International Journal of Scientific Research and Modern Technology (IJSRMT) ijsrmt.com
Volume 4, Issue 3, 2025
DOI: https://doi.org/10.38124/ijsrmt.v4i3.350

Sales Prediction using Ensemble Machine Learning Model

Mustapha Ismail[1]; Hafsat Muhammad Tukur[2]; Mamudu Friday[3]

[1,3] Department of Computer Science, Gombe State University, Gombe, Nigeria
[2] Student of Gombe state university, Nigeria

Publication Date: 2025/03/25

ABSTRACT

With increased competition in the supermarket industry, there is an increased need for higher-order predictive analytics to garner insight into consumer behavior for optimal sales strategies. Therefore, this research has presented a sales prediction using an ensemble machine learning approach by considering multiple algorithms: Random Forest, XGBoost, and Support Vector Machine, which further improve predictive accuracy and avoid possible overfitting. This paper presented a comprehensive data preprocessing and feature engineering, with the implementation of a stacking ensemble model, which resulted in excellent predictive performance. The stacking ensemble model achieved the best R^2 value of **0.9990** and the least mean absolute error. The results showed that machine learning techniques are very promising to improve sales prediction and provide a powerful tool for supermarkets in making better decisions, optimizing inventories, and conducting focused marketing. Hybrid models should be further explored in future research, with the addition of more external factors to improve the predictive accuracy.

Keywords: *Prediction, Ensemble Machine Learning, Decision-Making, Inventory Optimization and Market Strategies.*

I. INTRODUCTION

The term supermarket refers to a large retail store dealing in self-service offering a great variety of fresh, prepared, and sometimes even non-consumable merchandise, usually methodically arranged around single or multiple areas or aisles. Many also extend services such as pharmacies, online shopping, and home delivery services, making the supermarket central in modern consumer habits. With increased competition in the supermarket industry, supermarkets need to know their consumer behavior and purchasing patterns in order to remain competitive with others. According to Ramachandran (2023), among the innumerable applications which use the power of data, predictive analysis is one such tool that is revolutionary. Sales forecasting, as one of the main elements of predictive analysis, helps an organization develop foresight to decide on inventory management, pricing strategies, marketing campaigns, and resource allocations. With good predictions, companies will be able to make the most out of their investment by reducing inventory costs and enhancing sales and profitability without risking certain eventualities Varshini and Preethi (2022). There are a variety of ways in which an organization can perform and implement prediction for sales. Earlier managers usually make sales predictions randomly, professional managers, however, become hard to find and not always available (e.g., they can get sick or leave), sales predictions can be assisted by computer systems that can play the qualified managers' role when they are not available or allow them to make the right decision by providing potential sales predictions, such methodology can be carried out by the attempt to model the professional manager's skills in a computer program. (Aimufua et al., 2021). Traditional statistical forecasting/predicting methods like time series analysis have limitations in capturing real-world complexities (Wendy, 2018). The emergence of machine learning provides powerful techniques to uncover patterns from data and make accurate predictions (Kaplan et al., 2021), this has led to growing research on applying machine learning algorithms for sales prediction using historical sales data and additional variables. Previous sales prediction research has always relied on a single prediction model (Niya and Jasmine, 2021). However, a single model cannot be the most effective for all types of products, cannot unlock the secret of large data using machine learning, allowing retailers to better understand themselves and their competitors, adjust sales planning, and remain unstoppable (Raizada and Saini, 2021; Wisesa et al., 2021), There are several branches of Machine Learning Algorithms. Such as, Supervised Machine

Learning, Unsupervised Machine Learning, Semi-supervised Machine Learning and Reinforcement Machine Learning (Sharma et al., 2021), depending on the type of the models or techniques to be used. For retail sales forecasting, association rule mining algorithms are often used along with predictive modeling (Patangia et al., 2020). Techniques like Apriori algorithm help uncover relationships between purchased items. The generated association rules can inform predictive models. Comparative studies reveal that advanced algorithms like LSTM, XGBoost, random forests outperform classical statistical models like ARIMA and regression on retail sales data (Bajaj et al., 2020; Ranjitha and Spandana, 2021).

Supply and demand are two fundamental concepts of sellers and customers, predicting the demand accurately is critical for the organizations in order to be able to formulate the plans (Tom et al., 2021). Though, previous applications of machine learning techniques in sales forecasting in many businesses in order to predict future demand, optimize inventory, find key factors affecting sales, predict total sales, store combination and so on does not yield optimal results. However, predicting future sales for the whole business may not always be the case for a successful business operation, as there are some major factors that have impacts on the sales of some product. This research aims to analyze customer purchasing behavior in order to find most popular products, track their sales trend over time and predict future sales using ensemble model.

II. LITERATURE REVIEW

Previously several studies were conducted on this field, and it has become very popular nowadays, lots of researchers have implemented various algorithms for example (Patangia et al., 2020) investigate using machine learning approach with the goal to analyze customer purchase patterns and forecast sales to optimize inventory, marketing, and operations. They incorporate association rule mining and predictive modeling for sales forecasting. (Bajaj et al., 2020) also explore machine learning models to predict future sales of Big Mart outlets, Random Forest performs was the best with 93.53% testing accuracy among others like - Linear Regression, KNearest Neighbors and XGBoost. (Swami et al., 2020) applied machine learning techniques to predict total sales for every product and store combination for the next month given historical daily sales data. XGBoost with optimized hyperparameters is found to perform the best with the lowest RMSE compare to LSTM neural networks, and Autoregressive Integrated Moving Average (ARIMA) models. (Raizada and Saini, 2021) compared various supervised machine learning techniques for sales forecasting using data from 45 Walmart retail outlets and predict weekly sales based on factors like past sales, temperature, fuel price, holidays, etc. This paper concludes that ensemble methods such as Extra Trees and Random Forest are the most suitable options for short-term retail sales forecasting with limited historical data. Tom et al. (2021) proposed a study in which regression techniques were proposed for the prediction of supermarket sales, including linear regression, Ridge regression, Lasso regression, decision tree, and random forest, along with XGBoost; the results obtained depict that XGBoost gives the highest accuracy of 88.51% in the results. Also, Ranjitha and Spandana (2021) present a comparative analysis of machine learning techniques for sales prediction using Big Mart retail data. The goal is to forecast sales volume to assist inventory management. Four regression algorithms are implemented: Linear Regression, Polynomial Regression, Ridge Regression, and XGBoost. XGBoost gives the best results with the lowest errors, followed by Ridge Regression, while Polynomial Regression also outperforms Linear Regression. The authors of the work (Mohamed et al., 2022) deal with price prediction for seasonal commodities, such as Christmas gifts. The authors propose both machine learning and statistical approaches to solve this task. In this paper, four machine learning models are proposed: Support Vector Regression (SVR), Random Forest, Ridge Regression, and Linear Regression. They also propose an ARIMA model as a statistical approach. It was concluded that both machine learning and statistical models were suitable for predicting prices for seasonal retail goods. The paper by (Varshini and Preethi, 2022) implements models like XGBoost, random forest, ANN, and SVR on the BigMart retail sales dataset, doing exploratory analysis to identify correlations between sales with factors like price and visibility. The models were evaluated using RMSE, R2 score, and MAPE. Random forest and XGBoost are giving the best performance with the lowest errors. Kramar and Alchakov (2023), work titled “Time-series forecasting of seasonal data using machine learning methods algorithms” several machine learning methods were compared with time series forecasting of seasonal data using real data from a wastewater treatment plant. It was suggested that the approach by which the models will be trained using limited historical data of 15-17 days to forecast the next 2 days. The models will be retrained at every 24-hours with new data so that the accuracy could be maintained. Among these methods, XGBoost is the best in balancing accuracy, SARIMA and LSTM give good results but are slower, while Prophet and ETS also give reasonable accuracy. A study by (Ramachandran, 2023) used big data analytics to predict supermarket sales. The aim was to evaluate algorithms for the forecasting of sales for different products. The sales data was collected over three months period from a supermarket in India upon which Linear Regression, ARIMA, Random Forest, XGBoost, and LSTM models were used to implement and predict sales. The result showed that the LSTM model performed best by having the highest accuracy and R-squared; therefore, this might be a promising model in uncovering patterns of sales data. Limitations include small data and ethical considerations. A paper by (Hasan, 2024) explores the use of machine learning techniques to address seasonality and trend detection in predictive sales forecasting. In this study, three machine learning models were employed: Random Forest, Linear Regression, and Gradient Boosting. The results indicated that the Gradient Boosting model was invariably the best among the two, since it achieved the highest R2 score and lowest errors during testing. It does not go in-depth into more advanced deep learning techniques, such as LSTM networks, which have been proven to be effective in modeling long-term dependencies in sales data. Guru et al. (2024) present a study on product sales forecasting using machine learning algorithms. The researchers, focusing on the Big-Mart dataset, employed several machine learning techniques but emphasized the K-Nearest Neighbor algorithm. They developed the K-NN model for sales forecast prediction and further compared the performance with Linear, Polynomial, and Ridge regression methods. The results showed that in the K-NN model, the accuracy was very impressive, about 84.7%, whereas in other previous works, its results were less comparable. The work did not delve deep into the possibilities of other advanced machine learning techniques such as deep learning or ensemble learning. Vasuki et al. (2024) conducted a survey on the sales prediction using different machine learning algorithms such as linear regression, decision trees, random forests, gradient boosting machines, and neural networks. Though the paper does not specify the particular results with respect to each algorithm, the authors mention using metrics such as MAE and RMSE to evaluate model efficacy. Ensemble learning is a powerful technique that involves training multiple base models and combining their predictions to obtain a more accurate and robust final prediction (Zhou, 2012). Many researchers apply the ensembling machine learning approach for the prediction or forecasting of future sales. In the paper by Abdullahi et al. (2021) investigated the application of several machine learning algorithms to forecast sales in retail shops, comparing four algorithms: Random Forest, Extreme Gradient Boosting, Support Vector Machine for Regression, and an Ensemble Model. According to the studies, feature selection had a greater impact on the performance of the model compared to hyper-parameter optimization. The best approach is proposed by Hu, 2022, for sales prediction in e-commerce using a random forest, GBDT, and XGBoost algorithm. The obtained results prove that this combined model ensures better sales forecast results compared with separate machine learning models. (Alice et al., 2024) present a comparison of different machine learning models for sales forecasting. Focusing on ensemble learning techniques, the researchers implemented three gradient boosting models: XGBoost, LightGBM (LGBM), and CatBoost, along with a hybrid ensemble model using a Voting Regressor. The performance of each model was evaluated using accuracy metrics. Among the single models, CatBoost had the best result with a testing accuracy of 0.9860, closely followed by XGBoost at 0.9834. LGBM was the most efficient concerning computational speed but achieved a relatively lower testing accuracy of 0.9734. Surprisingly, the Voting Regressor ensemble model slightly outperformed all the individual models with a testing accuracy of 0.9870. Kamble et al. 2023, have proposed an ensemble machine learning approach to forecast sales in the e-commerce domain. In this work, a comparative analysis for some machine learning models such as linear regression, random forest, and XGBoost were carried out which predicts sales. The results showed that the ensemble model outperforms an individual machine learning algorithm on these performance measures. A noble approach by Koh et al. 2024, predicted the future sales of a big mart using Linear Regression, Decision Tree, Random Forest, XGBoost, Stacked Ensemble Model, and K-Nearest Neighbors (K-NN). In the results, the XGBoost Regression Model tuned with RandomizedSearchCV outperformed the other models. Also, Das et al. (2023) implement particularly Long Short-Term Memory (LSTM) and Convolutional Neural Networks (CNN) and two-step ensemble method, by combining DeepAR and XGBoost regression to enhance sell-in and sell-out forecasting in the retail industry; the results of the experiments demonstrate the effectiveness of the proposed ensemble approach.

