

Dynamic Pricing Prediction with Machine Learning Algorithm

Swati Kale¹, Prithviraj Patil², Darshan Patil³, Shivang Koul⁴

Dept of Electronics & Telecommunication

JSPM's Rajarshi Shahu College of Engineering, Tathawade, Pune, MS-India

Savitribai Phule Pune University, Pune

swadip.06@gmail.com prithvirajpatil1375@gmail.com darshanpatil8072@gmail.com gauravkoul1503@gmail.com

Abstract- Online prices can be the main thing you think about when buying because it's easy to understand. Dynamic pricing is a well-known idea. Because they are simple to understand, online costs can be the most important factor to consider when purchasing. It is not novel to consider dynamic pricing. It is extremely advantageous to many online businesses and is commonly used to increase sales and profits. This work is the outcome of continuous efforts to create a universal framework using powerful machine learning technologies. It assists clients in choosing the appropriate online store rather than the cheapest one. This approach can be used by online stores that are out of stock. Helping store owners set competitive rates by researching similar products supplied online is the aim.

Keywords— Predictive modeling, XGBoost, CatBoost, machine learning, dynamic pricing, and random forest regressor.

I. INTRODUCTION

A different approach to think about dynamic pricing is to change the prices based on what the consumer picks. Dynamic pricing is the act of altering inventory items by classifying customers based on the things they have selected and charging them various amounts for those products. Many industries are impacted by dynamic pricing, including retail, car, and other industries. air travel, mobile communication, electricity, and many more things[1]. The retail industry has grown as a result of increased access to customer demand data, the development of new technology that analyzes consumer behavior to help decide price more effectively, and the availability of decision support tools enabled by these technologies [2]. The effects noted in the mobile communication sector [3] can be attributed to heightened competition, reduced call expenses, and improved network architecture. Additionally, improved synchronization between manufacturing processes and inventory management decisions in the automobile sector leads to the development of a direct-to-consumer business concept. Internet accessibility has provided clients with two benefits: self-service access and time savings. Furthermore, rather than being a one-way highway that links customers and sellers, it serves as an open venue for debate and review exchange to enhance services. Dynamic pricing can only be a useful tool if certain requirements are satisfied. These include the cost of income, the availability of a segmented market, the likelihood of arbitrage being reduced, the willingness of customers to pay different prices, and fair play laws[4]. Dynamic pricing has become more and more commonplace worldwide as a result of its ability to streamline the purchasing and selling process. Dynamic pricing has a significant impact on many organizations, including travel, lodging, electric utilities, retail, online retail, wireless communication systems, athletics, autos, rental car businesses, and insurance [5].Combination auctions are a further dynamic pricing feature that is utilized in B2B exchange systems, e-

procurement, e-selling, and supply chain management. Marketing based on geography and history is included in this. As a result, dynamic pricing is becoming more and more significant in the e-commerce industry[11]. This is how the rest of the paper is structured. Section 2 provides a brief summary of current models to aid in understanding the dynamic pricing conceptual framework. The suggested models, which outline the procedure for creating dynamic pricing, are explained in Section 3. It will outline the various methods that this model employs. It will describe the many techniques this model uses. In Section 4, the results of the proposed models are presented together with an explanation of the data. They are compared to alternative dynamic pricing tactics used in retail settings. Section 5 presents the work's conclusion.

II. EXISTING MODELS

In today's environment, various models are employed to provide dynamic pricing. Some people are used to figuring out the cost of a wide range of things, while others are just concerned with estimating a certain price. The following is a list of the numerous techniques that are now used in pricing determination. Among these methods, their methodology developed a policy gradient-based reinforcement learning algorithm as well as employing a deep neural network to construct price policy and baseline neural network in order to reduce variance[8].This is accomplished using Bayesian model-based approach where MDP's transition and reward functions are modeled as distributions and action selection takes place by sampling[6].This methodology involves the utilization of statistical and machine learning models for predicting purchase decisions that are based on dynamic or adaptive pricing which exploits multiple data sources, leveraging web mining, big data technologies and machine learning[9].Using a Bayesian model-based strategy, the MDP's transition and reward functions are framed as distributions, and action selection is done through sampling[6].Deep neural networks were deployed to construct the price policy; this includes using a minimum amount of variation baseline network; a study that had its methodology include building reinforcement-learning algorithms based on policies involving gradient prices[8].The framework lays a strong foundation by amalgamating different data sources capturing visit attributes, visitor attributes, purchase history, web data, context understanding etc[9].

