



Product pricing solutions using hybrid machine learning algorithm

Anupama Namburu¹ · Prabha Selvaraj² · M. Varsha³

Received: 16 June 2022 / Accepted: 3 July 2022 / Published online: 25 July 2022
© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2022

Abstract

E-commerce platforms have been around for over two decades now, and their popularity among buyers and sellers alike has been increasing. With the COVID-19 pandemic, there has been a boom in online shopping, with many sellers moving their businesses towards e-commerce platforms. Product pricing is quite difficult at this increased scale of online shopping, considering the number of products being sold online. For instance, the strong seasonal pricing trends in clothes—where Brand names seem to sway the prices heavily. Electronics, on the other hand, have product specification-based pricing, which keeps fluctuating. This work aims to help business owners price their products competitively based on similar products being sold on e-commerce platforms based on the reviews, statistical and categorical features. A hybrid algorithm X-NGBoost combining extreme gradient boost (XGBoost) with natural gradient boost (NGBoost) is proposed to predict the price. The proposed model is compared with the ensemble models like XGBoost, LightBoost and CatBoost. The proposed model outperforms the existing ensemble boosting algorithms.

Keywords X-NGBoost · XGBoost · CatBoost · Product pricing · Ensemble algorithms

1 Introduction

E-Commerce platforms are popular over the last two decades as the people recognised the ease of buying products online without having to step out of their homes. In the recent past, there has been a boom in online shopping, with many sellers moving their businesses to these e-commerce platforms, Enberg [7]. Product pricing is quite difficult on e-commerce platforms, given the sheer number of products sold online.

Several factors influence the pricing of products, and thus, even pricing tends to be dynamic at times.

Yang et al. proposed an advanced business-to-business (B2B) and customer-to-customer(C2C) methods Yang and Xia [24] for complete supply chain, resulting in future prediction for market analysis, experimental aspects and other unknown facts of current facts. Predicting a strategy for prices in share market is highly demanded topic in today's world. Aljazzazen demonstrated the gap between the two methods skimming and penetration to show the optimal strategy for price prediction AlJazzazen [2]. Many investigations are performed to measure the effectiveness of various marketing plans to improve the industrial loyalty and individuality. Davicik described the facts of organising the campaigns Davicik and Sharma [6] at places with different products with respect to the prices that will help increase the marketing. By conducting the feedback and review session with the customers, behaviour and the characteristics of the end user purchase details were analysed by Li et al. [12]. With the analysis conducted, different strategies are framed and formulated by Xin Li (2021). Priester (2020) provided two ways by personalised dynamic pricing based on fairness and privacy policies Priester et al. [16]. This includes location- and history-based marketing. Minglun ren proposed three pric-

✉ Anupama Namburu
namburianupama@gmail.com

Prabha Selvaraj
prabha.s@vitap.ac.in

M. Varsha
varsha.m@vitap.ac.in

¹ School of Computer Science and Engineering, VIT-AP University, Beside AP Secretariat, Near Vijayawada, Andhra Pradesh 522237, India

² School of Computer Science and Engineering, VIT-AP University, Beside AP Secretariat, Near Vijayawada, Andhra Pradesh 522237, India

³ School of Computer Science and Engineering, VIT-AP University, Beside AP Secretariat, Near Vijayawada 522237, Andhra Pradesh, India

ing models by investigating backward and game theory Ren et al. [18]. This is framed based on the service cost assigned to the certain weights.

Many real-time examples are enclosed for different strategies of pricing and profit situations, which act as the attributes in the channel service. Survey on various consumers and end users helps to find the behaviour of each consumer, and this leads to analyse the impact of pricing decisions of the entire supply chain Wei and Li [22]. Theoretical analysis gives partial result in decision-making, and mathematical analysis will narrow down the decision but it is quite challenging, wei (2015). The author could find the loyalty to the channel that increases the purchase and favour manufacturer. Aliomar lino mattos found that cost-based approach is said to dominating when compared to the other value and competition-based approaches that leads to the ground work for future analysis, Mattos et al. [13]. This is done by the investigation made on 286 papers on total and from 31 journals that enhances the future opportunities. In addition to the business market strategies, psychological behaviours from individual throw a vital information in determining the cost of the products in markets, Hinterhuber et al. [9]. Partitioned pricing is estimated by the experimental studies to increase the price in the share market for the welfare and sustainability of the products, Brgin and Wilken [5] proposed by Burgin et al. (2021). Gupta (2014) developed a general architecture using machine learning models, Gupta and Pathak [8], that will help to predict the purchase price preferred by the customer. This can be included in the e-commerce platform and share market prediction. The research extended its functionalities to the next level by adapting personalised adaptive pricing. Since prediction of price depends on time-series data, LSTM model is used for predicting the product prices on a daily basis, Yu (2021). This model produced a notable result with high accuracy, Yu [25]. Spuritha (2021) suggested ensemble model that uses XGBoost algorithm for sales prediction, and the results are validated with low error rate using RMSE and MAPE values, Spuritha et al. [20]. Wiyati's (2019) finding revealed that the combining supply chain management with cloud computing has high performance and give reliable output, Wiyati et al. [23]. Winky compared the result with three performance measures for support vector machine model, Ho et al. [10].

