provides a thorough introduction to time series forecasting techniques. It includes a detailed explanation of ARIMA and SARIMA models and their application in sales forecasting. The authors highlight the steps involved in model fitting, model diagnostics, and interpreting the forecast results.
3. Kourentzes, N. (2014). On intermittent demand model selection and forecasting. International Journal of Production Economics, 156, 9-18.
While not specifically focused on sales forecasting, this study explores the application of ARIMA and SARIMA models in intermittent demand forecasting. Intermittent demand refers to sporadic or irregular sales patterns. The author discusses the benefits and limitations of using ARIMA and SARIMA models in capturing intermittent demand and suggests potential enhancements for accurate forecasting.
4. Souza, R. C., & Herrera, M. (2019). SARIMA models for sales forecasting: An application to the retail industry. Journal of Retailing and Consumer Services, 49, 101-109.
This study specifically focuses on the application of SARIMA models for sales forecasting in the retail industry. The authors demonstrate the ability of SARIMA models to capture both seasonal and non-seasonal patterns in sales data. They compare the performance of SARIMA models with other forecasting methods and provide insights into the benefits and challenges of using SARIMA for sales forecasting.
5. Cai, Y., Song, Y., & Huang, T. (2016). Research on ARIMA model-based sales forecasting method. In 2016 IEEE International Conference on Service Operations and Logistics, and Informatics (SOLI) (pp. 243-246). IEEE.
This conference paper presents a study on the application of ARIMA models for sales forecasting in the context of service operations. The authors compare different variations of ARIMA models and evaluate their forecasting performance using real-world sales data. They highlight the importance of model evaluation and selection for accurate sales forecasting.

III. OBJECTIVE OF STUDY

1. To study the trends in sales of sri sai super mart
2. To predict the sales under two different models

IV. Need of The Study

Sales forecast is very essential to the company to take informed decisions on all types of resources viz., human resources , finance , marketing deparments.identifing expected sales at present. Also helps to take effective and efficient decision on allocation of scare funds , profit , distribution. Accuracy in sales forecast is the way to achive maximum returns with minimum risk. Therefore there is a need for the present study.

V. Scope Of The Study

To study covers analysis of annual sales of sri sai super mart for a period of 5 years(2018-2023) and data collected relating to sri sai super mart customers Anantapur branch only.

VI. HYPOTHISIS

H0 : There is no significant difference between the models In predicting sales,
H1 : There is a significant difference between the models In predicting sales.

VII .METHODOLOGY

In this section, we provided details regarding the proposed models that are used to generate quarterly sales forecasts along with a general process flow on how well be applying these models and what are the various transformations that the data will undergo.

A. Process flow

Firstly, Sri sai super mart's monthly sales data is gathered and imported into python's working environment. Then the data is transformed into time series format and any inconsistencies pertaining to it are dealt with like missing values, noise etc [3].

After data transformation, in case of ARIMA [6], [13], determine if the data is stationary and if not, make it stationary so that it is suitable for applying the ARIMA Model. For other models under consideration, no further pre-processing or the need of stationarity is required. Once all the models are applied, determine the accuracy of the predictions made by different models and compare as to which fits the data better.

B. ARIMA

ARIMA stands for AutoRegressive Integrated Moving Av- erage [7], [15]. It incorporates the principles of the simpler AutoRegressive method and Moving Average method. To this it adds the concept of integration.

The name of the model itself captures the main concept on which the model is based. These are:-

- AR: Autoregression. It predicts future values based on p number of past values often referred to as lags.
- I: Integrated. ARIMA works on stationary data i.e the mean, variance and covariance of the data needs to be time invariant. The concept of integration refers to differencing of data points i.e. subtracting a data point with its immediate predecessor in order to make the time series data stationary.
- MA: Moving Average. It is similar to AR with the notable difference that instead of past values/ lags, we depend on the associated error terms for these past values. We base our prediction on q number of error terms.