III. DATA PROCESSING AND FEATURE ENGINEERING

The proposed model utilizes three distinct approaches: first, it identifies the target client groups; second, it calculates the appropriate pricing for them; and third, it assesses the possibility that they would make a purchase within that price range. Figure 1 displays the framework.

1) Data acquisition:

This is the most crucial and first step in the framework process. It requires combining data from various sources into one, fully comprehensive database. We used a portion of the data from internet marketplaces for the investigation. First we downloaded the dataset [10]. The dataset categories are depicted in table 1.

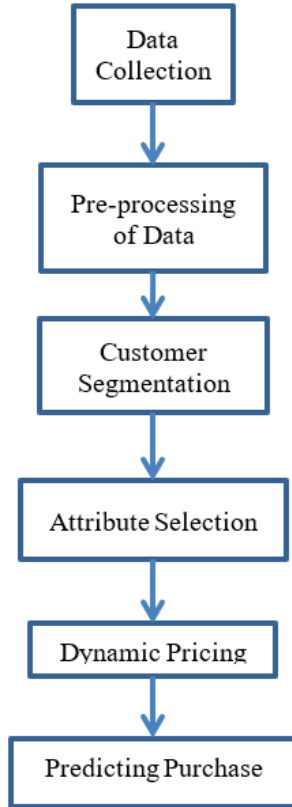


Fig. 1. Framework of System

2) Statistics Data:

This is the initial and most critical phase in the framework creation process. It requires merging info from various places into a single, integrated database. We used some of the information collected from online marketplaces for the investigation.

3) Exploratory data analysis:

1. Investigative data analysis was started.
2. The goal variable "Selling Price" on the plotting features was significantly left-skewed, which meant that the output likewise was becoming tilted left.
3. As a result, the target variable was normalized using a logarithmic transformation.
4. Following that, other features were developed, such as date-time and group-by features for analytical attributes, while label encoders were utilized to handle data that is categorical.
5. The XGBoost, LGBM, and CatBoost models were constructed following preprocessing.
6. Early stopping and parameter manipulation contributed to a high cross-validation score.
7. All of the models' outcomes were finally shown.

TABLE I. ATTRIBUTES OF THE DATA SET

Name	Description
Item	The product name.
Brand/Product	Brand to which the item is affiliated.
Product Type	The broader category to which the product is affiliated
Subcategory_1	The product falls into a more general category.
Subcategory_2	The item is part of a specific category that is two layers deep.
Product Evaluation	the evaluated rating that product purchasers have left behind.
Date	the day the product was offered for sale at that particular cost.
Selling_Price	The price at which the product was sold on the stated date.

IV. METHODOLOGY

The XGBoost, LightGBM, CatBoost, and random forest regressor ensemble learning methods are recommended for price solutions in this domain. Since relying merely on the output of a model trained with machine learning may be insufficient, ensembles lessons learned provide a systematic technique to integrating the prediction skills of many learners. Figure 2 shows the flow chart for the suggested models. The collected data is preprocessed to standardize it and add quantitative and category information. The suggested models XGBoost, LightGBM, and CatBoost are tested on this data set. The next subsection describes the designs.

A. XGBoost:

Instead of a search approach, the XGBoost uses both the first and second derivatives of the loss function. Ordering in advance and a large number of bits approach are employed to enhance the efficiency of the algorithm. The weak learners (decision trees) are suppressed by the regularization term in each iteration, preventing them from showing up in the final model. One-hot encoding is not a built-in feature of the XGBoost approach; it requires manual labor because it does not support categorical data. Because the technique can be parallelized and makes use of multi-core processors, it can be trained on very huge data sets. There is regularization, missing values, and parameter information for internal tree parameters. These are the primary concerns that XGBoost has identified. Tree clusters: XGBoost has determined these to be the main issues. Groups of trees: Iteratively, trees are formed one after the other. Cutting down on categorization errors is the aim. The output of the model is weighted each time n , taking into account its output at time $n-1$. Forecasts that are accurate are assigned lower weights, whereas forecasts that are not accurate are assigned larger weights. The technique known as "tree pruning" involves creating an over fitting tree and then removing its over fitting leaves in order to create more general models that may be classified using predetermined criteria. An outcome will be disregarded if it isn't better than. The figure 2 depicts process of XGBoost.

