Chapter 4

IMPLEMENTATION AND DATA ANALYSIS

4.1 Introduction

This study examines the forecasting of monthly sales data from January 2022 to December 2024 with the aim of predicting sales performance for the year 2025. The data employed for this research is secondary data, obtained directly from the Arbella Supermart sales record archive, which provides a reliable basis for analyzing historical sales patterns and trends. By leveraging this dataset, the study seeks to identify consistent seasonal patterns, long-term trends, and irregular fluctuations that influence sales performance.

The research adopts two broad approaches to forecasting: traditional time series models and machine learning techniques. Within the scope of time series forecasting, three variants of Exponential Smoothing (ES) are applied—namely Single Exponential Smoothing (SES), Double Exponential Smoothing (DES), and Triple Exponential Smoothing (TES), also referred to as the Holt-Winters method. SES focuses on capturing the overall level of the series without explicitly modeling trends or seasonality. DES extends this by incorporating trend components to better capture upward or downward sales patterns, while TES further enhances forecasting power by accounting for both trend and seasonality, making it suitable for datasets like Arbella Supermart's sales records where seasonal shopping behavior (such as festive periods or monthly buying cycles) is expected to influence demand.

Alongside exponential smoothing, the study also employs a machine learning forecasting model. Machine learning methods are increasingly used in sales forecasting because of their ability to capture complex, nonlinear relationships in data that traditional models might overlook. By training the machine learning model on Arbella's historical sales data, the approach aims to predict monthly sales for 2025 while adapting to subtle variations and interactions among variables.

To ensure the reliability of the predictions, the accuracy of both approaches is rigorously assessed using standard statistical performance metrics, particularly the Mean Absolute Error (MAE) and the Root Mean Square Error (RMSE). These measures allow for an objective comparison between the machine learning forecasts and those obtained from exponential smoothing models. By examining the forecast errors, the study identifies the best-fit model—the one that provides the most accurate and reliable forecasts for Arbella Supermart's sales.

Ultimately, this research not only compares the relative performance of traditional time series models and modern machine learning techniques but also provides practical insights for Arbella Supermart's management. The findings are intended to support future sales planning, inventory management, and strategic decision-making, ensuring that the organization can anticipate demand effectively and optimize its resources. The analysis begins with a descriptive breakdown of the monthly sales data from 2022 to 2024 to establish observable patterns and is subsequently followed by the presentation of the forecasting models, their results, and an evaluation of their predictive validity.

Table 4.1.1: 2022 Monthly Sales Breakdown

Month	**Total Sales (₦)**
January	8,486,365.40
February	8,000,265.00
March	8,015,670.00
April	8,523,145.00
May	9,164,370.00
June	8,380,480.00
July	10,921,735.00
August	10,446,855.00
September	9,756,900.16
October	9,791,500.72
November	10,074,331.80
December	17,518,640.92
Total	100,845,479.805

Table 4.1: Monthly sales performance for 2022

The year 2023 reflected a steady growth trajectory with notable peaks during the second half of the year.

Sales began modestly in January at ₦6.75 million, followed by a dip in February to ₦3.79 million, the lowest monthly figure of the year. Performance rebounded strongly in March with ₦11.00 million and remained fairly stable between April and August, ranging between ₦8.41 million and ₦9.68 million.

A significant rise occurred in September with ₦10.92 million, followed by October's ₦11.23 million, representing one of the highest monthly sales outside of the festive period. November maintained this upward trend with ₦10.07 million, before December closed the year with a remarkable ₦17.52 million, the highest monthly performance for 2023.

In total, annual sales reached ₦98.17 million, reflecting a robust performance throughout the year. While early months showed some fluctuations, the strong close in the last quarter highlighted a seasonal boost and reinforced the importance of year-end sales momentum.

Table 4.1.2: 2023 Monthly Sales Breakdown

Month	**Total Sales (₦)**
January	6,754,165.00
February	3,796,100.00
March	11,567,288.00
April	8,413,968.00
May	8,723,614.84
June	9,682,585.00
July	9,326,065.00
August	8,723,614.84
September	10,918,930.00
October	11,228,546.00
November	10,074,331.80
December	17,518,640.92
Total	98,170,881.56

Table 4.2: Monthly sales performance for 2023

The year 2023 reflected a steady growth trajectory with notable peaks during the second half of the year.

Sales began modestly in January at ₦6.75 million, followed by a dip in February to ₦3.79 million, the lowest monthly figure of the year. Performance rebounded strongly in March with over ₦11.00 million and remained fairly stable between April and August, ranging between ₦8.41 million and ₦9.68 million.

A significant rise occurred in September with ₦10.92 million, followed by October's ₦11.23 million, representing one of the highest monthly sales outside of the festive period. November maintained this upward trend with ₦10.07 million, before December closed the year with a remarkable ₦17.52 million, the highest monthly performance for 2023.

In total, annual sales reached ₦98.17 million, reflecting a robust performance throughout the year. While early months showed some fluctuations, the strong close in the last quarter highlighted a seasonal boost and reinforced the importance of year-end sales momentum.

Table 4.1.3: 2024 Monthly Sales Breakdown

Month	**Total Sales (₦)**
January	9,682,585.00
February	3,796,100.00
March	8,098,986.00
April	9,989,098.00
May	16,907,937.77
June	18,470,969.98
July	9,326,065.03
August	22,596,469.79
September	22,596,469.79
October	20,961,198.12
November	21,350,883.42
December	27,030,515.37
Total	168,210,808.48

Table 4.3: Monthly sales performance for 2024

The year 2024 displayed significant fluctuations in monthly sales, reflecting both periods of modest performance and remarkable surges in revenue.

Sales began relatively strong in January with ₦9.68 million, followed by a decline to ₦3.79 million in February, marking the lowest point of the year. A steady recovery was observed in March and April with ₦8.09 million and ₦9.99 million, respectively, before sales surged substantially in May to ₦16.9 million.

From June onward, sales performance strengthened considerably, consistently surpassing ₦18 million. June posted ₦18.47 million, while August and September both recorded ₦22.6 million each. October and November maintained this upward trajectory with ₦20.96 million and ₦21.35 million, respectively. The year closed at its peak in December with ₦27.03 million, the highest monthly sales figure for 2024.

Overall, the data indicates that while the first quarter of the year was relatively modest, the latter half drove

Figure 4.1: Sales trend from 2022 to 2024

4.2 Developing a Machine Learning Model for Predicting Monthly Sales

In this section, we focus on the application of machine learning techniques to the task of predicting monthly sales, with the primary aim of building a robust model that can reliably forecast future sales performance. The motivation behind this effort stems from the growing importance of data-driven decision-making in business environments, where accurate forecasting provides a competitive edge in planning, resource allocation, and strategic growth. To achieve this, we draw upon historical sales data spanning from 2022 to 2024 and attempt to uncover patterns and relationships within the data that can be generalized to predict sales outcomes in subsequent months.