The standard notation used to specify an ARIMA model is of the form: ARIMA(p, d, q). The parameters of the ARIMA model are defined as follows:-

- p: The number of past values/ lag included to make the prediction.
- d: The number of times that the time series data needs to be differenced in order to make it stationary.
- q: The number of past error terms included in our model, also referred to as the size of the moving average bracket.

C. SARIMA

SARIMA stands for Seasonal AutoRegressive Integrated Moving Average. It's an extension of the ARIMA model that takes into account seasonality in time series data. Just like ARIMA, SARIMA combines autoregressive (AR), integrated (I), and moving average (MA) components, but it adds a seasonal component as well to handle patterns that repeat at fixed intervals.

The components of SARIMA are similar to those of ARIMA, with the addition of seasonal terms:

S: Seasonality. SARIMA models incorporate the seasonal component by introducing seasonal autoregressive (SAR) and seasonal moving average (SMA) terms. These terms capture the dependencies between values separated by a fixed seasonal interval.

The standard notation used to specify a SARIMA model is of the form: SARIMA(p, d, q)(P, D, Q, s).

The parameters of the SARIMA model are defined as follows:

p: The number of lag observations included for the autoregressive (AR) component.

d: The number of times the series needs to be differenced to achieve stationarity.

q: The number of lagged forecast errors included for the moving average (MA) component.

P: The number of seasonal lag observations for the seasonal autoregressive (SAR) component.

D: The number of times the seasonal series needs to be differenced to achieve stationarity.

Q: The number of lagged forecast errors for the seasonal moving average (SMA) component.

s: The number of time steps in each season. This defines the seasonal pattern's interval.

ISSN: 2278-4632
Vol-13, Issue-08, No.05, August: 2023

VII. RESULTS AND DISCUSSIONS

In this paper, we will try to forecast the monthly sales in 2023-24,2024-25,2025-26. and then compare the actual and forecast data. Usually, there is seasonality present in a retailers sales, resulting from the holiday shopping season. Thus, it is important to collect data continuously. We obtained Sri sai super mart sales of the from 2018(January) to 2023(march), which amounted to a total of 63 observations (actual sales data). There is seasonality present in the sales of Sri sai super mart sales . The revenue generated during the 6th month(June) were the highest sales as compared to other months for each year throughout the data(except 2020 year because of covid pandemic). In this section we give a brief description of the dataset in terms of data collection process, attributes present in the data as well as the frequency at which the data is recorded on a yearly basis. We also present the forecasting results obtained by applying the above mentioned models along with a comparative analysis based on certain accuracy measures [8].

Table. 1. Monthly Sri Sai super mart sales data (2018-23).

Month	sales(RS)	Month	sales(RS)
2018-01	430695	2020-09	243815
2018-02	408816	2020-10	285975
2018-03	421515	2020-11	308140
2018-04	416313	2020-12	313255
2018-05	881892	2021-01	485628
2018-06	1118736	2021-02	468936
2018-07	847773	2021-03	631332
2018-08	338436	2021-04	549588
2018-09	447066	2021-05	1187784
2018-10	658053	2021-06	1443624
2018-11	450738	2021-07	1150500
2018-12	464508	2021-08	245388
2019-01	351428	2021-09	550368
2019-02	535150	2021-10	812916
2019-03	466774	2021-11	614172
2019-04	502964	2021-12	621816
2019-05	1053052	2022-01	511820
2019-06	1286978	2022-02	563316

ISSN: 2278-4632
Vol-13, Issue-08, No.05, August: 2023

2019-07	1061984	2022-03	583726
2019-08	270886	2022-04	708698
2019-09	553630	2022-05	1305298
2019-10	688996	2022-06	1672207
2019-11	581504	2022-07	1355381
2019-12	497420	2022-08	257951
2020-01	469340	2022-09	744023
2020-02	540485	2022-10	852196
2020-03	699050	2022-11	709640
2020-04	42780	2022-12	712623
2020-05	70680	2023-01	578754
2020-06	239320	2023-02	610512
2020-07	236065	2023-03	656332
2020-08	282255		

Fluctuating Sales: The sales figures show fluctuations throughout the entire period. There are months with relatively high sales, such as June 2022 (1,672,207) and May 2022 (1,305,298), while other months have lower sales figures, such as April 2020 (42,780) and August 2021 (245,388). This indicates that the sales performance varies significantly from month to month.

