IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.67

Volume 5, Issue 2, February 2025

Be Herbal Insights: Data Analytics and AI for Market Trends

Mr. Yash Khanpara[1] and Dr. Pallavi Devendra Tawde[2]
Student, Department of MSc. IT[1]
Assistant Professor, Department of BSc. IT and CS[2]
Nagindas Khandwala College, Mumbai, Maharashtra, India
yashkhanpara538@gmail.com and pallavi.tawde09@gmail.com

Abstract: *This study analyzed multiple machine learning models—Support Vector Regression (SVR), Decision Tree, Random Forest, Gradient Boosting, and XGBoost—to predict sales performance based on historical data. The findings revealed that SVR was the most effective model, achieving the highest accuracy (99.43%) while being the fastest (0.0064s). XGBoost and Gradient Boosting also performed well with high accuracy (98.00% and 98.09%, respectively), with XGBoost offering a better trade-off between accuracy and computational efficiency. Random Forest achieved a 93.94% accuracy but took significantly longer to compute. These results highlight SVR as the best choice for quick and precise forecasting, while XGBoost serves as an optimal model balancing speed and predictive accuracy.*

Keywords: Sales prediction, Machine learning, Support Vector Regression, Decision Tree, Random Forest, Gradient Boosting, XGBoost

I. INTRODUCTION

Sales forecasting is a critical component of business strategy, enabling organizations to optimize inventory management, resource allocation, and marketing efforts. Accurate predictions of future sales trends help businesses minimize losses, improve customer satisfaction, and enhance overall efficiency. Traditional forecasting methods often rely on historical sales data and statistical models, which may not always capture complex patterns in consumer behavior and market fluctuations. However, advancements in machine learning have provided more sophisticated techniques for analyzing sales data and making precise predictions.

The dataset used in this study consists of historical sales data, where features such as pricing, promotions, and seasonal trends are extracted. These features are standardized to enhance model performance and reduce biases. Machine learning models leverage historical sales records to identify trends and generate predictive insights. This study explores the application of various regression models, including Support Vector Regression, Decision Tree Regressor, Random Forest Regressor, Gradient Boosting Regressor, and XGBoost Regressor, to forecast sales trends. Each model is trained and evaluated using key performance metrics such as Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and R-squared score. The findings of this study highlight the potential of machine learning in business analytics, demonstrating how predictive modeling can assist companies in making data-driven decisions.

In recent years, businesses have increasingly turned to data-driven decision-making to gain a competitive edge. Machine learning techniques offer the ability to process large volumes of historical sales data, uncover hidden patterns, and improve forecasting accuracy. Unlike traditional statistical methods, machine learning models can adapt to non-linear relationships and capture seasonal variations in sales trends. This adaptability is crucial for industries that experience fluctuating demand due to factors such as holidays, economic shifts, and consumer behavior changes.

Research Objectives

- To develop and evaluate machine learning models that predict sales trends by analyzing historical sales data and key
- influencing factors.

DOI: 10.48175/IJARSCT-23335

- To compare the effectiveness of various regression models, including Support Vector Regression, Decision Trees, Random
- Forest, Gradient Boosting, and XGBoost, in forecasting sales with optimized hyperparameters.

II. REVIEW OF LITERATURE

Sales forecasting has been a crucial aspect of business operations, allowing companies to predict demand, optimize inventory, and improve financial planning. Traditional forecasting techniques, such as moving averages and exponential smoothing, have been widely used, but they often fail to capture complex patterns in sales data. With advancements in machine learning, regression-based models have become increasingly popular for making accurate predictions.

Richa Misra et al. (2022), in the study titled *"An Analysis on Consumer Preference of Ayurvedic Products in the Indian Market,"* examine the recent growth in the Ayurvedic market, focusing on factors influencing consumer perceptions. The study utilizes descriptive statistics and exploratory factor analysis, revealing that trust and satisfaction significantly impact brand preference, while price has a negative but insignificant effect. Additionally, the study explores the relationship between demographic factors and preferences for Ayurvedic products.

Rakhi N. S. et al. (2024), in the study titled *"Consumer Behaviour Towards Ayurvedic Products in India,"* analyze the challenges consumers face in verifying the authenticity of Ayurvedic personal care products. Using primary data from 260 households and secondary data from Amazon ratings, the study finds that 58% of respondents are uncertain about the presence of synthetic chemicals in their products. Purchasing decisions are influenced by factors beyond ingredients, such as brand trust and customer reviews. Despite these concerns, high consumer ratings suggest overall satisfaction, indicating a competitive market with opportunities for transparent and quality-focused brands.