Mohamed Ali Mohamed framed online trailer data set for the task prediction for the seasonal products, Mohamed et al. [14], with autoregressive-integrated-moving-average method with support vector machine to know the statistical-based analysis. Jose-Luis et al. have explained the use of ensemble model in real estate property pricing and showed improvement in estimating the dwelling price Alfaro-Navarro et al. [1] using random forest and bagging techniques. Jorge Ivan discussed method incremental sample with resampling (MINREM) P  rez Rave et al. [15] method for identifying

variables for inferential and predictive analysis by integrating machine learning and regression analysis. Grybauskas et al. have analysed about the influence of attributes on price revision of property, Andrius et al. [3]; Shahrel et al. [19], during pandemic and tested with different machine learning models.

To achieve high speed and accuracy, recently interesting ensemble algorithms based on gradient boost are added in the literature, namely Extreme gradient Boost (XGBoost) regressor, Light Gradient Boosting Machine (LightGBM) regressors, Category Boosting(CatBoost)Regressor and Natural gradient boosting (NGBoost). The XGBoost is a scalable algorithm and proved to be an challenger in solving machine learning problems. LightGBM considers the highest gradient instances using selective sampling and attains fast training performance. CatBoost achieves high accuracy with avoidance of shifting of prediction order by modifying the gradients. NGBoost provides probabilistic prediction using gradient boosting with conditioning on the covariates. These ensemble learning technologies provide a systematic solution for combining the predictive capabilities of multiple learners, because relying on the results of a single machine learning model may not be enough.

For pricing solutions, only the basic machine learning models are used, and the boosting algorithms are the way forward to achieve speed and efficiency. In the proposed work, XGBoost, LightGBM, CatBoost and X-NGBoost are used in order to train the huge online data set with speed and accuracy, the XGBoost is considered as the base learning model for the NGBoost, at every iteration the best gradient is estimated based on the score indicating the presumed outcome based on the observed features. XGBoost builds trees in order, so that each subsequent tree will minimise the errors generated in the previous tree but it fails to estimate the features. Hence, the NGBoost considers the outcomes of the XGBoost and provides the best estimate based on features further assisting in faster training and accuracy.

Thus, this paper is an attempt to offer business owners competitive pricing solutions based on similar products being sold on e-commerce platforms by applying ML algorithms with the main target being small-scale retailers and small business owners. Machine learning does not only just give pricing suggestions but accurately predict how customers will react to pricing and even forecast demand for a given product. ML considers all of this information and comes up with the right price suggestions for pricing thousands of products to make the pricing suggestions more profitable. The objectives of the work are listed below.

Objectives:

- Offer pricing solutions using ML trained on historical and competitive data while taking into consideration the date

Table 1 Attributes of the data set

Name	Description
Product	Name of the product
Product_Brand	Brand the product belongs to
Item_Category	The wider category of items the product belongs to
Subcategory_1	The subcategory the product belongs to—one level deep
Subcategory_2	Specific category the item belongs to—two levels deep
Item_Rating	The reviewed rating left behind by buyers of the products
Date	Date at which the product was sold at the specific price
Selling_Price	Price of the product sold on the specified date

time features, group-by of categorical variables statistical features.

- To help retailers make pricing decisions quicker and change prices more accordingly.
- To create predictable, scalable, repeatable and successful pricing decisions, business owners will always know the reason.
- The outcome of applying the pricing solutions can be used to get a better understanding of your customer engagement and customer behaviour.

2 Data preprocessing and feature engineering

We used the knowledge in the data domain to create features to improve the performance of machine learning models. Given that various factors are already available that influence pricing, the addition of date time features, statistical features based on categorical variables would help in predicting better results.

