Demand Forecasting in Retail Using a Time Series Model for Optimization

Dr. Sandhya Avasthi[1] SMIEEE and **Dr. Shweta Roy**[2] AMIETE, MIEEE
Department of Computer Science Engineering, ABES Engineering College, 19th Kmstone, NH 09, Ghaziabad 201009 UP India
[1]sandhya.avasthi@abes.ac.in, [2]shweta.roy@abes.ac.in

Abstract -- **Accurate demand forecasting is crucial for retailers to effectively manage their inventory, optimize pricing strategies, and enhance overall operational efficiency. This study explores the application of three widely used forecasting models, namely ARIMA, Exponential Smoothing, and Prophet models. The performance of these models is evaluated using historical sales data from a retail company. The results indicate the strengths and weaknesses, with varying degrees of accuracy depending on the specific characteristics of the analyzed sales data.This study provides insights into how widely or extensively these forecasting techniques can be applied in a retail setting. Knowledge of this degree provides retailers with the necessary insight into making better time and territorial decisions in real-time inventory, pricing strategies, and sales prediction; this process also improves their competitive advantage as the retail landscape changes minute by minute. Through time series analysis, retailers can leverage insights into consumer behavior and market fluctuations to become more responsive to changing conditions. This flexibility not only improves inventory stability but also enhances customer loyalty by ensuring product availability and satisfaction.**

Keywords: *Auto regressive integrated moving average, Time series, Demand forecasting, Prophet*

I. INTRODUCTION

THE retail market has increased significantly over the past years due to increased customer demand, and the expansion of businesses across the region. This growth brings a big challenge to inventory management, especially in balancing stock demand to avoid both overstocking and stock-outs. Mainly, it occurs in seasonal fluctuations, promotional events, and other marketing events necessary with an accurate demand forecast required effective retailing and customer satisfaction. Traditional forecasting methods have been used widely to forecast future demands [1].

Demand forecasting will be critical in inventory management, improvement in customer satisfaction and improvements in the production of operations. Businesses can make better strategic and operational decisions that minimize costs, maximize profits, and meet customer demand through prediction [2]. Traditionally, retailers can rely on historical data and trends to predict future demand, and continuously update their forecasts as new data becomes available [4]. Today, a variety of time series models have been created and utilized for retail demand forecasting for Moving Average, Exponential Smoothing, ARIMA, and more [3]. A predictive model has been developed to maintain the supply of goods and featuring the predictions of upcoming demand for the next few selected time period. The model involves being highly trained in the field of Machine Learning using various predefined models such as Linear Regression, GBT Model and even using the time series analysis for the defined period and using the outcome from the factor of inputs provided by the user to finally predict the upcoming demand.

By using the time series model, the retailer can analyze data easily to acquire demand forecast in retail and better decision-making to manage inventory efficiently. Retailers can also prepare the right products at the right time, which makes the customer satisfied and increases profitability [4]. This paper intends to suggest a comprehensive study to forecast demand in retail using techniques for demand forecasting including ARIMA, Exponential Smoothing, and Prophet, and forecasting historical trend data with R programming. We will examine different time-series forecasting models and compare their performance in predicting future retail demand. The methodology takes three main steps dataset collection and preprocessing, modeling, and forecasting. The research objectives are to study demand forecasting and the implementation of different time series models on experimental datasets.

II. LITERATURE REVIEW

Demand forecasting, in this context, refers to the prediction of the number of units of a commodity that would be sold, purchased, or required within some time. It has many strong operating advantages in stock optimization, and accurate delivery. It also helps in optimizing inventory levels and reducing excess stock, which can lead to financial losses [1]. It also improves the efficiency of supply chain responsiveness by ensuring that products are available when needed, thus increasing customer satisfaction [2]. The Time Series refers to collecting all the historical data points and looking into their patterns in historical records like inventories and other operations in retail. The traditional time series forecasting models that have been widely used for retail demand forecasting and e-commerce include ARIMA, exponential smoothing, and Prophet.

ARIMA (Auto Regressive Integrated Moving Average): This is a statistical model that uses historical data to predict future data based on autoregressive and moving average concepts. This is one of the most widely used time-series forecasting models because of its flexibility and ability to handle complex data patterns [5]. It has been characterized by three components: Auto-Regressive (AR), Integrated (I), and Moving Average (MA), all these are the terms that work together to help recognize the data patterns and allow for more accurate predictions. It operates best on stationary data [2], [6].

Exponential Smoothing: is a forecasting technique that gives decreasing weight to past observations. It is a type of moving average applied to forecasting sales and demand because of the higher accuracy in short-term predictions, which can easily be derived from the recent sales trend data [7]. There are several variations of exponential smoothing models, such as simple exponential smoothing, double exponential smoothing, and Holt-Winters exponential smoothing.

Prophet: is a forecasting library developed by Facebook's Core Data Science team especially designed for time-series data. It uses an effective model for forecasting long-term time-series forecasting sales where non-linear trends fit with yearly, weekly, and daily seasonality. It combines a generalized additive model with additional concepts for handling outliers and holidays in a manner that reduces the need for manual feature engineering compared to traditional time-series models [8], [9].