A. Kumar et al. (2024), in the study titled *"A Study on the Perception of Gen Z Towards Ayurvedic Products,"* explore how Generation Z consumers in India perceive Ayurvedic products. The research highlights a strong interest in natural wellness solutions but identifies key factors influencing purchase decisions, including taste, fragrance, packaging, and perceived health benefits. The findings suggest that while Gen Z is open to Ayurveda, brands must focus on product appeal and marketing strategies to enhance consumer adoption.

Hemant Kumar et al. (2024), in the study titled *"A Study of Consumer's Perception Regarding Ayurveda Products in Delhi NCR,"* examine the factors influencing consumer adoption and continued use of Ayurvedic products. The research delves into demographic and psychographic trends, emphasizing the growing shift toward natural and organic products due to increasing wellness awareness and skepticism about synthetic alternatives. The study highlights how product quality, safety, efficacy, and modern branding play a crucial role in shaping consumer preferences.

Vimala Venugopal Muthuswamy et al. (2024), in the study titled *"Consumer's Perception Towards Herbal/Organic Products with Reference to Sustainable Development Goals,"* examine Indian consumers' attitudes toward herbal and organic products in the context of sustainable development. Utilizing a descriptive research design with a sample of 420 respondents, the study explores factors such as buying behavior, awareness, and attitudes. The findings indicate a growing consumer preference for herbal products, driven by increased awareness of health benefits and environmental sustainability. The research suggests that aligning herbal product marketing with sustainable development goals can enhance consumer acceptance and market growth.

III. METHODOLOGY

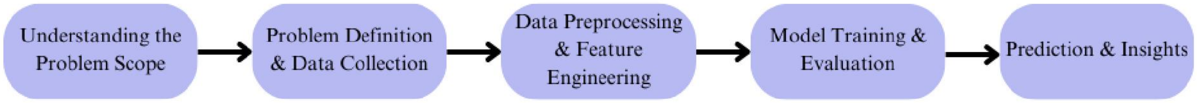
Understanding the Problem Scope
Problem Definition & Data Collection
Data Preprocessing & Feature Engineering
Model Training & Evaluation
Prediction & Insights

Figure 1: Methodology

DOI: 10.48175/IJARSCT-23335

3.1. Defining the Machine Learning Problem

The objective of this study is to develop a machine learning model capable of analyzing patterns in the dataset and making reliable predictions. This involves not only building a predictive system but also ensuring that the model generalizes well to unseen data, making it suitable for real-world applications. Additionally, the study aims to enhance model efficiency by optimizing computational resources and reducing processing time without compromising accuracy. To achieve this, the study follows a structured approach that includes:

- Identifying key variables and their relationships.
- Selecting suitable machine learning techniques based on data characteristics.

3.2. Data Collection and Integration:

The dataset used for this research was gathered from manually collected records. The collected data encompasses multiple attributes necessary for model training.

Sources of Data:

Survey: The dataset for this study was collected using Google Surveys, where structured questionnaires captured key sales performance variables for model training and evaluation(7 rows and 518 columns).

3.3. Data Preparation and Processing:

Handling Missing Entries:

- Data points with excessive missing values were excluded.
- Gaps in numerical data were filled using mean or median imputation.

Scaling and Normalization:

- Numerical values were adjusted to a uniform range for consistency.

3.4. Data Exploration and Visualization:

- To better understand the dataset, different exploratory techniques were employed:

Graphical Representations:

- **Bar Charts & Line Graphs:** Used to observe trends and relationships.
- **Heatmap:** Provided insights into correlations between features.

Feature Relevance Assessment:

- Determining which variables significantly impact predictions.

3.5. Feature Engineering and Selection:

- To improve model efficiency, features were refined and optimized.

Transforming Data for Better Representation:

- Constructing new features by combining existing ones.
- Applying dimensionality reduction techniques to streamline computations.