This research proposed an ensemble machine learning prediction model for more accurate sales prediction. By combining the predictions of the most used and most accurate individual machine learning algorithms from the previous study, we will delve on addressing the issue of overfitting through the use of cross-validation techniques by aiming at achieving the highest possible accuracy.

III. RESEARCH MATERIAL AND METHOD

Figure 1 below presented the proposed approach divided into eight segments, namely: Sales Dataset, Data pre-processing, Feature Engineering, Model selection, Ensemble Model, Training and testing, Model Evaluation, and Performance Analysis.

- *Data Collection*

The data source is a secondary data which was obtained from the sales of an Indian supermarket. The dataset was downloaded from Kaggle. It provides details of some products line sold at the supermarket for the consecutive 4 years from 2015 to 2018.

- *Data description*

Data collected comprises 1000 rows and 17 columns. Figure 2 shows the dataset information, column names, null and non-null entries and data types.

- *Data Pre-processing and cleaning*

The Data Pre-processing was done in several ways to ensure the datasets were clean to train the models. It involves cleaning, converting, and organizing raw data for analysis, identifying errors, addressing missing values, outliers, and transforming it into an appropriate format for analysis.

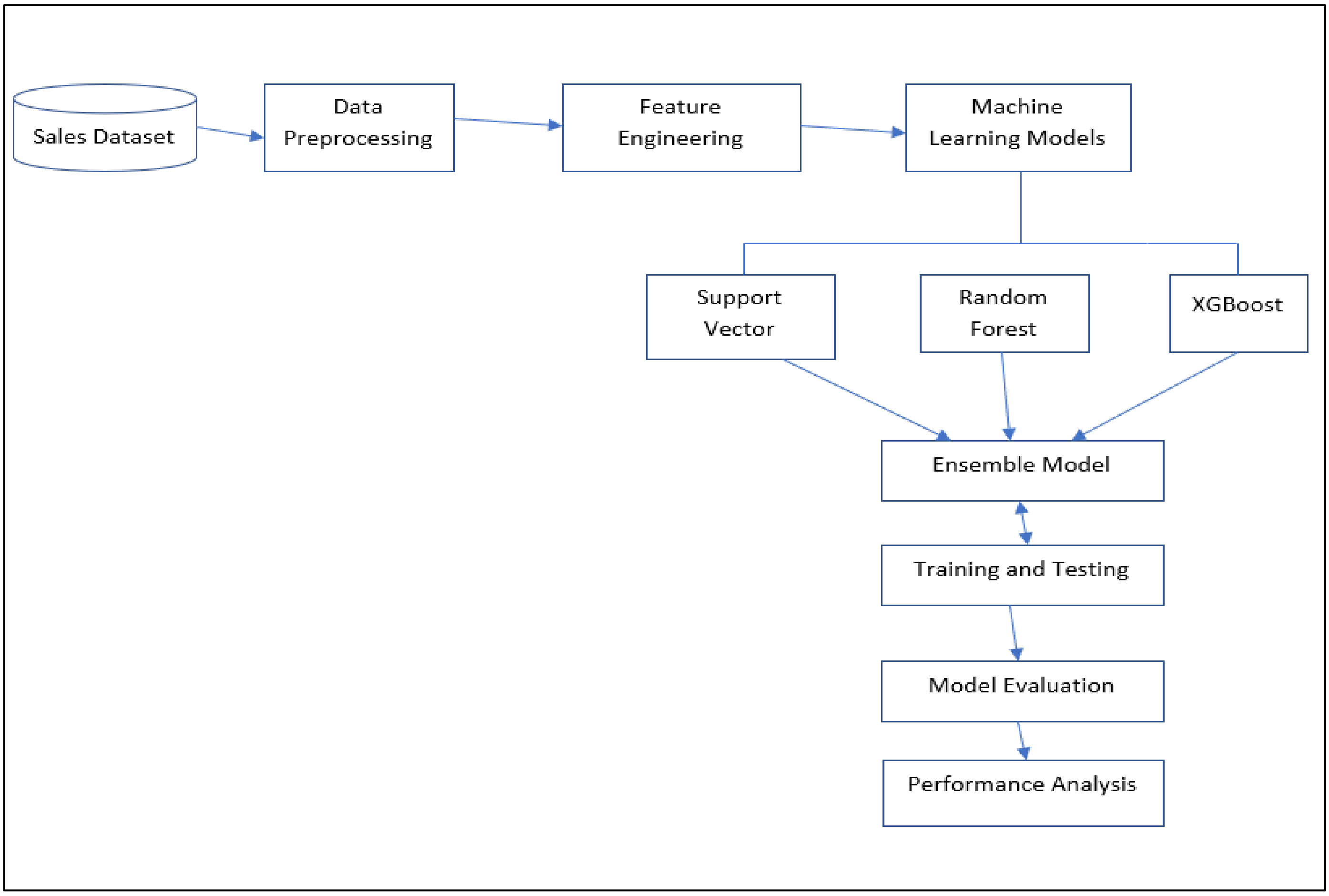

Fig 1 Framework for the Proposed System

```
In [7]: data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 17 columns):
 #   Column                   Non-Null Count  Dtype
---  ------                   --------------  -----
 0   Invoice ID               1000 non-null   object
 1   Branch                   1000 non-null   object
 2   City                     1000 non-null   object
 3   Customer type            1000 non-null   object
 4   Gender                   1000 non-null   object
 5   Product line             1000 non-null   object
 6   Unit price               1000 non-null   float64
 7   Quantity                 1000 non-null   int64
 8   Tax 5%                   1000 non-null   float64
 9   Total                    1000 non-null   float64
 10  Order_Date               1000 non-null   object
 11  Time                     1000 non-null   object
 12  Payment                  1000 non-null   object
 13  cogs                     1000 non-null   float64
 14  gross margin percentage  1000 non-null   float64
 15  gross income             1000 non-null   float64
 16  Rating                   1000 non-null   float64
dtypes: float64(7), int64(1), object(9)
memory usage: 132.9+ KB
```

Fig 2 Dataset Information

- *Feature Engineering*

In this step new attributes were derived from the existing attributes, in order to understand sales trend over time new attributes were created that might be of relevant, the new attributes are Day, Month, Year, Month name, Month year, Week of year and Season, these new attributes will help us detect sales trend over time.