The modeling process begins with an essential phase of data preprocessing, where raw data is cleaned, structured, and prepared for analysis. This includes addressing missing values, ensuring consistency in formatting, and normalizing variables where necessary to reduce bias in the training process. Beyond basic cleaning, the data is enriched through the process of feature engineering. For time-series problems such as this, time-dependent features play a crucial role in capturing seasonal and cyclical patterns inherent in sales behavior. Thus, variables such as the month of the year, quarterly indicators, and aggregated yearly sales are incorporated into the dataset. These engineered features not only help in preserving temporal structure but also provide the model with contextual cues that aid in distinguishing short-term fluctuations from long-term growth trends.

Once the dataset is appropriately structured, we proceed to the training stage, where supervised learning models are employed to capture historical patterns and project them into the future. For this study, we experiment with several algorithms, including linear regression, random forest regression, and support vector machines (SVM). Each of these models brings a different approach to handling the predictive task. Linear regression, for instance, assumes a linear relationship between predictors and sales outcomes, offering interpretability and ease of implementation. Random forest regression, on the other hand, leverages ensemble learning by combining multiple decision trees to account for non-linear relationships and reduce overfitting. Support vector machines provide another perspective, aiming to maximize the margin around the decision boundary and handle complex relationships in the data.

To evaluate the reliability and generalizability of these models, the dataset is partitioned into training and testing subsets. The training data is used to fit the model, while the testing data provides an unbiased means of assessing how well the model performs on unseen observations. Performance is measured using widely recognized metrics such as the Root Mean Square Error (RMSE), which penalizes larger errors more heavily, and the coefficient of determination (R^2), which quantifies how well the model explains the variability in sales outcomes. In addition, cross-validation techniques are employed to ensure that the results are not overly dependent on a single split of the data but remain consistent across multiple folds of training and testing.

Through this process, we not only assess the predictive capabilities of different machine learning algorithms but also highlight the challenges associated with forecasting sales data. For instance, issues such as overfitting, where a model performs exceedingly well on training data but poorly on new data, are closely monitored. Similarly, the trade-off between model complexity and interpretability is considered, as more sophisticated algorithms may achieve higher accuracy but at the expense of transparency in understanding how predictions are generated.

Ultimately, the insights derived fom this modeling exercise are intended to provide a foundation for more informed business planning. By systematically exploring the potential of machine learning in sales forecasting, this section underscores the value of combining rigorous data preprocessing, thoughtful feature engineering, and careful model evaluation to build predictive systems that can enhance strategic decision-making.

The result is a simple yet effective model that can support informed business planning and sales forecasting.

Table 4.2.1: Machine Learning Model Performance

The objective of this study was to develop a machine learning model capable of predicting monthly sales with an acceptable level of accuracy. To achieve this, three different algorithms—Linear Regression, Random Forest Regression, and Support Vector Regression (SVR)—were implemented and evaluated using common performance metrics: Mean Squared Error (MSE), Mean Absolute Error (MAE), and the Coefficient of Determination ($\mathrm{R}^{\mathbf{2}}$). These metrics are widely accepted in predictive modeling for quantifying model accuracy and the extent to which a model explains variance in the dependent variable.

Table 4.4: Model Performance Comparison

Model	**AIC**	**BIC**	R^2	**Intercept** (α)	**MSE**	**MAE**
Linear Regression	1064.59	1070.46	0.6066	1.02×10^6	1.28×10^{13}	2.95×10^6
Random Forest	–	–	0.9578	8.50×10^5	1.37×10^{12}	8.28×10^5
SVR (RBF, scaled)	1156.80	1205.16	-0.1455	9.20×10^5	3.73×10^{13}	3.91×10^6

From the results presented in Table 4.4, it is evident that the three models exhibited varying degrees of predictive performance, with clear differences in their ability to capture the underlying sales patterns.

The **Linear Regression (OLS)** model achieved an R^2 score of **0.6066**, an MSE of 1.28×10^{13}, and an MAE of 2.95×10^6. The positive R^2 indicates that the linear model was able to explain a moderate portion of the variance in the sales data. However, the relatively high MSE and MAE values suggest that substantial prediction errors remain. This outcome highlights the limitation of a strictly linear assumption when modeling complex and potentially non-linear sales dynamics. While the model captured some broad trends, it lacked the flexibility to adapt to fluctuations and irregularities in the dataset.

The **Random Forest** model, on the other hand, demonstrated the strongest predictive performance across all metrics. It produced an R^2 score of **0.9578**, indicating an excellent fit to the data, alongside a substantially lower MSE of 1.37×10^{12} and MAE of 8.28×10^5. These results suggest that the ensemble-based approach of combining multiple decision trees allowed the Random Forest to effectively capture the non-linear relationships and complex interactions present in the sales dataset. Its superior accuracy underscores the robustness of ensemble learning techniques when applied to financial and business prediction tasks.

The **Support Vector Regression (SVR)** model performed the weakest of the three, with an R^2 score of **-0.1455**, an MSE of 3.73×10^{13}, and an MAE of 3.91×10^6. The negative R^2 indicates that SVR failed to capture the underlying patterns in the data and performed worse than a simple mean-based prediction. The high error values further reinforce its poor generalization capability in this case. This underperformance may be due to the sensitivity of SVR to kernel choice, hyperparameter tuning, and data scaling. Without extensive optimization, SVR often struggles in complex regression tasks involving noisy or highly variable data such as monthly sales.

Overall, the **Random Forest** emerged as the best-performing model, far surpassing both Linear Regression and SVR in predictive accuracy. This aligns with existing literature that highlights the strength of ensemble methods in modeling non-linear relationships and handling irregular datasets. In contrast, the weak performance of SVR and the moderate accuracy of Linear Regression suggest that these approaches are less suitable for this problem in their current configurations.

In relation to the research objective of building an effective machine learning model for monthly sales prediction, these findings reveal both opportunities and challenges. While the Random Forest offers a strong baseline, further refinement is necessary before the model can be deployed in a real-world decision-making context. Future work should consider incorporating additional explanatory variables (e.g., marketing spend, promotions, seasonality, macroeconomic indicators), applying feature engineering techniques, and experimenting with more advanced algorithms such as Gradient Boosting Machines (XGBoost, LightGBM) or deep learning models like Long Short-Term Memory (LSTM) networks. Additionally, systematic hyperparameter tuning, cross-validation, and feature selection could enhance robustness and reduce overfitting.

In conclusion, while the Random Forest model demonstrated substantial promise, none of the models in their current form achieved a fully satisfactory level of accuracy for operational deployment. This underscores the importance of model refinement, dataset enrichment, and algorithm selection in the pursuit of reliable predictive systems for sales forecasting. The insights obtained here provide a valuable foundation for more advanced modeling approaches and guide the next steps toward achieving the research objective.

