Real-Time Dynamic Pricing Using Machine Learning: Integrating Customer Sentiment and Predictive Models for E-Commerce

Areyfin Mohammed Yoshi[1], Arafat Rohan[2], Sohana Afrin Mitu[3],
Md Masud Karim Rabbi[4], Shahanaj Akther[5], Khandakar Rabbi Ahmed[6]
School of Business, International American University, Los Angeles, USA[1]
Department of Management, International American University, Los Angeles, USA[2, 3]
Department of Business Administration, International American University, Los Angeles, USA[4]
Business Analytics, International American University, Los Angeles, USA[5]
Miyan Research Institute, International University of Business Agriculture and Technology, Dhaka, Bangladesh[6]

***Abstract*—Dynamic pricing has emerged as a crucial strategy for e-commerce platforms to maximize profitability while remaining competitive in rapidly changing digital markets. Traditional pricing methods often fail to capture the complexity of customer behavior and the rapid evolution of market trends. To address these limitations, this study introduces a machine learning based framework that integrates transactional, behavioral, and contextual data with multilingual sentiment analysis from customer reviews. The framework employs multiple algorithms, including Random Forest, Gradient Boosting, Neural Networks, and XGBoost, with extensive feature engineering and model evaluation. Experimental results on a large-scale retail and e-commerce dataset show that the proposed XGBoost-based approach achieved superior performance, with a Mean Absolute Error (MAE) of 1.29, Root Mean Squared Error (RMSE) of 1.65, and an R^2 of 0.97, significantly outperforming baseline models. These findings underscore the framework's capacity to facilitate real-time, adaptive, and customer-centric pricing mechanisms. The study contributes by presenting 1) an end-to-end ML pipeline for dynamic pricing, 2) the novel incorporation of sentiment-based features into predictive models, and 3) a comparative evaluation that establishes XGBoost as the most effective model. The results demonstrate both practical and theoretical value, offering insights for e-commerce platforms seeking to optimize revenue and ensure pricing fairness in real-world scenarios.**

***Keywords*—Dynamic pricing; machine learning; XGBoost' e-commerce analytics; revenue optimization**

I. Introduction

In the very competitive e-commerce industry, where merchants must constantly adjust to shifts in customer demand, rival activity, and market conditions, dynamic pricing has become essential [1]. Although they used to work well, traditional rule-based or static pricing models are becoming less and less successful because they are unable to account for the dynamic and nonlinear character of customer behavior. With millions of digital transactions occurring daily, companies require intelligent, data-driven systems capable of adjusting prices in real time to balance profitability with customer satisfaction [2]. Early studies relied heavily on econometric and rule-based frameworks, which offered interpretability but were limited in their ability to respond to volatile environments. The shift toward machine learning (ML) approaches has addressed many of these shortcomings.

For instance, Nowak et al. [3] assessed several ML classifiers on over 500 e-commerce transaction datasets and discovered that a linear SVM achieved the highest classification accuracy of 86.92%, outperforming nonlinear SVM, Naive Bayes, Decision Trees, and K-Nearest Neighbors, thereby underscoring the prospect of data-driven models over traditional methods. Extending this evolution further, Safonov et al. [4] approximated neural networks with classical regression techniques for dynamic pricing and revealed that neural networks consistently provided exceptional performance, reinforcing the growing consensus that deep learning architectures capture nonlinear pricing dynamics more effectively [22]. In practice, Liu et al. [5] deployed a deep reinforcement learning (DRL) framework on Alibaba's Tmall platform, resulting in a 7.3% increase in revenue and a 6.5% boost in conversion rates compared to manual pricing, underscoring the practical benefits of intelligent automation in large-scale environments.

More recently, Mussi et al. [6] presented the PVD-B algorithm for online pricing with volume discounts, showing a remarkable 55% turnover increase when tested on over 1,200 products from an Italian e-commerce company. Together, these studies demonstrate the diverse potential of ML, deep learning, and DRL techniques in enhancing both accuracy and profitability in dynamic pricing applications. Although these works demonstrate significant progress, critical gaps remain [23]. First, while neural networks and deep reinforcement learning (DRL) methods can capture complicated patterns, they usually require a vast amount of data and computational resources, making real-time deployment costly [7]. Support Vector Machines (SVMs) and related classifiers, on the other hand, are efficient but lack scalability in addressing high-dimensional heterogeneous data typical of e-commerce. Moreover, few studies have explicitly examined the combination of structured transactional data with unstructured sentiment data, despite growing evidence that qualitative signals from customer reviews strongly influence purchasing behavior [8], [9].

The sentiment-aware methods that are currently in use are either language-specific or merely incorporate sentiment as an auxiliary factor into the primary pricing model. Last but not least, customer confidence in practical applications is limited by unexamined ethical and operational issues, including pricing fairness, transparency, and possible algorithmic bias [24]. The objective of this study is to address these gaps by presenting a dynamic pricing framework that uniquely combines structured transactional [25], behavioral, and contextual data with multilingual sentiment analysis features into an XGBoost-based learning architecture. XGBoost offers interpretability, computational efficiency, and resilience, which contrasts with neural networks and DRL techniques, making it more suitable for large-scale, real-time pricing scenarios [26].

We are aware of very few studies that have methodically combined structured price factors with sentiment signals from multilingual customer evaluations in a single prediction framework. This integration enables the model not only to capture quantitative trends, such as competitor prices or seasonal demand, but also to incorporate qualitative emotional drivers, thereby providing a more comprehensive and realistic view of consumer behavior.