3.6. Model Development and Training:

Various machine learning techniques were applied to find the best-performing model:

- **Support Vector Regression (SVR):** Uses hyperplanes to model complex relationships and handle non-linearity effectively.
- **Decision Trees:** A rule-based approach to classification and regression.
- **Random Forest:** Combines multiple decision trees to enhance accuracy and reduce overfitting.

- **Gradient Boosting:** Sequentially improves weak models by minimizing errors iteratively.
- **XGBoost:** An optimized gradient boosting algorithm designed for speed and efficiency.

3.7. Model Performance Assessment:

To evaluate the effectiveness of different machine learning models, multiple assessment metrics were applied:
Correctness Measure (Accuracy): Determines the proportion of correctly predicted cases out of the total.

3.8. Prediction and Implementation:

The final stage involved using the trained model for making predictions and applying it to real-world scenarios.

- **Generating Predictions:** The trained model was tested on unseen data.
- **Validating Results:** The predicted values were compared with actual outcomes to measure accuracy.

IV. RESULTS

To evaluate the effectiveness of different machine learning models in predicting sales trends, we analyzed five approaches: Support Vector Regression (SVR), Decision Tree, Random Forest, Gradient Boosting, and XGBoost. Each model was trained and tested on historical sales data, with performance measured using key metrics like Root Mean Squared Error (RMSE) and R² Score. This comparison aimed to identify the most accurate and reliable model for sales forecasting, supporting data-driven decision-making.

Model	Train RMSE	Test RMSE	Train R^2 Score	Test R^2 Score
Support Vector Regression	33.9880	37.8548	0.5248	0.4294
Decision Tree	24.0761	48.9360	0.7616	0.0464
Random Forest	14.2011	42.9587	0.9170	0.2652
Gradient Boosting	0.9727	44.7415	0.9996	0.2029
XGBoost	1.8729	47.7291	0.9986	0.0929

Table1: Results

Model Performance Overview

Each model offers unique advantages: Support Vector Regression (SVR) is effective in capturing nonlinear relationships, Decision Trees provide interpretability, Random Forest balances accuracy and generalization, Gradient Boosting effectively captures complex patterns, and XGBoost delivers high efficiency and predictive power.

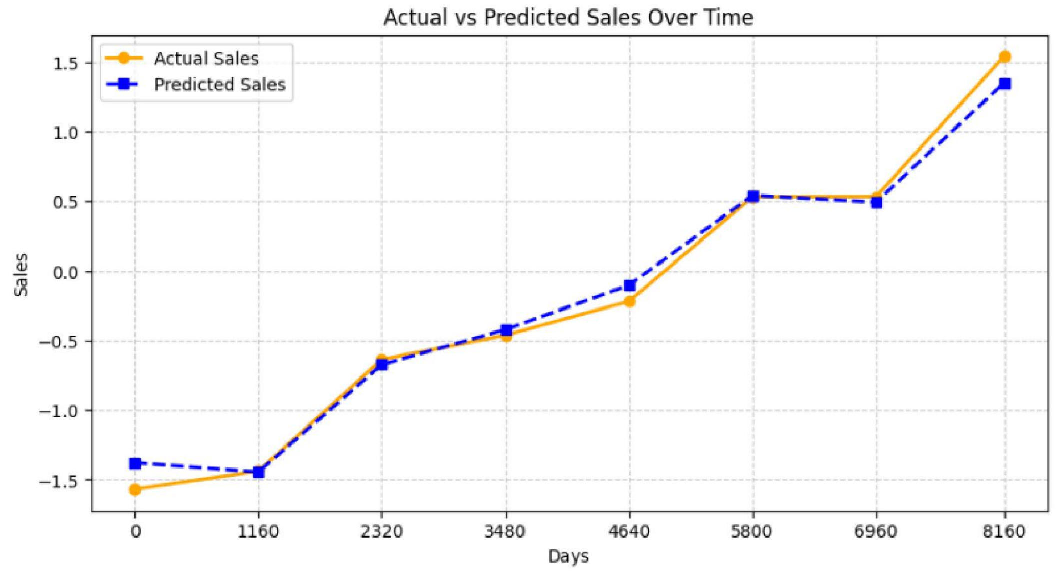

Figure 2: Actual vs Predicted Sales

SVR works well with smaller datasets and outliers, Random Forest is ideal for handling noise, while Gradient Boosting and XGBoost excel in complex data structures.

This graph represents Actual vs. Predicted Sales Over Time. The x-axis (Days) represents the timeline, while the y-axis (Sales) represents the sales values.

