STM JOURNALS

International Journal of
Algorithms Design and Analysis Review

ISSN: 2584-1866
Volume 2, Issue 1, 2024
January—June
DOI (Journal): 10.37591/IJADAR

https://journals.stmjournals.com/ijadar

Review IJADAR

# Time Series Sales Forecasting Using ARIMA Model

Kanishk Patil[1], Anuj Sawant[1], Shantanu Rane[1,*], Siddhant Pupulwad[1], D.K. Chitre[2]

**Abstract**

*Sales forecasting is a critical application in various industries and presents one of the most challenging problems worldwide. One method of prediction involves identifying patterns in historical data, where the outcome is known in advance and can be validated using more recent data. If a pattern consistently leads to the same outcome, it can be considered a genuine relationship. This method is highly flexible and can be utilized with diverse datasets, extending beyond climate data. Sales prediction involves numerous parameters such as the number of sales, production levels, associated costs, and time requirements, which are challenging to quantify and measure accurately. "Time Series Sale Forecasting Using ARIMA Model" presents a novel methodology for sales forecasting, specifically tailored to analyze seasonal sales data. The study extensively examines monthly sales data spanning multiple years, with a particular emphasis on identifying inherent seasonality within the dataset. An essential element of the methodology is the utilization of the Augmented Dickey-Fuller (ADF) test to confirm the stationarity of the time series data, establishing a strong basis for precise forecasting. At the core of the study is the sole application of the autoregressive integrated moving average (ARIMA) model for sales prediction. This model is renowned for its adaptability and flexibility, proving highly effective in capturing intricate patterns embedded within the sales data. By leveraging the ARIMA model's capabilities, the study demonstrates its efficacy in accurately predicting future sales trends, thereby enabling retail businesses to make well-informed decisions. The paper highlights the critical role of accurate sales forecasting in driving strategic decision-making processes within the retail sector. In an increasingly dynamic and competitive marketplace, precise sales predictions serve as indispensable tools for guiding business strategies, optimizing inventory management, and enhancing overall operational efficiency. Through the adept application of advanced analytical techniques and a comprehensive understanding of seasonal sales dynamics, this study offers valuable insights into sales prediction methodologies. By equipping retail enterprises with the means to anticipate market fluctuations and consumer trends, the research empowers businesses to navigate the ever-evolving retail landscape with confidence and strategic foresight.*

**Keywords:** Time series forecasting, machine learning, sales prediction, ARIMA, ADF, seasonal sales, retail businesses

***Author for Correspondence**
Shantanu Rane
E-mail: shantanurane3@gmail.com

[1]Student, Department of Computer Engineering, Terna Engineering College, Navi Mumbai, Maharashtra, India
[2]Assistant Professor, Department of Computer Engineering, Terna Engineering College, Navi Mumbai, Maharashtra, India

Received Date: April 18, 2024
Accepted Date: April 20, 2024
Published Date: May 30, 2024

**Citation:** Kanishk Patil, Anuj Sawant, Shantanu Rane, Siddhant Pupulwad, D.K. Chitre. Time Series Sales Forecasting Using ARIMA Model. International Journal of Algorithms Design and Analysis Review. 2024; 2(1): 17–27p.

## INTRODUCTION

Sales prediction is quite a complex process and is a challenging task for all researchers around the globe. It demands expertise in multiple different disciplines. Predictions related to the atmospheric parameters play a crucial role in various applications. Accurate prediction of the Sales parameters can be a quite strenuous task due to the highly dynamic nature of the atmosphere. Prediction is being done on time series data. A time series represents a sequence consisting of observed values of some entity that are measured at various points in time, which complicates the process. Due to advancements in data collection methods, vast quantities of data have been accumulated, rendering manual processing practically unfeasible. This is exactly where automated time series analysis stuff comes into the picture and steps up to benefit from modern computing mechanisms. This write-up focuses on using the ARIMA model for the prediction of retail store sales. We will compare the performance of ARIMA with other models like SARIMAX and Linear Regression to figure out which one proves to be better for the task of predicting sales within a given seasonal time frame. We have gathered a dataset that spans from January 2022 to December 2023 from various retail electronic stores. In this study, we are concentrating solely on the parameter "sales" where we plan to apply ARIMA methodology to predict the sales.