Demand forecasting in the retail sector is challenging because of various reasons like seasonality, promotions, economy, and emerging trends and disruptions. The primary challenges are related to promotions, seasonality, economic factors, and due to emerging trends and disruptions. Sales events, discounts, and new product launches are examples of events that largely affect demand and should thus be included in the related models of demand forecasting. Accurately forecasting demand during promotions poses an enormous challenge for retailers [10]. Seasonality often occurs over a year in business patterns due to the changing of seasons, holidays, and other recurring events. Accordingly, seasonality helps to determine that some of such business decisions occurred, like inventories and staffing [11]. Economic factors and macroeconomic indicators such as unemployment rates, consumer sentiment, and inflation affect consumer spending, hence demand. The impact of emerging trends and disruptions such as the COVID-19 pandemic have disrupted demand patterns and caused significant challenges to traditional approaches to forecasting [12]. While most of the emerging literature on the pandemic has focused on disease progression, a few have focused on consequent regulations and their impact on individual behavior.

The contributions of this paper include a quantitative behavior model of fear of COVID19, impact of government interventions on consumer behavior, and impact of consumer behavior on consumer choice and hence demand for goods. It brings together multiple models for disease progression, consumer behavior and demand estimation– thus bridging the gap between disease progression and consumer demand. We use panel regression to understand the drivers of demand during the pandemic and Bayesian inference to simplify the regulation landscape that can help build scenarios for resilient demand planning. We illustrate this resilient demand planning model using a specific example of gas retailing. We find that demand is sensitive to fear of COVID-19 – as the number of COVID-19 cases increase over the previous week, the demand for gas decreases - though this dissipates over time. Further, government regulations restrict access to different services, thereby reducing mobility, which in itself reduces demand.

III. METHODOLOGY

This section outlines the processes that were used to gather, prepare, and analyze the data, selecting and implementing different forecasting models, performing predictions, and evaluating the results. Figure 1 shows an overview workflow of demand forecasting in retail.

Dataset and preprocessing: The data of this study come from a Kaggle Global Super Store Dataset for online retail business. It consists of 51,290 records of four years (2011 to 2014) in order of historical sales data. This dataset will give a sales trend and patterns for predicting demand sales that are developed by using the predictive model. The analysis will seek the most important variables that have an impact on sales performance and enable

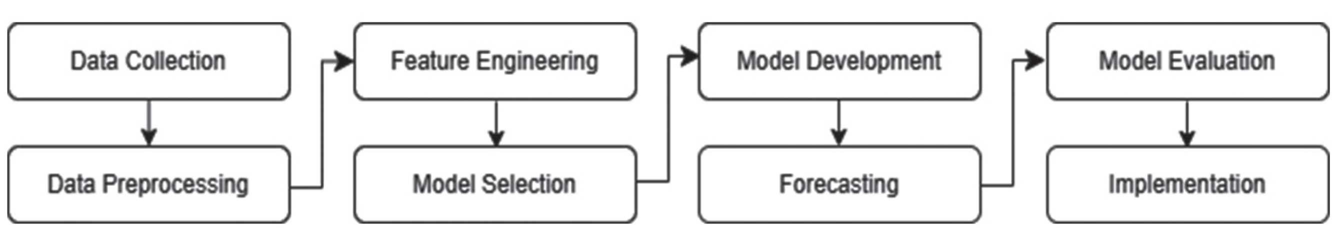

Figure 1. Demand forecasting workflow in Retail demand forecasting.

more accurate predictions of the insights into how consumers' behavior affects demand.The description of the column of the Global Super Store Dataset is described in Table 1.

TABLE 1-- DATASET ATTRIBUTES AND DESCRIPTION

Column Name	Description
Row ID	A unique ID for each row in the dataset.
Order ID	A unique Order ID placed by a customer.
Order Date	The date when the order was ordered by the customer.
Ship Date	The date when the order was shipped to the customer.
Ship Mode	The mode through which the product is shipped.
Customer ID	Unique ID for the customer who made the purchase.
Customer Name	The name of the customer making the purchase.
Segment	The market segment, such as consumer, corporate, or home office.
City	The city where the customer is situated.
State	The state where the customer is located.
Country	The country where the customer lives.
Postal Code	The postal code for the shipping address of the customer.
Market	The market in which the customer operates, such as consumer goods, technology, etc.
Region	The geographical region where the customer lives, such as North America, Europe, etc.
Product ID	Unique ID for each product.
Category	The general type of product (office supplies, furniture, technology, etc.).
Sub-Category	A specific subclass within the product category (*e.g.*, chairs, phones, etc.).
Product Name	The name of the product purchased by the customer.
Sales	The total value of sales in the sales order for the product (usually in units of currency).
Quantity	The number of products sold in the order.
Discount	The discount % applied to the product during the sale.
Profit	The profit earned from the sale of the product.
Shipping Cost	The cost incurred to ship the product to the customer.
Order Priority	The priority level of the order (*e.g.*, high, medium, low).

Several preprocessing techniques were performed on the dataset such as type conversion, cleaning, and checking missing values. The Date needs to be converted into "DD/MM/YYYY" format which can be used for grouping order-date with the correct format. The dataset needs to be checked for the missing records that cause incorrect analysis results.

Exploratory Data Analysis (EDA): is the first step in any data science project because it helps to understand the characteristics of the data, identify patterns, and reveal any potential issues or anomalies. The goal of the EDA in this study is to explore the sales data and identify any trends, seasonality, or other patterns that could be useful for demand forecasting [21]. Figure 2 shows the unique order-date format is "DD/MM/YYYY", and one needs to change the format to "YYYY/MM/DD" for better standardization, clarity, and sorting. This change follows the ISO 8601 standard, helping avoid confusion and making it easier to chronologically order the dates**.** Exploring dataset shows missing value for each column in the dataset. The column "Postal Code" has 41,296 missing values but other columns have not. If you are considering "Post Code" in the analysis, you need to replace missing values with Mean, Medium, and Mode or Remove row for high-quality data leads to more accurate and reliable analysis. Time series plot (Figure 2) shows a strong annual seasonality pattern, as well as an overall upward trend in sales over time.