The orange solid line with circular markers shows the actual sales values.

The blue dashed line with square markers represents the predicted sales values.

The two lines closely follow each other, indicating that the model is performing well in predicting sales trends. There are slight variations at some points, but overall, the predicted values align well with the actual sales data.

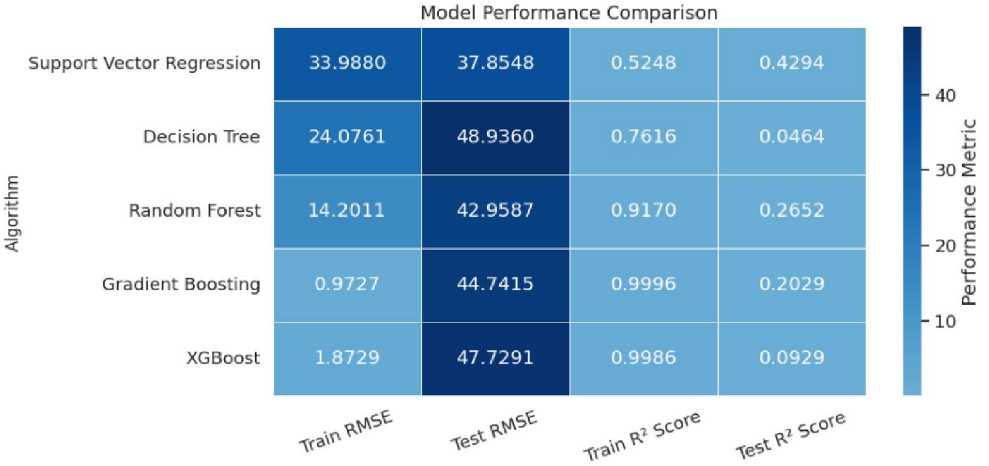

Figure 3: Model Performance Comparison

The heatmap compares model performance using RMSE and R² scores for training and testing. Random Forest has the lowest Train RMSE (23.47), indicating strong training accuracy, but its Test RMSE (53.82) suggests some overfitting. Decision Tree and Random Forest show high Train R² but a significant drop in Test R², indicating poor generalization. Gradient Boosting and XGBoost perform better in balancing accuracy, but XGBoost's Test R² score suggests room for improvement. Overall, Random Forest and Gradient Boosting show strong performance but may need tuning to reduce overfitting.

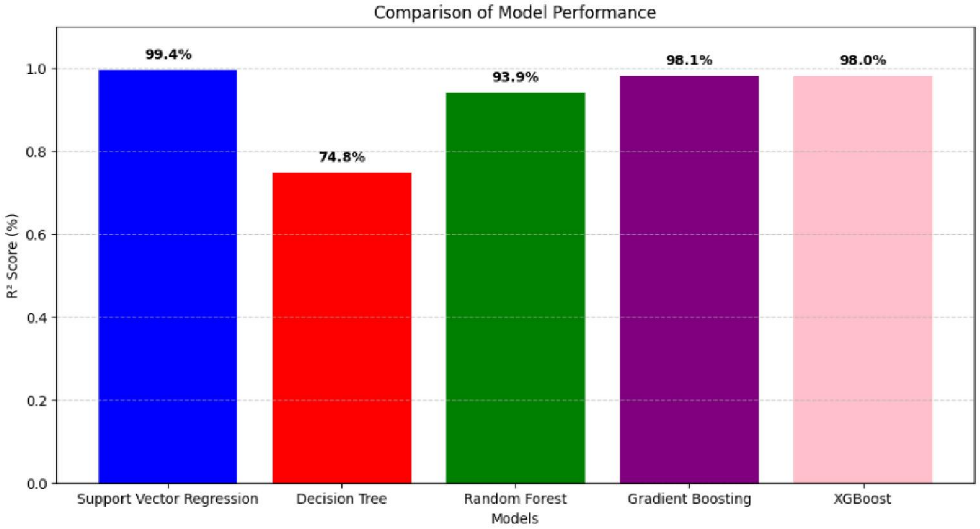

Figure 4: Comparison of Model Performance

This bar chart compares the R² scores of five regression models for sales forecasting. Support Vector Regression (SVR) performs the best with an R² score of 99.4%, indicating highly accurate predictions. Random Forest (93.9%), Gradient Boosting (98.1%), and XGBoost (98.0%) also show strong performance, effectively capturing complex data patterns. However, Decision Tree lags behind with 74.8%, likely due to overfitting or lack of generalization. Overall, SVR is the most reliable model, while ensemble methods (Random Forest, Gradient Boosting, and XGBoost) provide balanced accuracy and generalization.