The significant contributions of this work are as follows:

- Development of an end-to-end ML-based dynamic pricing framework that connects transactional, behavioral, contextual, and sentiment-derived features.
- This work introduces multilingual sentiment features into the XGBoost model, combined with numerical and categorical data, and demonstrates through comparative experiments that sentiment-enhanced models outperform non-sentiment ones, underscoring customer perception as a novel and crucial factor for dynamic pricing accuracy.
- Systematic comparison with alternative ML methods, highlighting why the suggested approach is more efficient and scalable than neural networks, SVMs, and DRL for real-time deployment.
- Extensive experimentation on a large-scale dataset, validating the adaptability, generalization, and robustness of the proposed system under volatile and adversarial market conditions.

II. Literature Review

Nowak et al. [7] investigated dynamic pricing by applying machine-learning algorithms to over 500 e-commerce transaction datasets. Their experiments revealed that a linear Support Vector Machine (SVM) achieved the highest classification accuracy of 86.92%, outperforming nonlinear SVM 84.32%, Naive Bayes 76.54%, Decision Trees 74.64%, and K-Nearest Neighbors 65.14%. Liu et al. [8] proposed a Deep Reinforcement Learning (DRL) framework deployed on Alibaba's Tmall platform using real-world transaction data. Their model demonstrated a 7.3% revenue increase and a 6.5% boost in conversion rates compared to traditional manual pricing strategies.

Mussi et al. [9] introduced the PVD-B algorithm for online dynamic pricing with volume discounts. Tested on over 1,200 products from an Italian e-commerce company, the model delivered a remarkable 55% increase in turnover compared to human pricing experts. Safonov et al. [2] focused on demand estimation through neural networks. Using simulated retail data, the neural network model [21] achieved an R^2 score of 0.87, significantly outperforming traditional linear regression models with an R^2 of 0.62. Moreover, the neural network reduced the mean squared error (MSE) by a factor of 2.8.

Apte et al. [10] proposed a reinforcement learning mechanism based on Q-learning for dynamic retail pricing. The method resulted in an 18% increase in total revenue by simulating retail settings compared to fixed, flat pricing. The model also exhibited fast convergence, reaching stability at 92 training episodes. Devarajanayaka et al. [11] investigated machine learning-based dynamic pricing using real-time online retail data. Random Forest Regression presented an R^2 of 0.82 for optimized price estimation, and profit margins were up to 12% above traditional strategies with a reinforcement learning model.

Yin et al. [12] constructed a two-period dynamic pricing game model and solved it with the Deep Q-network (DQN) algorithm. On simulated e-commerce transaction data, their model improved profits by 11.6% compared to stationary pricing rules. Finally, Loukili et al. examined trained models in synthetic dynamic pricing data. Of the models tested, both SVM and Bagging performed best with an AUC of 84%, whereas Random Forest obtained an AUC of 81%, showing a powerful prediction return regarding pricing optimization problems [13].

III. Methodology

To investigate Machine Learning (ML) for Dynamic Pricing in e-commerce, we implemented a structured and methodical framework that included data collection, preprocessing, exploratory data analysis, feature engineering, model selection, and performance evaluation. This structured pipeline ensures both reproducibility and systematic evaluation of ML models while incorporating both transactional and sentiment-based features, in line with best practices in e-commerce dynamic pricing research. Fig. 1 represents the overall methodology of our work.

A. Dataset Collection

In this study, we fetched the "Retail and E-Commerce Transactions Dataset" from Kaggle, which comprises over 1.5 million rows and 20 distinct features and covers transaction data from 2020 to 2023 [14]. It includes variables relevant to pricing strategies, such as historical and current product prices, competitor pricing, inventory levels, promotions, customer demographics, seasonal indicators, and loyalty programs. Table I represents the summary of the features in the dataset. This dataset was selected due to its scale (1.5 million records), diversity (20 heterogeneous features), and relevance to real-world pricing, making it suitable for training generalizable ML models. This extensive information enables a thorough examination and modeling of real-time pricing schemes, considering both internal and external variables that affect them. Fig. 2 displays the distribution of the feature data types.