This work investigated into leveraging the ARIMA model for time series sales forecasting, targeting specifically the monthly sales data that extends over multiple years and presents evident seasonality traits. The flexibility and adaptability of the ARIMA model make it quite suitable for a wide range of applications, allowing for accurate predictions, whether it is stock prices, exchange rates, or even sales volumes. Unlike previous studies that considered multiple models, this research solely relies on the ARIMA model for forecasting, attempting to streamline the process while ensuring high-accuracy outcomes. The methodology in this project includes data preprocessing and ADF testing for stationarity, followed by the application of the ARIMA model for sales forecasting. By providing an in-depth overview of the research approach and methodology, this write-up aims to offer beneficial insights into the field of sales forecasting through the lens of time series analysis.

The process spans data preprocessing, model testing, and the application of ARIMA for sales forecasting, highlighting its efficiency in handling seasonal sales patterns. With this research study, we aim to equip businesses with the essential tools they require to make informed decisions and remain competitive in the continuously evolving retail landscape.

## LITERATURE REVIEW

Nikita Malik [1] documented "Sales Prediction Model for Big Mart" explores the use of machine learning algorithms to predict sales for Big Mart, a large shopping center.

It underscores the significance of data analysis in predicting future demand and enhancing inventory management optimization. key metrics such as the R-squared, root mean square error (RMSE), and Mean Absolute Error (MAE) are utilized for assessing model performance. Python and libraries like NumPy and Matplotlib are used for implementation. The document also demonstrates discussions on correlations between various item weights, outlet establishment year, and item visibility with sales. It concludes with insights into sales predictions for different retail outlet locations and seriously underpins the importance of efficient data mining for business development.

Ziru Zhang [2] highlighted the importance of comparing different analytical forecasting models for sales prediction. By utilizing historical sales data from Kaggle, research demonstrates the impact of linear regression and ARIMA models on sales forecasting. The analysis reveals that while the multiple factorial linear regression model achieved a lower RMSE of 21.135, the ARIMA model had an RMSE of 23.221. This indicates varying degrees of fit between the two models. The results highlight the importance of choosing the most appropriate model according to the unique characteristics of the data and the goals of the analysis. Such insights are crucial for guiding future research and improving the accuracy of sales forecasts using time series models like ARIMA.

Ronnie D. Caytiles [3]—sales forecasting is a pivotal application in industries, presenting significant scientific and technological challenges. One effective approach to prediction involves identifying patterns in historical data and verifying them on more recent data. The ARIMA model is utilized in this paper to predict tractor sales for ten years (2003–2014) using data from Mahindra Tractors Company. Through analysis of various plots and graphs, the ARIMA model demonstrates its effectiveness in predicting sales for the next 5 years. This unique study highlights the importance of accurate sales forecasting and the applicability of the ARIMA model in predicting complex sales phenomena.

**Autoregressive (AR) Process**

In an AR(p) model, the future value of a variable is predicted to be a linear combination of its p previous observations, along with a random error and a constant term.

Mathematically, the AR(p) model can be expressed as:

$$y_t = c + \sum_{i=1}^{p} \varphi_i \, y_{t-i} + \varepsilon_t = c + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \ldots\ldots\ldots\ldots + \varphi_p y_{t-p} + \varepsilon_t$$

**Moving Average (MA) Process**

The current deviation from the mean is influenced by past deviations. In an AR(p) model, regression is performed against past values of the series, while in an MA(q) model, past errors serve as the explanatory variables.

The MA(q) model is given by:

$$y_t = \mu + \sum_{j=1}^{q} \theta_j \, \varepsilon_{t-j} + \varepsilon_t = \mu + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots\ldots\ldots\ldots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

Here, μ represents the mean of the series (where, j=1, 2, ..., q), $\theta_j$ are the model parameters, and q is the order of the model. Essentially, a moving average model is akin to a linear regression of the current observation of the time series against the random shocks of one or more prior observations.

**Stationary Series (Integrated)**

To utilize the Box-Jenkins approach for modeling a time series, the series must exhibit stationarity. In practical terms, a series is considered stationary if it tends to fluctuate around a fixed level. From a statistical perspective, a stationary process is presumed to be in a consistent state of statistical equilibrium, meaning that the probability distribution of xt remains constant for all t. To attain stationarity in a series, regular differencing (RD) is required.

(1st order) ∇xt = (1 – B)xt = xt – xt-1

(2nd order) ∇2xt = (1–B)2xt = xt –2xt-1+(stupid)xt-2

"B" represents the backward shift operator. Rarely, more than two regular differencing would ever be necessary.

Carina Intan Permatasaria [4] employed the ARIMA models forecasting newspaper demand, intending to reduce returns, avoid missing sales, and manage supply more effectively. Using data from a newspaper company in Surakarta, the ARIMA (1, 1, 0) model is identified as the most suitable for predicting newspaper demand, based on its ability to minimize the mean absolute percentage error (MAPE). This research underscores the importance of using advanced forecasting models like ARIMA, to adapt to changing consumer behaviors, and optimally inventory management in the newspaper industry.