- *Data Transformation and Partitioning*

The dataset has a mix of categorical and numerical features, it is necessary to transform it before feeding the data into the machine learning model. Column transformer is set to apply transformations to specific columns while leaving others unchanged. One-hot encoding is applied to the categorical feature (Item) and the numerical features remain unchanged. After the encoding columns were created for all the crops in the dataset and the other columns remain as it was. The dataset was split into two parts: **80% training and 20 % testing.**

- *Machine Learning Models*

Three individual modeling techniques were chosen and implemented in this stage based on their performance accuracy in predicting sales, the interpretability of the results and the computational requirements from the previous research. The models used are:

- *Random Forest (RF)*

Random Forest is a kind of ensemble learning that can be used for both classification and regression. Its basic working principle lies in creating several decision trees and then combining their predictions with a view to getting more accurate and robust results.

- *Extreme Gradient Boosting (XGBoost)*

Extreme Gradient Boosting (XGBoost) is a highly acclaimed and widely adapted machine learning algorithm that has proven its efficiency while dealing with large datasets and has achieved state-of-the-art performance in several tasks involving classification and regression. Advantages for Regression Tasks: High Predictive Accuracy, Flexibility, Feature Ranking, and Robustness. On the other hand, disadvantages are Complexity, Interpretability, and Computational Resources.

- *Support Vector Machine*

The Support Vector Regressor is one such algorithm in the realm of machine learning that provides regression by the determination of the best possible hyperplane available in a given predefined margin of tolerance. It has extensive applications in time series forecasting and financial data analyses, among numerous others, since the method works properly in both conditions, linear and nonlinear relations. Advantages of SVR in the context of regression tasks are: High predictive accuracy, Flexibility, Robustness to overfitting, and Scalability. Limitations include: Parameter sensitivity, Computational cost, Interpretability. Applications in regression tasks: With the capability to capture linear and nonlinear patterns, robustness to outliers, and the flexibility in modeling complex relationships of data, SVR is a useful technique that can be applied to a range of regression tasks, such as financial forecasting, sales predictions, and environmental modeling.

- *Ensemble Learning Models*

This is a learning method in which different models are trained on diverse algorithms and hyper-parameters, and their predictions combined to avoid overfitting and biases. The application domains include finance, healthcare, marketing, and image recognition, among others. Other categories of ensemble learning included: Bagging, Adaboost, Stacking, and Blending. Our work used stacked generalization ensemble technique where several models are trained and their predictions are combined with the help of another model, usually called a meta-model. The meta-model was trained by taking the predictions of the base models as input features. The goal of this stacking is to take advantage of the strengths of different models and enhance predictive performance.

- *Metrics of Evaluation*

The models were evaluated on three key metrics, namely **Mean Squared Error, Mean Absolute Error, and Coefficient of Determination.**

$MSE = (1/n) * \Sigma(Y_i - \hat{Y}_i)^2$ }- }- }- }- } 1

$MAE = (1/n) * \Sigma|Y_i - \hat{Y}_i|$ }- }- }- }- - 2

$R^2 = 1 - (SS_{res} / SS_{tot})$ }- }- }- }- - 3

Where:

$SS_{res} = \Sigma(Y_i - \hat{Y}_i)^2$ (Sum of squares of residuals)

$SS_{tot} = \Sigma(Y_i - \bar{Y})^2$ (Total sum of squares)

Y_i = actual value

$\hat{Y}_i$ = predicted value

$\bar{Y}$ = mean of actual values

IV. RESULTS AND DISCUSSION

The correlation matrix shows the correlation coefficients among various variables which gives the quantitative measure of the linear relationship between pairs of variables, with coefficients ranging from -1 to 1. Here, a heatmap was created using the Seaborn library to display the correlation matrix of a dataset as shown in **figure 3 below**.

This heatmap contains the correlation coefficients colored from blue to red, representing negative to positive, with annotations for clarity.

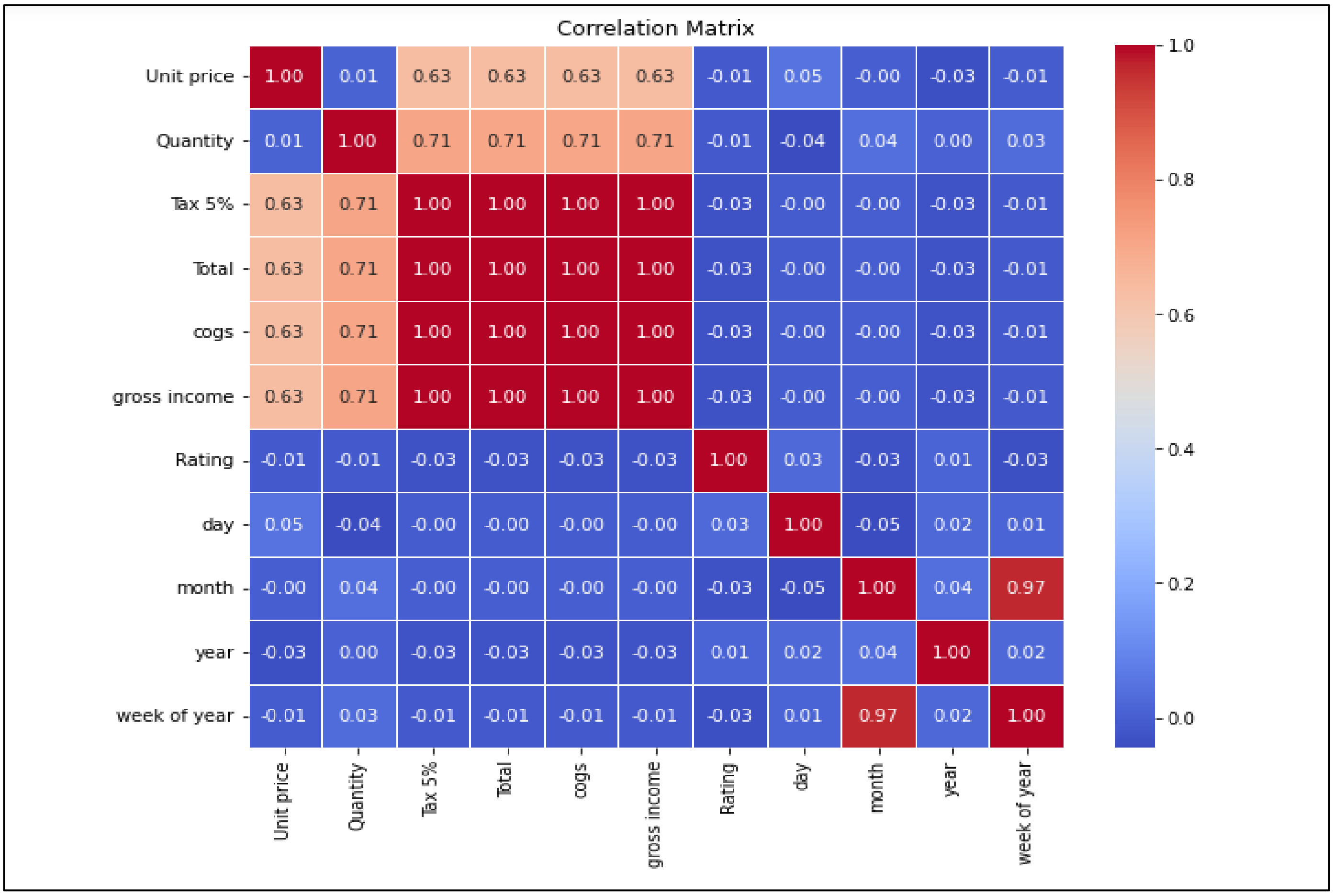

Fig 3 Correlation Matrix

Figure 4 shows the distribution in gender. We could also see from this pie chart that the number of sale count does not vary significantly between the gender. While **Figure 5 gives the customer type distribution**. In this case too, the sales counts among the two types of customers, namely: **Member and Normal** does not show a significant variation in number.