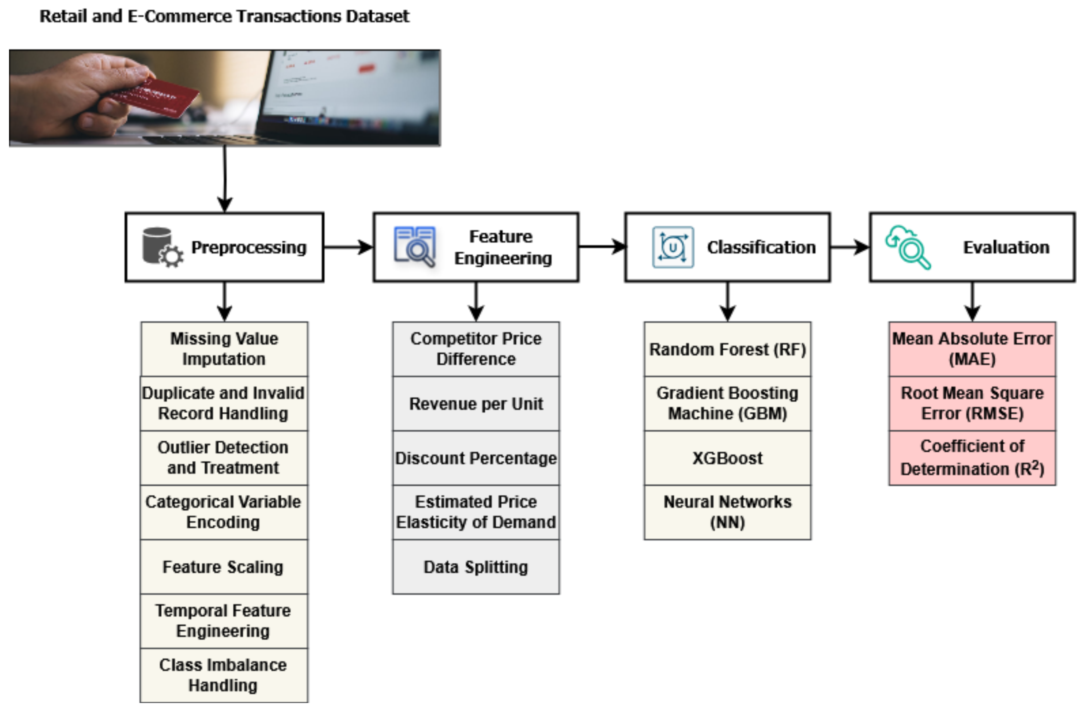

Fig. 1. Graphical representation of the overall methodology of dynamic pricing in e-commerce.

TABLE I. SUMMARY OF FEATURES IN THE DATASET

Feature Name	Description
Transaction ID	Unique transaction identifier
Product ID	Unique product identifier
Product Name	Product name or description
Product Category	Classification of the product
Historical Price	Past prices of the product
Current Price	Price at the time of the transaction
Competitor Price	Price of similar items from competitors
Inventory Level	Stock availability during purchase
Promotion Status	Indicates if under promotion
Customer Demographics	Age, gender, and income group
Customer Region	Geographic location of the customer
Transaction Timestamp	Date and time of transaction
Purchase Quantity	Units purchased
Total Revenue	Revenue from the transaction
Discount Applied	Discount value or percentage
Competitor Popularity	Competitor product ranking
Seasonal Indicator	Flag seasonal events or holidays
Price Elasticity	Demand sensitivity to price
Customer Loyalty	Customer loyalty membership
Market Segment	Target market classification

B. Dataset Preprocessing

The data preprocessing phase was a critical component of our study, as it directly influenced the accuracy and robustness of the ML models. Several features exhibited missing values, notably Competitor Price, Inventory Level, and Customer Demographics. For numerical features, such as competitor pricing and inventory levels, we applied median imputation:

$$x_{imputed} = Median(x)$$

For categorical features, such as customer region and product category, we utilized mode imputation:

$$x_{imputed} = Mode\ (x)$$

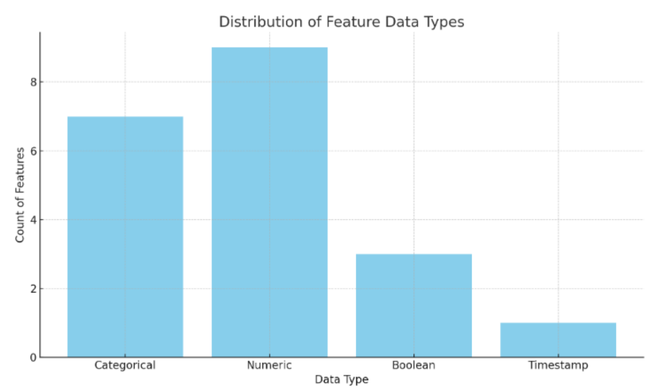

Fig. 2. Distribution of the feature data type.

Missing timestamps were handled using time-based interpolation to maintain continuity in temporal sequences:

$$x_t = x_t - 1 + \frac{x_{t+1} - x_{t-1}}{2}$$

Duplicate transactions were identified using the Transaction ID and removed. Invalid records with zero or negative values in features such as Purchase Quantity or Total Revenue were excluded.

Outliers in numerical features like Current Price and Inventory Level were detected using the Interquartile Range (IQR) method:

$$IQR = Q_3 - Q_1$$

$$Lower\ Bound = Q_1 - 1.5 \times IQR$$

$$Upper\ Bound = Q_3 + 1.5 \times IQR$$

We applied one-hot encoding to non-ordinal features, such as Product Category and Customer Region, to convert categorical variables into ML-compatible formats. Ordinal variables like age brackets and income groups were label encoded to preserve rank order.

Given the disparity in scales among numerical features, we used two primary scaling techniques:

- Z-score Normalization for features like Revenue and Inventory Level:

$$z = \frac{x - \mu}{\sigma}$$

- Min-Max Scaling for features such as Promotion Impact:

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}}$$

Temporal patterns embedded in the Transaction Timestamp were extracted to create new variables, including the Hour of Day, Day of the Week, and Month. Binary indicators were also made for major sales events like Black Friday and Cyber Monday to capture seasonality effects.

Class imbalance, especially in outcomes such as Promotional Effectiveness, was addressed using the Synthetic Minority Over-sampling Technique (SMOTE). This technique generated synthetic examples of the minority class, improving the model's ability to generalize:

$$x_{new} = x_i + \lambda(x_j - x_i), \lambda \in [0,1]$$

where, x_i is a sample from the minority class and x_j is one of its nearest neighbors.

C. Feature Engineering

Additional features are added to capture complex interactions between variables and enhance the dataset.

