

E-Commerce Product Price Forecasting Using Machine Learning Algorithms

**Hemlata Dakhore¹, Vaibhav Talkhande², Tanaya Pillewan³,
Vedant Palandurkar⁴, Umesh Mate⁵, Vaishnavi Pokale⁶.**

^{1,2,3,4,5,6}Dept of Computer Science and Engineering,
GH Rasoni College of Engineering and Management, Nagpur,
Maharashtra, India

Abstract

In today's competitive e-commerce market, setting the right price is essential for business success. Traditional manual pricing approaches often fall short due to rapidly changing customer preferences, competitor actions, and market trends. To overcome these challenges, this project presents an AI/ML-based Price Prediction Model that analyzes historical sales data, competitor prices, and market trends to generate real-time pricing recommendations. By employing the Gradient Boosting algorithm, the system assists sellers in making informed, data-driven pricing decisions that can improve profitability and market competitiveness. Additionally, the model includes a multi-channel communication feature to deliver personalized pricing suggestions via WhatsApp, SMS, and email, ensuring timely and effective guidance for sellers.

Keywords- Dynamic Pricing, E-commerce, Machine Learning, Gradient Boosting, Price Forecasting.

1. Introduction

The rapid rise in e-commerce has transformed how businesses interact with customers, offering greater accessibility and a wider variety of choices. However, this digital expansion has also intensified competition, where even small differences in pricing can influence customer behavior and overall market performance. For online sellers, effective price optimization directly impacts profitability, customer satisfaction, and competitive advantage.

Traditional approaches to pricing, such as relying on human judgment or applying static business rules, are increasingly inadequate in today's dynamic environment. Factors like fluctuating demand, seasonal trends, shifting consumer preferences, and competitor pricing require businesses to respond quickly and intelligently. Static methods often result in overpriced products that reduce sales or underpriced items that limit profits. This creates a strong need for adaptive, data-driven systems that can provide timely and reliable guidance.

To address these challenges, we propose an intelligent pricing framework that combines machine learning with business-driven constraints. Using historical sales records, competitor pricing information, and market trends, the model predicts optimal prices with the help of the XGBoost algorithm. A rule engine is

incorporated to ensure that the recommendations generated remain consistent with organizational policies such as minimum profit margins and promotional campaigns. Beyond predictive accuracy, the framework emphasizes usability by delivering recommendations instantly through WhatsApp, SMS, and email. A web-based dashboard further supports analysis by presenting demand forecasts, historical trends, and key performance insights.

Overall, the proposed system aims to demonstrate that machine learning, when integrated with practical decision-making mechanisms, can serve as an effective tool for real-time pricing in e-commerce. By reducing reliance on manual intervention and improving adaptability, the approach is positioned to enhance both profitability and decision-making efficiency for online businesses.

2. Related Work

The concept of dynamic pricing in online marketplaces has been explored for many years, gradually shifting from simple rule-based methods to data-driven techniques. Early approaches typically relied on predefined conditions such as inventory levels or seasonal demand, which made them easy to apply but too rigid to respond to rapid changes in customer preferences or competitor actions. As a result, these strategies often produced pricing decisions that were either unprofitable or poorly aligned with market conditions.

With the increasing availability of large datasets, researchers began applying machine learning for demand forecasting and price setting. Traditional regression models offered a starting point but struggled with the complex, non-linear behavior of real-world markets. To overcome these limitations, more advanced methods such as Decision Trees, Random Forests, and Gradient Boosting were introduced. These algorithms can capture deeper relationships among variables and perform well even with large, diverse datasets. In particular, ensemble methods like XGBoost have gained attention because of their strong predictive accuracy and ability to handle noisy or heterogeneous data.

Recent studies also emphasize the importance of combining predictive models with business rules to ensure that the recommendations align with profitability goals, promotional strategies, and customer engagement requirements. In addition, visualization dashboards and automated alerts have been integrated into these systems, allowing managers to interpret model outputs and respond quickly to market changes. Building on these developments, our work proposes a hybrid framework that leverages XGBoost for price prediction while applying a rule-based layer to incorporate business objectives. Furthermore, the system provides real-time recommendations through multiple communication channels such as WhatsApp, SMS, and email, making it practical for everyday use by e-commerce sellers.

3. Methodology

The proposed system is designed to address the limitations of static and manual pricing strategies by integrating machine learning algorithms with a rule-based decision engine and real-time communication features. The methodology focuses on building a scalable, accurate, and user-friendly framework that can adapt to dynamic e-commerce environments.