2.1 Exploratory data analysis

1. Began with Exploratory Data Analysis.
2. On plotting features, the target variable ‘Selling Price’ was highly left-skewed, due to which the output also was getting left-skewed.
3. Thus, a logarithmic transformation to normalise the target variable was applied.
4. Next, creation of new features like date time features, group-by of categorical variables for statistical features; categorical variables were handled through label encoder.
5. After preprocessing, the models were built: XGBoost, LGBM, CatBoost and X-NGBoost.
6. Manipulation of parameters helped give rise to a good cross-validation score through early stopping.
7. Finally, the results of all the models were presented.

2.2 Data set

The data set used carries various features of the multiple products sold on an e-commerce platform. The attributes used for working the model are given in Table 2. The data set has two parts: the training data set sample shown in Table 2 and the testing data set shown in Table 3. The data set is collected from kaggle, named as E-Commerce_Participants_Data. It contains the 2453 records of training data and 1052 testing data.

2.3 Preprocessing

On plotting a distribution plot of the data set in Fig. 1, it was observed that the target variable, that is the Selling Price attribute, was highly skewed.

Since the data do not follow a normal bell curve, a logarithmic transformation is applied to make them as “normal” as possible, making the statistical analysis results of these data more effective.

In other words, applying the logarithmic transformation reduces or eliminates the skewness of the original data. Here, the data are normalised by applying a logarithmic transformation. The result of the normalised variable is shown in Fig. 2.

2.4 Attributes of the data set

2.5 Statistical features added

- Date Features
Seasonality heavily influences the price of a product. Adding this feature would help get better insight on how prices change based on not just the date but financial quarter, part of the week, the year and more. The data set sample after adding the date features is shown in Table 4.
- Group-by category unique items feature
Products that belong to the same category generally follow similar pricing trends. Grouping like products helps with better pricing accuracy. The data set sample after

Table 2 Training data preview

S.no.	Product	Product_Brand	Item_Category	subcategory_1	subcategory_2	Item_Rating	Date	Selling_Price
0	P-2610	B-659	Bags wallets belts	Bags	Handbags	4.3	2/31/2017	291.0
1	P-2453	B-3078	Clothing	Women's clothing	Western wear	3.1	7/1/2015	897.0
2	P-6802	B-1810	Home décor festive needs	Show pieces	ethnic	3.5	1/12/2019	792.0
3	P-4452	B-3078	Beauty and personal care	Eye care	h2opluseyecare	4.0	12/12/2014	837.0
4	P-8454	B-3078	Clothing	Men's clothing	T-shirts	4.3	12/12/2013	470.0

Table 3 Testing dataset preview

S. no.	Product	Product_Brand	Item_Category	Subcategory_1	Subcategory_2	Item_Rating	Date
0	P-11284	B-2984	Computers	Network components	Routers	4.3	1/12/2018
1	P-6580	B-1732	Jewellery	Bangles Bracelets armlets	Bracelets	3.0	20/12/2012
2	P-5843	B-3078	Clothing	Women's clothing	Western wear	1.5	1/12/2014
3	P-5334	B-1421	Jewellery	Necklaces chains	Necklaces	3.9	1/12/2019
4	P-5586	B-3078	Clothing	Women's clothing	Western wear	1.4	1/12/2017