Yuxuan Hans [5] presented a forecasting method for pharmaceutical sales using an ARIMA-LSTM model. It combines the ARIMA and LSTM models for sales forecasting. The study details the construction and training of both models, emphasizing data preprocessing and model establishment. Moreover, the ARIMA-LSTM model is constructed and trained, showcasing the integration of both approaches for improved forecasting accuracy.

Zhang M [6] encompassed various critical aspects concerning time series analysis, statistical prediction, and financial econometrics. It investigated into the fundamental principles underlying time series modeling, with a particular emphasis on the ARIMA model's modeling methodology and its utilization in variance analysis and seasonal data testing.

 

Wang Z and Lou Y [7] investigated into the complexities of forecasting hydrological time series, acknowledging the myriad factors affecting their accuracy. It recognizes the limitations of traditional forecast models in addressing these challenges. In response, the paper introduces a novel forecasting model that integrates wavelet de-noising and a hybrid approach merging ARIMA and LSTM methodologies.

Wang Q [8] investigated into the evolution of environmental monitoring technology and its implications for analyzing meteorological trends. It explores how advancements in science, technology, and the economy have contributed to enhancing environmental monitoring methods, especially concerning meteorological big data analysis.

Han J, Fu Y, and Zhang H [9] explored the importance of precise market demand forecasting and operational planning for superstores in the digital age. It may discuss how rapid digitalization emphasizes the need for accurate insights into market trends to enhance competitiveness. Additionally, it may touch upon the use of time series models like ARIMA to forecast sales volumes and prices, as well as the development of mathematical planning models to optimize replenishment and pricing strategies.

Kogekar AP, Nayak R, and Pati UC [10] investigated into the importance of monitoring and forecasting water quality, especially in regions like India where water challenges are prevalent. It explores the existing studies on time series-based models utilized for water quality prediction, with a focus on models such as autoregressive integrated moving average (ARIMA), seasonal ARIMA (SARIMA), and Prophet. These models have been effectively applied in previous research to accurately forecast water quality parameters and Water Quality Index (WQI).

## PROPOSED WORK

The proposed strategy for time series sales forecasting utilizing the ARIMA model consists of a couple of key phases. In the beginning, the data is imported and inspected quite thoroughly, followed by a cleaning process to manage the missing values, filter out any incorrect data present, and convert the related columns. Following this, let us embark on some feature engineering by extracting the months and doing some calculations to find the total sales. Next up, the data is grouped by month, summary statistics are computed, and the results are showcased for some deeper insights into sales trends over time. This all-encompassing approach guarantees that the dataset is all set for accurate and insightful sales forecasting using the ARIMA model.

### Input Data

The input data is in the form of movie store sales records from an electronic retail store fetched from Kaggle; it comprises 6 columns, to be precise. This dataset includes pivotal sales data, including order IDs, product particulars, quantities, prices, timestamps, and purchase addresses as shown in Figure 1. Its structured format facilitates a comprehensive analysis of sales trends and patterns, allowing for insights into the popularity of products and pricing strategies.

### Data Preprocessing

#### *Data Import and Inspection*

- *Importing libraries:* NumPy, Pandas, Matplotlib, and seasonal decomposition functions from Statsmodels are essential for data manipulation, visualization, and time series analysis.
- *Reading the dataset:* Loading the dataset from the specified path into a Pandas DataFrame (all_data) allows us to work with the data effectively.

#### *Data Cleaning*

- *Handling missing values:* Identifying plus handling missing values is crucial to ensure the quality and integrity of the dataset. By dropping rows with missing values, also storing them in nan_df, we ensure that the analysis is based on complete data.

 