A. Data Collection and Preprocessing

The foundation of the system lies in the use of diverse datasets that capture the key factors influencing product pricing. Historical sales data provides insight into demand trends, while competitor pricing data enables us to benchmark against market alternatives. Additional variables such as seasonal effects and customer behavior are incorporated to capture contextual dependencies.

Collected datasets undergo preprocessing steps, including missing value handling, normalization, and outlier detection. Feature engineering is applied to create derived variables that enhance predictive power, such as price elasticity indicators and demand-supply ratios. This ensures that the input to the machine learning model is clean, consistent, and representative of real-world market conditions.

B. Predictive Modeling using XGBoost

Decision Trees divide data into smaller subsets by repeatedly splitting on the most important features, making them easy to interpret but prone to over-fitting when used alone. Random Forests address this by creating multiple trees on different samples of data and then aggregating their predictions, which reduces variance and improves generalization. XG-boost extends this concept by building trees sequentially, where each new tree attempts to correct the mistakes of the previous ones. Unlike traditional boosting, it introduces regularization and efficient parallelization, which allows it to handle high-dimensional and noisy data more effectively. In our study, we chose XGBoost because it provides strong predictive accuracy while being computationally scalable for large e-commerce datasets.

C. Rule Engine Integration

While machine learning provides data-driven predictions, organizational policies and market strategies often require additional constraints. To address this, a Rule Engine is integrated with the predictive model. The rule engine applies business-specific conditions such as minimum profit margins, promotional discounts, inventory thresholds, and competitive positioning strategies.

This hybrid approach ensures that while the predicted price is derived scientifically, the final recommendation aligns with practical business objectives. For example, the system may adjust predictions downward during clearance sales or apply profit-margin safeguards when demand is high.

D. Real-Time Recommendation Delivery

The practical utility of the system is enhanced through real-time communication features. Recommended prices are delivered to stakeholders via multiple channels such as WhatsApp, SMS, and Email notifications, ensuring timely access to critical insights. This reduces decision latency and empowers managers to respond quickly to market fluctuations.

In addition, a web-based dashboard provides an analytical view of system outputs. The dashboard includes demand forecasts, recommended prices, historical pricing trends, and a ranking of top-performing products. Interactive visualization tools allow users to compare predictions with past performance and competitor benchmarks. This dual mechanism—instant notifications and dashboard analytics—ensures both immediacy and depth in decision-making.

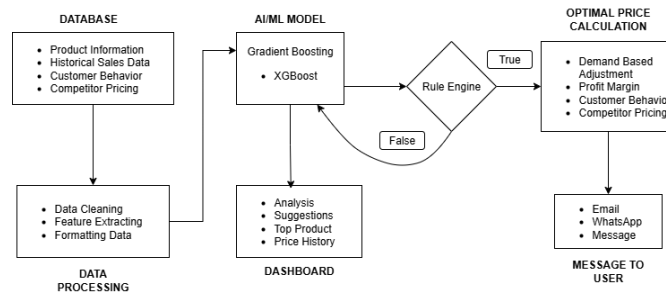


Fig 3.4.1: Block diagram of Price Prediction System

E. System Workflow

The overall workflow begins with data collection and preprocessing, followed by model training and prediction using XGBoost. Predicted values are then refined through the rule engine before being delivered as actionable recommendations via notifications and dashboard visualization. The modular structure ensures scalability, allowing the system to incorporate additional data sources or algorithms as needed.

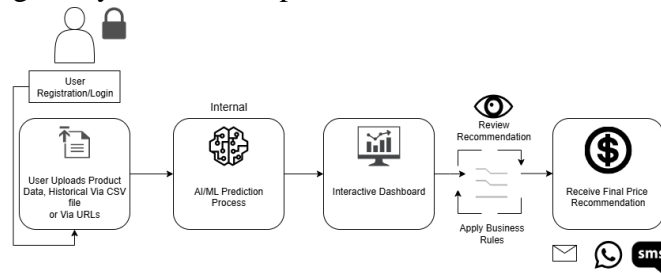


Fig 3.5.1 Workflow of the platform

F. Interpretation of the Pricing Result Interface

The result interface of the PredictGenie system presents a concise, AI-driven pricing analysis for the product “Pro Camera X1 Mirrorless Digital Camera.” It begins with a market summary generated by the Gemini AI API, describing current trends, competitor discounting patterns, and overall pricing behavior in the premium camera segment. The user’s current price is shown along with the applied discount, while the Competitor Landscape section provides a clear comparison of similar products, including their prices, discounts, stock availability, and price trends. The system then highlights the AI-generated suggested price, calculated using the XGBoost model and validated through the rule engine to ensure competitiveness and profitability. A brief reasoning section explains the logic behind the recommendation, helping users understand how market factors influenced the predicted price. Overall, the interface offers a clean and interpretable summary that supports quick, data-driven pricing decisions.