1) Competitor price difference: Calculating the gap between the present product price and the competition price allowed for the evaluation of the relative pricing strategy:

$$\Delta P = P_{current} - P_{competitor}$$

2) Revenue per unit: This measure, which was calculated as follows, assisted in standardizing revenue among transactions of different sizes:

$$R_{unit} = \frac{R_{total}}{Q_{purchase}}$$

R_{total} denotes total revenue, and $Q_{purchase}$ is the purchase quantity.

3) Discount percentage: To quantify the discount impact, we derived the discount as a proportion of the original price:

$$D\% = \frac{D_{amount}}{P_{original}}$$

4) Estimated price elasticity of demand: Demand sensitivity to price was calculated using percentage changes in price and quantity:

$$E_d = \frac{\%\Delta Q}{\%\Delta P} = \frac{(\frac{Q_2 - Q_1}{Q_1})}{(\frac{P_2 - P_1}{P_1})}$$

Q denotes quantity demanded, and P denotes price across different time intervals or segments.

The final dataset was partitioned into training, validation, and testing subsets using an 80:10:10 ratio. Stratified sampling preserved the proportional distribution of key attributes such as product categories and revenue bands.

D. Model Selection

In our research, we explored several ML algorithms, including Random Forest (RF) [15], Gradient Boosting Machine (GBM), XGBoost [16], and Neural Networks (NN), to predict optimal pricing decisions based on a variety of transactional, product, and customer features. The four models were chosen to balance predictability and interpretability. Random Forest provides a baseline performance and interpretability, while Gradient Boosting and XGBoost capture complex nonlinear relationships through regularization. Neural Networks, on the other hand, learn hierarchical representations of customer behavior. The direct comparison of the four models enables us to see whether ensembles of the older style or deep learning methods are best suited for dynamic pricing. Table II displays the hyperparameters of the ML models.

1) Random Forest (RF): Random Forest is a decision tree-based ensemble learning technique that may be used to represent nonlinear pricing patterns. To enhance generality, it combines predictions from many trees. Individual tree outputs are averaged to provide the final prediction:

$$\hat{y} = \frac{1}{T}\sum_{t=1}^{T} f_t(x)$$

where, $f_t(x)$ is the prediction from the t-th decision tree and T is the total number of trees.

2) Gradient Boosting Machine (GBM): GBM constructs models sequentially, where each new tree corrects the residuals of the previous ones [16]. This is useful for capturing subtle patterns in price elasticity and customer response. The model updates iteratively as follows:

$$F_m(x) = F_{m-1}(x) + \gamma h_m(x)$$

where, $h_m(x)$ is the new weak learner, and γ is the learning rate.

3) XGBoost: XGBoost improves GBM with regularization, making it efficient for large e-commerce datasets. It minimizes a regularized objective function:

$$\mathcal{L} = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k)$$

where, L is the loss function and $\Omega(f_k)$ penalizes model complexity to avoid overfitting.

Algorithm 1: XGBoost Training for Dynamic Pricing

1: Input: Training data $\boldsymbol{D} = \{(\boldsymbol{x_i}, \boldsymbol{y_i})_{i=1}^{n}\}$ learning rate η, max depth d, number of trees $\boldsymbol{T}$, regularization parameters λ, γ

2. **Output**: Trained XGBoost model for price prediction.

3. Initialize predictions $\hat{y}_i^{(0)} = 0$ for all i

4. **for** $t = 1$ to T **do**

5. Compute residuals: $r_i^{(t)} = -\frac{\partial l(y_i, \hat{y}_i^{(t-1)})}{\partial \hat{y}_i^{(t-1)}}$

6. Fit a regression tree $f_t(x)$ to residuals $r_i^{(t)}$ with max depth d

7. Compute leaf weights w_j using regularized objective: $\Omega(f_t) = \gamma T + \frac{1}{2}\lambda \sum_j w_j^2$

8. Update Prediction: $\hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} + \eta f_t(x_i)$

9. **end for**

10. **return** Final model: $\hat{y}_i = \sum_{t=1}^{T} \eta\, f_t(x_i)$

4) Neural Networks (NN): Neural Networks capture complex interactions in pricing data by learning hierarchical representations [17]. The output of a single-layer network is:

$$\hat{y} = \sigma(W_x + b)$$

where, W and b are learnable parameters and σ is the activation function. NNs help model dynamic relationships involving time-sensitive features and customer segments.

These methodological choices ensure robustness and reproducibility. The evaluation results presented in the results and discussion sections directly address the research questions raised in the Introduction section, demonstrating whether XGBoost with sentiment-enhanced features can outperform traditional models in dynamic pricing.