Figure 3 shows the missing value for each column in the dataset. The column "Postal Code" has 41,296 missing values but other columns have not. If you are considering "Post Code" in the analysis, you need to replace missing values with Mean, Medium, and Mode or Remove row for high-quality data leads to more accurate and reliable analysis. Figure 4 clearly shows a strong annual seasonality pattern, as well as an overall upward trend in sales over time.

Checking for Forecasting Model: The modeling steps are described here. The logic step-by-step of choosing a forecasting model involves several critical stages, starting with data preprocessing to ensure the dataset is clean and choosing the best-suitable model for the forecasting. This process includes Algorithm-evaluating models, such as ARIMA, Exponential Smoothing, and Prophet.

Step 1: Initialize Parameters
Install and load the required R libraries: "**tidyverse**", "**forecast**", "**readxl**", "**scales**", and "**prophet**", to handle data manipulation, forecasting, and visualization.

- **tidyverse**: For data manipulation and visualization.
- **Forecast**: For ARIMA and ETS modeling.
- **readxl**: For importing Excel files.
- **Scales**: For formatting axis labels (like converting sales into millions).
- **Prophet**: For time-series forecasting using the Prophet model.

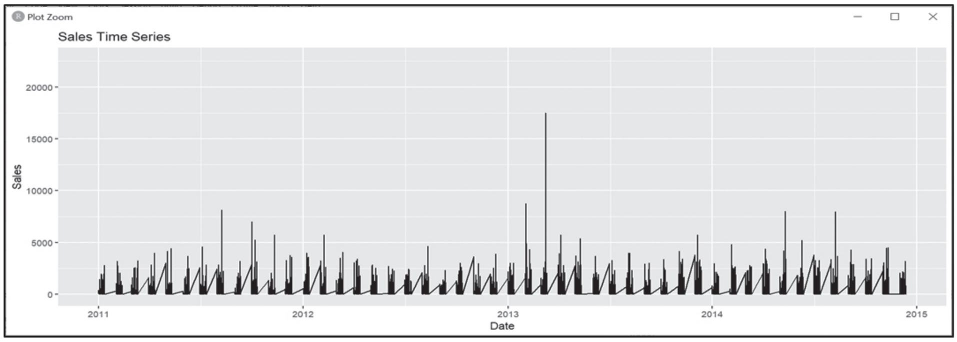

Figure 2. Time-series plot.

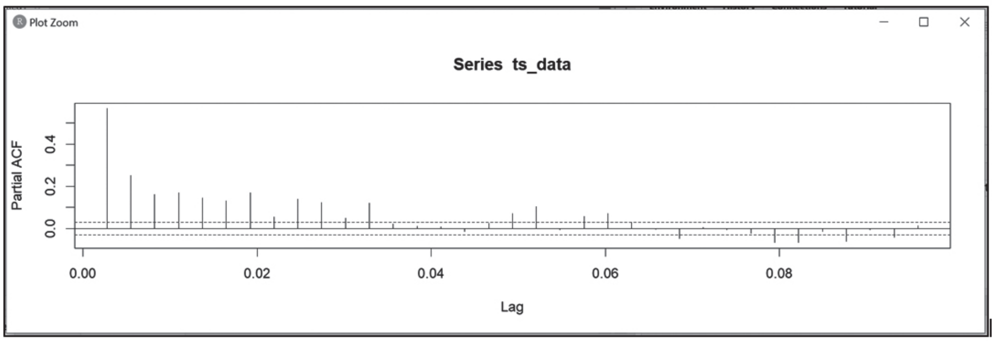

Figure 3. ACF and PACF plot.

Step 2: Load Data

Define the file path to the dataset that contains sales data and read it using the "**read_excel**" function.

Step 3: Preprocess Data

Verifying the "Order date" column and converting its column to a **standard date format (%d-%m-%Y)**

Extract the year and month from the "Order Date" column and create a new column.

Step 4: Aggregate Sales Data by Month

Group the data by Year Month to calculate the total sales by each month, storing it in a new data frame monthly sales.

Step 5: Prepare Time Series Object

Construct a time-series object (ts_data) data with frequency (12 monthly data) and start year and month (January 2011).

Step 6: Fit Models and Generate Forecasts:

- Automatically identify and fit the best ARIMA model using **auto.arima** and forecast sales for the next 12 months.
- Apply the Exponential Smoothing (ETS) method to build a forecasting model and generate forecasts for the next 12 months.
- Prepare data by renaming columns (Date → ds, Sales → yhat).
- Create and fit a Prophet forecasting model using the reformatted data and extend the date range by 12 months using "**make_future_dataframe".**

Step 7: Visualize the Forecast

- Use autoplot for ARIMA/ETS and ggplot for Prophet to display forecast results, including a line for predicted values (that for Prophet), a shaded ribbon for confidence intervals (yhat_lower and yhat_upper).
- Customize the plot with Title: "Sales Forecast for the Next 12 Months", X-axis: "Order Date" and Y-axis: "Sales (in Million)". Format the y-axis to display sales in millions using label_number with appropriate scaling and suffix.

ARIMA Model: The ARIMA (Autoregressive Integrated Moving Average) model is a popular and widely used time series forecasting model. It is a class of models that capture linear dependencies of the current value on the past value of the time series. ARIMA model is defined by three parameters (p, d, q) [13].

- p is the number of autoregressive terms,
- d is the number of non-seasonal differences needed for stationarity,
- q is the size of the moving average window.