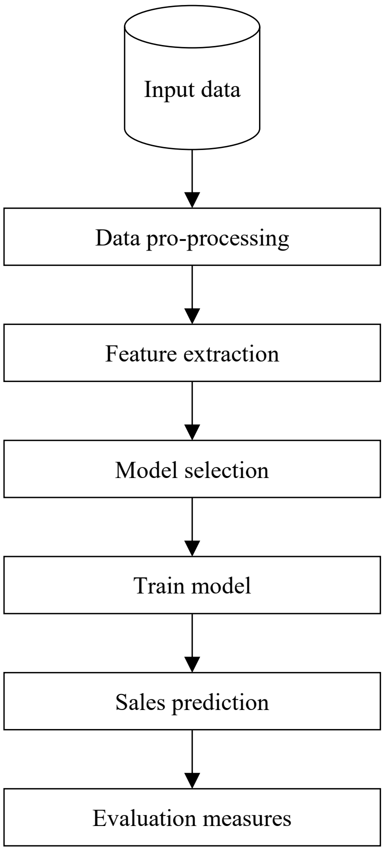


**Figure 1.** Workflow diagram.

- *Filtering incorrect data:* Removing rows with ‘Order Date’ starting with ‘Or’ helps remove potential errors or anomalies in the data, ensuring that only valid records are considered.
- *Data type conversion:* Converting columns like ‘Quantity Ordered’ and ‘Price Each’ to numeric data types also ‘Order Date’ to datetime format enables us to perform mathematical operations, and date-based analysis seamlessly.

**Feature Engineering**

- *Extracting month:* Creating a new column ‘Month’ by extracting the month from the ‘Order Date’ allows us to analyze sales trends every month, which is essential for time series forecasting.
- *Calculating total sales:* Total sales are computed for each row by ‘Quantity Ordered’ and ‘Price Each’ provides a fundamental metric to help us understand the revenue generated from each transaction.
- *Aggregation and summary:* Grouping and Aggregating: The data is grouped by ‘Month’ and calculates summary statistics like maximum, minimum, mean, and total sales which aggregates the data into a more manageable form. These aggregated statistics serve as key inputs to understanding sales patterns and trends over time.
- *Renaming columns:* The renaming of the columns of the aggregated DataFrame enhances clarity and interpretability, making it easier to understand the summary statistics.
- *Result display:* Present the summary of the monthly sales statistics in a tabular format so it allows stakeholders to grasp key insights quickly such as maximum, minimum, average, and total sales for each month.

Essentially, this data preprocessing pipeline transform is complex. This paves the way for misunderstandings and less clarity makes the dataset readable for accuracy and makes sales forecasting using ARIMA model insights.

***Model Selection***

*Test for stationarity*

- Based on the output of the Augmented Dickey-Fuller (ADF) test.ADF
  - *Statistics:* -3.543097636322846.
- The ADF statistic is below the critical values, indicating stationarity.
  - *P-value:* 0.006946834175587418.
- With a p-value less than 0.05, the null hypothesis of non-stationarity is rejected in favor of stationarity.
- *Number of lags:* 0. This indicates the number of lags used in the regression.
- *Number of observations used:* 24. This is the number of data points used in the analysis.
- *Critical values:*
  - *1%:* -3.7529275211638033
  - *5%:* -2.998499866852963
  - *10%:* -2.6389669754253307
- All critical values are surpassed by the ADF statistic, further indicating stationarity.

In summary, the ADF test results suggest that the time series is stationary, as indicated by the ADF statistic being lower than the critical values and the p-value being less than 0.05. Therefore, differencing may not be necessary, and the *data can be directly used for ARIMA modeling*.

The 'auto_arima' function from the 'pmdarima' library assists in identifying the most suitable parameters for an ARIMA model and provides a fitted ARIMA model as output as shown in Figure 2.

The output indicates the stepwise search process to minimize AIC, resulting in the selection of the ARIMA (0, 0, 0) (0, 1, 0) [6] model as the best model with the lowest AIC of 113.377 as shown in Figure 3.

- *ARIMA (0, 0, 0) (0, 1, 0) [6]:* This ARIMA model has non-seasonal parameters (p=0, d=0, q=0) and seasonal parameters (P=0, D=1, Q=0) with a seasonal period of 6.
- *Total fit time:* The total time taken for the fitting process is 1.885 seconds.