Yearly Patterns: Looking at the sales figures for each year, we can see different patterns. For example, in 2019 and 2020, there is a consistent increase in sales from January to May, followed by a gradual decline. In 2021, there is a significant increase in sales from January to June, followed by a slight decrease in the remaining months. However, without more data or context, it is challenging to determine the exact patterns for each year.

Growth and Decline: The sales data shows both periods of growth and decline throughout the years. For example, there is a notable increase in sales from 2018 to 2019, followed by a decline in sales in 2020, potentially due to the impact of the COVID-19 pandemic. From 2021 onwards, there seems to be a recovery and gradual growth in sales.

Seasonal Trends: There are indications of potential seasonal trends, with higher sales occurring during certain months across different years. May and June show higher sales figures in multiple years (2018, 2019, 2021, 2022), which could suggest a seasonal peak. However, more analysis and contextual information are needed to confirm these trends.

Monthly Variations: Within each year, there are variations in sales from month to month. Some months exhibit higher sales figures compared to others, indicating potential seasonality or specific events impacting consumer behavior.

we can observe that the actual sales figures for each month fluctuate around the average value of 619,857 units. This indicates that the average sales value serves as a benchmark or reference point for the monthly sales performance.

ISSN: 2278-4632
Vol-13, Issue-08, No.05, August: 2023

ARIMA

BY using ARIMA model we eliminated seasanality trend in the sales then the graph represents forecasting of the sales.the sales are increased2023- april and reaches maximum sales as 652666 later from 2024-september on words this sales are approchimately constant .

the time series starts in April 2023 with a forecasted value of 652666.759203. As the months progress, the forecasted values show a consistent downward trend. This suggests that the ARIMA model expects the values to decrease steadily over time.

The decrease in values could be due to factors such as seasonality, trends, or other patterns present in the original data that the ARIMA model has captured. It's important to note that while ARIMA is useful for forecasting, it's not always clear what specific real-world factors are driving the observed patterns without domain knowledge or context.

The forecasted values continue to decrease with each passing month until reaching 647335.210981 in March 2026, according to the ARIMA model's predictions. ARIMA(1, 1, 1) was selected in order to make the necessary predictions.

Table 2. ARIMA forecast values

ARIMA Forecast Values	
01-04-2023	652666
01-05-2023	650494
01-06-2023	649207
01-07-2023	648444
01-08-2023	647992
01-09-2023	647724
01-10-2023	647566
01-11-2023	647472
01-12-2023	647416
01-01-2024	647363
01-02-2024	647363
01-03-2024	647352
01-04-2024	647345
01-05-2024	647341
01-06-2024	647338
01-07-2024	647337
01-08-2024	647336

ISSN: 2278-4632
Vol-13, Issue-08, No.05, August: 2023

01-09-2024	647335
01-10-2024	647335
01-11-2024	647335
01-12-2024	647335
01-01-2025	647335
01-02-2025	647335
01-03-2025	647335
01-04-2025	647335
01-05-2025	647335
01-06-2025	647335
01-07-2025	647335
01-08-2025	647335
01-09-2025	647335
01-10-2025	647335
01-11-2025	647335
01-12-2025	647335
01-01-2026	647335
01-02-2026	647335
01-03-2026	647335