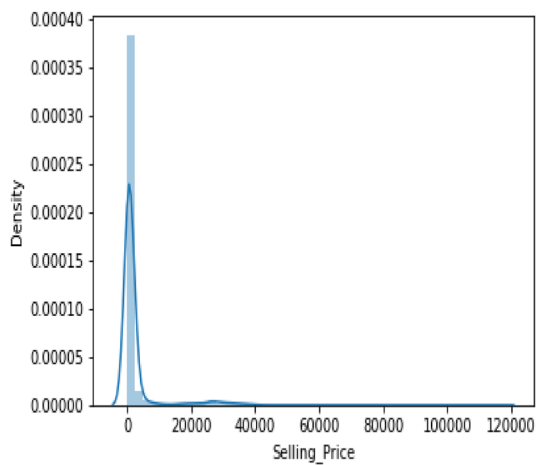


Fig. 1 Target variable before preprocessing

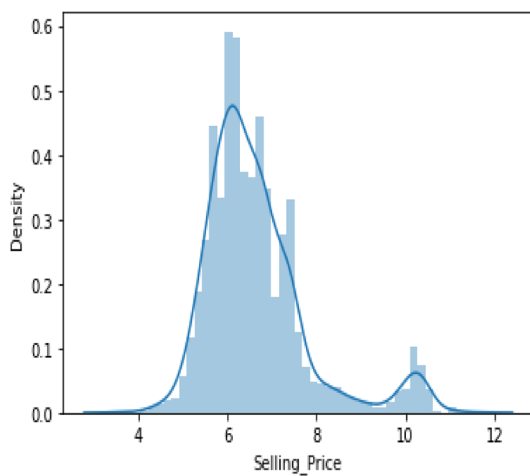


Fig. 2 Target variable after preprocessing

adding the group-by category unique items feature is shown in Table 5.

2.6 Categorical features added

Most algorithms cannot process categorical features and work better with statistical features. Thus, one-hot coding is done using label encoders to add the below features as shown in Tables 6, 7 and 8.

- Item_Category
- Subcategory_1
- Subcategory_2

With one-hot coding the categorical values are converted into a new categorical column with assigned binary value of 1 or 0 to those columns. Each one-hot coding represents an integer value representing a binary vector.

Table 4 Dataset after the addition of statistical features

S. no.	Product	Product_Brand	Item_Category	subcategory_1	subcategory_2	Item_Rating	Month	Day	DayofYear	Week	Quarter
0	P-2610	B-659	Computers	Network components	Routers	4.3	2	3	34	5	1
1	P-2453	B-3078	Jewellery	Bangles bracelets armlets	Bracelets	3.0	7	1	182	27	3
2	P-6802	B-1810	Clothing	Women's clothing	Western wear	1.5	1	12	12	1	1
3	P-4452	B-3078	Jewellery	Necklaces chains	Necklaces	3.9	12	12	346	50	4
4	P-8454	B-3078	Clothing	Women's clothing	Western wear	1.4	12	12	246	50	4

Table 5 Dataset after the addition of statistical features

Is_month_start	Is_month_end	Unique_Item_category_per_product_brand	Unique_Subcategory_1_product_brand	Unique_Subcategory_2_product_brand
False	False	1	1	1
True	False	13	29	100
False	False	1	1	1
False	False	13	29	100
False	False	13	29	100

Table 6 Dataset after the addition of categorical features

S. no.	Product	Product_Brand	Item_Category	subcategory_1	subcategory_2	Item_Rating	Selling_Price	month	Day	Day of Year	Week	Quarter
0	P-2610	B-659	9	11	159	4.3	5.676754	2	3	34	5	1
1	P-2453	B-3078	17	139	387	3.0	6.800170	7	1	182	27	3
2	P-6802	B-1810	38	119	118	1.5	6.675823	1	12	12	1	1
3	P-4452	B-3078	12	40	155	3.9	6.731018	12	12	346	50	4
4	P-8454	B-3078	17	86	344	1.4	6.154858	12	12	246	50	4

Table 7 Dataset after the addition of categorical features

Is_month_start	Is_month_end	Unique_Item_category_per_product_brand	Unique_Subcategory_1_product_brand	Unique_Subcategory_2_product_brand
False	False	1	1	1
True	False	13	29	100
False	False	1	1	1
False	False	13	29	100
False	False	13	29	100

Table 8 Dataset after the addition of categorical features

Std_rating_per_product_brand	Std_rating_Item_category	Std_rating_Subcategory_1	Std_rating_Subcategory_2
NaN	1.105474	1.094790	1.018614
1.197104	1.205383	1.227087	1.186726
1.202082	1.192523	1.094301	1.028398
1.197104	1.215595	0.636396	NaN
1.197104	1.205383	1.146774	1.123001

2.7 Graphical overview of processed data set

Heat map

A graphical representation of the correlation between all the attributes is shown in Fig. 3. Here, the correlation values are denoted by the colours where red bears the most correlation and blue the least.

3 Proposed methodology

In this section, the models XGBoost, LightGBM, CatBoost and X-NGBoost ensemble learning techniques are proposed for pricing solution. These ensemble learnings provide a systematic solution for combining the predictive capabilities of multiple learners, because relying solely on the results of a machine learning model may not be enough. The flow chart of the proposed models is shown in Fig. 4.

The data set is preprocessed to add the statistical and categorical features and performing the normalisation. The proposed models XGBoost, LightGBM, CatBoost and the X-NGBoost are executed on the data set. The models are explained in the following subsection.