The ARIMA model can be extended to include seasonal components to handle time-series data that exhibit seasonal patterns [14]. The benefit of the ARIMA model [15], [16] can handle non-stationary time series data, can model both short-term and long-term patterns, and is a very general model that can be adapted to many different time series.

Exponential Smoothing Model: Exponential Smoothing is another popular time series forecasting technique. The Holt-Winters exponential smoothing method is an extension of simple exponential smoothing that can handle time series with trend and seasonality.

The Holt-Winter's method uses three parameters:
1. α (Coefficient of level smoothing or base value): The parameter α determines the weight of the past values.
2. β (Coefficient of trend smoothing or trend value): The parameter β determines the degree of the recent trend values.
3. γ (Coefficient of seasonality smoothing or seasonal component): The parameter γ indicates the coefficient for the seasonal smoothing.

Exponential is simple to implement, requires fewer parameters compared to ARIMA, and can handle non-stationary data well.

Prophet Model: The Prophet model, developed by Facebook, is designed to handle an additive time-series model to analyze and predict trends in time-series data. It can manage multiple seasonality trends (such as daily, weekly, and yearly cycles), holidays, and other special events. It is user-friendly and particularly useful for quickly testing ideas and generating reports when the data has strong repeating patterns [21], [22].

Key Features of the Prophet:
- Seasonality and Holidays: Prophet automatically identifies seasonal patterns and includes holiday effects, making it suitable for business applications such as sales forecasting [23].
- Ease of Use: The model requires less preprocessing of data as compared to LSTM, which demands significant tuning and preparation of data [21].
- Performance: Prophet has shown better performance of accuracy metrics such as Mean Absolute Error (MAE) in forecasting complex and non-linear trends [22].

The basic equation that represents the Prophet model is:

$$y(t) = g(t) + s(t) + h(t) + \varepsilon_t \tag{1}$$

where $y(t)$ is the observed value at time t and $g(t)$ is the trend function that models non-periodic changes. In equation 1, $s(t)$ is the seasonality component that models periodic changes (*e.g.*, daily, weekly, yearly), $h(t)$ represents the effects of holidays or special events and ε_t is the error term that accounts for any noise in the observations.

Model Evaluation Metrics: To evaluate the performance of the forecasting models, we can use various metrics.

Mean Absolute Error (AME): Measures the average absolute difference between the actual and forecasted values. It's a measure of overall accuracy but does not consider the direction (whether the forecast was over or under the actual value:

$$\text{AME} = \frac{1}{n}\sum_{i=1}^{n} |\text{Actual}_i - \text{Forecast}_i| \tag{2}$$

The advantages are that it is easy to compute and gives an overall view of forecast error. The disadvantage is that it can be misleading if you have values with different scales, as it doesn't consider the size of the errors relative to the values themselves. The main use case is, it is helpful when you want a general idea of how far off the forecasts are from the actual values on average.

Mean Absolute Percentage Error (MAPE): Measures the average percentage error of the forecasts. It's a relative measure, meaning it compares the magnitude of the forecast errors relative to the actual values:

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n} \left|\frac{\text{Actual}_i - \text{Forecast}_i}{\text{Actual}_i}\right| \times 100 \tag{3}$$

The advantage is that it gives a percentage error that is easy to understand and compares across different datasets, even if they have different scales. It's sensitive to small actual values—if an actual value is close to zero, MAPE can become very large or undefined. It also doesn't handle negative values well (this can be an issue for certain types of data like profit and loss, where both positive and negative values exist). MAPE is widely used for understanding the accuracy of a model in a more interpretive way, especially when comparing models across datasets with different magnitudes.

Root Mean Squared Error (RMSE): Measures the average squared difference between the actual and forecasted values. Squaring the errors penalizes larger errors more heavily, making RMSE sensitive to large forecast errors:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (\text{Actual}_i - \text{Forecast}_i)^2} \tag{4}$$

RMSE is widely used because it has the same units as the data itself, making it intuitive and useful for measuring the overall magnitude of the forecast error. Like MAPE, RMSE can be disproportionately influenced by outliers (large errors). It is also sensitive to the scale of the data, so it can be less useful when comparing models on datasets with different units or scales. RMSE is helpful when you want to understand how many errors, on average, are made by your model, with larger errors having a more significant impact.

R-squared (R2): Measures the proportion of the variance in the dependent variable that is predictable from the independent variables. It is often used to assess how well the forecasting model has captured the underlying patterns in the data:

$$R\text{-}root = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\text{Residual}_i)^2} \tag{5}$$

where Residuals are:

$\text{Residual}_i = Actual_i - Forecast_i$

R-squared is a common metric for assessing the performance of the model's error distribution and ensures that the model fits well to the data without overfitting or underfitting. Like RMSE, it is influenced by outliers and doesn't provide insight into the direction of the forecast errors. R-root is in diagnosing whether your model is consistently underestimating or overestimating the values, as it is based on the residuals. Through this comprehensive analysis, we can determine the most suitable forecasting model for the given retail demand data based on its performance and the specific business requirements [18], [19], [20].

III. RESULT AND EVALUATION

This section presents an analysis of time-series data results by using ARIMA, Exponential Smoothing, and Prophet models fitted on the Global Super Store Dataset for demand forecasting in retail. The key findings of the ARIMA model result are:

After applying the ARIMA models to the dataset, Figure 4 depicts the sales forecast for the next 12 months of the time series. The *x*-axis indicates the period (monthly), and the *y*-axis indicates the sales amount in millions. The blank line represents past sales performance from 2011 to 2015 which is noticeable patterns, such as seasonality or trends, are essential for understanding future forecasts. And the blue line indicates the forecasted sales in the next 12 months till 2016. Upward trending implies growth. However, if it is going downwards, that may indicate the possibility of a fall in sales. It has cyclic patterns that occur at specific intervals; these are seasonal patterns. Detecting these patterns can also help one determine peak periods of sales and prepare for them.