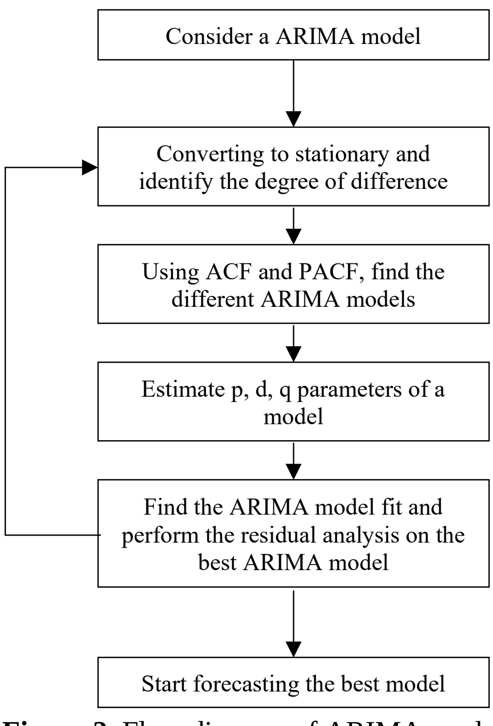


**Figure 2.** Flow diagram of ARIMA model.

 

```
Performing stepwise search to minimize aic
 ARIMA(1,0,1)(0,1,1)[6] intercept   : AIC=inf, Time=0.25 sec
 ARIMA(0,0,0)(0,1,0)[6] intercept   : AIC=115.275, Time=0.02 sec
 ARIMA(1,0,0)(1,1,0)[6] intercept   : AIC=inf, Time=0.51 sec
 ARIMA(0,0,1)(0,1,1)[6] intercept   : AIC=inf, Time=0.20 sec
 ARIMA(0,0,0)(0,1,0)[6]             : AIC=113.377, Time=0.02 sec
 ARIMA(0,0,0)(1,1,0)[6] intercept   : AIC=inf, Time=0.25 sec
 ARIMA(0,0,0)(0,1,1)[6] intercept   : AIC=inf, Time=0.13 sec
 ARIMA(0,0,0)(1,1,1)[6] intercept   : AIC=inf, Time=0.35 sec
 ARIMA(1,0,0)(0,1,0)[6] intercept   : AIC=115.996, Time=0.04 sec
 ARIMA(0,0,1)(0,1,0)[6] intercept   : AIC=116.549, Time=0.05 sec
 ARIMA(1,0,1)(0,1,0)[6] intercept   : AIC=117.224, Time=0.06 sec

Best model:  ARIMA(0,0,0)(0,1,0)[6]
Total fit time: 1.885 seconds
```

SARIMAX Results

| | | | |
|---|---|---|---|
| **Dep. Variable:** | y | **No. Observations:** | 24 |
| **Model:** | SARIMAX(0, 1, 0, 6) | **Log Likelihood** | -55.688 |
| **Date:** | Wed, 17 Apr 2024 | **AIC** | 113.377 |
| **Time:** | 07:07:39 | **BIC** | 114.267 |
| **Sample:** | 01-01-2022 | **HQIC** | 113.499 |
| | - 12-01-2023 | | |
| **Covariance Type:** | opg | | |

| | coef | std err | z | P>\|z\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| **sigma2** | 28.4946 | 14.125 | 2.017 | 0.044 | 0.810 | 56.179 |

| | | | |
|---|---|---|---|
| **Ljung-Box (L1) (Q):** | 1.44 | **Jarque-Bera (JB):** | 1.06 |
| **Prob(Q):** | 0.23 | **Prob(JB):** | 0.59 |
| **Heteroskedasticity (H):** | 1.02 | **Skew:** | 0.31 |
| **Prob(H) (two-sided):** | 0.98 | **Kurtosis:** | 1.99 |

**Figure 3.** Auto-selection of ARIMA model.

- *The SARIMAX* results include:
  - *Log-Likelihood:* The log-likelihood value is -55.688.
  - *BIC:* Bayesian Information Criterion (BIC) is 114.267.
  - *HQIC:* Hannan-Quinn Information Criterion (HQIC) is 113.499.
- *Covariance Type:* The type of covariance matrix used is “opg” (outer product of gradients).
- *coef, std err, z, P>|z|, [0.025 0.975]:* These values represent the coefficients, standard errors, z-scores, p-values, and confidence intervals for the model parameters.
- *sigma2:* The estimated variance of the residuals (sigma squared) is 28.4946.
- *Ljung-Box (L1) (Q):* The Ljung-Box statistic for lag 1 is 1.44, with a p-value of 0.23.
- *Jarque-Bera (JB):* The Jarque-Bera statistic is 1.06, with a p-value of 0.59.
- *Heteroskedasticity (H):* The Heteroskedasticity statistic is 1.02, indicating homoscedasticity.
- *Skew:* The skewness of the residuals is 0.31.
- *Kurtosis:* The kurtosis of the residuals is 1.99.

These results provide insights into the goodness of fit and diagnostic tests for the selected Auto-ARIMA model.