3.1 Gradient boost models

- XGBoost, LightGBM, CatBoost and NGBoost are ensemble learning techniques.
- Ensemble learning provides a systematic solution for combining the predictive capabilities of multiple learners, because relying solely on the results of a machine learning model may not be enough.
- The aggregate output of multiple models results in a single model.
- In the boost, the trees are built in order, so that each subsequent tree aims to reduce the errors of the previous tree.
- The residuals are updated, and the next tree growing in the sequence will learn from them.

3.2 Comparison of existing algorithms

The comparison of the Boosting Algorithms Bentéjac et al. [4] is presented in Table 9 with respect to splits of trees, missing values, leaf growth, training speed, categorical feature handling and parameters to control over fitting.

3.3 XGBoost

XGBoost, Jabeur et al. [11], is a boosting integration algorithm obtained by combining the decision trees and gradient lift algorithm. The XGBoost makes use of the first and sec-

ond derivatives of the loss function instead of search method. It uses pre-ordering and node number of bits technique to improve the performance of the algorithm. With the regularisation term, in every iteration the weak learners (decision trees) are suppressed and are not part of the final model. The objective function o of the XGBoost is given as:

$$O^{(t)} = \sum_{i=1}^n [l(y_i, \hat{y}_i^{(t-1)}) + p_i f_t(x_i) + \frac{1}{2} q_i f_t^2(x_i)] + \Omega(f_t) \quad (1)$$

where i is the i^{th} sample and t represents the iteration, y_i represents the i^{th} sample's real value and $\hat{y}_i^{(t-1)}$ is the predicted outcome of $(t-1)^{th}$ iteration, p_i and q_i are the first and second derivatives and $\Omega(f_t)$ represents the term of regularisation.

XGBoost algorithm does not support categorical data; therefore, one-hot encoding is required to be done manually—no inbuilt feature present. The algorithm can also be trained on very large data sets because the algorithm can be parallelised and can take advantage of the capabilities of multi-core computers. It internally has parameters for regularisation, missing values and tree parameters. Few important concerns of XGBoost are listed below. Tree ensembles: Trees are generated one after the other, and with each iteration attempts are made to reduce the mistakes of classification.

At any time n , the results of the model are weighted based on the results at the previous time $n-1$. The results of correct predictions are given lower weights, and the results of incorrect predictions are given higher weights. Tree pruning is a technique in which an overfitted tree is constructed and then leaves are removed according to selected criteria. Cross-validation is a technique used to compare the partitioning and non-partitioning of an over-fitted tree. If there is no better result for a particular node, then it will be excluded.

3.4 LightBoost

LightBoost, Sun et al. [21], is also based on decision tree and boosting learning gradient framework. The major difference between the LightBoost and XGBoost is the former uses a histogram-based algorithm or stores continuous feature values in discrete intervals to speed up the training process. In this approach, the floating values are digitised into k bins of discrete values to construct the histogram. LightGBM offers good accuracy with categorical features being integer-encoded. Therefore, label encoding is preferred as it often performs better than one-hot encoding for LightGBM models. The framework uses a leaf-wise tree growth algorithm. Leaf tree growth algorithms tend to converge faster than depth algorithms and are more prone to overfitting.