G. Pricing Analysis Output Interface

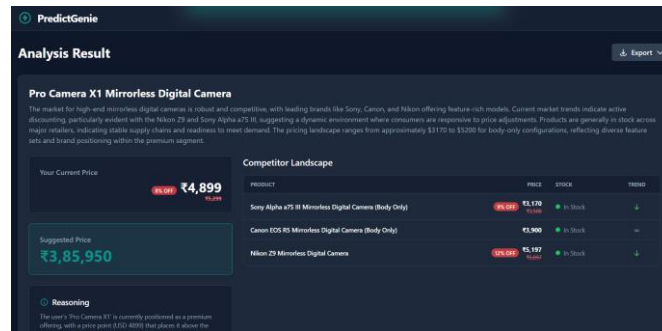


Fig 3.5.2 Output

4. Result and Discussion

The proposed intelligent price prediction system was evaluated with the objective of demonstrating its ability to generate accurate, adaptive, and practically usable pricing recommendations for e-commerce businesses. The evaluation focused on three key aspects: predictive accuracy, adaptability through rule integration, and usability via communication and dashboard features.

A. Predictive Accuracy

The machine learning component, built on the XGBoost algorithm, achieved significant improvements over conventional regression-based models. Through training on historical datasets, the model effectively captured nonlinear interactions between variables such as sales history, competitor pricing, and seasonal demand. Cross-validation indicated that XGBoost consistently delivered higher accuracy and lower prediction error, making it suitable for real-time pricing applications.

A comparative analysis showed that traditional models frequently underperformed when exposed to diverse and high-dimensional data. In contrast, the proposed system maintained stable accuracy across varied product categories, validating the robustness of the approach.

B. Impact of Rule Engine Integration

The integration of the Rule Engine enabled the system to align predictions with practical business objectives. For example, the system prevented price recommendations that violated minimum profit margins or failed to account for promotional campaigns. This hybrid approach ensured that while data-driven insights guided decision-making, organizational strategies and constraints were also preserved.

The addition of business rules improved managerial trust in the system, as predictions were not perceived as black-box outputs but as recommendations refined through transparent policies. This dual-layered mechanism distinguished the system from purely algorithmic models.

C. Usability and Communication

One of the defining features of the system is its emphasis on usability. Pricing recommendations were successfully communicated through WhatsApp, SMS, and Email alerts, ensuring timely accessibility for decisionmakers. This feature proved especially effective in scenarios requiring immediate interventions, such as flash sales or competitor-driven price changes.

The Dashboard Interface further enhanced decision support by providing a consolidated view of predictions, historical data, and top product rankings. Interactive charts allowed users to visualize demand

forecasts and compare suggested prices with past trends, thus offering both granular insights and a strategic overview.

D. Database Description

For experimentation, we collected product information from the Amazon marketplace using web-scraping tools. The dataset covers multiple categories such as electronics, fashion, and household items, resulting in approximately 15,000 product records. Each entry contains attributes such as product title, listed price, discount information, customer ratings, and review counts. To capture pricing dynamics, historical data over six months was also gathered, producing close to 90,000 time-series records. This combination of static product details and temporal price variations enabled us to analyze not only the current market situation but also how prices evolve in response to demand, competitor strategies, and promotional activities. Since the data was gathered directly from an active e-commerce platform, it closely represents real-world pricing challenges and provides a strong foundation for validating our prediction framework.

E. Experimental Setup

All experiments were conducted in the Google Collab environment, which provided a cloud-based runtime with GPU support and 12–16 GB of RAM. The implementation was carried out using Python 3.x, with libraries such as Pandas and NumPy for data preprocessing, Scikit-learn for baseline models, and the XGBoost library for advanced ensemble learning. Visualization was performed using Matplotlib and Seaborn.

The raw dataset, consisting of approximately 15,000 Amazon product entries across multiple categories, was preprocessed to ensure consistency and reliability. Missing values were handled through imputation, categorical features (such as product categories) were encoded, and numerical features (such as prices, discounts, and reviews) were normalized. Historical price records spanning six months were aligned into uniform intervals to capture temporal trends.