***Train Model***

- *Split data:* The training set contains data from January 2021 to September 2022, comprising 9 observations. The testing set includes data from October 2022 to January 2023, consisting of 15 observations, totaling 24 data points.
- *SARIMAX model selection:* The SARIMAX model with parameters (0, 1, 0) for the non-seasonal part and (0, 1, 0, 6) for the seasonal part is chosen for fitting on the training data.
- *Model fitting:* The SARIMAX model is fitted to the training data, with the ‘Average Sale’ column as the dependent variable.
- *Model summary:* The summary of the SARIMAX model results shows statistical metrics:
  - *Log-Likelihood:* -8.209
  - *AIC:* 18.419, AIC score to judge how good a particular order model is.
  - *BIC:* 17.517
  - *Ljung-Box (L1) (Q):* 2.89
  - *Jarque-Bera (JB):* 0.40

Other parameters such as coefficients, standard errors, z-values, and confidence intervals are also provided. These metrics aid in assessing the model's fit to the training data and diagnosing its performance.

***Sales Prediction***

The dataset contains 24 data points. We split the data into Train and Test data. Thus 15 points for training and 9 points for testing. Thus, the prediction plot of the model displays the prediction and average sales for the testing dataset for the last 9 months as shown in Figures 4 and 5.

```
# Split data into train/test sets
train = monthly_avg_sales.iloc[:len(monthly_avg_sales)-15]
test = monthly_avg_sales.iloc[len(monthly_avg_sales)-15:]
```

Plot of test dataset and prediction model.

Plotting sales graph for the full dataset from January 2022 to December 2023 and Forecasting sales with ARIMA model for next year 2024.

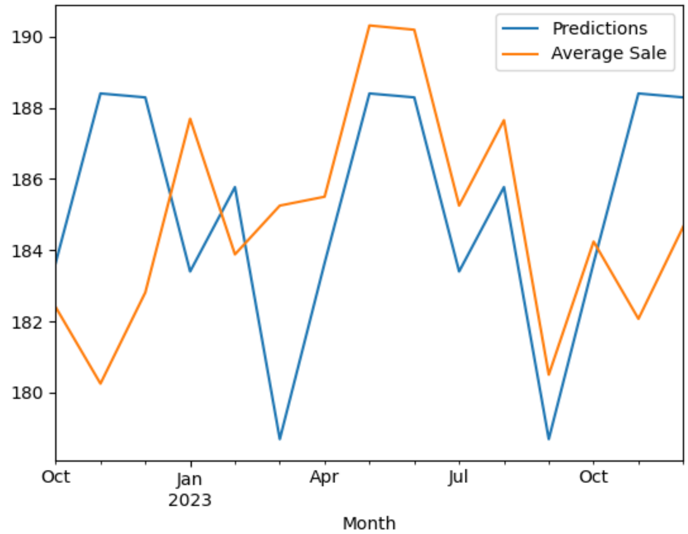


**Figure 4.** Plot of test dataset and prediction model.

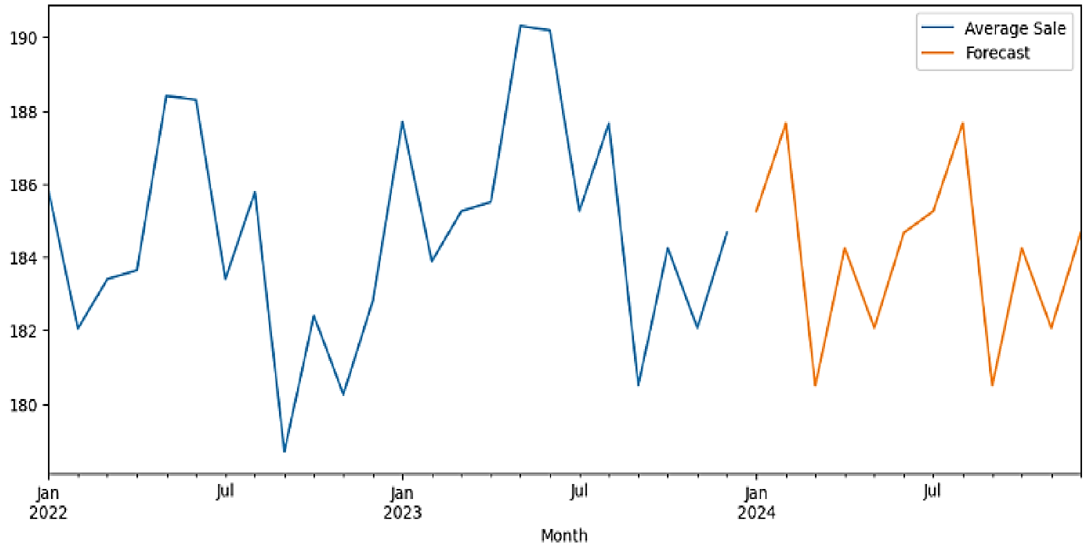


**Figure 5.** Plotting sales graph for full dataset from January 2022 to December 2023 and forecasting sales with ARIMA model for next year 2024.