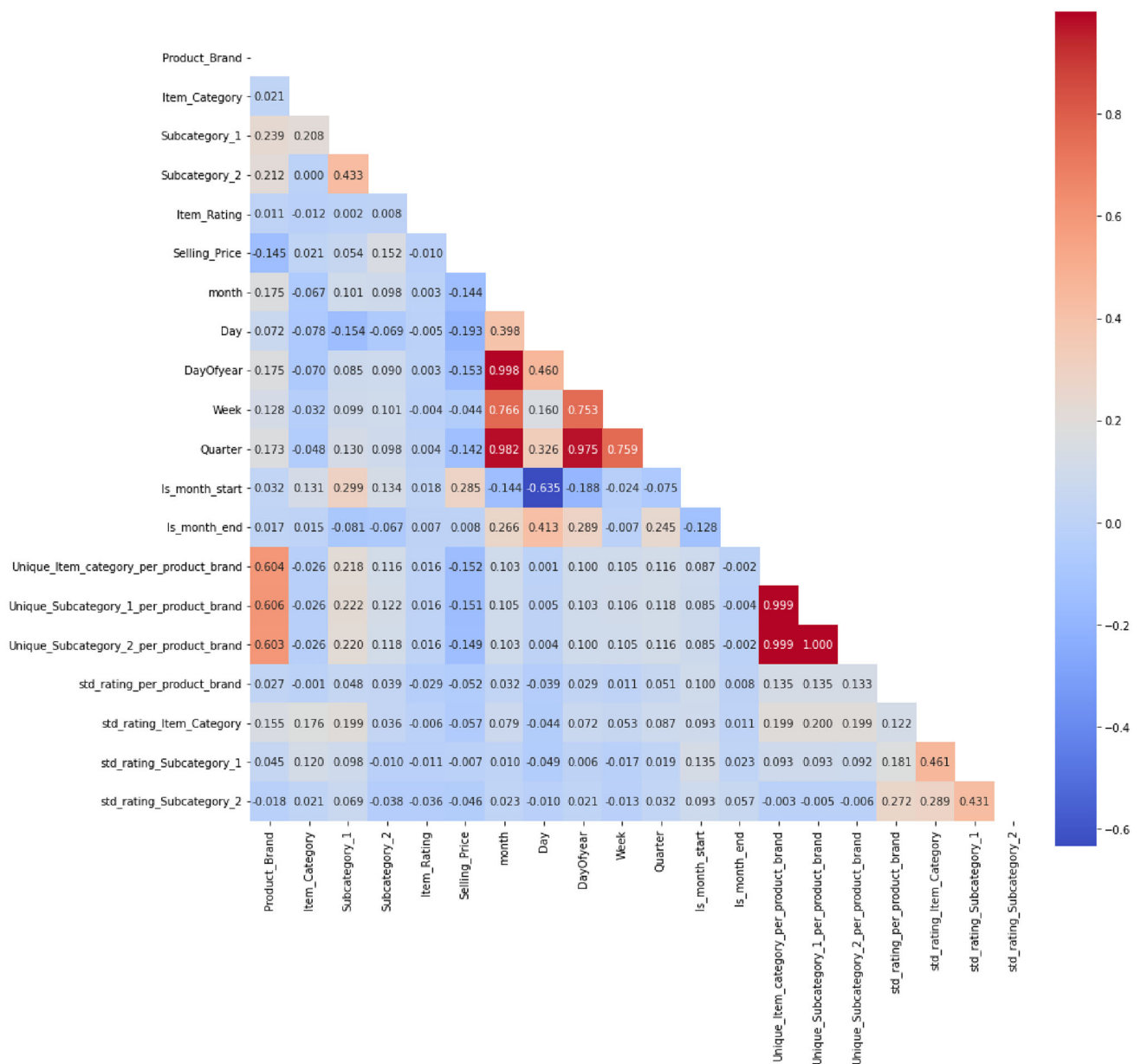


Fig. 3 Heat map of correlation between the attributes

Fig. 4 The flow chart of the proposed model

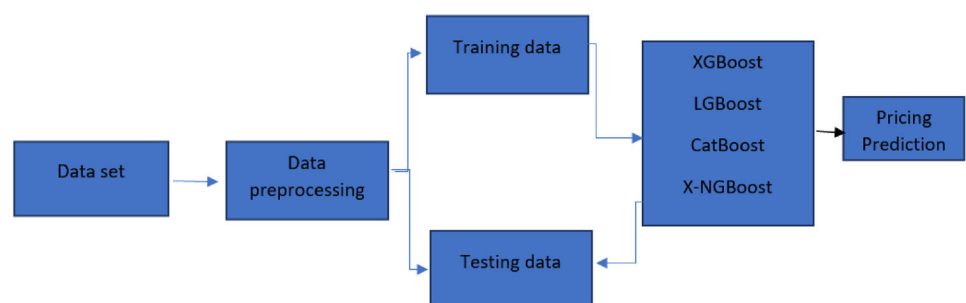


Table 9 Comparison table of boosting algorithms

Function	XGBoost	LightGBM	CatBoost
Splits	It does not use any weighted sampling technique, which makes its division process slower than GOSS and MVS.	It provides gradient-based one-sided sampling (GOSS), which uses instances with large gradients and random samples with small gradients to select segmentation	It provides a new technique called minimal variance sampling (MVS), where weighted sampling occurs at the tree level rather than the split level
Missing value	Missing values will be allocated to the side that reduces the loss in each split	The missing values will be assigned to the side that reduces the loss in each division	It has “Min” and “Max” modes for processing missing values
Leaf Growth	It splits to the specified max_depth and then starts pruning the tree backwards by deleting splits which have no positive gain, because splits without loss reduction can be followed by splits with loss reduction	Use the best first tree growth because you select the leaves that minimise growth loss, allowing unbalanced trees to grow because overfitting can occur when data are small	Grow a balanced tree and at each layer in the tree, the feature segmentation pair that produces the least loss is selected and used for all nodes in this level
Training speed	Slower than CatBoost and LightGBM	Faster than CatBoost, XGBoost	Faster than XGBoost
Categorical feature handling	It does not have a built-in method for classifying features—the user has to do the encoding.	The method is to sort the categories according to the training objectives. Categorical_features: Used to specify the features that we consider when training the model	Combines one-hot encoding and advanced media encoding. One_hot_max_size: Perform active encoding for all functions with multiple different values
Parameters to control overfitting	- Learning rate - Maximum depth - Minimum child weight	- Learning rate - Maximum depth - Number of leaves - Minimum data in leaf	- Learning rate - Depth - L2 leaf reg

XGBoost uses GOSS (Gradient-Based One-Sided Sampling) with an concept that not all data points contribute in the same way to training. Data points with small gradients tend to be better trained; therefore, it is more effective to focus on data points with larger gradients. By calculating its contribution to the loss changes, LightGBM will increase the weight of the samples with small gradients.

3.5 Category boosting (CatBoost)