Figure 5 depicts the sales forecast for the next 12 months of the time series by making use of Exponential smoothing. The *x*-axis represents the period (monthly), and the *y*-axis represents the sales amount in millions. Data from 2011 to 2015 are

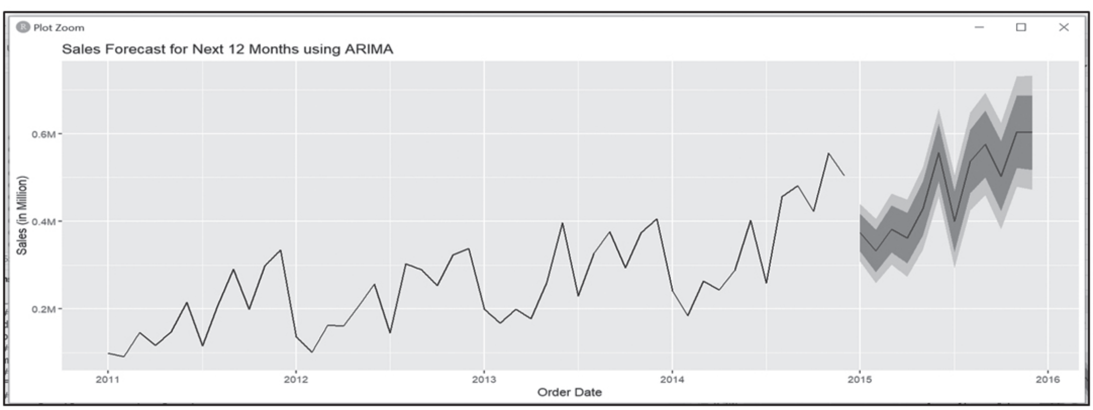

Figure 4. Sales forecast for the next 12 months using the ARIMA model.

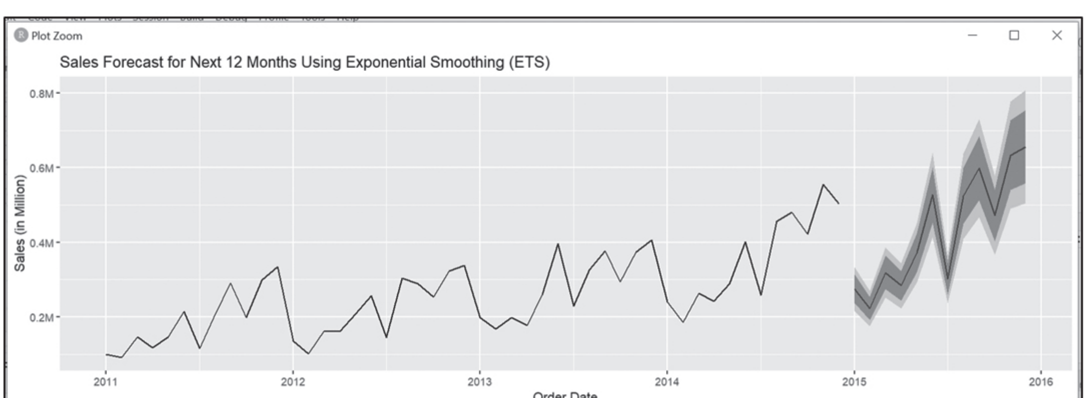

Figure 5. Sales forecast for the next 12 months using the Exponential Smoothing model.

shown as overall trends and seasonal patterns of historical sales performance which is very important in making proper forecasts. The blue line indicates the forecast from 2015 to 2016. The graph might display seasonality, where certain times of the year have higher or lower sales. These trends allow companies to plan for inventory and marketing.

Figure 6 depicts the sales forecast for the next 12 months of the time series by applying the Prophet model. The *x*-axis represents the period (monthly), and the *y*-axis represents the sales amount in millions. The lines show general trends and seasonal patterns in sales that are important for making accurate forecasts. The width of the shaded area indicates increasing uncertainty about the sales predictions as we move further into the future. The close fit of the black line, historical data, and the beginning of the blue line, forecasted data, suggests that the Prophet model is capturing the trend and seasonality in the historical data well. However, the widening confidence interval points out that the uncertainty in the forecast is increasing, a common phenomenon as we move further into the future.

The key findings of the analyzed sales distribution across by-month results are shown in Fig. 7. The box plot displays the key highlight of trends and patterns for sales distribution across by month. The x-axis represents the months from Jan to Dec, while the y-axis shows the sales values. Each box illustrates the inter-quartile range (IQR) for sales within that month, the whiskers span to capture the sales values within 1.5 times the IQR. Outliers, which represent the values of sales far apart from the rest, appear as individual points beyond the whiskers. This visualization helps detect the variability in sales, trends and anomalies throughout the year. The key findings of identifying seasonal patterns across time are shown in Figure 8.