 

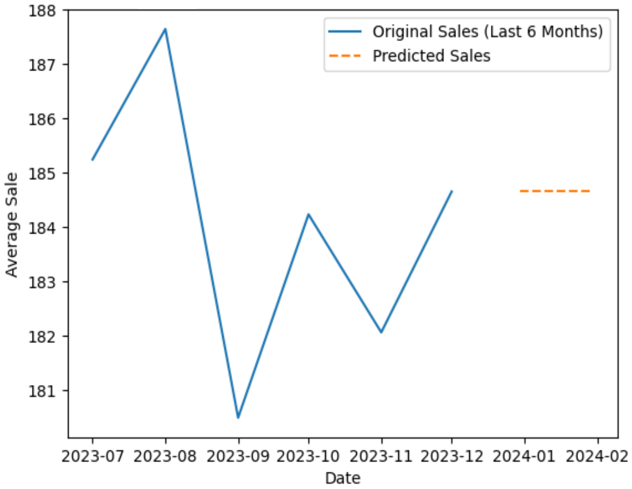


**Figure 6.** Original vs. predicted sales for the last 6 months.

### Future Sales Forecasting

An ARIMA model is fitted to the monthly average sales data using ARIMA from statsmodels.tsa.arima.model. The model is initialized with an order of (0,0,0) indicating no differencing or autoregressive terms. The fitted model (model2) is used to predict future sales. The prediction method is called with start= len(df) and end= len(df)+31 to predict 31 days beyond the length of the existing data. Predictions are assigned to pred. A date range for the future dates is created using pd.date_range starting from '2023-12-30' to '2024-01-30'. This range is assigned to index_future_dates. The index of the predicted values (pred) is set to the future date range (index_future_dates).

*Plotting:* Matplotlib is imported. The original sales data for the last six months (last_six_months_sales) is plotted along with the predicted sales (pred). Different line styles are used for clarity as shown in Figure 6.

## RESULT AND ANALYSIS

Performance measures for sales forecasting typically include metrics such as accuracy, mean squared error (MSE), and RMSE. These metrics assess the performance of sentiment analysis models in correctly classifying text data into positive and negative sentiment categories as shown in Figure 7.

Accuracy is a widely used metric for evaluating the effectiveness of classification models. It quantifies the percentage of accurately classified instances relative to the total instances in the dataset. Specifically, it calculates the ratio of true predictions (both true positives and true negatives) to the total number of predictions as shown in Tables 1 and 2.

$$Accuracy = \frac{Number\ of\ correct\ predictions}{Total\ number\ of\ predictions}$$

Mean squared error is a frequently employed metric for assessing the efficacy of a forecasting or regression model. It computes the average of the squared differences or deviations between predicted and actual values.

$$MSE = \frac{1}{N} \sum_{i=1}^{N} (\hat{y}_i - y_i)^2$$

Root mean squared error (RMSE) is a widely utilized metric for assessing the effectiveness of a forecasting or regression model, especially when the errors are assumed to follow a normal distribution.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (\hat{y}_i - y_i)^2}$$

 

**Table 1.** Evaluation of ARIMA model.

| ML model | Accuracy (%) | MSE | RMSE |
|---|---|---|---|
| ARIMA | 97.85% | 15.83 | 3.97 |

**Table 2.** Comparison across models.

| ML models | Linear regression | SARIMAX | ARIMA |
|---|---|---|---|
| Accuracy (%) | 71.84 | 69.14 | 97.85 |