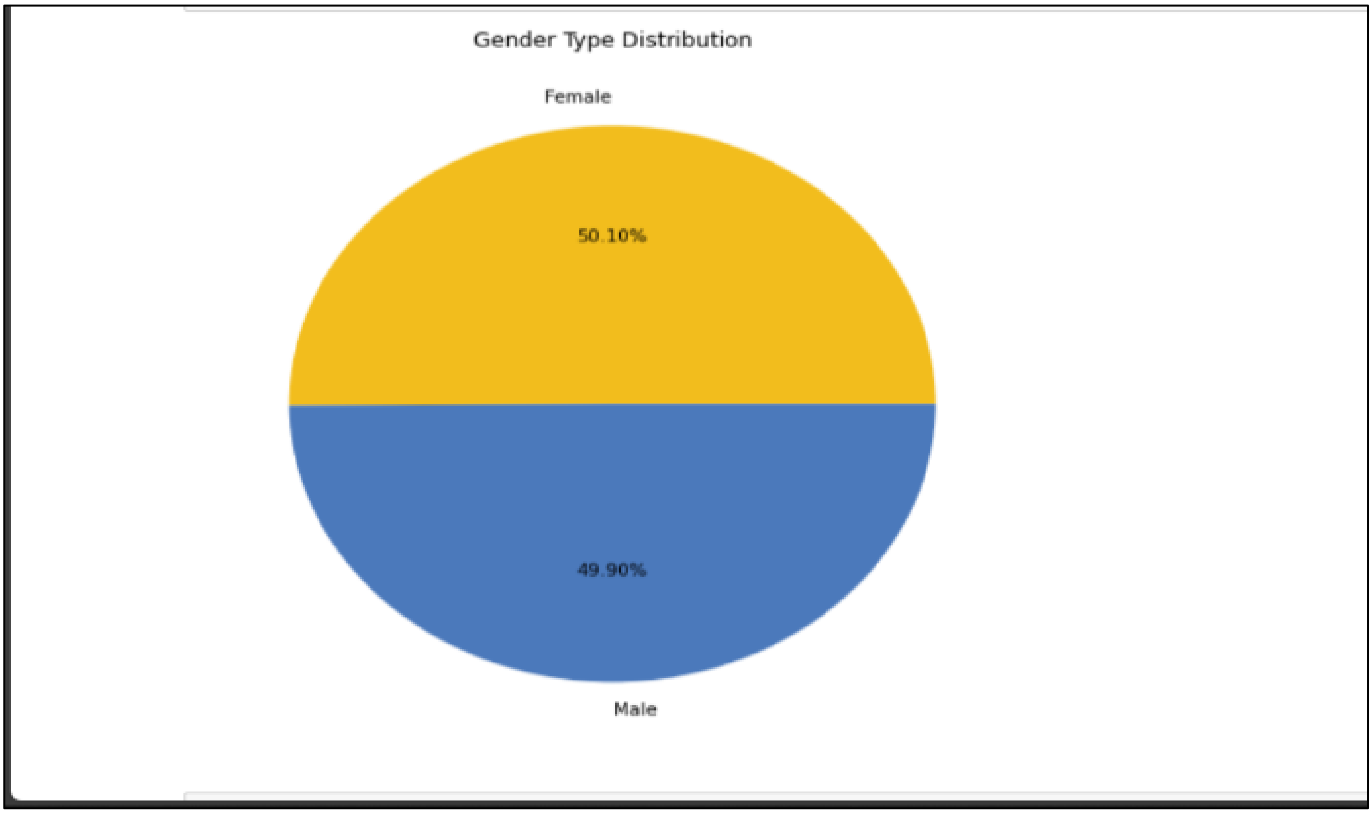

Fig 4 Gender Types Distribution

Figure 6: Sales distribution by product line. The total sales by product line show that Food and beverages have the highest percentage of 17.38%, followed by Sport and travel with the percentage of 17.07, followed by Electronic and accessories with the percentage of 16.82, followed by Fashion accessories with the percentage of 16.81, followed by Home and life style with the percentage of 16.68, and the last one with the lowest percentage is Health and beauty with the percentage of 15.23.