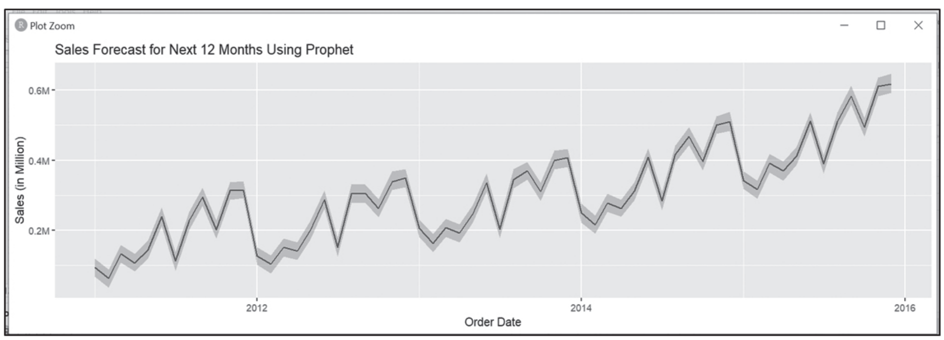

Figure 6. Sales forecast for the next 12 months using the Prophet model.

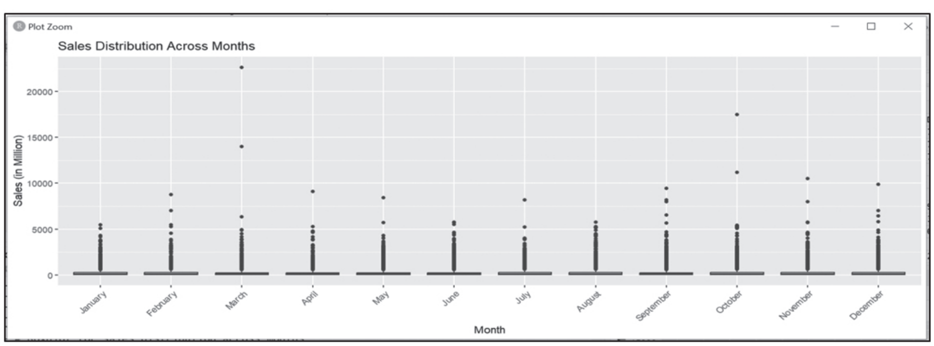

Figure 7**.** Box plot across by month.

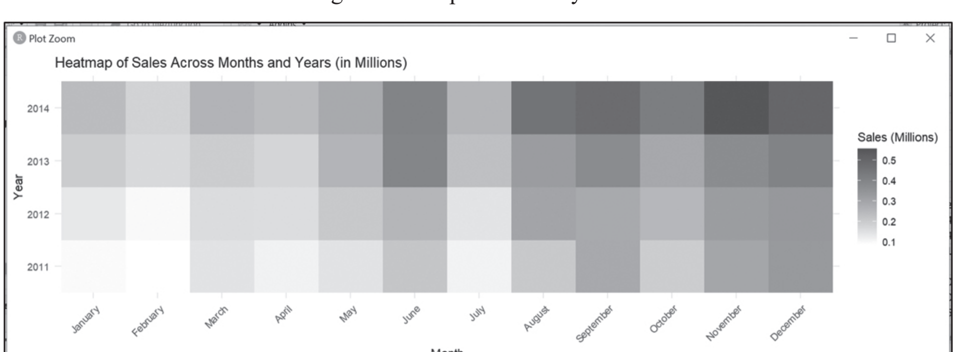

Figure 8. Heatmap sales across time.

The heatmap displays monthly sales data, with deeper blue shades representing greater sales and lighter shades signifying lower sales. The color bar is between 0.1 and 0.5 million sales. The heat map shows that sales peaked during November and December. Thus, it indicates seasonal increased sales during November and December of the year. Therefore, retailers need to prepare stock in that period to avoid going out of stock. The key findings for understanding the long-term growth of sales results are shown in Fig. 9.

The plot of the moving average of sales (Fig. 10) reveals a steady rise between 2011 and 2015, thus indicating long-term growth. While short-term fluctuations are evident, the moving average smooths out these variations to reveal overall sales progression. Thus, periods of notable growth, such as between 2013 and 2014, can be marked with a steeper sales increase. This visualization presents an easy way to understand the long-term performance while showing seasonal patterns for further analysis.

IV. MODEL PERFORMANCE COMPARISON

To determine the most effective model for retail demand forecasting, we evaluated the three models using key performance metrics: Mean Absolute Error (AME), Mean Absolute Percentage Error (MAPE), Root Mean Squared Error (RMSE), and R-squared.

Table 2--THE PERFORMANCE METRICS OF THREE MODELS

Model	*AME*	*MAPE*	*RMSE*	*R-squared*
ARIMA Metrics	7.316117	1.426464	9.414287	-4.317755
Exponential Smoothing Metrics	7.405794	1.516789	9.466270	-4.476307
Prophet Metrics	4.092523	8.198102	5.166865	5.687250

Thus, Prophet usually overcomes ARIMA and Exponential Smoothing in most of the metrics. It has the smallest Absolute Mean Error (AME), and Root Mean Square Error (RMSE) values, as well as the highest R-squared value, which means that this model is the most accurate for forecasting in this scenario. The MAPE for ARIMA is the lowest, while its RMSE is competitive. Therefore, ARIMA would appear that the general accuracy of the highest model. The R-squared of the ARIMA model is larger than the R-squared of the Exponential Smoothing; hence, the ARIMA model can better describe the variation in the data. Overall, Exponential Smoothing performs worse on all. Ultimately, Prophet and ARIMA both show strong performance, with Prophet excelling in most metrics and ARIMA having notable accuracy in MAPE and RMSE. Different combinations of *p, d,* and *q* values that were tested and their corresponding AIC valueswere recorded while evaluating ARIMA model. The optimal ARIMA model can then be chosen based on the lowest AIC (values are given in Table 3).

Figure 9. Cumulative sales plot.