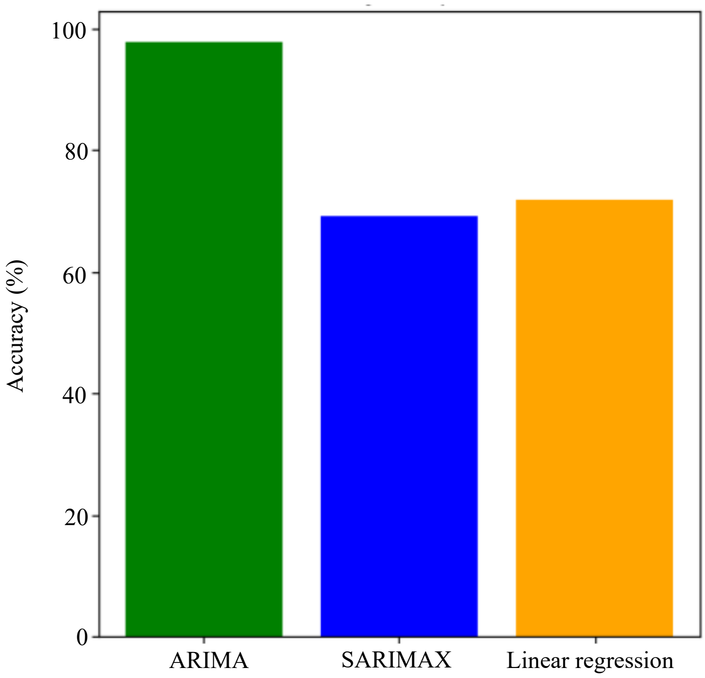


**Figure 7.** Accuracy comparison model.

## CONCLUSION AND FUTURE WORK

In this paper, we have implemented "Time Series Sales Forecasting Using ARIMA Model" which presents a comprehensive approach to sales forecasting in the context of seasonal sales data. It emphasizes the importance of accurate sales prediction for effective decision-making in retail businesses. The research focuses on leveraging the ARIMA model, which combines autoregressive, differencing, and moving average components to capture underlying patterns in sales data. By employing the ARIMA model exclusively for forecasting, the study aims to streamline the process while maintaining high accuracy. The methodology includes data preprocessing, ADF testing for stationarity, and the application of the ARIMA model for sales forecasting, highlighting its efficacy in handling seasonal sales patterns. The results of the study show that the ARIMA model, when compared with other models like SARIMAX and linear regression, performs well in predicting sales within seasonal time frames. The research contributes valuable insights to the field of sales forecasting using time series analysis, equipping businesses with tools to make informed decisions and stay competitive in the retail industry.

## REFERENCES

1. Malik N, Singh K. Sales prediction model for Big Mart. Parichay: Maharaja Surajmal Institute Journal of Applied Research. 2020;3:22–32.
2. Zhang Z. Sales prediction based on Arima time series and multifactorial linear model. Highlights in Science, Engineering and Technology. 2023 Mar 16;38:1–8. DOI: 10.54097/hset.v38i.5680.
3. Shakti SP, Hassan MK, Zhenning Y, Caytiles RD, Iyenger NC. Annual automobile sales prediction using Arima model. Int J Hybrid Inf Technol. 2017 Jun;10(6):13–22. DOI: 10.14257/ijhit.2017.10.6.02.

 

4. Permatasari CI, Sutopo W, Hisjam M. Sales forecasting newspaper with Arima: A case study. AIP Conf. Proc. 2018 Feb 9;1931(1):030017. DOI: 10.1063/1.5024076.
5. Han Y. A forecasting method of pharmaceutical sales based on Arima-LSTM model. 2020 5th International Conference on Information Science, Computer Technology and Transportation (ISCTT); 2020 Nov 13; Shenyang, China. IEEE; 2020. pp. 336–339. DOI: 10.1109/ISCTT51595.2020.00064.
6. Zhang M. Financial time series frequent pattern mining algorithm based on time series Arima model. 2023 International Conference on Networking, Informatics and Computing (ICNETIC); 2023 May 29; Palermo, Italy. IEEE; 2023. pp. 244–247. DOI: 10.1109/ICNETIC59568.2023.00057.
7. Wang Z, Lou Y. Hydrological time series forecast model based on wavelet de-noising and Arima-LSTM. 2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC); 2019 Mar 15; Chengdu, China. IEEE; 2019. pp. 1697–1701. DOI: 10.1109/ITNEC.2019.8729441.
8. Wang Q. Time series simulation method of meteorological elements based on Arima model. 2021 2nd International Conference on Big Data & Artificial Intelligence & Software Engineering (ICBASE); 2021 Sep 24; Zhuhai, China. IEEE; 2021. pp. 358–362. DOI: 10.1109/ICBASE53849.2021.00073.
9. Han J, Fu Y, Zhang H. Optimization research on vegetable sales in supermarkets based on greedy algorithm and Arima time series model. 2024 IEEE 3rd International Conference on Electrical Engineering, Big Data and Algorithms (EEBDA); 2024 Feb 27; Changchun, China. IEEE; 2024. pp. 979–985. DOI: 10.1109/EEBDA60612.2024.10485713.
10. Kogekar AP, Nayak R, Pati UC. Forecasting of water quality for the river Ganga using univariate time-series models. 2021 8th International Conference on Smart Computing and Communications (ICSCC); 2021 Jul 1; Kochi, Kerala, India. IEEE; 2021. pp. 52–57. DOI: 10.1109/ICSCC51209.2021.9528216.

 