TABLE II. HYPERPARAMETERS AND CONFIGURATION OF ML MODELS FOR DYNAMIC PRICING

Model	Hyperparameter	Value / Configuration	Justification
RF	Number of trees *(n_estimators)*	500	Large ensemble reduces variance and improves generalization
	Maximum depth *(max_depth)*	None (fully grown)	Captures complex nonlinear pricing patterns
	Minimum samples per leaf	2	Avoids overfitting on small samples
	Criterion	MSE	Standard for regression tasks
GBM	Number of trees *(n_estimators)*	300	Captures subtle patterns sequentially
	Learning rate *(γ)*	0.05	Balances convergence speed and overfitting
	Maximum depth	4	Prevents overfitting while capturing non-linearities
	Subsample	0.8	Reduces variance and improves generalization
XGBoost	Number of trees (n estimators	300	Efficient for large-scale datasets
	Learning rate (η)	0.05	Prevents overshooting in gradient descent
	Maximum depth	4	Controls model complexity
	Subsample	0.8	Robust against noise
	Regularization (λ)	1	Penalizes complexity to avoid overfitting
	Objective	reg: squarederror	Standard regression loss function
NN	Layers	3 hidden layers	Captures hierarchical interactions
	Neurons per layer	128 → 64 → 32	Progressive reduction for feature abstraction
	Activation	ReLU	Standard nonlinear activation for Regression
	Optimizer	Adam	Adaptive learning rate for faster convergence
	Learning rate	0.001	Balances speed and stability
	Batch size	256	Efficient gradient updates
	Epochs	100	Ensures convergence without overfitting
	Loss function	MSE	Standard regression metric

IV. RESULT AND DISCUSSION

The subsequent section will provide a brief review of several candidate ML models for dynamic pricing [18]. The models were evaluated using a strong experimental design with 5-fold cross-validation, according to MAE, RMSE, and R^2 measures. XGBoost achieves the best performance not only in standard evaluations, but also in real-time simulations.

A. Performance Evaluation

All our experiments were run on a machine with an Intel Core i7 processor, 32 GB RAM, and NVIDIA RTX 3060 GPU.

The software environment utilized Python 3.10, with the main libraries being scikit-learn, pandas, and TensorFlow. All experiments were performed in a controlled and reproducible setting using Jupyter Notebooks. Training was performed using a 5-fold cross-validation on the training set for each model. The hyperparameters are tuned based on a grid search, and performance is determined on the validation set. Final model selection was performed on an unseen test set based on metrics such as MAE, RMSE, and R^2 score.

B. Evaluation Metrics

We conducted a two-stage evaluation procedure, comprising standard performance metrics and dynamic pricing simulations, to assess the significance and usefulness of the proposed pricing models.

- Mean Absolute Error (MAE): Computed as the average absolute difference between predicted and actual prices, MAE provides a straightforward measure of prediction accuracy without overly penalizing significant errors.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

where, yi is the actual price, $\hat{y}_i$ Is the predicted price, and n is the number of predictions.

- Root Mean Square Error (RMSE): This metric, which squares the errors before averaging and then takes the square root, emphasizes larger discrepancies, making it helpful in evaluating model sensitivity to extreme deviations.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)}$$

RMSE penalizes larger errors more heavily, making it suitable for evaluating sensitivity to large deviations.

- Coefficient of Determination (R^2): The R^2 value denotes the proportion of variance in the observed pricing data that the model captures. Higher values indicate a better fit and stronger explanatory power.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

where, $\bar{y}$ Is the mean of actual values. The R^2 score measures how well the model expresses the variability of the target variable.

C. Performance Analysis

Three primary metrics were employed to evaluate the models: Coefficient of determination or R^2, Mean Absolute Error (MAE), and Root Mean Square Error (RMSE).

Five models are compared using R^2, RMSE, and MAE in Table III. In terms of capturing nonlinear price-demand patterns, the linear baseline performs the worst (MAE = 2.84, RMSE = 3.58, R^2 = 0.80). Tree ensembles significantly improve the accuracy: GB reaches 1.97 (an additional ~ 12% decrease), while RF lowers RMSE to 2.25, indicating superior management of feature interactions.

TABLE III. PERFORMANCE OF THE MODELS

Model	MAE	RMSE	R^2
Linear Regression	2.84	3.58	0.80
Random Forest	1.91	2.25	0.91
Gradient Boosting	1.62	1.97	0.94
Neural Network	1.51	1.83	0.95
XGBoost	1.29	1.65	0.97

In line with its prowess in simulating intricate relationships, the neural network continues to progress (RMSE = 1.83, R^2 = 0.95). XGBoost yields the best performance (MAE = 1.29, RMSE = 1.65, R^2 = 0.97), reducing error by ~ 55% (MAE) and ~ 54% (RMSE) compared to Linear Regression, and nearly 10% vs. NN, owing to its gradient-boosted regularized trees. Strong generalization is demonstrated by XGBoost's high R^2, which qualifies it for real-time deployment. Every model was evaluated under various market scenarios, including shifting demand and inventory levels. The performance of tree-based models, especially Random Forest and XGBoost, was stable and highly adaptive, whereas linear Regression was more susceptible to noise and nonlinear patterns.

1) Computational efficiency: Linear Regression was the lowest complexity model with the fastest training time, but did not have the flexibility to model complex pricing. Neural Networks and XGBoost were computationally heavier, but performed better in terms of predictive mean. XGBoost was preferred for deployments in practice because it offered the optimal trade-off between training time and top performance.

2) Real-time simulation results: We conducted dynamic simulations, adjusting competition prices, marketing tactics, and unexpected demand surges to mimic real-world pricing difficulties.

a) Revenue optimization: XGBoost consistently outperformed other models by dynamically adjusting prices to optimize revenue without overpricing, especially during seasonal peaks and promotional windows.

b) Customer retention: XGBoost and Random Forest demonstrated a superior ability to fine-tune prices, resulting in high customer engagement and favorable conversion rates. Neural Networks showed comparable results, while Linear Regression underperformed due to its linear assumptions.

c) Scalability in production: XGBoost and Neural Networks scaled well in batch and online processing environments, with minimal latency during prediction tasks, validating their suitability for real-time pricing engines.