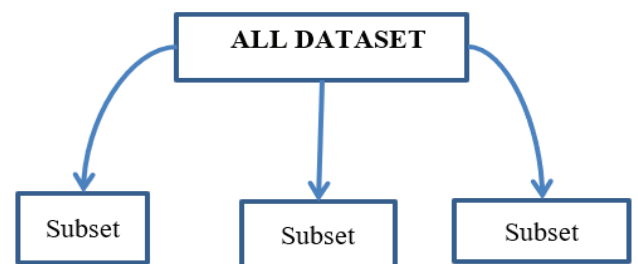


Fig. 2. Process of XGBoost

B. Light Boost :

The main distinction between XGBoost and Light Boost is that the latter maintains continuous feature values at discrete intervals, whereas the former accelerates training using a histogram-based method. To create the histogram using this method, the floating values are digitized and placed into k discrete value bins. LightGBM uses integers to encode categorical attributes, which results in a high degree of precision. For LightGBM models, label encoding is therefore recommended because it frequently outperforms one-hot encoding. A leaf-wise tree-growing method is used by the framework.

C. Category boosting (CatBoost):

It is widely employed to handle a broad variety of business challenges and works well with many data types. No particular data preprocessing is required to transform categories into numbers. We use a variety of mathematical data and intrinsic categorization characteristics to convert categorization values into numerical representation. It can handle enormous quantities of data and requires minimal RAM. The goal values in each scenario are obtained solely from observable historical data. The following steps are worked on by CatBoost: Creating subsets from the records; converting numbers to categorization labels. Converting the numerical attributes from the category CatBoost makes use of random decision trees, in which the splitting criteria are applied consistently at every tree level. In addition to being balanced and less prone to over fitting, this type of tree can greatly accelerate prediction during testing.

D. Random forest Regressor:

One of the main differences between LightBoost and XGBoost is that while the later speeds up training using histogram-based method, the former retains continuous feature values in discrete intervals. To create a histogram with this method, real numbers are converted into integers and placed in k different value bins. Depth algorithms are much slower to converge and less susceptible to over fitting than leaf methods for tree growth. Allocate separate portions of data sets for training and testing purposes. Use training datasets for random forest regression model. This model would be given input features which affect target variable such as product pricing. Fine-tune the hyper parameters of Random Forest Model to enhance its performance. Grid search or random

search techniques can be employed here. Use the model to forecast the best prices for products in real-time after it has been trained and validated. To obtain pricing estimates, enter the state of the market, the prices of competitors, and other pertinent information. The mean squared error (MSE) is a useful tool for determining how your data branches out from each node. You use the Random Forest Algorithm to address regression problems. The formula as per equation 1.

$$\frac{1}{N} \sum_{i=1}^N (f_i - y_i)^2 \quad (1)$$

The working of the Random Forest Regressor is shown in the figure 3.

V. RESULTS AND DISCUSSION

The proposed model gives us a reasonable cost solution while meeting the stated objectives. Using the recommended ensemble models, decision tree techniques like LightGBM, XGBoost, CatBoost, and random forest regressor are put into practice. One advantage of these models is that they can estimate significant attributes with prediction. Feature-based scores reinforce the decision trees in the model and improve the prediction result. In addition to XGBoost, the proposed approach also includes changes to the natural gradient predictive model to further boost the score. Therefore, it would be the optimal model to use among all the boosting ensemble models to generate pricing solutions for the datasets that are being considered. The ensemble techniques based on boosting can be expanded to forecast pricing solutions for products across several e-commerce platforms and recommend the best alternative for the customer. Considering the sheer number of products available on e-commerce platforms and the considerable variety within each category, there is a large range of products.

Table 2 shows predicted price of product and corresponding mean square error.