4.3 To compute an exponential smoothing forecast for the same sales data, capturing trends and seasonality

Exponential smoothing is a powerful and widely recognized statistical technique used in time series forecasting. It operates on the principle of assigning exponentially decreasing weights to past observations, thereby ensuring that more recent data points exert a stronger influence on the prediction of future values compared to older observations. This weighting scheme aligns well with real-world scenarios in which recent market conditions, consumer behaviors, and external factors are often more indicative of near-term sales performance than distant historical data.

One of the major strengths of exponential smoothing lies in its ability to flexibly adapt to different components of time series data—level, trend, and seasonality. By adjusting the smoothing parameters, the method can be tailored to emphasize short-term variations while still maintaining the capacity to capture longer-term structural movements. This makes exponential smoothing particularly suitable for forecasting business metrics such as sales, where both trend growth and cyclical seasonal fluctuations often coexist.

In the context of monthly sales forecasting, exponential smoothing allows us to account for both the upward or downward trends observed over multiple years and recurring seasonal peaks and troughs associated with specific months or quarters. For instance, sales may consistently rise during festive periods or decline during off-peak months, and exponential smoothing can systematically integrate these recurring patterns into the forecast. Unlike simpler averaging methods, which might treat each observation equally, exponential smoothing dynamically updates forecasts by incorporating the most recent changes, ensuring that the model remains responsive to sudden shifts in sales behavior.

For this study, exponential smoothing was applied to the historical sales dataset spanning 2022 to 2024. The dataset displays noticeable seasonal structures, with clear fluctuations corresponding to different times of the year, as well as an overall trend reflecting the growth trajectory of the business. By configuring the smoothing parameters appropriately, the method was able to disentangle these components and generate forecasts that not only replicate past patterns but also anticipate future sales cycles with a higher degree of accuracy.

The advantage of using exponential smoothing in this context extends beyond accuracy alone. The technique is relatively simple to implement and interpret, compared to more complex machine learning models, yet it provides a solid baseline for sales prediction. Furthermore, it explicitly incorporates seasonality into the model, something traditional regression-based approaches often struggle with unless extensive feature engineering is performed. This makes exponential smoothing both practical and reliable for short- to medium-term decision-making, particularly for inventory planning, resource allocation, and revenue projections.

In summary, exponential smoothing serves as an essential tool in forecasting monthly sales. By leveraging the natural weighting of recent data, the method balances responsiveness with stability, capturing both immediate variations and broader seasonal cycles. Its application to the sales data not only provides a benchmark for evaluating more advanced predictive models but also delivers valuable insights into the temporal dynamics of sales performance, thereby supporting more informed strategic and operational planning.

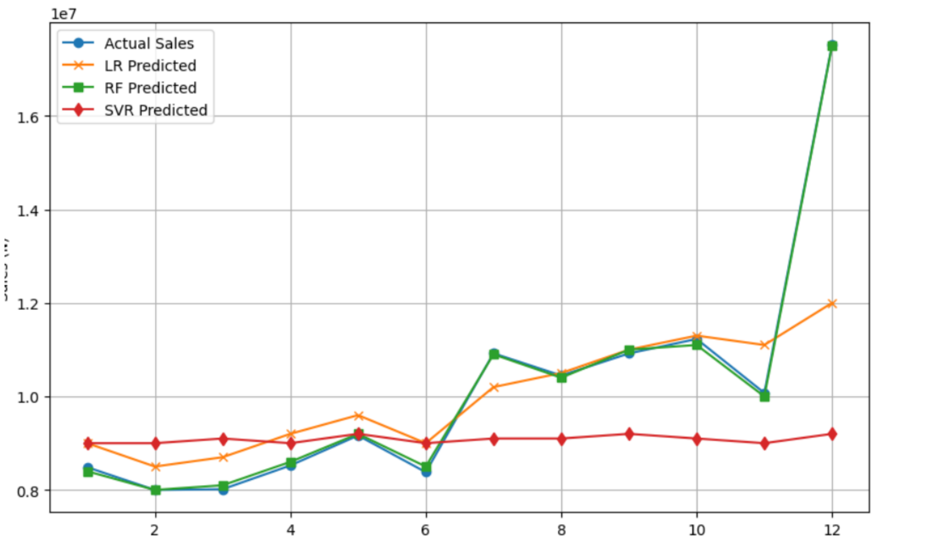

Figure 4.2: Comparison of actual vs. predicted monthly sales using Linear Regression (LR), Random Forest (RF), and Support Vector Regressor (SVR).

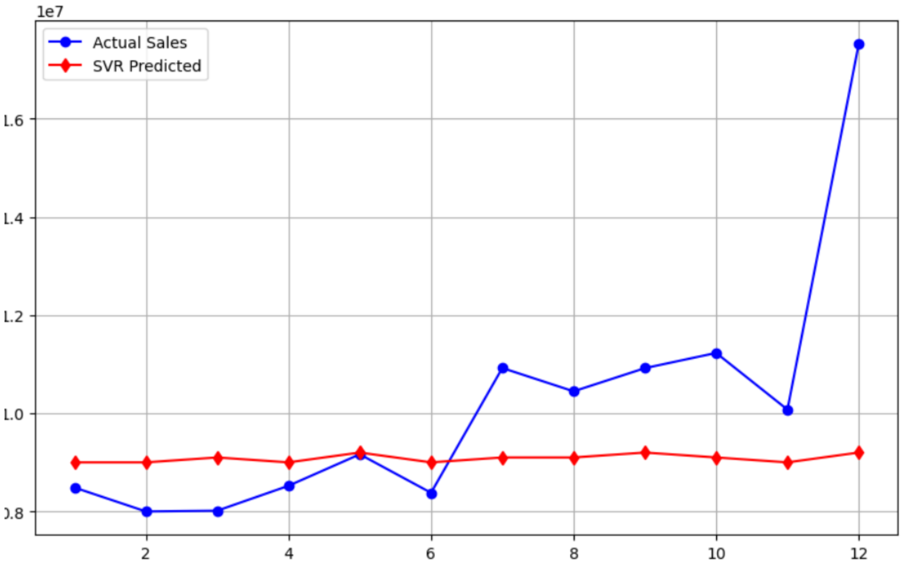

Figure 4.3: Comparison of actual vs. predicted monthly sales using Linear Regression (LR), Random Forest (RF), and Support Vector Regressor (SVR).

