ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 11| Issue: 10| October 2025|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2025: 8.691 || ISI Value: 1.188

PRICE OPTIMIZATION IN RETAIL USING PREDICTIVE ANALYSIS

Shruti Vivek Gavali
Department of Data Science, Dr. D. Y. Patil College, Pimpri, Pune, India

ABSTRACT

In today's competitive retail market, dynamic pricing has become a crucial component of successful business strategies. This paper explores how predictive analytics and machine learning can be effectively utilized to optimize retail prices and maximize profits. Traditional static pricing methods fail to adapt to the fast-changing market conditions influenced by seasonality, competition, and consumer behavior. The study implements and compares three supervised machine learning models – Linear Regression, Random Forest, and Support Vector Regression (SVR) – to predict optimal pricing using real retail data. After extensive training, tuning, and validation, the Random Forest model emerged as the most accurate, achieving an R^2 score of 0.9897, indicating a near-perfect fit. This research highlights how data-driven approaches enhance decision-making accuracy and pave the way for real-time dynamic pricing in modern retail environments.

KEYWORDS: *Dynamic Pricing, Price Optimization, Predictive Analysis, Machine Learning, Random Forest, Linear Regression, Support Vector Regression, Retail Analytics.*

INTRODUCTION

Pricing plays a pivotal role in determining a company's success in the retail and e-commerce sector. Consumers today are more informed, price-sensitive, and influenced by promotional strategies and digital accessibility. In such a dynamic market, relying solely on traditional pricing methods—where prices remain fixed for extended periods—is inefficient. Dynamic pricing, powered by predictive analytics and artificial intelligence, allows businesses to adjust prices based on real-time factors such as demand, competition, seasonality, and inventory levels. Machine learning models can analyze historical and live data to identify trends, predict consumer behavior, and recommend optimal price points. This research focuses on applying predictive models to retail datasets to enhance profitability while maintaining competitiveness.

LITERATURE REVIEW

Multiple studies have examined the application of machine learning for price prediction and demand forecasting. Rahman et al. (2024) found that Random Forest models outperform traditional regression models in predicting optimal prices due to their ability to capture complex interactions among features. Das et al. (2024) proposed Gradient Boosting Machines (GBM) for dynamic pricing, emphasizing model tuning for accuracy. Perumallapalli (2014) highlighted the potential of reinforcement learning in continuously improving pricing strategies based on real-time data feedback. Other research integrates clustering algorithms for customer segmentation (Sarkar et al., 2023), allowing personalized pricing strategies tailored to purchasing patterns. Together, these studies form the foundation for this research, which evaluates three predictive models to determine their practical efficiency in price optimization.

PROBLEM STATEMENT

The main challenge for retailers is determining optimal prices that maximize profit without losing competitiveness. Static pricing strategies fail to respond to market dynamics, resulting in missed revenue opportunities and inefficient inventory management. This research aims to design and evaluate predictive models capable of analyzing historical retail data to recommend data-driven pricing decisions that adapt to fluctuating market conditions.

OBJECTIVES

1. To apply predictive analytics for optimizing product prices using real-world retail datasets.
2. To compare the performance of Linear Regression, Random Forest, and SVR in predicting price and demand.
3. To identify the most accurate and reliable model using MAE, RMSE, and R^2 evaluation metrics.
4. To simulate pricing strategies using the top-performing model and analyze potential revenue improvement.
5. To propose a practical framework for implementing predictive analytics in retail decision-making.

METHODOLOGY

The methodology outlines the systematic approach followed in this research to predict product demand and identify optimal pricing strategies in the retail sector. The study uses predictive analytics and supervised machine learning models to achieve accurate price optimization results.

1.Data Collection and Pre-processing

The dataset used for this study was sourced from Kaggle, a well-known public repository for data science project. It contains various features such as unit price, competitor prices weekday,

ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 11| Issue: 10| October 2025|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2025: 8.691 || ISI Value: 1.188

weekend, holiday, and month. The target variable for prediction is quantity sold (qty).

Data pre-processing was carried out to ensure that the dataset was clean and ready for model training. The following steps were performed:

Data Cleaning: Missing or null values were checked and handled appropriately to maintain data quality.

Data Type Conversion: Features such as date and categorical variables were converted into numerical or encoded formats suitable for model input.

Feature Selection: Only relevant features that influence product demand and pricing were selected to improve model efficiency.

Normalization: Scaling techniques were applied where necessary to bring all features to a similar range, especially for models like Support Vector Regression (SVR) that are sensitive to data scale.