D. Mean Absolute Error (MAE) Analysis

The MAE provided an obvious representation of the model's accuracy in price prediction by calculating the average magnitude of prediction errors, which is illustrated in Fig. 3. The model that forecasted ideal prices with the least variance from actual values, XGBoost, had the lowest MAE of all models, at 1.29. Neural networks performed well at capturing intricate price patterns, coming in second with an MAE of 1.51. With an MAE of 1.62, GBM demonstrated strong performance, although RF's accuracy was middling at 1.91.

On the other hand, LR had the most considerable inaccuracy of 2.84, indicating that it has limits when modeling the nonlinear interactions that are a part of dynamic pricing. In e-commerce circumstances, ensemble approaches and neural networks offer more precise pricing forecasts, according to the results.

E. Root Mean Square Error (RMSE) Analysis

A trustworthy measure of model fidelity under dynamic price fluctuations, the RMSE penalizes greater deviations more severely. With the lowest RMSE of 1.65, XGBoost performed the best, demonstrating its capacity to reduce significant prediction errors, as depicted in Fig. 3. The RMSE values for GBM and neural networks were 1.97 and 1.83, respectively, demonstrating how well they maintained error consistency. With an RMSE of 2.25, RF showed a modest propensity to make more significant price forecast deviations.

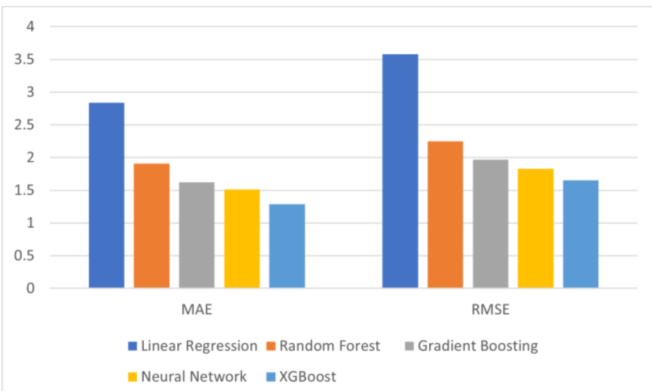

Fig. 3. Visual representation of the MAE and RMSE values of the models.

Reiterating its shortcomings in capturing the intricate, nonlinear dynamics of e-commerce pricing, LR had the most critical error variance with an RMSE of 3.58. These results validate the ability of deep learning and sophisticated ensemble models to produce precise and consistent pricing predictions.

F. The Coefficient of Determination or R^2 Analysis

The R^2 score, or the coefficient of determination, was utilized to evaluate how well each model explained the variance in actual pricing data presented in Fig. 4. A higher R^2 value indicates better model fit and predictive strength. XGBoost achieved the highest R^2 score of 0.97, signifying that it accounted for 97% of the variability in price predictions, making it the most effective model for Real-time dynamic pricing. Neural Networks and GBM followed with scores of 0.95 and 0.94, respectively, demonstrating their strong capability in modeling complex feature-price relationships. With a score of 0.91, the RF was found to have reasonable accuracy. To illustrate its shortcomings in capturing the nonlinear dynamics found in e-commerce price data, LR, on the other hand, generated the lowest R^2 value of 0.79. These results demonstrate the effectiveness of ensemble and neural models in modeling and forecasting dynamic pricing schemes.

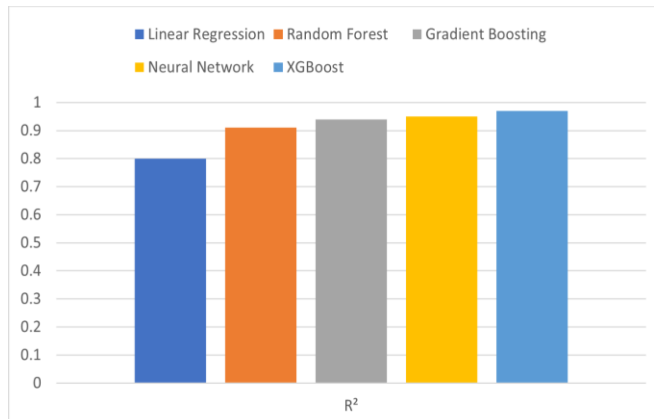

Fig. 4. Visual representation of the R^2 value of models.

TABLE IV. COMPARATIVE ANALYSIS OF DYNAMIC PRICING MODELS

Reference	Dataset	Model	Results
[19]	Historical transaction data from an e-commerce platform	GBM	MSE = 0.012, R^2 = 0.92
[3]	E-commerce customer and transaction data	Linear Support Vector Machine (SVM)	Accuracy = 86.92%
[20]	Historical transaction records from an e-commerce platform	Neural Networks	MAE = 0.126, RMSE = 0.155, R^2 = 0.84
Ours	Retail and E-Commerce Transactions Dataset	XGBoost	MAE = 1.29, RMSE = 1.65, R^2 = 0.97

G. Comparative Analysis

In e-commerce, dynamic pricing has historically relied on transactional and behavioral data [3], [19], [20], frequently overlooking customer sentiment, which has a significant influence on price sensitivity and purchase decisions. Multilingual sentiment variables have not been methodically integrated into prediction models in any previous research. By integrating sentiment signals with numerical and categorical data inside an XGBoost framework, our study closes that gap. Our results show that incorporating multilingual sentiment features significantly enhances predictive performance compared to models trained solely on transactional data. As illustrated in Table IV, our presented XGBoost model performed the lowest error rates (MAE = 1.29, RMSE = 1.65) and the highest explanatory power (R^2 = 0.97), surpassing existing approaches such as GBM (R^2 = 0.92) [19], SVM (Accuracy = 86.92%) [3], and Neural Networks (R^2 = 0.84) [20].