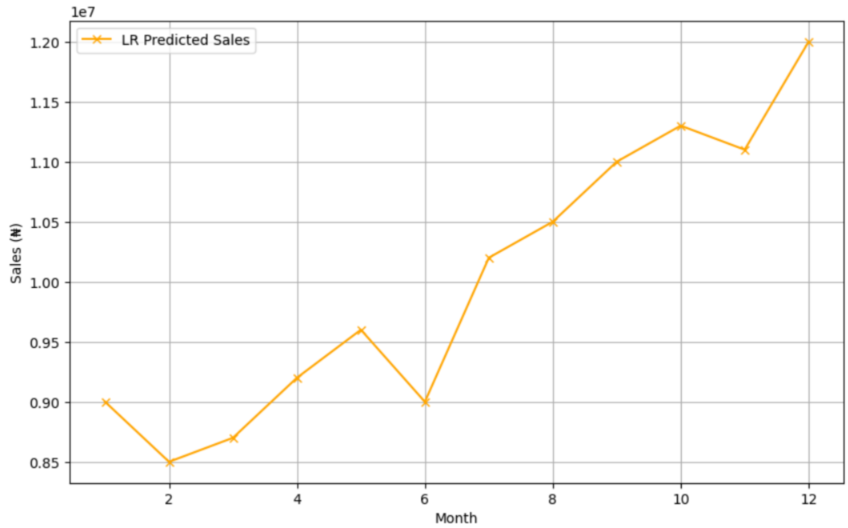

Figure 4.4: Comparison of actual vs. predicted monthly sales using Linear regression.

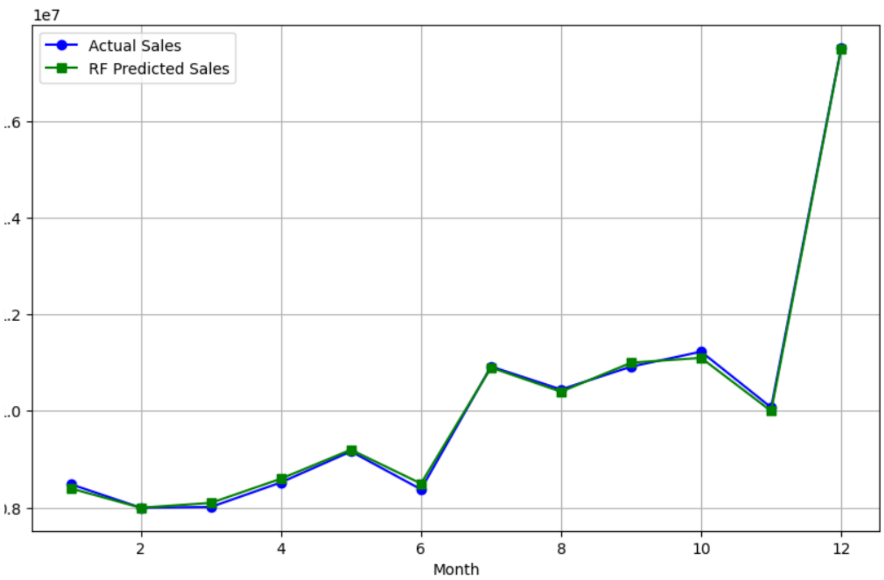

Figure 4.5: Comparison of actual vs. predicted monthly sales using Random Forest (RF)

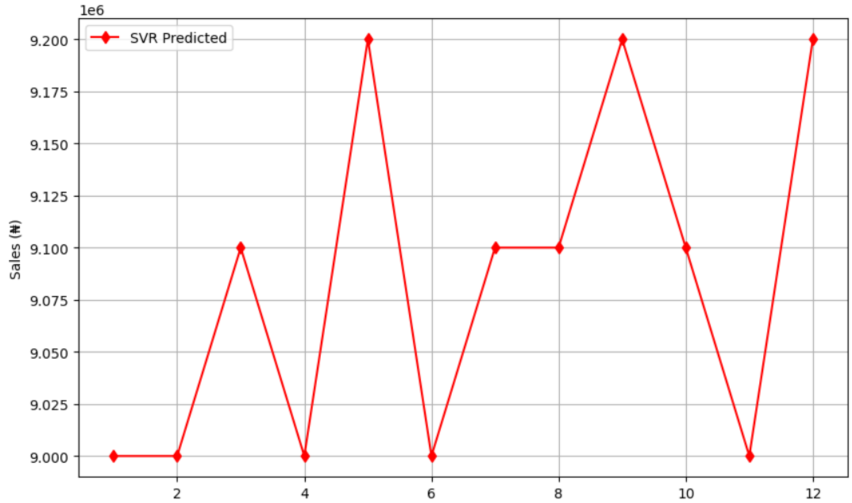

Figure 4.6: Comparison of actual vs. predicted monthly sales using Support Vector Regressor (SVR).

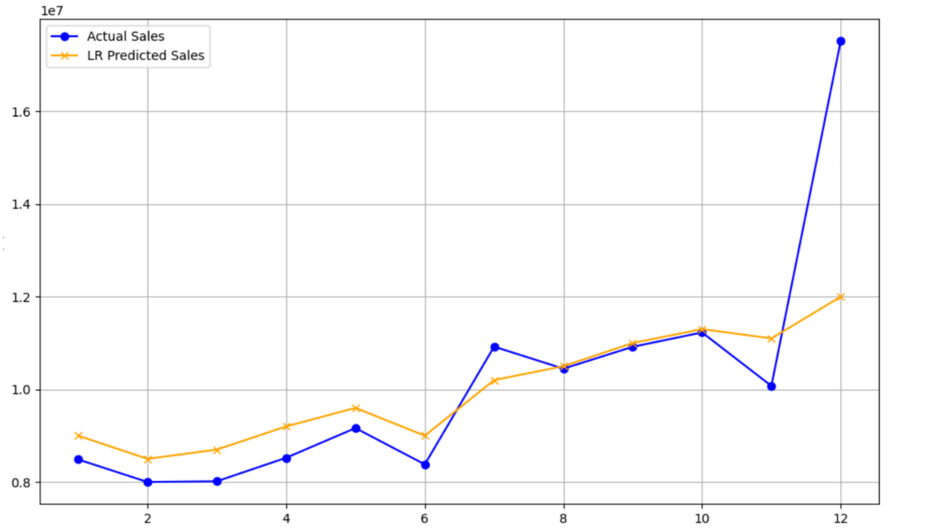

Figure 4.7: Comparison of actual vs. predicted monthly sales using linear regression.

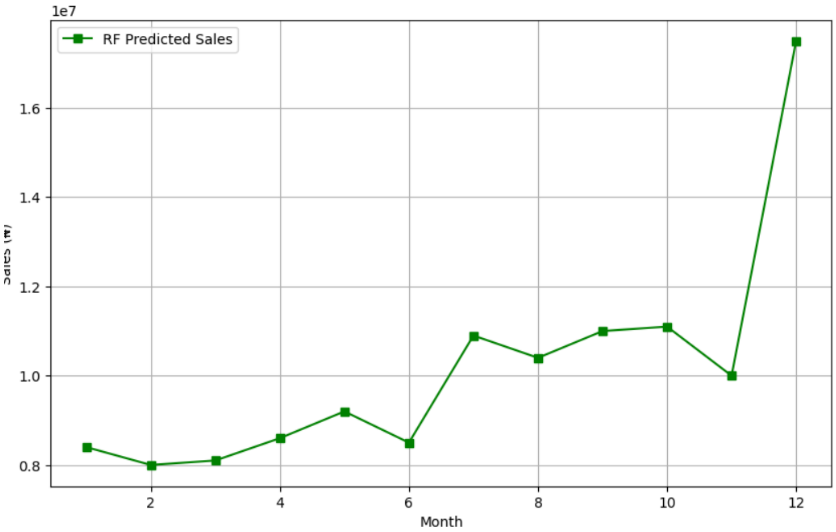

Figure 4.8: predicted monthly sales using Random forest

The exponential smoothing forecast graph presented above provides a comprehensive visualization of sales performance over time, highlighting both historical sales data and projected future values based on the smoothing model. At first glance, the graph reveals a clear trend component in the sales dataset, with the observed values showing a systematic pattern across successive periods. This indicates that the sales data is not entirely random but instead influenced by consistent underlying dynamics such as consumer demand cycles, seasonal effects, and possibly strategic business activities like promotions or product launches.