Category Boosting approach, Punmiya and Choe [17], mainly focuses on the permutation and target-oriented statistics. It works well with multiple categories of data and is widely applied to many kinds of business challenges. It does not require explicit data preprocessing to convert categories to numbers. We use various statistical data combined with internal classification features to convert classification values into numbers. It can deal with large amounts of data while using less memory to run. It lowers the chance of overfitting leading to more generalised models. For categorical encoding, a sorting principle called Target Based with Previous (TBS) is used. Inspired by online learning, we get training examples sequentially over time. The target values for each example are based solely on observed historical records. We use a combination of classification features as additional classification features to capture high-order dependencies. CatBoost works on the following steps:

- Forming the subsets from the records
- Converting classification labels to numbers
- Converting the categorical features to numbers

CatBoost uses random decision trees, where splitting criteria are uniformly adopted throughout the tree level. Such a tree is balanced, less prone to over fitting, and can significantly speed up prediction during testing.

4 X-NGBoost

The primary aim of the work is to design a novel technique based on AI for predicting the product price based on XGBoost model and NGBoost algorithm, called X-NGBoost technique. The flow chart of the model is shown in Fig. 5

In the proposed model, the preprocessed data set is processed with XGBoost adapted with natural probability prediction algorithm (as the base learning model). The initial training data are fitted to the XGBoost to start the model. Further, the hyperparameters of XGBoost are chosen with trial and error and are optimised by Bayesian optimisation. Based on the score estimated, the optimised parameters are obtained from the Bayesian optimisation. The proposed optimised predictive models effectively enhance the accuracy.

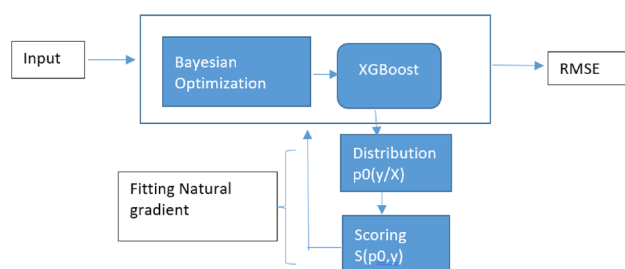


Fig. 5 X-NGBoost algorithm

Table 10 Root-mean-square error of the algorithms

Model	RMSE training	RMSE testing
XGBoost	6.62	7.48
LightGBM	6.62	7.44
CatBoost	5.91	6.96
X-NGBoost	4.23	5.34

To compare the model, the XGBoost, LightBoost and CatBoost algorithms are trained and implemented on the same data set. The comparison of algorithms is made based on root-mean-square error (RMSE) quantified with meters. RMSE is the prominent measurement in case of calculating sensitive error.

5 Results and discussion

The proposed model abides by the objectives discussed and provides us with the appropriate pricing solution. The proposed ensemble models of decision tree methods like LightGBM, XGBoost, and CatBoost are implemented. The advantage of these models is that the features of importance are estimated with prediction. Feature's based score is used to boost the decision trees with in the model improving the prediction result. In the proposed model along with the XGBoost, the natural gradient predictive model is adapted as to further enhance the score.