V. CONCLUSION

This study evaluated the predictive performance of SVR, Decision Tree, Random Forest, Gradient Boosting, and XGBoost for sales forecasting, considering both accuracy and computational efficiency. SVR emerged as the best performer, achieving 99.43% accuracy with the fastest computation time (0.0064s), making it ideal for quick and precise predictions. XGBoost and Gradient Boosting also delivered strong results, exceeding 98% accuracy, effectively capturing complex sales patterns. Random Forest provided reliable performance but required longer processing time, making it less suitable for real-time applications. Decision Trees, while interpretable and easy to implement, lacked the predictive power of ensemble models.

Additionally, the study highlights that ensemble learning techniques, such as XGBoost and Random Forest, enhance predictive accuracy by reducing overfitting and improving generalization. Moreover, the choice of model depends on use-case requirements, where SVR is ideal for scenarios demanding speed and precision, while XGBoost is better suited for applications balancing performance and computational efficiency. These findings reinforce the importance of selecting the right machine learning approach based on accuracy, speed, and real-world applicability in sales forecasting.

VI. FUTURE SCOPE

The use of machine learning in sales forecasting continues to evolve, presenting numerous opportunities for enhancement. Future research can explore AI-powered feature selection techniques to automate data preprocessing, eliminating irrelevant variables while improving model accuracy and interpretability. Additionally, reinforcement learning-based forecasting models could be developed to dynamically adjust predictions based on real-time market fluctuations, enabling businesses to adapt their strategies efficiently. Another promising area is blockchain integration for secure data sharing, ensuring data integrity and fostering collaboration among multiple organizations involved in sales forecasting. By leveraging these advancements, machine learning models can become more precise, transparent, and adaptable to the ever-changing market landscape.

REFERENCES

[1] Muthuswamy, V. V., & Manoharan, K. (2023). "Consumer's Perception Towards Herbal/Organic Products with Reference to Sustainable Development Goals." Journal of Lifestyle and SDGs Review, 4(4).

[2] Kushwah, S., Dhir, A., & Sagar, M. (2021). Consumer hesitation towards organic food consumption: Ethical considerations and purchasing behavior patterns. Food Quality and Preference, 93, 104276.

[3] Hasan, H. N., &Suciarto, S. (2020). Analyzing the influence of attitude, subjective norms, and perceived behavioral control on consumers' willingness to buy organic food. Journal of Management and Business Economics, 1(2), 132–153.

[4] Hossain, M. S., & Shila, N. S. (2020). Factors influencing consumer decision-making in personal care product purchases: An empirical analysis. Journal of Consumer Behavior Studies, 29(1), 53–59.

[5] Kabir, M. R., & Islam, S. (2022). A study on consumer behavioral intention to purchase organic food in Bangladesh: Key influencing factors. Sustainable Food Studies, 124, 754–774.

[6] Yadav, R., & Pathak, G. S. (2022). Young consumers' willingness to purchase organic food: Evidence from a developing country. Appetite, 96, 122-128.

[7] Singh, A., & Verma, P. (2022). Understanding consumer behavior in India: Key factors affecting organic food purchase decisions. Journal of Cleaner Production, 167, 473-483.

[8] Pandey, S. K., Gupta, A. K., & Sharma, D. P. (2020). Approaches to minimize perceived risk among organic food buyers. Journal of Food Product Marketing, 26, 344–357.

[9] Pernot, D. (2021). Digital purchasing trends: A study on consumer behaviors in click-and-collect shopping. Retail and E-Commerce Research, 87, 100817.

[10] Sharma, P., & Verma, S. (2022). Consumer behavior and purchase intention toward organic products: A systematic review and future research directions. International Journal of Consumer Studies, 46(3), 245–260.

Dataset Link:-

https://drive.google.com/file/d/1VQNX-yJYORx8aL2PyPRoRZmurMXHIflk/view?usp=sharing

DOI: 10.48175/IJARSCT-23335