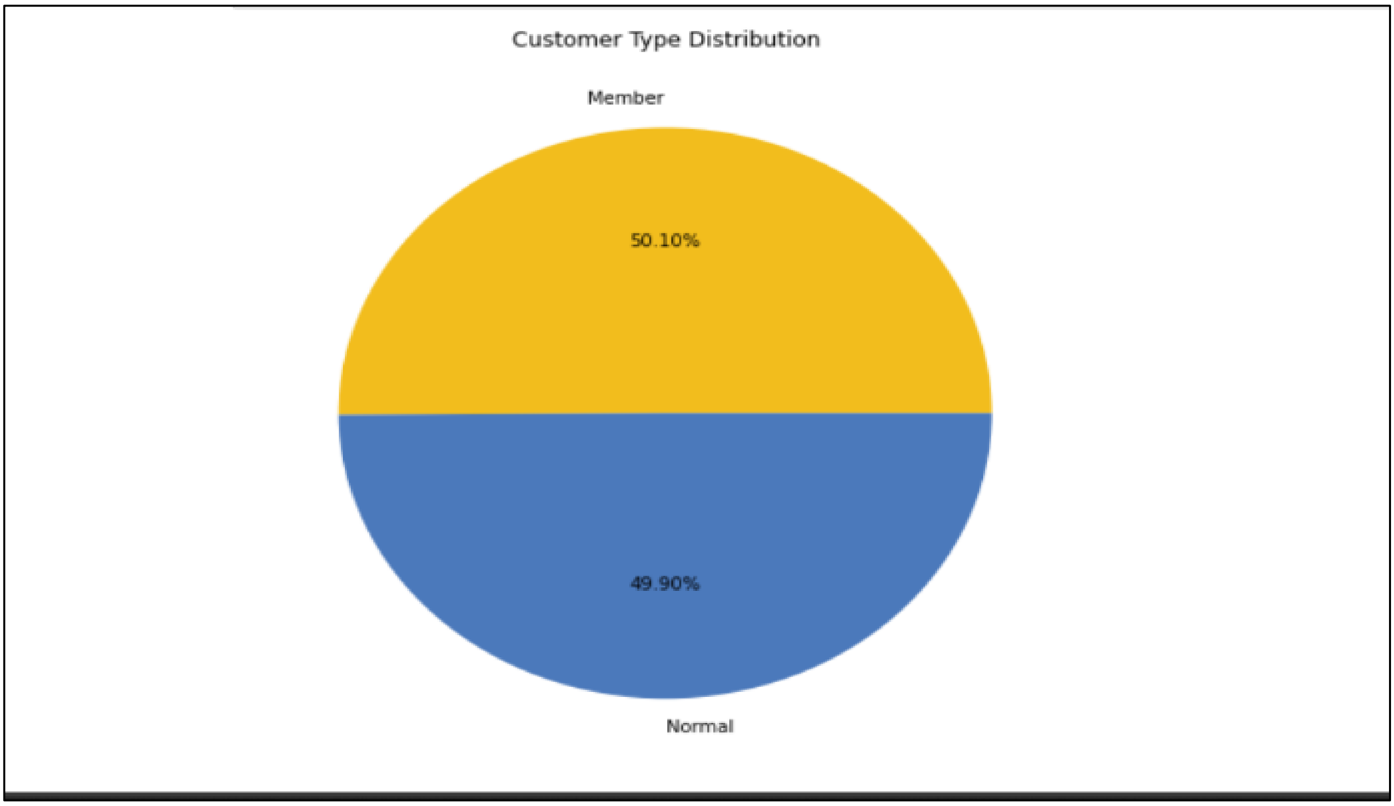

Fig 5 Customer Type Distribution

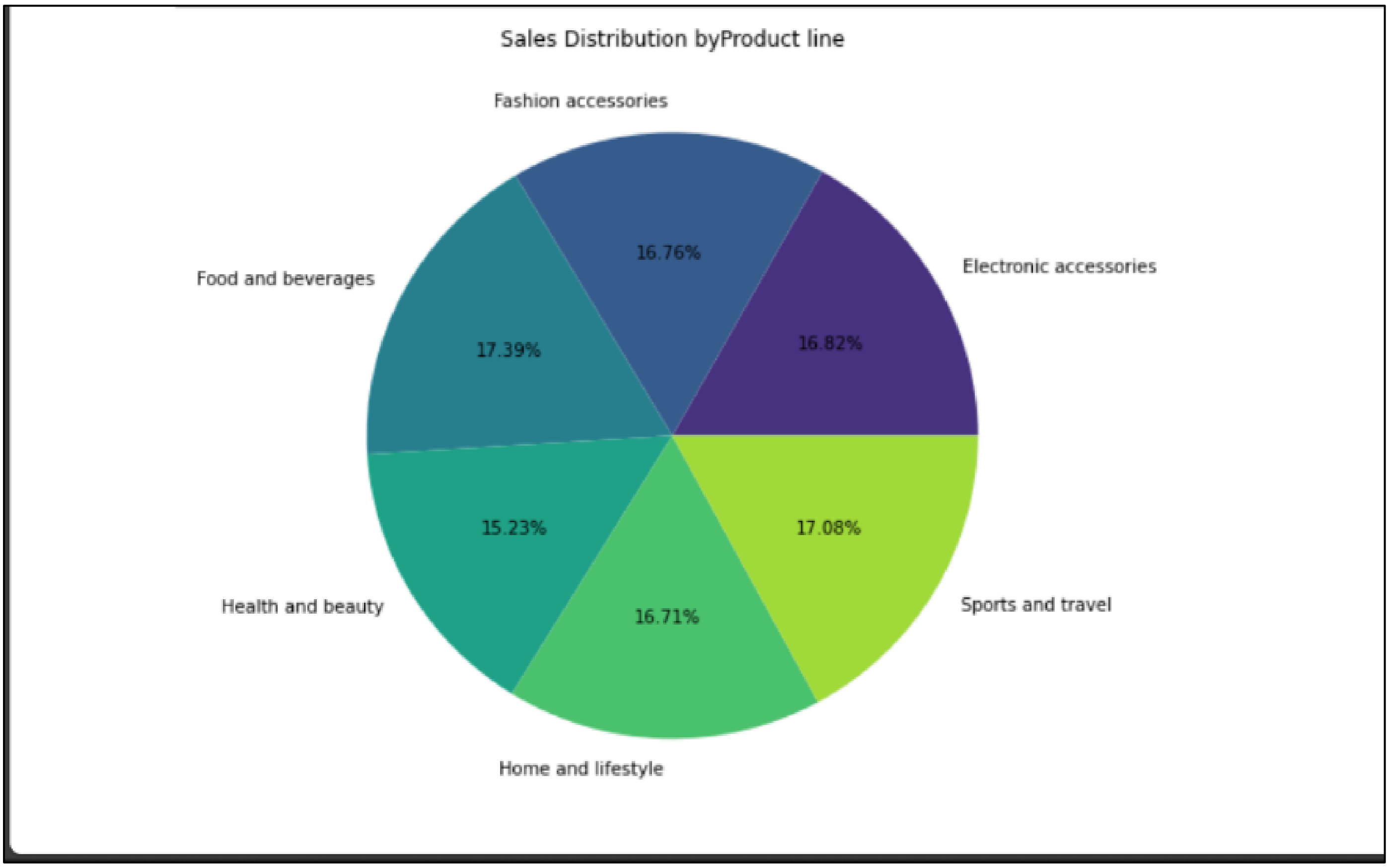

Fig 6 Sales Distribution by Product Line

Figure 7 demonstrate the sales count in each season, Autumn has the highest sales count followed by winter, then monsoon and the last one with the lowest sales count is summer. From this figure we can conclude that there is a significant change in the sales in each month across the year and with this the retailers can easily target the season that has the highest sales count and plan inventory management ahead the season. While **Figure 8** shows the sales trend across the months throughout the years.