This performance gain demonstrates that sentiment-enriched models are more capable of capturing nonlinear demand–price relationships and customer perceptions that drive purchasing behavior. Compared with prior models that relied exclusively on structured transactional data, our integration of multilingual sentiment features introduces an additional behavioral dimension. The advancement over Chowdhury et al. [20], where a purely neural network–based model produced $R^2 = 0.84$, reveals that sentiment acts as a complementary signal that enriches feature richness beyond what deep architectures can acquire with numerical data alone.

Moreover, while El et al. [19] demonstrated the strength of gradient boosting with $R^2 = 0.92$, our results show that combining boosting with sentiment inputs enables the model to generalize further, yielding a substantial 5.4% improvement in variance explanation. This distinction supports our claim of novelty, as no prior dynamic pricing model in e-commerce has demonstrated such integration of multilingual sentiment analysis. Limitations include dependence on the quality of sentiment extraction, the dataset's specificity to retail and e-commerce, and the computational demands of training XGBoost with enhanced features. Future studies could expand to other domains and optimize the model for real-time applications.

V. Conclusion

This study demonstrates the effectiveness of machine learning (ML) techniques in developing dynamic pricing strategies for e-commerce. We developed a robust framework that captures the complex, nonlinear dynamics of real-world pricing by establishing a comprehensive data-driven pipeline encompassing data collection, preprocessing, feature engineering, model selection, and performance evaluation. Of all the models tested, XGBoost consistently performed better than the others, demonstrating remarkable prediction accuracy and dependability with the lowest RMSE (1.65), lowest MAE (1.29), and highest R^2 score (0.97). The study showed the effectiveness of ML and ensemble techniques in e-commerce pricing engines, highlighting their potential in real-time deployment, revenue optimization, and client retention. Our approach outperforms previous research regarding accuracy and generalization, thanks to substantial feature engineering and improved model tuning. This reinforces the viability of machine learning-based solutions in dynamic pricing scenarios and lays the framework for future research into online learning models, reinforcement learning, or hybrid strategies that include real-time feedback systems for continuous optimization. Ultimately, e-commerce platforms seeking to enhance their pricing agility, maximize revenue, and maintain competitiveness in a dynamic market will find significant value in the study's conclusions. The findings indicate that sentiment signals can significantly enhance transactional characteristics, underscoring the importance of consumer perception in determining effective dynamic pricing. This creates several research opportunities. First, more performance advantages may be found by investigating deep neural architectures that simultaneously represent textual sentiment data and transactional data. Second, sentiment dynamics over time (e.g. shifts in consumer tone during promotional campaigns or seasonal events) should be incorporated for finer-grained prediction. Finally, industry-level studies could validate the practical impact of sentiment-aware dynamic pricing systems on profitability and consumer satisfaction.

VI. Disclosure and Conflict of Interest

The author declares that there are no conflicts of interest related to this research. Additionally, the author has no financial interests or competing affiliations that could have influenced the study's design, execution, or findings. This manuscript is the author's original work and has not been previously published or submitted for review to any other journal or conference.

References

[1] L. Almahadeen, M. K. Rathod, S. F. Sulaiman, K. Deepika, G. R. Teltumbade, and R. Saravanakumar, "Artificial intelligence and dynamic pricing strategies: Enhancing competitiveness in e-commerce," in 2025 3rd International Conference on Smart Systems for Applications in Electrical Sciences (ICSSES). IEEE, 2025, pp. 1–5.

[2] O. Famoti, C. P.-M. Ewim, O. Eloho, T. P. Muyiwa-Ajayi, O. N. Ezechi, and H. E. Omokhoa, "Revolutionizing customer experience management through data-driven strategies in financial services," *International Journal of Advanced Multidisciplinary Research and Studies*. 2025a, vol. 5, no. 1, pp. 948–957, 2025.

[3] M. Nowak and M. Pawłowska-Nowak, "Dynamic pricing method in the e-commerce industry using machine learning," *Applied Sciences*, vol. 14, no. 24, p. 11668, 2024.

[4] K. Safonov, "Neural network approach to demand estimation and dynamic pricing in retail," arXiv preprint arXiv:2412.00920, 2024.

[5] J. Liu, Y. Zhang, X. Wang, Y. Deng, and X. Wu, "Dynamic pricing on e-commerce platform with deep reinforcement learning: A field experiment," *arXiv preprint arXiv*:1912.02572, 2020.

[6] M. Mussi, G. Genalti, A. Nuara, F. Trovo, M. Restelli, and N. Gatti, "Dynamic pricing with volume discounts in online settings," *arXiv preprint arXiv:*2211.09612, 2022.

[7] Y. Zou, N. Qi, Y. Deng, Z. Xue, M. Gong, and W. Zhang, "Autonomous resource management in microservice systems via reinforcement learning," *arXiv preprint arXiv*:2507.12879, 2025.

[8] Y. Sun, K. Sekiguchi, and Y. Ohsawa, "Optimizing sentiment analysis in product descriptions: effects on customer purchase intentions," *Information Technology and Management*, pp. 1–24, 2025.