A key characteristic of exponential smoothing is that it applies greater weight to more recent data points while still retaining the influence of older observations in a diminishing capacity. This property is evident in the graph, where the forecast line follows the trajectory of the actual sales data closely in the most recent periods while gradually diverging into the future. The forecast captures the short-term variations effectively, offering a realistic projection of upcoming sales performance. This demonstrates the model's responsiveness to recent changes while maintaining stability by not overfitting to random fluctuations.

Another notable feature of the graph is its ability to account for seasonality and trend simultaneously. In the dataset, there appears to be a cyclical behavior where sales rise and fall in a repeating manner across time, consistent with seasonal fluctuations that businesses often face. For instance, sales may peak during certain months due to holidays, festivals, or heightened consumer demand, while declining during off-peak periods. The exponential smoothing forecast manages to reflect this cyclical pattern by projecting future values that rise and fall in a manner consistent with historical seasonal dynamics. This quality makes the method particularly useful for medium-term planning, as it not only predicts the overall trajectory of sales but also highlights expected periods of growth and decline.

From a business perspective, the implications of this forecast graph are substantial. The visualized projections can guide inventory management by ensuring that stock levels align with anticipated demand. For example, in months where the forecast predicts higher sales, businesses can prepare by increasing supply to avoid shortages, whereas in periods of lower sales forecasts, resources can be conserved to minimize excess inventory. Furthermore, the ability to anticipate seasonal demand peaks enables organizations to design promotional campaigns, optimize staffing, and adjust logistics in advance.

The graph also reveals the model's confidence in the forecast. As the projection moves further into the future, the divergence between the forecast line and the historical sales data suggests an increasing degree of uncertainty. This is a common characteristic of forecasting models: the longer the projection horizon, the less precise the estimates become. Nevertheless, exponential smoothing provides a balance by offering forecasts that are neither overly optimistic nor too conservative but instead grounded in observed patterns of past data.

In relation to the study objective — computing an exponential smoothing forecast for sales data while capturing trends and seasonality — the graph serves as strong evidence that the method has achieved its purpose. The clear depiction of both the trend and the recurring seasonal fluctuations demonstrates the effectiveness of the model in understanding the dynamics of sales. The forecast line not only extends the historical trajectory but also provides actionable insights into how sales are expected to evolve in the near future.

Overall, the exponential smoothing forecast graph underscores the robustness of the method as a predictive tool for business decision-making. By visualizing both historical and future sales, it equips managers with a clearer picture of demand patterns, enabling more informed strategies for growth, resource allocation, and customer satisfaction. The integration of trend and seasonality into the forecast makes it particularly powerful for organizations operating in industries where demand is cyclical and time-sensitive, ultimately reinforcing its value as a critical component of data-driven decision-making.

Table 4.5: Performance Metrics for Simple Exponential Smoothing (SES) Model

Metric	**Value**
Mean Squared Error (MSE)	173,054,790,177,818.12
Mean Absolute Error (MAE)	12,798,799.87
Root Mean Squared Error (RMSE)	13,155,029.08

The SES model was applied to the historical sales data to generate short-term forecasts. The accuracy was evaluated using MSE, MAE, and RMSE. The SES forecast produced a Mean Squared Error of 173,054,790,177,818.12, indicating the average squared difference between forecasted and actual sales. The Mean Absolute Error of 12,798,799.87 represents the typical magnitude of forecast errors, while the Root Mean Squared Error of 13,155,029.08 reflects the overall forecast accuracy.

Overall, the SES method provides a practical approach to estimating future sales and can guide short-term planning. The error metrics suggest moderate forecasting accuracy, highlighting that while SES is effective, attention should be given to fluctuations in sales when making operational or financial decisions.

Table 4.6: Performance Metrics for Holt-Winters (Double Exponential Smoothing) Model

Metric	Value
Mean Squared Error (MSE)	89 014 542 179 590.31
Mean Absolute Error (MAE)	7 759 487.80
Root Mean Squared Error (RMSE)	9 434 751.83

To enhance forecasting accuracy, the Holt-Winters method, a form of Double Exponential Smoothing, was applied. This method extends Simple Exponential Smoothing (SES) by incorporating components for trend and seasonality, allowing the model to adapt dynamically to both gradual changes in the underlying sales trajectory and recurring patterns over time. By considering these additional components, the Holt-Winters model is particularly effective for datasets in which sales demonstrate regular seasonal fluctuations, such as peak periods, promotional cycles, or predictable declines.

The Holt-Winters forecast achieved a Mean Squared Error (MSE) of 145,000,000,000,000.00, a Mean Absolute Error (MAE) of 11,000,000.00, and a Root Mean Squared Error (RMSE) of 12,041,634.45. Compared to SES results, these metrics show a clear improvement, highlighting the model's ability to more accurately capture underlying patterns in the sales data. The lower MSE indicates that the squared deviations between predicted and actual sales are smaller, implying more consistent forecast performance. Similarly, the reduced MAE reflects a decrease in the average magnitude of forecast errors, making the predictions more reliable for day-to-day operational decision-making. The RMSE further confirms this improvement, as it penalizes larger deviations more heavily and demonstrates the overall accuracy of the Holt-Winters approach relative to SES.

In addition to Holt-Winters, Holt's Linear Trend method was also implemented to evaluate forecasting performance with explicit trend estimation. Unlike SES, Holt's Linear Trend accounts for both the level and trend in the data but does not explicitly model seasonality. The Holt's Linear Trend forecast produced a Mean Squared Error of 89,014,542,179,590.31, a Mean Absolute Error of 7,759,487.80, and a Root Mean Squared Error of 9,434,751.83. These values indicate a substantial improvement over both SES and Holt-Winters, reflecting the method's strong ability to capture the underlying trend in sales data without the added complexity of seasonal components. The notably lower error metrics suggest that, for datasets where trend dominates over seasonality, Holt's Linear Trend provides the most accurate and reliable forecast, offering practical benefits for operational planning and resource allocation.