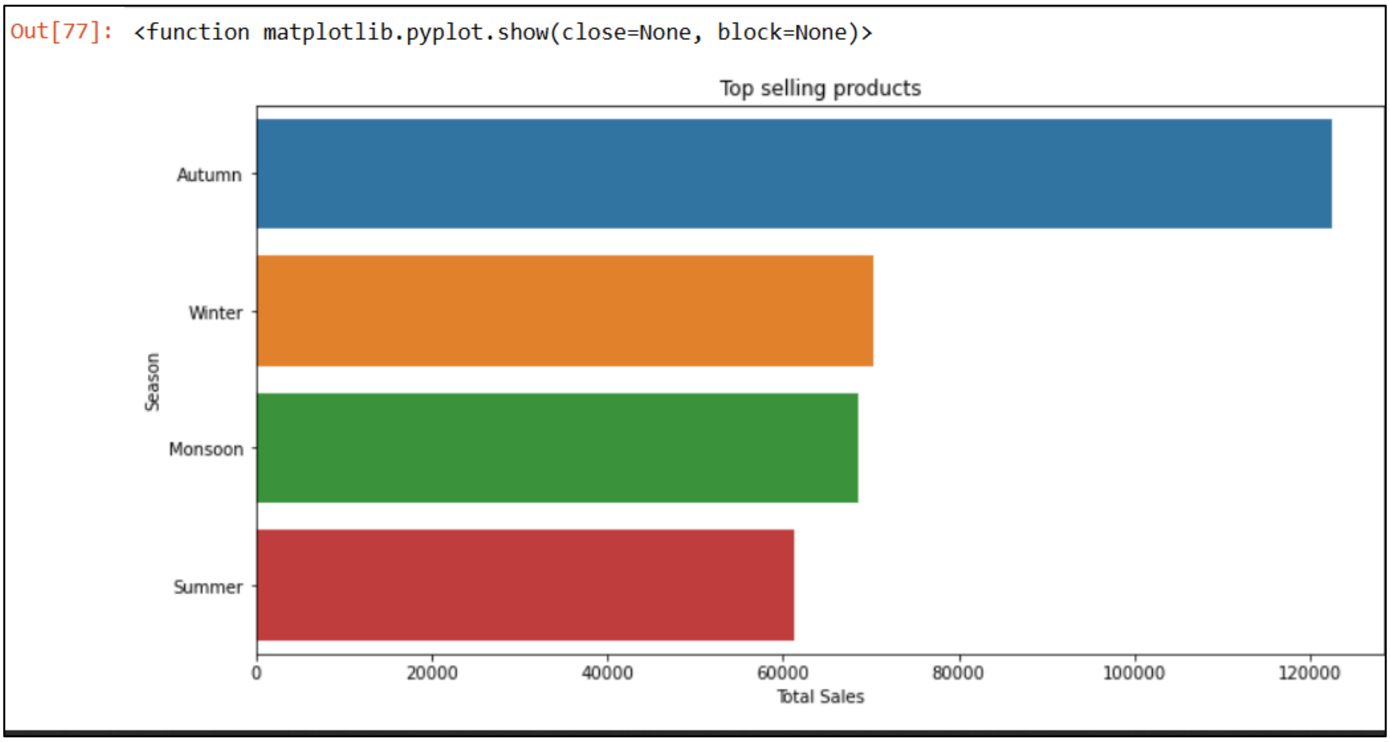

Fig 7 Seasonal Sales Count

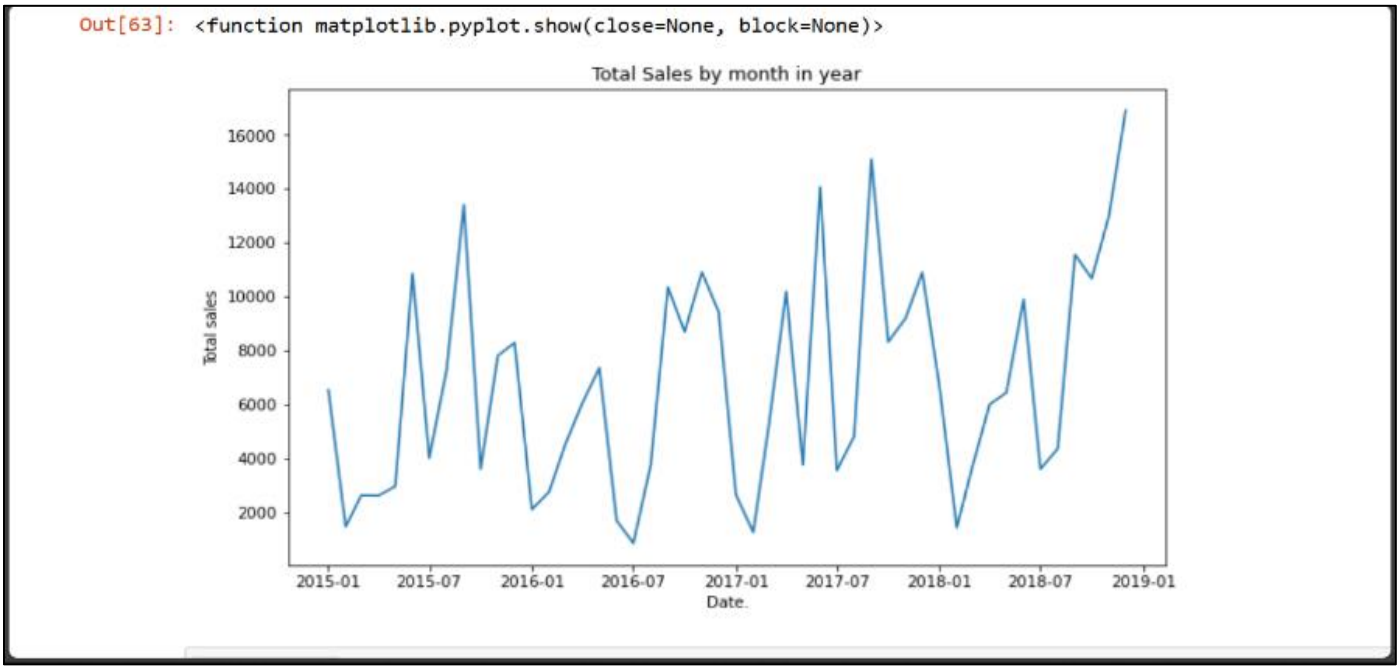

Fig 8 Sales Trends Over Time

➢ *Performance of Individual Models*

Various individual model and ensemble methods performances were compared in this work using different evaluation metrics. K-Fold Cross-Validation was done to ensure the results obtained will be robust and would enhance predictive accuracy. The best model among others with a higher R2 and lower MAE and MSE metrics is to be chosen.

➢ *Hyperparameter Tuning*

Hyperparameter tuning is essential for optimizing model performance. The primary technique used in this study is Grid Search. Grid search is an exhaustive search technique that evaluates all possible combinations of hyperparameters specified in a grid (Meddage et al., 2024). It ensures finding the optimal combination even though it can be computationally expensive.

Table 1 Summary of the Models' Performance Metrics

Model	MAE	MSE	R^2
Support vector	**108.653**	**31274.17**	**0.5193**
Random Forest	**4.85543**	**56.38510**	**0.9840**
XGBoost	**7.48054**	**103.3136**	**0.9984**
Stacking	**5.12214**	**63.51036**	**0.9990**