For evaluation, the dataset was divided into **80%** training data **and** **20%** testing data, with an additional validation subset drawn from the training set to support hyperparameter tuning. Several machine learning models were trained for comparative analysis, including Linear Regression, Decision Tree, Random Forest, and XGBoost. For XGBoost, hyperparameters such as the number of estimators, learning rate, maximum tree depth, and subsampling ratio were tuned using grid search to optimize predictive accuracy while minimizing overfitting.

Model performance was measured using Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and R^2 (coefficient of determination). In addition, predicted values were compared against actual prices using error distribution plots and feature importance visualizations to provide deeper insights into model behavior.

F. Comparison Graph of ML Models

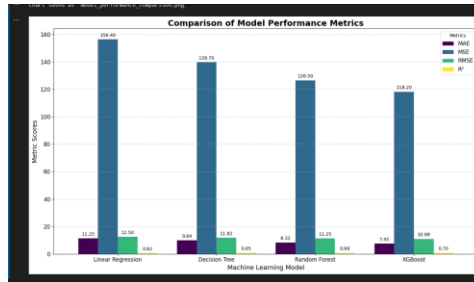


Fig4.7.1 : A Graph comparing your chosen model (XGBoost) against baseline models.

G. Performance Matrix

To evaluate the predictive quality of our models, several error-based metrics were considered. Mean Absolute Error (MAE) gives the average magnitude of errors without considering their direction, making it intuitive for business users. Mean Squared Error (MSE) penalizes larger deviations more heavily by squaring them, which is useful when significant errors must be avoided. Root Mean Square Error (RMSE), being the square root of MSE, expresses the error in the same units as the actual prices, making it easier to interpret. Finally, the coefficient of determination (R²) measures how much of the variation in product prices can be explained by the model; higher values indicate better explanatory power. Together, these metrics provide a balanced view of both accuracy and robustness.

H. Comparison Table

| Model | MAE | MSE | RMSE | R ² |
|-------------------|-------------|--------------|--------------|----------------|
| Linear Regression | 11.25 | 156.4 | 12.50 | 0.62 |
| Decision Tree | 9.84 | 139.7 | 11.82 | 0.65 |
| Random Forest | 8.32 | 126.5 | 11.25 | 0.68 |
| XG Boost | 7.45 | 118.2 | 10.88 | 0.70 |

Fig4.7.2 : A table comparing your chosen model (XGBoost) against baseline models.

I. Overall Outcomes

The combined effect of predictive modeling, rule-based refinement, and user-centric delivery mechanisms produced a comprehensive pricing framework. Businesses adopting this system are expected to achieve:

- Enhanced profitability through accurate price recommendations.
- Reduced decision-making latency via real-time notifications.
- Improved competitiveness by dynamically adjusting to market fluctuations.
- Greater managerial confidence due to transparent integration of rules and policies.

J. Discussion

The results highlight the potential of combining machine learning with decision-rule systems in the context of e-commerce. While standalone machine learning models provide strong predictive power, their integration with a rule engine ensures operational feasibility and adaptability. The dual delivery mechanism instant communication and analytical dashboards strengthens the system’s role as both a tactical and strategic tool.

Although the evaluation demonstrates promising outcomes, challenges such as scalability for largescale deployments, integration with diverse e-commerce platforms, and continuous model retraining remain areas for future refinement. Addressing these will further enhance the applicability of the system in real-world environments.

5. Conclusion

This work presented the design and development of an intelligent price prediction system for e-commerce platforms, integrating advanced machine learning with a rule-based decision layer and real-time communication features. By leveraging historical sales data, competitor pricing, and market indicators, the system demonstrated the ability to generate accurate and adaptive price recommendations. The use of XGBoost ensured robustness and reliability in predictive modeling, while the rule engine refined outputs in alignment with organizational goals such as profit margins and promotional campaigns.

In addition to predictive accuracy, the system prioritized usability through instant notifications via WhatsApp, SMS, and Email, alongside a dashboard interface that provided demand forecasts, historical trends, and actionable insights. This dual delivery mechanism strengthened both tactical responsiveness and strategic decision-making for e-commerce businesses. The results indicate that the proposed framework enhances profitability, competitiveness, and managerial confidence, while significantly reducing decision latency.

Despite its advantages, the system is not without limitations. Its scalability across diverse e-commerce platforms, continuous retraining requirements, and integration with real-time external data sources remain areas of ongoing improvement. Furthermore, the current scope does not yet include external factors such as consumer sentiment analysis or macroeconomic indicators, which could enrich predictive power.