[9] M. Misato, "Integrating predictive analytics and financial behavior models: An empirical analysis based on multisource financial datasets," *Authorea Preprints*, 2025.

[10] M. Apte, K. Kale, P. Datar, and P. Deshmukh, "Dynamic retail pricing via q-learning – a reinforcement learning framework for enhanced revenue management," *arXiv preprint arXiv:*2411.18261, 2024.

[11] K. M. Devarajanayaka, S. S. Banu, D. J. Desai, V. TV, M. R. Palav, and S. K. Dash, "Machine learning-based pricing optimization for dynamic pricing in online retail," *Journal of Informatics Education and Research*, vol. 4, no. 3, 2024.

[12] C. Yin and Y. Han, "Dynamic pricing model of e-commerce platforms based on deep reinforcement learning," *Computer Modeling in Engineering Sciences*, vol. 127, no. 1, pp. 291–307, 2020.

[13] M. Loukili, F. Messaoudi, and M. El Ghazi, "Machine learning-driven dynamic pricing strategies in e-commerce," *ResearchGate Preprint*, 2023.

[14] P. Das, T. Pervin, B. Bhattacharjee, M. R. Karim, N. Sultana, M. S. Khan, M. A. Hosien, and F. Kamruzzaman, "Optimizing real-time dynamic pricing strategies in retail and e-commerce using machine learning models," *The American Journal of Engineering and Technology*, vol. 6, no. 12, pp. 163–177, 2024.

[15] N. N. I. Prova, "Advanced machine learning techniques for predictive analysis of health insurance," in *2024 Second International Conference on Intelligent Cyber Physical Systems and Internet of Things (ICoICI). IEEE*, 2024, pp. 1166–1170.

[16] N. Prova, "Detecting ai generated text based on nlp and machine learning approaches," *arXiv preprint arXiv:*2404.10032, 2024.

[17] R. M. Bhagat, J. M. Raut, B. V. Bahoria, S. W. Dhengare, P. B. Pande, Y. S. Pethe, N. Shelke, and V. S. Vairagade, "Strength prediction of fly ash-based sustainable concrete using machine learning techniques: an application of advanced decision-making approaches," *Multiscale and Multidisciplinary Modeling, Experiments and Design*, vol. 8, no. 1, pp. 1–19, 2025.

[18] Z. Mei, X. Bi, D. Li, W. Xia, F. Yang, and H. Wu, "Dhhnn: A dynamic hypergraph hyperbolic neural network based on variational autoencoder

for multimodal data integration and node classification," *Information Fusion*, p. 103016, 2025.

[19] R. El Youbi, F. Messaoudi, and M. Loukili, "Machine learning-driven dynamic pricing strategies in e-commerce," in *2023 14th International Conference on Information and Communication Systems (ICICS). IEEE,* 2023, pp. 1–5.

[20] M. S. Chowdhury, M. S. Shak, S. Devi, M. R. Miah, A. Al Mamun, E. Ahmed, S. A. S. Hera, F. Mahmud, and M. S. A. Mozumder, "Optimizing e-commerce pricing strategies: A comparative analysis of machine learning models for predicting customer satisfaction," *The American Journal of Engineering and Technology,* vol. 6, no. 09, pp. 6–17, 2024.

[21] R. El Youbi, F. Messaoudi, and M. Loukili, "Machine Learning-driven Dynamic Pricing Strategies in E-Commerce," *2023 14th International Conference on Information and Communication Systems (ICICS), Irbid, Jordan*, 2023, pp. 1-5, doi: 10.1109/ICICS60529.2023.10330541.

[22] F. Messaoudi, M. Loukili, and M. E. Ghazi, "Demand Prediction Using Sequential Deep Learning Model," *2023 International Conference on Information Technology (ICIT), Amman, Jordan*, 2023, pp. 577–582, doi: 10.1109/ICIT58056.2023.10225930.

[23] M. Loukili, F. Messaoudi and M. E. Ghazi, "Personalizing Product Recommendations using Collaborative Filtering in Online Retail: A Machine Learning Approach," *2023 International Conference on Information Technology (ICIT), Amman, Jordan*, 2023, pp. 19–24, doi: 10.1109/ICIT58056.2023.10226042.

[24] M. Loukili and F. Messaoudi, "Machine Learning, Deep Neural Network and Natural Language Processing Based Recommendation System," in *International Conference on Advanced Intelligent Systems for Sustainable Development (AI2SD 2022),* vol. 637, J. Kacprzyk, M. Ezziyyani, and V. E. Balas, Eds., Springer, Cham, 2023, pp. 7–17. doi: 10.1007/978-3-031-26384-2_7.

[25] L. Almahadeen, M. K. Rathod, S. F. Sulaiman, K. Deepika, G. R. Teltumbade and R. Saravanakumar, "Artificial Intelligence and Dynamic Pricing Strategies: Enhancing Competitiveness in E-Commerce," *2025 3rd International Conference on Smart Systems for applications in Electrical Sciences (ICSSES), Tumakuru, India*, 2025, pp. 1-5, doi: 10.1109/ICSSES64899.2025.11009613.

[26] Y. Wu, L. Liu, J. Bae, K. H. Chow, A. Iyengar, C. Pu, and Q. Zhang, "Demystifying learning rate policies for high accuracy training of deep neural networks," *2019 IEEE International conference on big data (Big Data) (pp. 1971–1980). IEEE*, 2019, December.