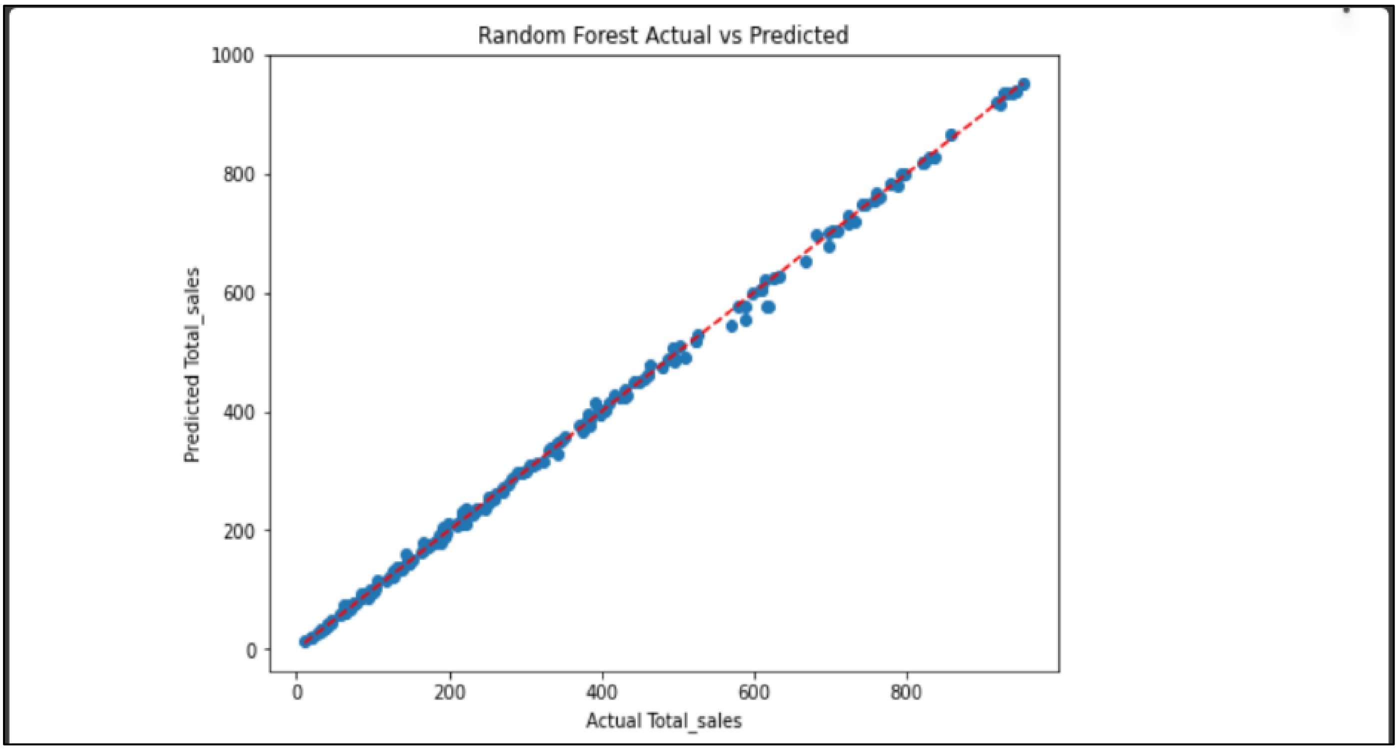

Fig 9 Random Forest Actual and Predicted Values

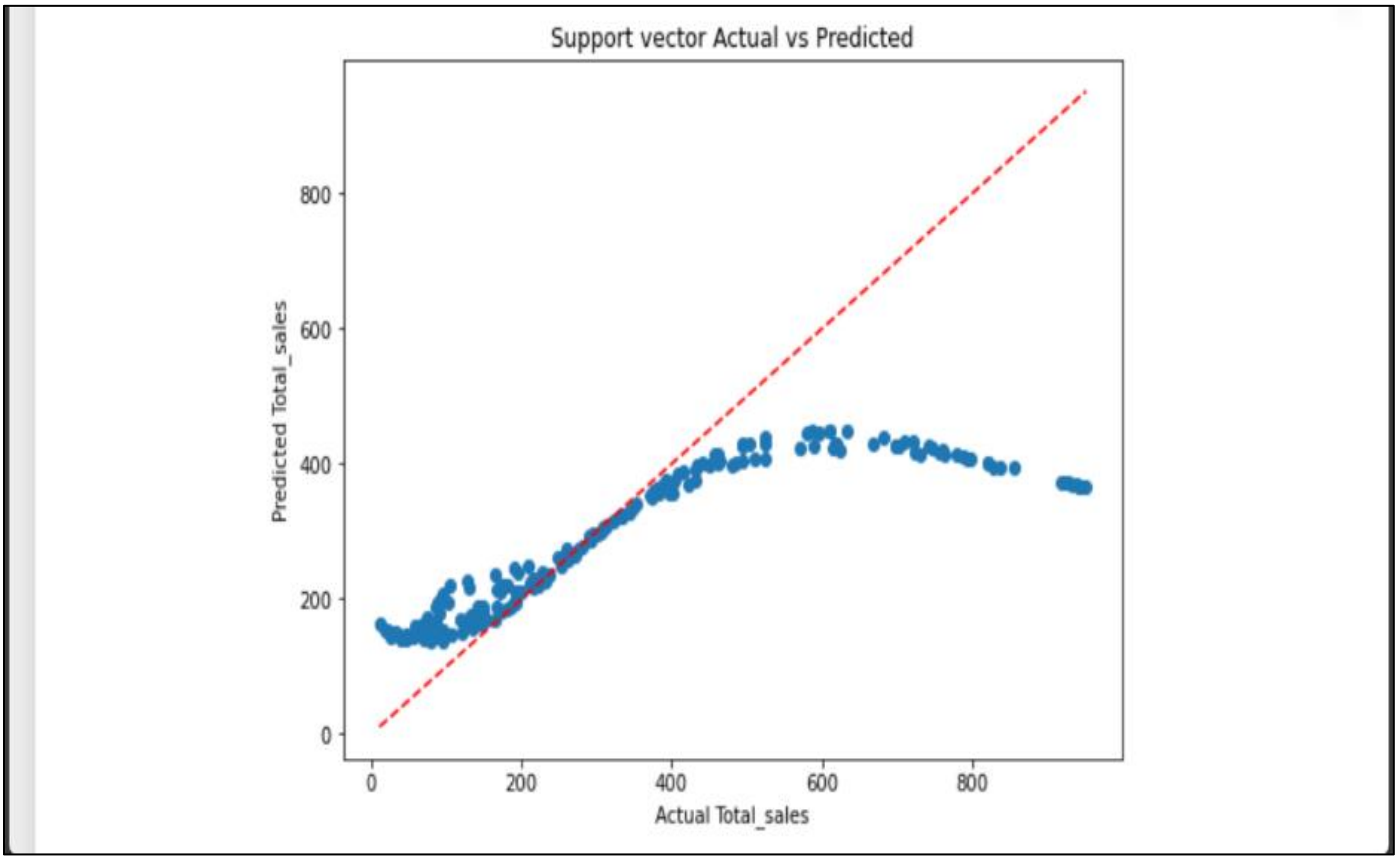

Fig 10 Support Vector Actual and Predicted Values

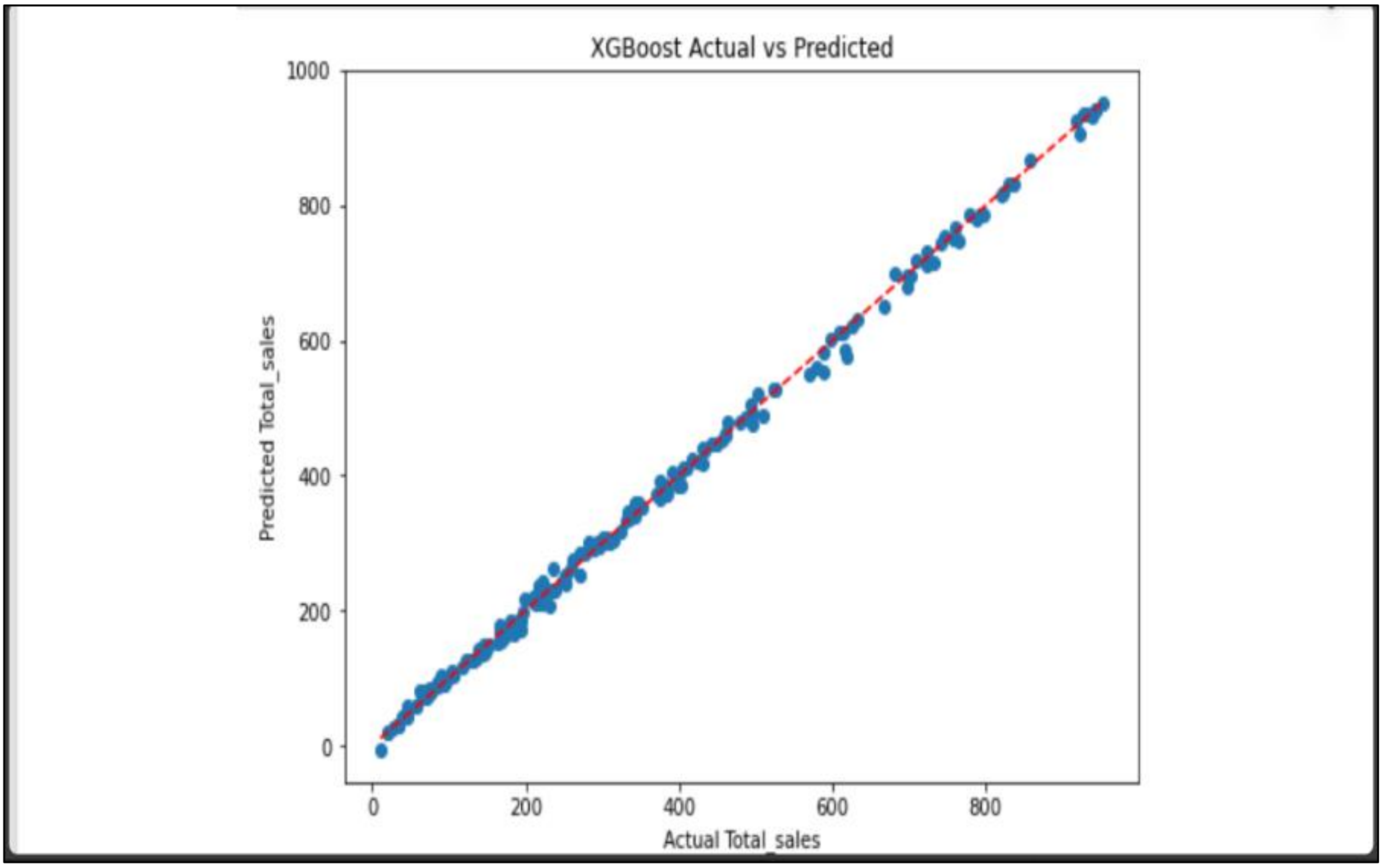

Fig 11 XGBoost Actual and Predicted Values

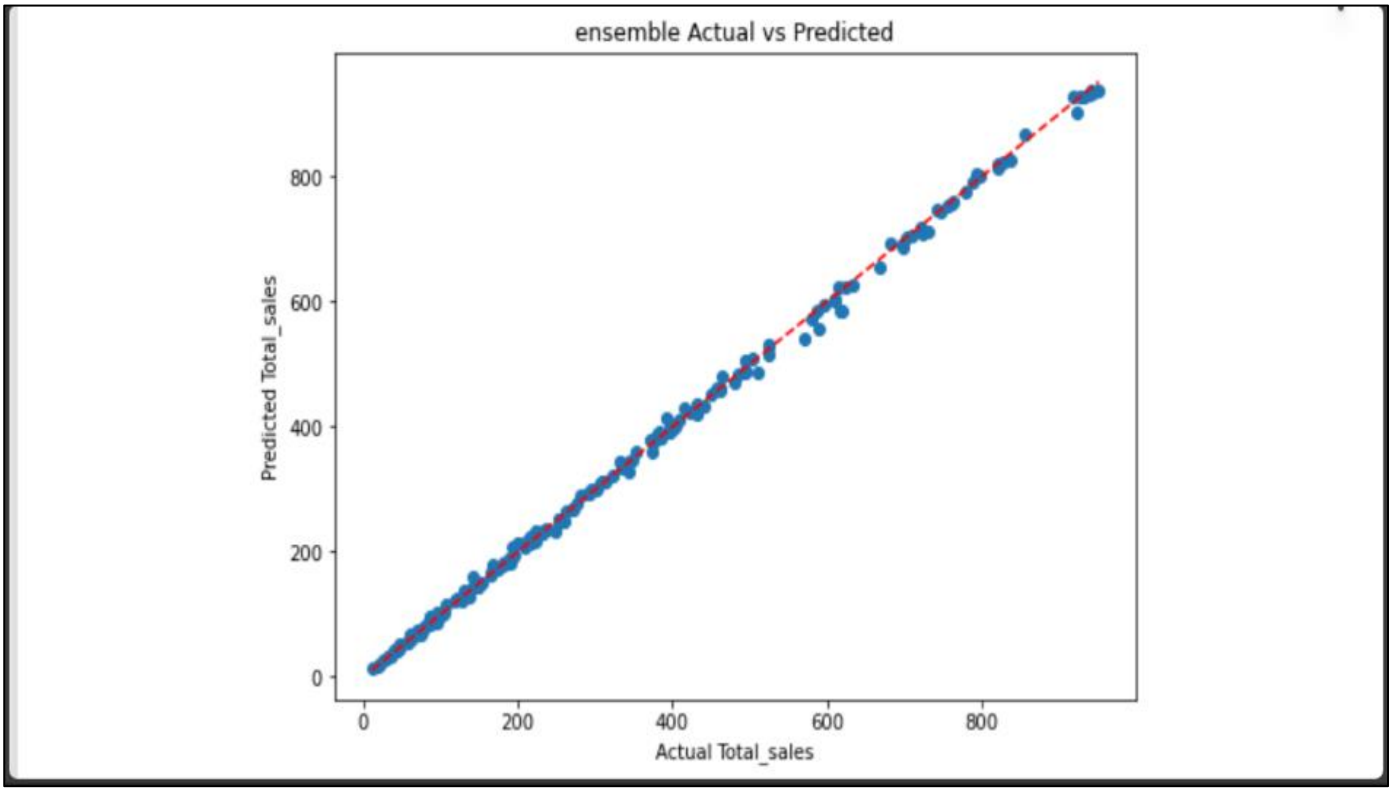

Fig 12 Stacking Actual vs Predicted Values

➢ *Cross-Validation (CV) Technique*

To check the generalization capabilities and resiliency, the models were evaluated with k-fold cross-validation. The data were divided into k subsets of approximately equal size. The model was trained on k-1 subsets, and using the remaining subset as test data the, process was repeated k times, with each subset serving as a test set exactly once. The average performance over these k iterations provided a very reliable estimate of the model's performance on new, unseen data following a five-fold cross-validation approach.