TABLE II. PRICE PREDICTION USING RANDOM FOREST

Category	Size	Quantity	Date	Price	MSE
Kurta	S	14	03-03-2024	5306/-	53779.8136
Top	M	23	03-03-2025	8717/-	52779.7134
Set	L	65	03-03-2026	24635/-	53779.8134
Ethnic Dress	XL	107	03-03-2027	40453/-	53775.56789
Bottom	M	456	03-03-2028	175195/-	53459.8145
Blouse	L	228	03-03-2029	86412/-	53599.9876
Saree	S	987	03-03-2030	37414/-	53725.1257
Kurta	S	14	03-03-2025	5348/-	53779.1475
Kurta	S	14	03-03-2026	5358/-	53779.1467
Top	M	23	03-03-2026	8726/-	53779.8652
Top	M	23	03-03-2027	8786/-	53779.7654
Set	L	65	03-03-2028	24830/-	53779.1678
Set	L	65	03-03-2029	24850/-	53779.3278
Ethnic Dress	XL	107	03-03-2029	40874/-	53779.0876
Ethnic Dress	XL	107	03-03-2025	40965/-	53779.1653
Bottom	M	456	03-03-2026	174192/-	53779.1432
Bottom	M	456	03-03-2025	174052/-	53779.9203
Saree	S	987	03-03-2026	377034/-	53779.5643
Saree	S	987	03-03-2025	375043/-	53779.7852
Blouse	L	228	03-03-2026	87096/-	53779.5856
Blouse	L	228	03-03-2024	84158/-	53779.1485

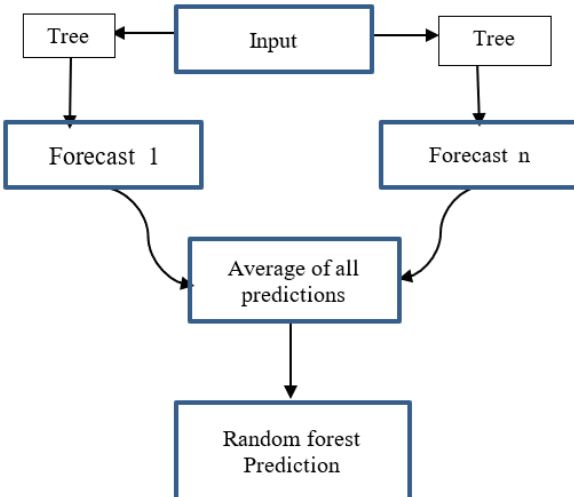


Fig. 3. Working of Random forest Regressor

Figure 4 depicts sample of sales form developed at the user end.

Total Price: 786

Sales Form

Date: 01/10/2024

Category: Set

Size: M

Quantity: 1

Submit

Fig. 4. GUI generated at user end.

Figure 5 shows the graphical presentation of predicted price using Random Forest.

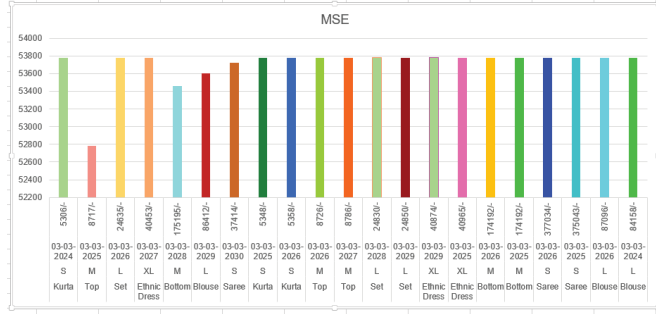


Fig. 5. Graphical presentation of Table 2

Table 3 depicts the price prediction using Decision Tree algorithm.

TABLE III. PRICE PREDICTION USING DECISION TREE

Category	Size	Quantity	Date	Price	MSE
Kurta	S	14	03-03-2024	5288/-	52788.845
Top	M	23	03-03-2025	8917/-	53878.8134
Set	L	65	03-03-2026	24700/-	54899.97
Ethnic Dress	XL	107	03-03-2027	40450/-	53989.6780
Bottom	M	456	03-03-2028	175200/-	54459.8145
Blouse	L	228	03-03-2029	86412/-	53599.9876
Saree	S	987	03-03-2030	37414/-	53725.1257
Kurta	S	14	03-03-2025	5348/-	53779.1475
Kurta	S	14	03-03-2026	5238/-	51779.1467
Top	M	23	03-03-2026	8645/-	54779.8652
Top	M	23	03-03-2027	8717/-	51779.7654
Set	L	65	03-03-2028	23450/-	51779.1678
Set	L	65	03-03-2029	23450/-	51776.3278
Ethnic Dress	XL	107	03-03-2029	40225/-	50778.0876
Ethnic Dress	XL	107	03-03-2025	40965/-	53879.1653
Bottom	M	456	03-03-2026	174292/-	52769.1432
Bottom	M	456	03-03-2025	174292/-	54789.9203
Saree	S	987	03-03-2026	377034/-	5488.5643
Saree	S	987	03-03-2025	375185/-	52456.7852
Blouse	L	228	03-03-2026	87096/-	53779.5856
Blouse	L	228	03-03-2024	86125/-	51677.1485

Price prediction using random forest is better than decision tree. Figure 6 shows the graphical presentation of predicted price using Decision Tree.