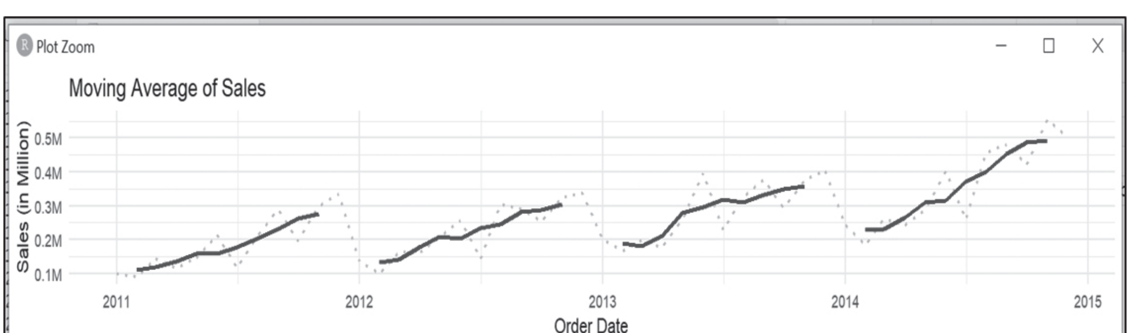

Figure 10. Moving average plot.

TABLE 3 -- THE RESULT OF THE ARIMA MODEL WITH DIFFERENT (p, d, q)

Order No	*ARIMA (p, d, q)*	*AIC*	*Order No*	*ARIMA (p, d, q)*	*AIC*
1	(0,0,0)	23,045.34	17	(2,0,0)	22,966.80
2	(0,0,1)	22,973.49	18	(2,0,1)	22,865.13
3	(0,0,2)	22,971.95	19	(2,0,2)	22,860.18
4	(0,0,3)	22,965.17	20	(2,0,3)	22,821.22
5	(0,1,0)	23,466.80	21	(2,1,0)	23,146.54
6	(0,1,1)	22,854.18	22	(2,1,1)	22,833.18
7	(0,1,2)	22,846.06	23	(2,1,2)	22,805.91
8	(0,1,3)	22,833.21	24	(2,1,3)	22,772.27
9	(1,0,0)	22,966.39	25	(3,0,0)	22,956.20
10	(1,0,1)	22,869.99	26	(3,0,1)	22,848.48
11	(1,0,2)	22,862.91	27	(3,0,2)	22,812.66
12	(1,0,3)	22,847.60	28	(3,0,3)	22,787.41
13	(1,1,0)	23,278.48	29	(3,1,0)	23,088.15
14	(1,1,1)	22,848.43	30	(3,1,1)	22,828.73
15	(1,1,2)	22,843.59	31	(3,1,2)	22,802.02
16	(1,1,3)	22,814.44	32	(3,1,3)	22,809.10

The ARIMA modeling results suggest that the ARIMA (2, 1, 3) model is the most suitable for forecasting the given time series, as it yields the lowest AIC value of 22,772.27. This model configuration indicates that the time series requires first differencing ($d = 1$) to achieve stationarity, utilizes two lagged values ($p = 2$) for the autoregressive part, and incorporates three previous forecast errors ($q = 3$) in the moving average component. The low AIC value suggests that this model achieves an optimal fit while avoiding unnecessary complexity. Therefore, the ARIMA (2, 1, 3) model is the best candidate for future predictions based on its performance in comparison to other configurations tested.

V. CONCLUSION

The study indicates that ARIMA, Exponential Smoothing, and Prophet models are highly efficient in terms of retail demand forecasting. In general, the models have strong points, but Prophet was considered the most efficient in general performance, especially for seasonal and promotional variation datasets. The type of characteristic of the sales data should influence the selection of the forecast model. Therefore, the frequent seasonality and other special occasions characterize the trends of interest, Prophet will likely be a better fit. However, for relatively straightforward trends where such external disturbances are minimal, ARIMA or Exponential Smoothing can be useful also. It also highlights the importance of using the right forecasting model to suit the nature of data and the needs of a specific retailer. The selection of the right model can be of great help in optimizing business activities, improving customer satisfaction, and gaining competitive advantage in an increasingly complex and dynamic retail market.

REFERENCES

[1] Z. H. Kilimci *et al.*, "An improved demand forecasting model using deep learning approach and proposed decision integration strategy for supply chain," *Complexity*, p. 9067367, Jan. 2019, doi: 10.1155/2019/9067367.

[2] "Forecasting sales in the supply chain: Consumer analytics in the big data era - ScienceDirect." Accessed: Dec. 06, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0169207018301523?via%3Dihub

[3] P. Gupta *et al.*, "Implementation of demand forecasting – A comparative approach," *J. Phys. Conf. Ser.*, Vol. 1714, no. 1, p. 012003, Jan. 2021, doi: 10.1088/1742-6596/1714/1/012003.

[4] "Demand seasonality in retail inventory management - ScienceDirect." Accessed: Dec. 06, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0377221714002690?via%3Dihub

[5] N. Liul *et al.*, "Sales Forecasting for Fashion Retailing Service Industry: A Review," *Math. Probl. Eng.*, vol. 2013, pp. 1–9, 2013, doi: 10.1155/2013/738675.

[6] A. K. Pundir *et al.*, "Machine learning for revenue forecasting: A case study in retail business," *Proc. 11th IEEE Annual Information Technology, Electronics and Mobile Communication Conf.*, Nov. 2020, pp. 0201–0207. doi: 10.1109/IEMCON51383.2020.9284819.

[7] K. J. Atmaja and I. B. G. Anandita, "Sales forecasting system using single exponential smoothing," Vol. 4, no. 36, 2021.