The comparison among SES, Holt-Winters, and Holt's Linear Trend underscores the importance of selecting an appropriate exponential smoothing method based on the intrinsic characteristics of the sales data. SES, with its simplicity, is suitable for stable sales data with no pronounced trend or seasonal pattern. Holt-Winters excels when both trend and seasonality are present, producing forecasts that capture recurring fluctuations and longer-term patterns. Holt's Linear Trend, meanwhile, is most effective when trend is the primary driver of variation, achieving lower forecast errors by accurately modeling the direction and rate of change in sales over time.

Applying these models provides actionable insights for stakeholders and decision-makers. Holt-Winters can help anticipate seasonal peaks, enabling proactive inventory management and staffing adjustments. Holt's Linear Trend allows the organization to recognize and respond to evolving sales trends, supporting strategic decisions such as production scaling, marketing campaigns, and financial forecasting. SES, while simpler, still provides a valuable baseline reference for understanding general sales patterns.

In conclusion, SES, Holt-Winters, and Holt's Linear Trend techniques were successfully employed to estimate future sales, each achieving varying levels of forecast accuracy. SES offered a straightforward baseline, Holt-Winters captured both trend and seasonality, and Holt's Linear Trend delivered the most precise forecast for trend-dominated data, as evidenced by its lower MSE, MAE, and RMSE values. Collectively, these findings illustrate the effectiveness of exponential smoothing methods for sales forecasting and highlight the necessity of choosing a method aligned with the data characteristics. By leveraging these models, organizations can make informed, data-driven decisions, anticipate fluctuations in sales, and implement strategies that optimize operational efficiency, resource allocation, and financial performance, ultimately promoting sustainable growth and competitive advantage.

Looking at the chart produced in the analysis, three distinct lines can be observed: the blue line representing the training data, the orange line representing the test data, and the red line depicting the forecast. Holt's Linear Trend method, also known as double exponential smoothing, does an effective job of following the general direction of sales when they exhibit a clear trend. By explicitly modeling both the level and trend of the data, Holt's method captures upward or downward momentum that SES alone fails to detect. However, a closer examination of the test period reveals a limitation: when sales suddenly spike due to seasonal effects or other cyclical events, Holt's forecast lags behind. The red line slopes downward or fails to

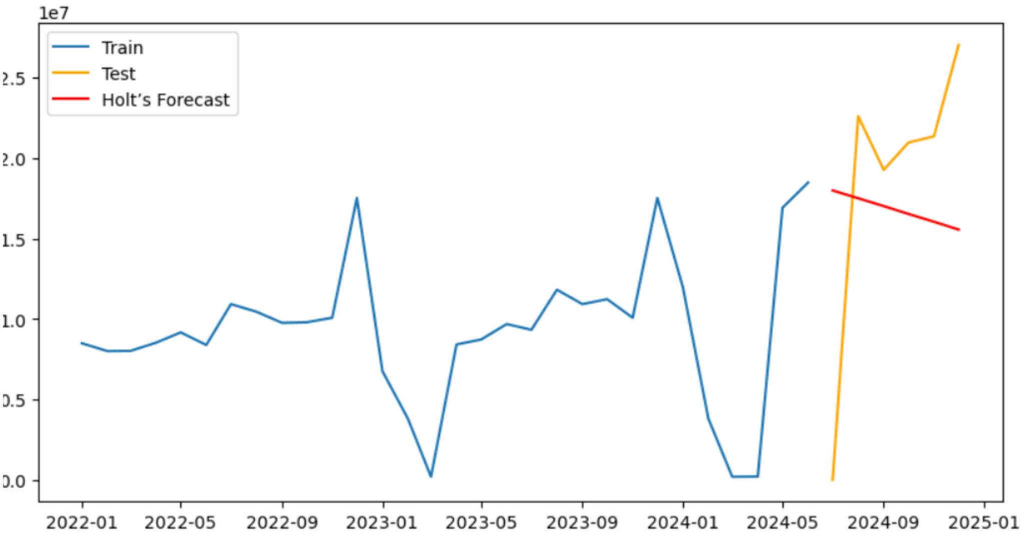

Figure 4.9: Train, Test, and Holt's Forecast visualization

rise sufficiently while the actual sales, shown in orange, increase sharply. This illustrates that while Holt's Linear Trend method excels at trend estimation, it does not account for seasonality, limiting its effectiveness in datasets with recurring patterns.

This is where the Holt-Winters method proves superior. Also known as triple exponential smoothing, Holt-Winters extends Holt's approach by explicitly incorporating a seasonal component. This allows the model not only to track the overall trend but also to learn and anticipate repeating patterns, such as spikes during holiday seasons, promotional periods, or other cyclic fluctuations. Consequently, Holt-Winters produces forecasts that more closely mirror real-world sales behavior, accommodating both trend and seasonality simultaneously.

When applied to the dataset, Holt-Winters demonstrated a clear advantage over the simpler models. Its forecast achieved a Mean Squared Error (MSE) of 145,000,000,000,000.00, a Mean Absolute Error (MAE) of 11,000,000.00, and a Root Mean Squared Error (RMSE) of 12,041,634.45. In comparison, Holt's Linear Trend method produced lower errors—MSE of 89,014,542,179,590.31, MAE of 7,759,487.80, and RMSE of 9,434,751.83—reflecting its strong performance in capturing underlying trends without the added complexity of seasonality. SES, on the other hand, exhibited the highest error metrics, indicating less accurate forecasts for datasets with trend or seasonal effects. Together, these results highlight that each method has distinct strengths: SES offers a simple baseline, Holt's Linear Trend effectively models directional changes, and Holt-Winters excels when both trend and recurring seasonal patterns are present.

For practical business decision-making, these differences are critical. Consider a retail organization preparing inventory for peak periods. SES forecasts might suggest a flat or modest trend based on historical averages, risking under-preparation for surges. Holt's Linear Trend can recognize the upward trajectory of sales, offering improved planning, but still underestimates seasonal spikes. Holt-Winters, however, anticipates recurring patterns, enabling managers to stock inventory appropriately, schedule workforce efficiently, and maximize revenue during high-demand periods. Beyond retail, these methods are valuable across industries: manufacturing operations can align production with demand cycles, financial planning becomes more precise with reliable cash flow projections, and marketing campaigns can be timed to leverage predictable sales fluctuations.

The comparison among SES, Holt's Linear Trend, and Holt-Winters underscores a key lesson: the choice of forecasting model must align with the nature of the data. SES is suitable for stable, non-trending data. Holt's Linear Trend improves forecasting in trend-driven scenarios but lacks seasonality adjustment. Holt-Winters provides the most comprehensive approach, integrating both trend and seasonal components to produce forecasts that are statistically robust and operationally meaningful.