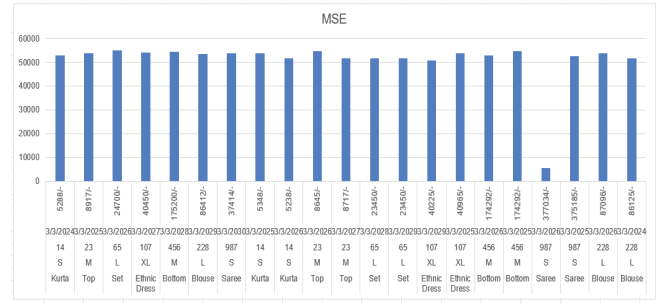


Fig. 6. Graphical presentation of Table 3

VI. CONCLUSION

A suggested framework has been developed by exploiting powerful techniques of data mining, machine learning and statistical methods to predict an online customer's buying behaviour through dynamic pricing that selects the appropriate price range for them. An e-commerce company tested this system using large datasets which shows good results for a full adoption of the approach. A suggested framework has been created based on the powerful ways of data mining, machine learning and statistical methodologies to predict purchasing behaviours of an online customer by selecting the right price range for them which is through dynamic pricing. Such a system was tested using sizable datasets by one e-commerce firm, and the results are sufficient to warrant its full implementation. Error rates have been reduced as well as better prices determined that benefit both organization and customer. This general architecture can be modified to suit specific applications in various online industries. Furthermore, it is expected that this research will continue with further discussion on progress outcomes.

REFERENCES.

- [1] Cummings, T. "Everything you need to know about dynamic pricing", accessed from <http://www.csmonitor.com/Business/Saving-Money/2013/1104/Everything-you-need-to-know-about-dynamic-pricing> on 19th June 2014.
- [2] Elmaghraby, W., & Keskinocak, P. "Dynamic pricing in the presence of inventory considerations: Research overview, current practices, and future directions", *Management Science*, 2003
- [3] Karpowicz, A.; Szajowski, K, "Double Optimal Stopping Times and Dynamic Pricing Problem: Description of the Mathematical Model", 2007.
- [4] Chen, Y.; Wang, F. A, " Dynamic Pricing Model for E-Commerce Based on Data Mining", *Proceedings of the 2009 Second International Symposium on Computational Intelligence and Design*, Changsha, China, 12–14 December 2009.
- [5] Kanishka Misra, Eric M. Schwartz, J. Abemethy "Dynamic Online Pricing with Incomplete Information Using Multi armed Bandit Experiments", *Marketing science* (Providence, R.I.) 2019.
- [6] Wei Han "A Dynamic Pricing Algorithm by Bayesian Q-learning", *Second International Conference on Computer Modeling and Simulation*, 2010.
- [7] Jiayi Liu, Yidong Zhang, Xiaoqing Wang, Yuming Deng, Xingyu Wu "Dynamic Pricing on E-commerce Platform with Deep Reinforcement Learning", *arXiv.org*, 2019.
- [8] Shiyu Chen, Lingxiang Li, Zhi Chen, Shaoqian Li "Dynamic Pricing for Smart Mobile Edge Computing: A Reinforcement Learning Approach", *IEEE Wireless Communications Letters*, 2021.
- [9] Rajan Gupta, Chaitanya Pathak "A Machine Learning Framework for Predicting Purchase by Online Customers based on Dynamic Pricing", Elsevier B.V, 2014.
- [10] <https://www.kaggle.com/datasets/thedevastator/unlock-profits-with-e-commerce-sales-data/?select=Amazon+Sale+Report.csv>

- [11] Poh, L.Z. Connie T.; Ong, T.S. Goh, M.K.O “Deep Reinforcement Learning-Based Dynamic Pricing for Parking Solutions”, Licensee MDPI, Basel, Switzerland, 2023.
- [12] Josef Bauer, Dietmar Jannach, “Optimal pricing in e-commerce based on sparse and noisy data”, MDPI, 2017.