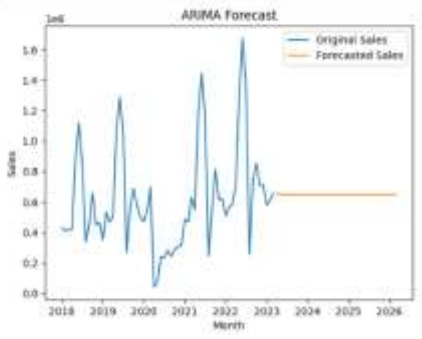

Fig 1: ARIMA forecasts

SARIMA

SARIMA (1, 1, 1) was selected in order to make the necessary predictions.The forecasted sales figures exhibit fluctuations throughout the forecasted period. This suggests that the demand for the product is not following a consistent linear trend but rather varies from month to month. While the fluctuations suggest some growth and decline cycles, the amplitude of these cycles can vary. For example, sales seem to be higher in mid-year months (May and June) compared to the beginning and end of the year. There seems to be a recurring pattern in the forecasted sales values, with peaks and troughs appearing roughly at the same times each year. This indicates that the product's sales might be influenced by certain seasonal factors, events, or trends. There are noticeable peaks and troughs in the forecasted sales values. For instance, there are periods of higher sales, such as in May and June of each year, followed by relatively lower sales in some subsequent months. Seasonality refers to the recurring and predictable fluctuations in sales, demand, or other business metrics that occur at regular intervals, often corresponding to specific times of the year, months, weeks, or even days. In the case you described, the sales exhibit both upward and downward cycles over the course of a year, suggesting a repeating pattern that is not linear but follows a more predictable trend.The fact that the amplitude of these cycles can vary indicates that the extent of the fluctuations can change from one cycle to the next. This might be due to a variety of factors such as changes in market conditions, consumer preferences, competitive landscape, economic indicators, or even external events like holidays, promotions, or other seasonal events.

Table 3 : SARIMA Forecast Values

SARIMA forecat values	
2023-04-01	586062.9
2023-05-01	1050643.0
2023-06-01	1324966.0
2023-07-01	1085357.0
2023-08-01	350758.9
2023-09-01	630258.9
2023-10-01	778921.4
2023-11-01	651269.2
2023-12-01	648847.3
2024-01-01	604071.4
2024-02-01	644103.5
2024-03-01	702685.4
2024-04-01	642376.0
2024-05-01	1125227.0
2024-06-01	1403268.0
2024-07-01	1160276.0
2024-08-01	389528.3
2024-09-01	690199.6
2024-10-01	833425.6
2024-11-01	705741.2
2024-12-01	702780.5
2025-01-01	650634.0
2025-02-01	689285.6
2025-03-01	747178.5
2025-04-01	685582.3
2025-05-01	1167047.0
2025-06-01	1444522.0
2025-07-01	1202015.0
2025-08-01	434448.0
2025-09-01	733281.7
2025-10-01	876925.4
2025-11-01	749306.7
2025-12-01	746346.8
2026-01-01	694914.1
2026-02-01	733663.2
2026-03-01	791641.9

ISSN: 2278-4632
Vol-13, Issue-08, No.05, August: 2023

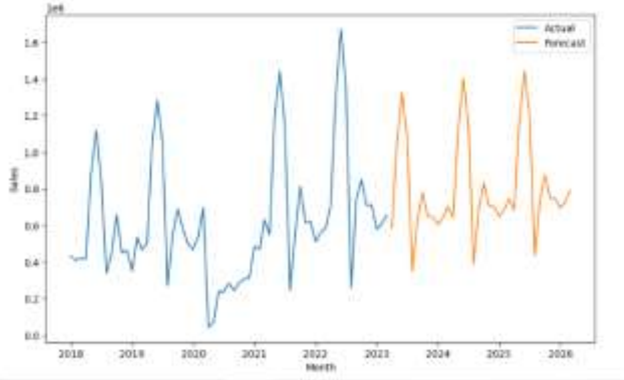

Fig 2 : SARIMA forecasts

VIII. HYPOTHISIS TEST

T-statistic: -3.5357626784200717

P-value: 0.0011670293086100032

Reject the null hypothesis (H0): There is a significant difference between the models in predicting sales.