Train-Test Split: The dataset was divided into 80% training data and 20% testing data to evaluate the model performance on unseen samples.

2. Model Architecture

This research focuses on the comparison of three supervised machine learning algorithms — Linear Regression, Random Forest, and Support Vector Regression (SVR) to identify the best model for price optimization.

1.Linear Regression

A basic statistical model that assumes a straight-line relationship between price and demand. It provides a simple baseline for comparison.

2.Random Forest

An ensemble learning method that uses multiple decision trees and combines their outputs to make robust and accurate predictions. It handles non-linear data effectively and reduces overfitting.

3.Support Vector Regression (SVR)

A kernel-based regression technique that maps input data into a higher-dimensional space to capture complex patterns. It helps in identifying non-linear relationships between pricing and demand.

Each of these models was chosen because they represent different levels of model complexity from simple (Linear Regression) to advanced (SVR and Random Forest). This allows for a comprehensive evaluation of predictive performance for retail pricing scenarios.

3.Performance Metrics Overview

Model performance in price prediction is evaluated using three key metrics:

- RMSE (Root Mean Squared Error): Measures the average magnitude of prediction errors, giving more weight to large errors. Lower RMSE indicates higher accuracy.
- MAE (Mean Absolute Error): Calculates the average absolute difference between predicted and actual values. It's simple, robust, and less affected by outliers.
- R^2 (Coefficient of Determination): Shows how well the model explains variation in the data. Values closer to 1 indicate a stronger and more accurate fit.

4.Implementation Environment

The implementation and analysis were performed using Python programming language in the Jupyter Notebook environment. The following libraries and tools were used throughout the project:

Pandas and NumPy for data cleaning, manipulation, and analysis.

Matplotlib and Seaborn for data visualization and graphical representation of model results.

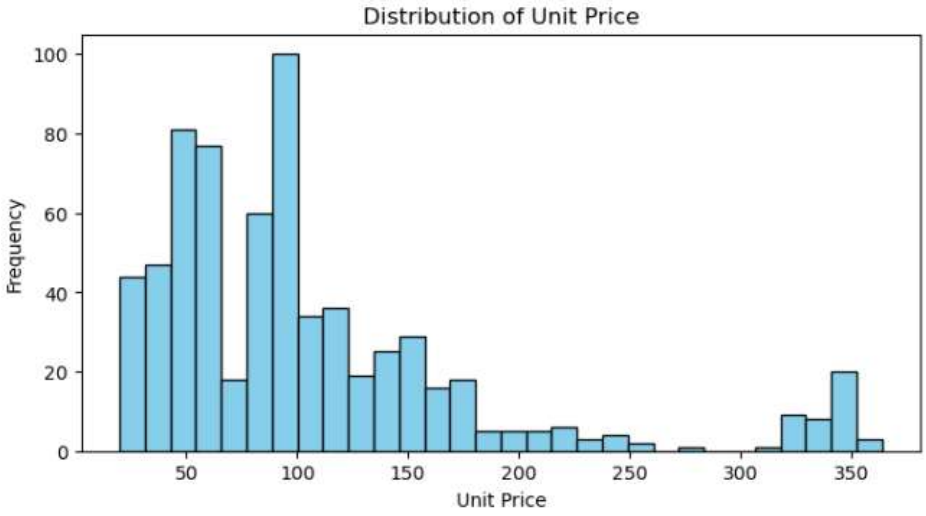
Scikit-learn (sklearn) for implementing machine learning algorithms like Linear Regression, Random Forest, and SVR.