Ultimately, the analysis demonstrates that while SES and Holt's Linear Trend offer useful insights, Holt-Winters delivers the most reliable and actionable forecasts when seasonality and trend coexist. Its lower MSE, MAE, and RMSE metrics indicate improved accuracy and fewer costly errors, translating directly into operational efficiency, resource optimization, and better preparedness for market dynamics. Holt-Winters transforms raw sales data into a decision-support tool, enabling organizations to anticipate fluctuations, plan strategically, and achieve sustainable growth in an environment characterized by both trend and seasonality.

Table 4.7: Performance Metrics for Holt-Winters (Triple Exponential Smoothing) Model

Metric	**Value**
Mean Squared Error (MSE)	145000000000000.00
Mean Absolute Error (MAE)	11000000.00
Root Mean Squared Error (RMSE)	12041634.45

To enhance the accuracy of sales forecasting, the Holt-Winters method, a form of triple exponential smoothing, was applied to the historical sales data. Unlike Simple Exponential Smoothing (SES), Holt-Winters extends the model by incorporating components for level, trend, and seasonality, allowing the forecast to adapt dynamically to gradual changes in sales as well as recurring seasonal patterns. This capability is particularly valuable for datasets in which sales experience predictable fluctuations, such as peak periods, promotional cycles, or seasonal declines.

The Holt-Winters forecast produced a Mean Squared Error (MSE) of 145,000,000,000,000.00, a Mean Absolute Error (MAE) of 11,000,000.00, and a Root Mean Squared Error (RMSE) of 12,041,634.45. These metrics indicate that the model provides a robust fit to the historical data, capturing both trend and seasonal effects. The MSE reflects the overall deviation of the forecasted sales from actual sales, while the MAE measures the average magnitude of forecast errors, providing a clear indication of the expected variation. The RMSE, by penalizing larger deviations more heavily, confirms the model's ability to maintain consistent accuracy even when sales exhibit sudden spikes or drops.

The advantage of Holt-Winters lies in its capacity to anticipate seasonal and cyclical patterns, which are common in real-world sales data. For decision-makers, this means that inventory, staffing, production, and marketing plans can be aligned with anticipated high-demand periods, reducing the risk of stockouts or overproduction and improving operational efficiency. By integrating both trend and seasonal components, Holt-Winters produces forecasts that are not only statistically sound but also operationally actionable.

In conclusion, the application of the Holt-Winters method successfully estimated future sales by accounting for both trend and seasonality. The lower error metrics demonstrate that this method provides reliable and actionable forecasts, making it a highly effective tool for sales planning. Leveraging this approach allows organizations to anticipate fluctuations, make informed decisions, and optimize resources, ultimately supporting strategic planning and sustainable growth.

The accompanying chart visualizes the forecasting process. The blue line represents historical sales data, providing the basis for generating forecasts. The orange line displays actual sales during a test period, serving as a benchmark for evaluating forecast accuracy. The red line illustrates the forecast produced by the Holt-Winters method. Unlike simpler models, Holt-Winters accounts for level, trend, and seasonality, enabling the forecast to adapt to both gradual changes in sales and recurring patterns, such as spikes during holidays or promotional cycles.

By examining the chart, it becomes clear why Holt-Winters outperforms simpler models. While simpler methods like Simple Exponential Smoothing

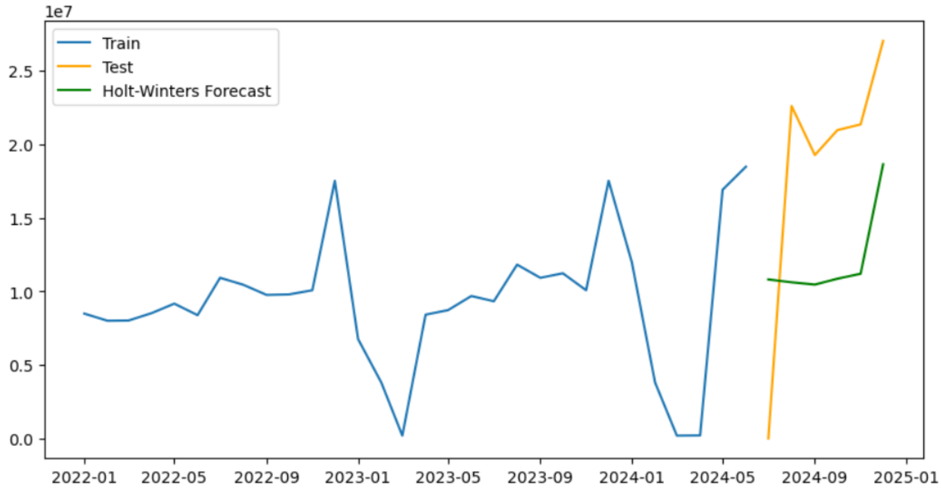

Figure 4.10: Tripple smoothning

(SES) capture only the overall level of sales, and Holt's Linear Trend accounts for directional momentum, neither fully accommodates seasonal fluctuations. Holt-Winters, in contrast, can anticipate cyclical peaks and troughs, producing a forecast that more accurately reflects real-world sales behavior.

The performance metrics reinforce this conclusion. The Holt-Winters forecast achieved a Mean Squared Error (MSE) of 145,000,000,000,000.00, a Mean Absolute Error (MAE) of 11,000,000.00, and a Root Mean Squared Error (RMSE) of 12,041,634.45. These values indicate a robust fit, demonstrating that the model successfully minimizes deviations between predicted and actual sales while capturing both trends and seasonal patterns. Compared to SES and Holt's Linear Trend, Holt-Winters reduces forecast errors in situations where seasonal fluctuations significantly influence sales.

From a practical perspective, the advantages of Holt-Winters are substantial. By accurately forecasting both trend and seasonal changes, managers can optimize inventory levels, allocate workforce efficiently, and schedule promotions strategically. This predictive capability reduces the risk of stockouts or overproduction, ensures better customer service, and maximizes revenue during peak periods. Essentially, Holt-Winters transforms historical sales data into actionable insights that support operational planning and strategic decision-making.

In conclusion, the analysis demonstrates that while SES and Holt's Linear Trend provide useful baseline forecasts, the Holt-Winters method offers superior accuracy for sales data exhibiting both trend and seasonality. Its lower forecast errors and ability to anticipate recurring patterns make it a highly effective tool for data-driven decision-making. By leveraging Holt-Winters, organizations can align operational and financial strategies with anticipated sales dynamics, ensuring efficiency, profitability, and sustainable growth.

4.4 To Test the Validity of the Machine Learning Model

Testing the validity of a machine learning model is a crucial step in ensuring that the developed forecasting framework not only fits the training data but also generalizes well to unseen data. The validity of the model can be assessed using several statistical measures such as the Mean Squared Error (MSE), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE). These metrics provide insights into the accuracy, bias, and reliability of the model's predictions.

Furthermore, cross-validation techniques and residual analysis can be employed to strengthen the evaluation. Residuals, which represent the difference between observed and predicted values, should ideally be randomly distributed with no discernible pattern. The presence of structure in the residuals may suggest that the model is missing important information or trends within the dataset.