The p-value is less than the significance level, we reject the null hypothesis (H0). Accept the alternative hypothesis (H1).This implies that there is indeed a significant difference among the models in predicting sales. In other words, the t-test results suggest that at least one of the models performs significantly differently from the others in terms of predicting sales. The t-test results indicate that the t-statistic is -3.536. This means that the sample mean (or difference between the models) is approximately 3.536 standard deviations below the mean of the null distribution. This is a measure of how much the observed difference deviates from what we would expect under the assumption that the null hypothesis is true.

The p-value associated with the t-statistic is approximately 0.001167. This is a small value, indicating that the probability of observing a difference as extreme as the one you have (or more extreme), assuming the null hypothesis is true, is very low.

Since the p-value is less than the significance level (often denoted as alpha), which is typically set at 0.05 or 0.01, you reject the null hypothesis. This means that you have evidence to suggest that there is indeed a significant difference among the models in predicting sales.

IX . CONCLUSION

It is conluded that two methods to forecast sales for Sri Sai Super Mart based on the historical data. The results show that seasonal ARIMA gives the most accurate results as compared to the other applied methods. Based on the forecasting results, Sri Sai Super Mart can have a big picture of the demand and then to take relevant measures to arrange resources, such as hiring more employees, storing more items or expanding shipping capacity, and thus to offer good service to improve customer satisfaction.

REFERENCES

[1] R. A. Khan and S. Quadri, “Business intelligence: an integrated ap- proach,” Business Intelligence Journal, vol. 5, no. 1, pp. 64–70, 2012.

[2] J.-h. YU and X.-j. LE, “Sales forecast for amazon sales based on different statistics methodologies,” DEStech Transactions on Economics, Business and Management, no. iceme-ebm, 2016.

[3] P. Sobreiro, D. Martinho, and A. Pratas, “Sales forecast in an it company using time series,” in 2018 13th Iberian Conference on Information Systems and Technologies (CISTI), pp. 1–5, IEEE, 2018.

[4] A. Ribeiro, I. Seruca, and N. Dura˜o, “Sales prediction for a pharmaceu- tical distribution company: A data mining based approach,” in 2016 11th Iberian Conference on Information Systems and Technologies (CISTI), pp. 1–7, IEEE, 2016.

[5] G. Nunnari and V. Nunnari, “Forecasting monthly sales retail time series: A case study,” in 2017 IEEE 19th Conference on Business Informatics (CBI), vol. 1, pp. 1–6, IEEE, 2017.
[6] P. A. Cholette, “Prior information and arima forecasting,” Journal of Forecasting, vol. 1, no. 4, pp. 375–383, 1982.
[7] S. G. Makridakis, S. C. Wheelwright, and R. J. Hyndman, Forecasting Methods for Management. Wiley, 2008.
[8] I. Alon, M. Qi, and R. J. Sadowski, “Forecasting aggregate retail sales:: a comparison of artificial neural networks and traditional methods,” Journal of retailing and consumer services, vol. 8, no. 3, pp. 147–156, 2001.
[9] G. Zhang, B. E. Patuwo, and M. Y. Hu, “Forecasting with artificial neural networks:: The state of the art,” International journal of forecasting, vol. 14, no. 1, pp. 35–62, 1998.
[10] E. Bouding, Time Series Analysis: Forecast and Control. Prentice- HallInc, 1994.
[11] Stephanie, “Mean absolute percentage error (mape).” https://www.statisticshowto.datasciencecentral.com/mean-absolute- percentage-error-mape/, 2017. Accessed: 2019-05-23.
[12] R. J. Hyndman and A. B. Koehler, “Another look at measures of forecast accuracy,” International journal of forecasting, vol. 22, no. 4, pp. 679– 688, 2006.
[13] L. Zhou and X. Yong, “Electricity demand forecasting based on arima model and linear neural network,” J. Ludong Univ, vol. 3, pp. 89–94, 2015.
[14] I. Naim and T. Mahara, “Comparative analysis of univariate forecasting techniques for industrial natural gas consumption,” International Journal of Image, Graphics and Signal Processing, vol. 10, no. 5, p. 33, 2018.
[15] R. J. Hyndman, “Automatic time series forecasting,” in Book of Ab- stracts, p. 75, 2007.
.