GridSearchCV from sklearn for hyperparameter tuning of SVR

5.Visualizations and Interpretation

5.1 Distribution of Unit Price

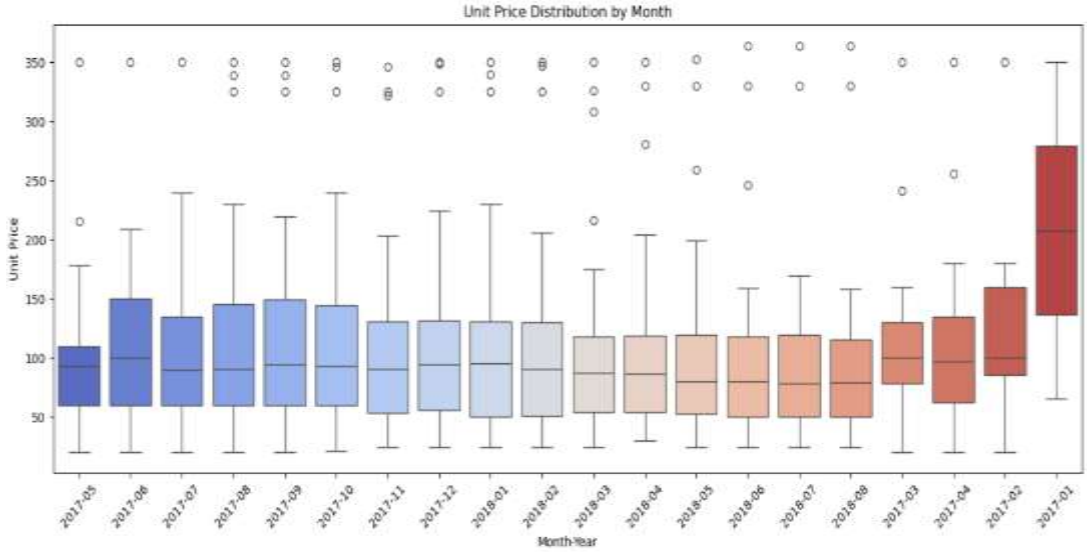
This Fig1 shows how product prices are distributed in the dataset. Most products fall within a moderate price range, with few high-priced outliers.



ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 11| Issue: 10| October 2025|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2025: 8.691 || ISI Value: 1.188

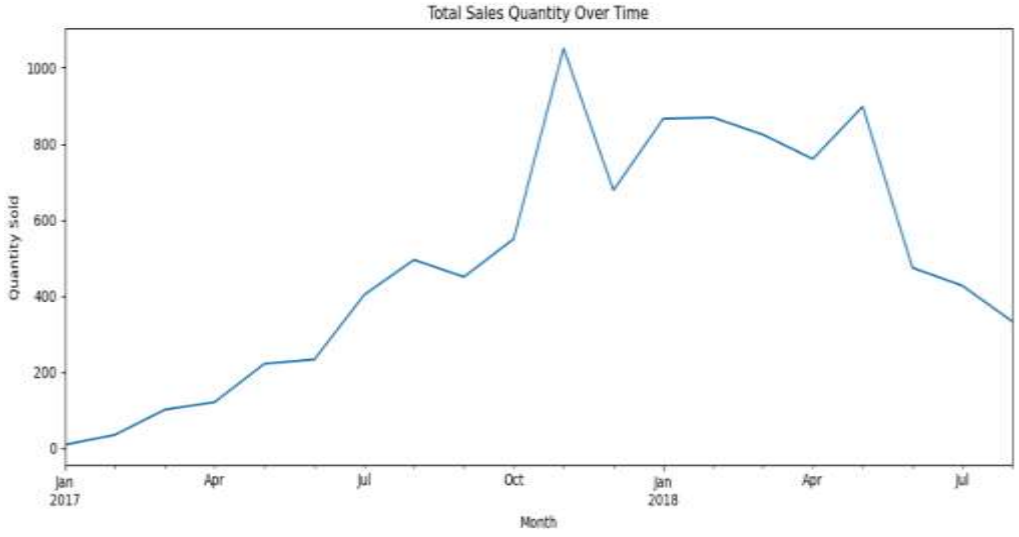
5.2 Unit Price Distribution by Month

This box plot illustrates monthly variations in product prices. Median prices stay stable, with some months showing higher average prices.



5.3 Total Sales Quantity Over Time

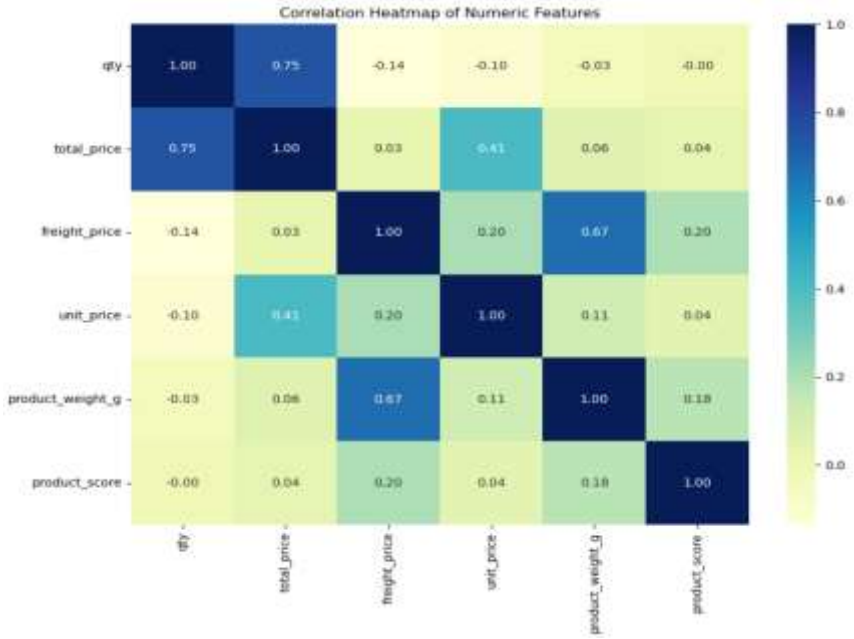
The line chart represents total sales over time, showing demand trends across different months.



ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 11| Issue: 10| October 2025|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2025: 8.691 || ISI Value: 1.188

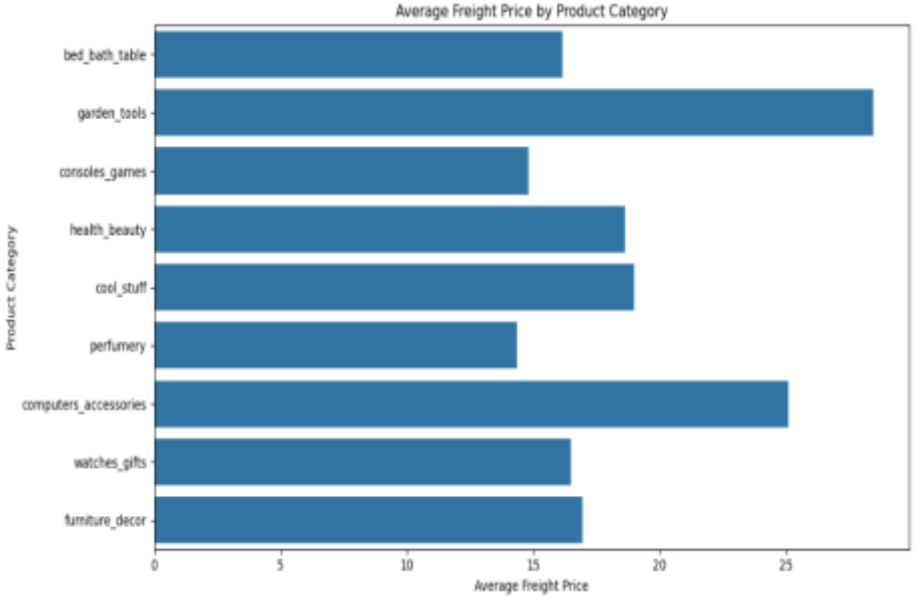
5.4 Correlation Heatmap of Numeric Features

The heatmap visualizes relationships among key numeric variables such as quantity, price, freight, and product weight.



5.5 Bar Plot of Average Freight Price by Product Category

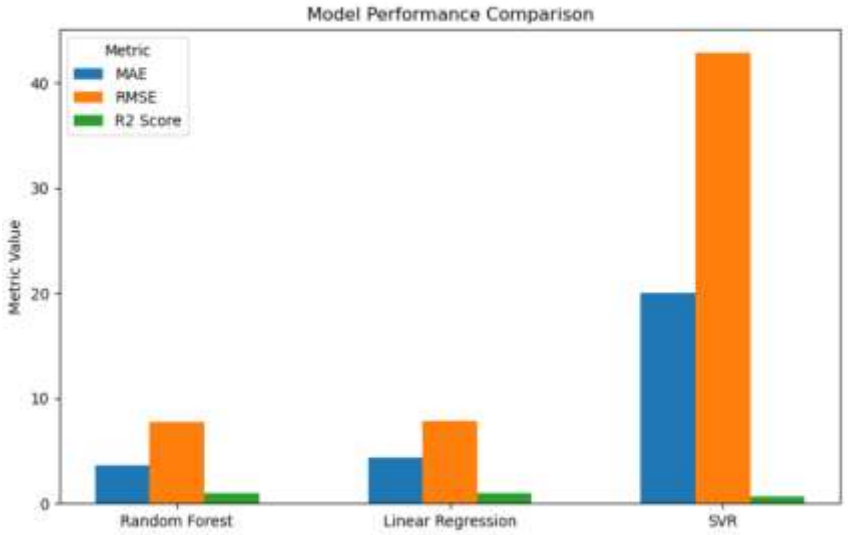
This bar chart shows how average freight costs vary across product categories.



ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 11| Issue: 10| October 2025|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2025: 8.691 || ISI Value: 1.188

5.6Model Performance Comparison

This bar graph compares the performance of Linear Regression, Random Forest, and SVR models using MAE, RMSE, and R^2 metrics.



Insights and Interpretation

The visual analysis of the graphs provides several key insights. The distribution of unit prices and monthly price variations shows that most products are low to mid-priced, with occasional high-priced outliers, indicating price-sensitive consumer behavior and the influence of seasonal promotions on pricing. Total sales over time reveal clear seasonal peaks, suggesting that demand fluctuates predictably during festive or promotional periods. Correlation analysis highlights that sales quantity is strongly linked to total price, while freight costs increase with product weight, emphasizing the impact of logistics on overall profitability. Freight price comparisons across product categories further confirm that heavier or bulkier items incur higher shipping costs, affecting final retail prices. Finally, the model performance comparison indicates that the Random Forest model provides the highest accuracy, outperforming Linear Regression and SVR in terms of MAE, RMSE, and R^2, making it the most reliable method for predicting sales or pricing trends in this dataset. Collectively, these graphs illustrate how pricing, seasonal demand, product characteristics, shipping costs, and advanced modeling techniques interact to shape sales outcomes and business strategy.

RESULTS AND DISCUSSION

The comparison between models revealed that Random Forest achieved the highest predictive accuracy, outperforming both Linear Regression and SVR. Random Forest achieved MAE = 3.66, RMSE = 7.72, and $R^2 = 0.9897$. Linear Regression achieved similar but slightly lower accuracy, while SVR struggled with non-linear data, resulting in high error metrics.

Table 1: Model Accuracy Metrics

Model	MAE	RMSE	R^2 Score
Linear Regression	4.39	7.86	0.9893
Random Forest	3.66	7.72	0.9897
Support Vector Regression (SVR)	20.05	42.91	0.6835

The Random Forest model's ensemble approach enables it to handle diverse data relationships effectively. Visualizations such as correlation heatmaps confirmed strong links between sales quantity, total price, and freight cost, showing how logistics and seasonal demand affect profitability.

FUTURE SCOPE

Future research could explore reinforcement learning to automate price adjustment strategies based on market feedback. Additionally, integrating unstructured data like social media sentiment and customer reviews using natural language processing (NLP) could improve model interpretability. Real-time pricing systems, powered by deep learning, may further enable adaptive, autonomous retail operations.

CONCLUSION

This research establishes that machine learning, particularly Random Forest, provides a robust approach to price optimization

ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 11| Issue: 10| October 2025|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2025: 8.691 || ISI Value: 1.188

in retail. By leveraging predictive analytics, retailers can move from static, assumption-based strategies to adaptive, data-driven frameworks. The outcomes validate that integrating AI-powered systems enhances profit margins, customer satisfaction, and market competitiveness.

ACKNOWLEDGMENT
The author thanks Dr. D.Y. Patil College for guidance and support.

CONFLICT OF INTEREST
The author declares no conflict of interest.

REFERENCES

1. *Das, P., Pervin, T., Bhattacharjee, B., et al. (2024). Optimizing Real-Time Dynamic Pricing Strategies in Retail and E-Commerce Using Machine Learning Models. The American Journal of Engineering and Technology, 06(12), 163-177.*
2. *Jana, A. K. (2021). Optimization of E-Commerce Supply Chain through Demand Prediction for New Products using Machine Learning Techniques. Journal of Artificial Intelligence, Machine Learning and Data Science, 1(1), 565-569.*
3. *Perumallapalli, R. K. (2014). Machine Learning Algorithms for Dynamic Pricing Optimization in Retail. International Journal of Research and Analytical Reviews (IJRAR), 1(4), 642-649.*
4. *Rahman, M. A., Modak, C., Mozumder, M. A. S., et al. (2024). Advancements in Retail Price Optimization: Leveraging Machine Learning Models for Profitability and Competitiveness. Journal of Business and Management Studies, 6(3), 103-110.*
5. *Sarkar, M., Ayon, E. H., Mia, M. T., et al. (2023). Optimizing E-Commerce Profits: A Comprehensive Machine Learning Framework for Dynamic Pricing and Predicting Online Purchases. Journal of Computer Science and Technology Studies, 5(4), 186-193.*