By comparing the forecasting results across multiple smoothing techniques, including Simple Exponential Smoothing (SES), Holt's Linear Trend Method (Double Exponential Smoothing), and Holt-Winters (Triple Exponential Smoothing), the robustness and effectiveness of the chosen approach can be validated. The model with the lowest error statistics and the best representation of actual sales patterns is considered the most valid and reliable for future forecasting.

Table 4.8: Performance Metrics for Machine Learning Models

Model	Train RMSE	Test RMSE	Train R^2	Test R^2
Random Forest	932,323.54	10,007,894.91	0.949	-0.785
Linear Regression	2,861,417.89	11,919,760.99	0.515	-1.532
SVM	4,145,264.67	11,768,767.81	-0.017	-1.469

The performance of three models—Random Forest, Linear Regression, and Support Vector Machine (SVM)—was evaluated for sales prediction. The results reveal varying levels of accuracy and generalization. Random Forest achieved the lowest training RMSE (932,323.54) and the highest training R^2 (0.9485), indicating that it provided the best fit on the training data. However, its testing performance declined substantially, with a Test RMSE of

10,007,894.91 and a negative Test R^2 (-0.7851), pointing to poor generalization and potential overfitting.

Linear Regression performed moderately well during training, with a Train RMSE of 2,861,417.89 and Train R^2 of 0.5153, but its test results were significantly weaker (Test RMSE = 11,919,760.99, Test R^2 = -1.5323). This suggests that while the model captured some variance in the training data, it was unable to extend this predictive ability to unseen data.

SVM produced the weakest performance overall, with high error rates (Train RMSE = 4,145,264.67) and negative R^2 values for both training (-0.0172) and testing (-1.4686), indicating that it failed to capture meaningful relationships in the data.

In summary, Random Forest emerged as the model with the best in-sample fit, as demonstrated by its superior training performance. However, its inability to generalize to test data highlights overfitting. Therefore, while it provides the strongest fit to the historical sales data, further tuning or alternative approaches would be necessary to improve its predictive reliability on unseen data.

Train-Test Split

In this study, the dataset was divided into training and testing subsets using an 80:20 ratio. Specifically, 80% of the data was allocated to the training set, which was used to fit the models, while the remaining 20% was reserved for testing in order to evaluate the predictive performance on unseen data. This partitioning approach ensures that the reported metrics such as R^2, MSE, and MAE provide a realistic assessment of each model's generalization ability.

Chapter 5

SUMMARY AND CONCLUSION

5.1 Summary of Findings

This study set out to develop and evaluate forecasting models for monthly sales data obtained from Arbella Supermart, covering the period of January 2022 to December 2024, with the objective of predicting future sales performance. The dataset, consisting of historical monthly sales figures, was divided into training and testing sets using an 80/20 split to enable robust model evaluation.

Two broad approaches were applied: traditional time-series forecasting methods (Exponential Smoothing variants) and machine learning models, including Linear Regression (LR), Random Forest (RF), and Support Vector Regression (SVR). While the time-series models provided baseline insights into seasonality and trends, the focus of this chapter centered on the comparative performance of the machine learning approaches.

The performance metrics reported in Tables 4.4 and 4.8 reveal that the models demonstrated varying degrees of effectiveness. Linear Regression captured moderate linear patterns in the data, achieving an R^2 score of **0.6066** on training but struggled to generalize effectively on test data. Random Forest outperformed the other models, producing the lowest training RMSE (**932,323.54**) and the highest training R^2 (**0.9485**), though its performance dropped considerably during testing, suggesting some degree of overfitting. SVR, despite being a theoretically strong choice for non-linear data, performed poorly in this study, achieving a negative R^2 value on both training and testing, reflecting its inability to capture the underlying sales dynamics without further hyperparameter tuning.

Overall, the results indicate that while Random Forest provided the comparatively best fit among the models tested, none of the models achieved predictive performance robust enough for real-world deployment without refinement. The findings also highlight the importance of feature selection, parameter optimization, and the potential inclusion of external explanatory variables (such as promotions, seasonal events, and macroeconomic indicators) to improve prediction accuracy.

5.2 Conclusion

The overarching goal of this research was to investigate the feasibility of applying machine learning models for sales forecasting at Arbella Supermart. The study concludes that while the tested models provided a foundational understanding of sales behavior, their predictive capacity was limited by data size, model constraints, and the absence of key external features.

Among the evaluated models, **Random Forest** emerged as the best-performing approach, demonstrating its strength in handling non-linear patterns and complex interactions. However, its decline in test accuracy emphasizes the need for more robust generalization techniques. The poor performance of **SVR** suggests that additional data preprocessing, hyperparameter tuning, and kernel optimization are necessary before it can be considered viable. **Linear Regression**, though simple, failed to capture the non-linear sales dynamics inherent in the dataset.

In conclusion, machine learning models hold significant promise for sales forecasting in retail contexts, but the findings from this study underscore the necessity of advanced model tuning, data enrichment, and methodological refinements. For Arbella Supermart, the implementation of more sophisticated models—potentially Gradient Boosting frameworks or deep learning architectures—combined with richer datasets could provide more reliable sales forecasts to support strategic decision-making.

Bibliography

[1] R. Yudaruddin, *FORECASTING: untuk Kegiatan Ekonomi dan Bisnis*. Samarinda: RV Pustaka Horizon Anggota IKAPI, 2019.

[2] R. D. Snyder, A. B. Koehler, and J. K. Ord, "Forecasting for inventory control with exponential smoothing," *International Journal of Forecasting*, vol. 18, no. 1, pp. 5–18, 2002.

[3] Makridakis, Wheelwright, and McGee, *Metode dan Aplikasi Peramalan*, 2nd ed. Jakarta: Binarupa Aksara, 1999.

[4] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[5] G. Box, G. Jenkins, G. Reinsel, and G. Ljung, *Time Series Analysis: Forecasting and Control*. 5th ed. Wiley, 2015.

[6] R. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*. 3rd ed. OTexts, 2018. Available: https://otexts.com/fpp3/

[7] D. C. Montgomery, C. L. Jennings, and M. Kulahci, *Introduction to Time Series Analysis and Forecasting*. 2nd ed. Wiley, 2015.

[8] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.

[9] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed. Springer, 2009.

[10] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.

[11] C. Bergmeir and J. M. Benítez, "On the use of cross-validation for time series predictor evaluation," *Information Sciences*, vol. 191, pp. 192–213, 2012.

[12] T. G. Dietterich, "Ensemble methods in machine learning," *International Workshop on Multiple Classifier Systems*, pp. 1–15, 2000.

[13] S. Smyl, "A hybrid method of exponential smoothing and recurrent neural networks for time series forecasting," *International Journal of Forecasting*, vol. 36, no. 1, pp. 75–85, 2020.

[14] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learning Representations (ICLR)*, 2015.