V. CONCLUSION AND FUTURE WORK

The findings indicate that leveraging machine learning for sales prediction can lead to more informed decision-making in supermarkets. By employing advanced predictive models, businesses can better understand market trends, optimize their inventory levels, and enhance customer satisfaction through targeted marketing strategies. Key conclusions from the study include the significant enhancement in predictive accuracy of sales models through ensemble methods, especially Stacking, compared to individual algorithms. The XGBoost algorithm is particularly noted for its robustness in sales prediction, effectively capturing underlying patterns without overfitting. Future studies should explore hybrid models that combine multiple machine learning techniques and consider external factors influencing sales to improve predictive accuracy and invest in improving the quality and comprehensiveness of data collected from various sources, including point-of-sale systems and customer loyalty programs.

REFERENCES

[1]. Abdullahi, M. I., Aimufua, G. I. O., Muhammad, U. A. (2021). Application of sales forecasting model based on machine learning algorithms. In Proceedings of the 28th iSTEAMS Intertertiary Multidisciplinary Conference (pp. 205-216). *American International University West Africa, The Gambia*. https://doi.org/10.22624/AIMS/iSTEAMS-2021/V28P17

[2]. Aimufua, G. I. O., Abdullahi, M. I., & Muhammad, U. A. (2021). Application of sales forecasting model based on machine learning algorithms. In Proceedings of the 28th iSTEAMS Intertertiary Multidisciplinary Conference (pp. 205-216). *American International University West Africa, The Gambia*. https://doi.org/10.22624/AIMS/iSTEAMS-2021/V28P17

[3]. Alice, K., Andrabi, S. H. u. H., & Jha, S. (2024). Sales Forecasting Based on Ensemble Learning. Unpublished manuscript.

[4]. Bajaj, P., Ray, R., Shedge, S., & Vidhate, S. (2020). Sales Prediction using Machine Learning Algorithms. *International Research Journal of Engineering and Technology*, 7(6), 3619-3625.

[5]. Das, V., Mao, T., Geng, Z., Flores, C., Pelloso, D., & Wang, F. (2023). Enhancing sell-in and sell-out forecast ting using ensemble machine learning method. Unpublished manuscript.

[6]. DPP Medagge, Isuri Fonseka, D Mohotti, K Wijesooriya, CK Lee. (2024). An explainable machine learning approach to predict the compressive strength of graphene oxide-based concrete. *Construction and Building Material* 449 (13834), 6,.

[7]. Guru, P., Sathyapriya, J., Rajandran, K. V. R., Bhuvaneswari, J., & Parimala, C. (2024). Product sales forecasting and prediction using machine learning algorithm. *International Journal of Intelligent Systems and Applications in Engineering*, 12(4s), 355-366.

[8]. Hasan, M. R. (2024). Addressing seasonality and trend detection in predictive sales forecasting: A machine learning perspective. *Journal of Business and Management Studies*, 6(2), 100-109. https://doi.org/10.32996/jbms.2024.6.2.10

[9]. Hu, Y. (2022). Sales Prediction Based on State-of-art Machine Learning Scenarios. *In FFIT 2022: 2022 6th International Symposium on Frontiers in Information Technologies.* https://doi.org/10.4108/eai.28-10-2022.2328458

[10]. Kamble, T., Vardhan, H., Ghuge, M., Shelar, Y., Rana, R., & Machale, T. (2023). Ensemble Machine Learning Models to Forecast Sales. *Proceedings of the Third International Conference on Innovative Mechanisms for Industry Applications* (ICIMIA 2023), 1056-1061. doi:10.1109/ICIMIA60377.2023.10426565

[11]. Kaplan, A., Mark, J. F., & Mastoridis, P. (2021). Artificial Intelligence/Machine Learning in Respiratory Medicine and Potential Role in Asthma and COPD diagnosis. *The journal of Allergy and Clinical immunology in Practice,* DOI: 10.1016/j.jaip.2021.02.014

[12]. Koh, Y. W., Joseph, M. H., & Sivakumar, V. (2024). Big Mart sales prediction using machine learning. *EAI Endorsed Transactions on Internet of Things*, 10. doi:10.4108/eetiot.6453

[13]. Kramar, V., & Alchakov, V. (2023). Time-series forecasting of seasonal data using machine learning methods. *Algorithms*, 16(5), 248. https://doi.org/10.3390/a16050248

[14]. Mohamed, M. A., El-Henawy, I. M., & Salah, A. (2022). Price Prediction of Seasonal Items Using Machine Learning and Statistical Methods. *Computers, Materials & Continua,* DOI:10.32604/cmc.2022.020782.

[15]. Niya, N.J. and Jose, J. (2021). Sale Prediction using Linear Regression Model. *International Journal of Creative Research Thoughts*, 9(3), 1430-1433.

[16]. Patangia, S., Mohite, R., Shah, K., Kolhe, G., Mokashi, M., & Rokade, P. (2020). Sales Prediction of Market using Machine Learning. *International Journal of Engineering Research and Technology*, 9(9), 708-713.)

[17]. Ramachandran, K.K. (2023). Predicting Supermarket Sales with Big Data Analytics: A Comparative Study of Machine Learning Techniques. *International Journal of Data Analytics*, 1(1), 1-11.

[18]. Raizada, S., & Saini, J.R. (2021). Comparative Analysis of Supervised Machine Learning Techniques for Sales Forecasting. *International Journal of Advanced Computer Science and Applications*, 12(11), 102-110.

[19]. Ranjitha, P., & Spandana, M. (2021). Predictive Analysis for Big Mart Sales Using Machine Learning Algorithms. *2021 International Conference on Intelligent Computing and Control Systems (ICICCS)*, 1416-1421.

[20]. Sharma, p., Khater, S., and Vashisht, V. (2021). Sales Forecast of Manufacturing Companies using Machine Learning navigating the Pandemic like COVID – 19. *2021 2nd International Conference on Computation, Automation and KnowledgeManagement (ICCAKM)*, 2021, pp.1 - 5.
[21]. Swami, D., Shah, A. D., & Ray, S. K. B. (2020). Predicting Future Sales of Retail Products using Machine Learning. *arXiv preprint arXiv*:2008.07779.
[22]. Tom, M., Raju, N., Issac, A., James, J., & R, R.S. (2021). Supermarket sales prediction using regression. *International Journal of Advanced Trends in Computer Science and Engineering*, 10(2), 1153-1157.
[23]. Varshini S & Preethi, D. (2022). An analysis of machine learning algorithms to predict sales. *International Journal of Science and Research*, 11(6).
[24]. Vasuki, M., Amalraj Victorie, T., & Karishmaa Shamim, R. K. (2024). Survey on prediction of sales using machine learning techniques. *International Research Journal of Modernization in Engineering Technology and Science*, 6(5). https://doi.org/10.56726/IRJMETS56619
[25]. Wisesa, O., Adriansyah, A., & Khalaf, O.I. (2021). Prediction Analysis Sales for Corporate Services Telecommunications Company using Gradient Boost Algorithm. *2021 IEEE Conference on Big Data and Analytics (ICBDA)*, 101-106.
[26]. Wendy, T. (2018). Why use Machine Learning instead of Traditional Statistics. *Towards Data Science.*
[27]. Zhou, Z. H. (2012). Ensemble methods: foundations and algorithms. Chapman and Hall/CRC. https://doi.org/10.1201/b12207