[8] X. Qi *et al.*, "From known to unknown: knowledge-guided transformer for time-series sales forecasting in Alibaba," Sep. 22, 2021, *arXiv*: arXiv:2109.08381. doi: 10.48550/arXiv.2109.08381.

[9] "Customer reviews for demand distribution and sales nowcasting: a big data approach", *Annals of Operations Research*. Accessed: Dec. 06, 2024. [Online]. Available: https://link.springer.com/article/10.1007/s10479-016-2296-z

[10] C. S. Bojer *et al.*, "Retail promotion forecasting: A comparison of modern approaches," in *Advances in Production Management Systems. Towards Smart Production Management Systems*, Vol. 567.

[11] F. Ameri *et al.* Eds., *Proc. IFIP Advances in Information and Communication Technology*, Vol. 567, Springer 2019, pp. 575–582. doi: 10.1007/978-3-030-29996-5_66.

[12] "Seasonality: What it means in business and economics, Examples," *Investopedia*. Accessed: Dec. 06, 2024. [Online]. Available: https://www.investopedia.com/terms/s/seasonality.asp

[13] S. Hoda *et al.*, "Consumer Demand Modeling During COVID-19 Pandemic," *Proc. IEEE International Conference on Bioinformatics and Biomedicine*, Seoul, Korea (South): IEEE, Dec. 2020, pp. 2282–2289. doi: 10.1109/BIBM49941.2020.9313281.

[14] M. R. Abonazel and A. I. Abd-Elftah, "Forecasting Egyptian GDP using ARIMA models," *Rep. Econ. Finance*, Vol. 5, no. 1, pp. 35–47, 2019, doi: 10.12988/ref.2019.81023.

[15] G. M. N. A. Geetha, *Time-series modelling and forecasting: modelling of rainfall prediction using ARIMA model*.

[16]. M. Paulescu *et al.*, "Time Series Forecasting," in *Weather Modeling and Forecasting of PV Systems Operation*, Eds., London: Springer, 2013, pp. 181–201. doi: 10.1007/978-1-4471-4649-0_6.

[17] S. Siami-Namini, N. Tavakoli, and A. S. Namin, "A Comparative Analysis of Forecasting Financial Time Series Using ARIMA, LSTM, and BiLSTM," Nov. 21, 2019, *arXiv*: arXiv:1911.09512. doi: 10.48550/arXiv.1911.09512.

[18] T. Srivastava, "12 Important model evaluation metrics for machine learning everyone should know (Updated 2025)," *Analytics Vidhya.* Accessed: Dec. 06, 2024. [Online]. Available: https://www.analyticsvidhya.com/blog/2019/08/11-important-model-evaluation-error-metrics/

[19]. N. F. Aurna *et al.*, "Time series analysis of electric energy consumption using autoregressive integrated moving average model and Holt Winters model," *TELKOMNIKA Telecommun. Comput. Electron. Control*, Vol. 19, no. 3, p. 991, Jun. 2021, doi: 10.12928/telkomnika.v19i3.15303.

[20] "Demand forecasting with four-parameter exponential smoothing" *ScienceDirect,* Accessed: Dec. 06, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0925527316301839?via%3Dihub

[21] "Very short-term electricity demand forecasting using adaptive exponential smoothing methods | IEEE Conference Publication | *IEEE Xplore*." Accessed: Dec. 06, 2024. [Online]. Available: https://ieeexplore.ieee.org/document/7086716

[22] H. Verma, A. Singh, S. Avasthi and T. Sanwal, "AI-based agriculture application for crop recommendation and guidance system for farmers", *Proc. IEEE Int'l Conf. Communication, Computer Sciences and Engineering,* May 2024, pp. 401-406.

[23] S. Avasthi, S.L. Tripathi, T. Sanwal and V. Dwivedi, "Multi-Channel sales forecasting in a Kenyan retail environment", *Proc. IEEE 2nd Int'l Conf. Computational Intelligence, Communication Technology and Networking,* Feb 2025, pp. 141-146.

Dr. Sandhya Avasthi (SM, IEEE) completed Ph.D. in the area of Machine Learning and Text Mining from Amity University, Noida. She is an Associate Professor in the Department of Computer Science Engineering, ABES Engineering College (Dr Abdul Kalam Technical University), Ghaziabad, India. She did her M.Tech in Computer Science and Engineering from UP Technical University, Lucknow and B.E. in Computer Science & Engineering from Dr BR Ambedkar University, Agra. She has over 19 years of teaching experience and is an active researcher in the field of machine learning and data mining. Her research interests include Natural Language Processing, Information Extraction, Information Retrieval, Data Science and Business Intelligence. She has published numerous research articles in refereed international journals, conference proceedings and book chapters. She is also contributing as a reviewer in Springer Journal (JIMS, JAIHC), IEEE conferences, Hindawi and Aging International Journal. She is associated as a member ACM and is continuously involved in different professional activities along with academic work.

Dr. Shweta Roy, AMIETE, MIEEE is working as Senior Assistant professor (Selection Grade), Department of Computer Science and Engineering, ABESEC, Ghaziabad, UP. She did her B.Tech in Computer science and Engineering from Magadh University, Bodhgaya, Bihar, India. She did her Masters in computer science and engineering from Abdul Kalam Technical University, Lucknow, UP., India.

She obtained PhD from Alfalah University, Faridabad, Haryana, India. She has over 19 years of teching experience. She is doing her research in the field of deep learning algorithms to obtain new set of classifiers for the share market. She has zeal to teach new technologies and subjects. She has around 12 research papers published in national and international journals. She has also written book chapters in her research domain.