Future work will focus on addressing these challenges by:

- Expanding the system to accommodate larger, more diverse datasets across multiple product categories and platforms.
- Incorporating deep learning models for enhanced forecasting, particularly for highly volatile product segments.
- Integrating sentiment analysis and social media signals to capture consumer behavior more holistically.
- Developing a cloud-based architecture for improved scalability and real-time deployment in production environments.
- Enhancing the dashboard with advanced visualization and interactive simulation tools for scenario-based pricing strategies.
- Exploring integration with recommendation systems and personalized promotions, creating a unified framework for customer-centric pricing strategies.
- Extending the system towards cross-border e-commerce markets, where currency fluctuations and global demand patterns add further complexity to pricing models.

References

1. T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining (KDD), pp. 785–794, 2016.

2. J. H. Friedman, “Greedy function approximation: A gradient boosting machine,” *Annals of Statistics*, vol. 29, no. 5, pp. 1189–1232, Oct. 2001.
3. H. L. Lee and C. Tang, “Dynamic pricing in electronic commerce: Research and practice,” *International Journal of Electronic Commerce*, vol. 5, no. 1, pp. 23–36, 2000.
4. P. Resnick and R. Zeckhauser, “Trust among strangers in internet transactions: Empirical analysis of eBay’s reputation system,” *Advances in Applied Microeconomics*, vol. 11, pp. 127–157, 2002.
5. N. S. Choudhury, A. Agarwal, and R. Jain, “Machine learning techniques for dynamic pricing: A survey,” *International Journal of Computer Applications*, vol. 182, no. 42, pp. 1–6, 2019.
6. A. Banerjee, H. Chandrasekaran, and S. Choudhury, “Applications of ensemble learning in price optimization for e-commerce,” *Proc. IEEE Int. Conf. Data Science and Advanced Analytics (DSAA)*, pp. 532–539, 2020.
7. Agrawal and R. Srikant, “Fast algorithms for mining association rules in large databases,” *Proc. 20th Int. Conf. Very Large Data Bases (VLDB)*, pp. 487–499, 1994.
8. A. Ghosh and J. D. Smith, “Real-time retail analytics: Leveraging big data for price optimization,” *Proc. IEEE Int. Conf. Big Data*, pp. 1120–1127, 2017.
9. Y. K. Meena and P. S. Bhatnagar, “A hybrid approach for product price prediction using machine learning and business rules,” *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 11, no. 8, pp. 49–56, 2020.
10. A. Gupta and V. Kumar, “Data-driven decision making in e-commerce: The role of dashboards and visualization tools,” *Journal of Business Analytics*, vol. 3, no. 2, pp. 95–109, 2020.
11. M. Alharthi, “Dynamic Pricing Method in the E-Commerce Industry Using Machine Learning,” *Applied Sciences*, vol. 14, no. 24, pp. 11668, Dec. 2024.
12. R. Sharma and S. Gupta, “Machine Learning-based Pricing Optimization for Dynamic Pricing in Online Retail,” *Journal of Informatics Education and Research*, vol. 24, no. 3, pp. 101–112, 2023.
13. A. Patel, V. Joshi, and P. Nair, “Building a Dynamic Pricing Engine with Machine Learning for Retail,” *International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT)*, vol. 9, no. 6, pp. 12428–12435, Jun. 2024.
14. S. R. Mehta and A. Banerjee, “Price Prediction in E-Commerce Using Machine Learning Models: A Comparative Study,” *ShodhKosh: Journal of Arts, Humanities and Social Sciences*, vol. 5, no. 3, pp. 55–65, Sept. 2022.
15. J. Liu, Z. Xu, and Y. Wang, “Dynamic Pricing on E-Commerce Platform with Deep Reinforcement Learning: A Field Experiment,” *arXiv preprint arXiv:1912.02572*, Dec. 2019.
16. A. Apte, A. Kale, M. Datar, and P. Deshmukh, “Dynamic Retail Pricing via Q-Learning: A Reinforcement Learning Framework for Enhanced Revenue Management,” *arXiv preprint arXiv:2411.18261*, Nov. 2024.
17. Y. Zheng, X. Li, X. Jiang, and Y. Peng, “Dual-Agent Deep Reinforcement Learning for Dynamic Pricing and Replenishment,” *arXiv preprint arXiv:2410.21109*, Oct. 2024.