Therefore, from the results in Table 10 it can be concluded that, for the data set, X-NGBoost algorithm gives the appropriate pricing solution. Thus, it would be the appropriate model among the boosting ensemble model for providing pricing solutions for the considered dataset. The boosting-based ensemble techniques further can be extended to predict pricing solutions for products of multiple e-commerce sites and suggesting the appropriate option for the costumer. Large products variety considering the multitude of products sold on e-commerce platforms and the window variety within each category.

6 Conclusion and future work

Since the e-commerce platform has been growing rapidly over the last decade, it has become the norm for many small businesses to sell their products on such a platform to begin with as opposed to setting up shop.

Considering the limited budget of such small business owners, having a tool that is easily accessible and gives them a basic idea of how to price their products effectively would prove to be very useful. Enhancing such businesses in the long run would help improve the economy while promoting local industries to establish and develop themselves.

The results obtained indicate that the X-NGBoost algorithm is suitable for providing appropriate pricing solutions with the lowest error rates. Furthermore, the implementation of ensemble techniques resulted in an output that was reliable and efficient—possibly even having a lower error rate and therefore providing users with a fruitful result and a greater degree of satisfaction.

This can further be extended by integrating it with an e-commerce platform to enable the provision of dynamic results. Additionally, the sales data can be analysed for understanding the other contributing factors towards effective pricing.

Data availability The data set used for the model implementation is available at kaggle named as E-Commerce_Participants_Data. it contains 2453 records of training data and 1052 testing data.

Declarations

Conflict of interest The authors have no conflicts of interest in any matter related to the paper.

References

1. Alfaro-Navarro JL, Cano EL, Alfaro-Cortés E, et al (2020) A fully automated adjustment of ensemble methods in machine learning for modeling complex real estate systems. *Complexity* 2020
2. AlJazzazen SA (2019) New product pricing strategy: skimming vs. penetration. In: *Proceedings of FIKUSZ symposium for young researchers*, Óbuda University Keleti Károly Faculty of Economics, pp 1–9
3. Andrius G, Vaida P, Alina S (2021) Predictive analytics using big data for the real estate market during the covid-19 pandemic. *J Big Data* 8(1)
4. Bentéjac C, Csörgő A, Martínez-Muñoz G (2021) A comparative analysis of gradient boosting algorithms. *Artif Intell Rev* 54(3):1937–1967
5. Bürgin D, Wilken R (2021) Increasing consumers' purchase intentions toward fair-trade products through partitioned pricing. *J Bus Ethics* 1–26
6. Davcik NS, Sharma P (2015) Impact of product differentiation, marketing investments and brand equity on pricing strategies: a brand level investigation. *Eur J Mark*

7. Enberg J (2021) Covid-19 concerns may boost ecommerce as consumers avoid stores
8. Gupta R, Pathak C (2014) A machine learning framework for predicting purchase by online customers based on dynamic pricing. *Procedia Comput Sci* 36:599–605
9. Hinterhuber A, Kienzler M, Liozu S (2021) New product pricing in business markets: the role of psychological traits. *J Bus Res* 133:231–241
10. Ho WK, Tang BS, Wong SW (2021) Predicting property prices with machine learning algorithms. *J Prop Res* 38(1):48–70
11. Jabeur SB, Mefteh-Wali S, Viviani JL (2021) Forecasting gold price with the xgboost algorithm and shap interaction values. *Ann Oper Res* 1–21
12. Li X, Xu M, Yang D (2021) Cause-related product pricing with regular product reference. *Asia Pac J Oper Res* 38(04):2150,002
13. Mattos AL, Oyadomari JCT, Zatta FN (2021) Pricing research: state of the art and future opportunities. *SAGE Open* 11(3):21582440211032,168
14. Mohamed MA, El-Henawy IM, Salah A (2022) Price prediction of seasonal items using machine learning and statistical methods. *CMC* 70(2):3473–3489
15. Pérez Rave JI, Jaramillo Álvarez GP, González Echavarría F (2021) A psychometric data science approach to study latent variables: a case of class quality and student satisfaction. *Total Quality Manag Bus Excel*
16. Priester A, Robbert T, Roth S (2020) A special price just for you: effects of personalized dynamic pricing on consumer fairness perceptions. *J Rev Pricing Manag* 19(2):99–112
17. Punmiya R, Choe S (2019) Energy theft detection using gradient boosting theft detector with feature engineering-based preprocessing. *IEEE Trans Smart Grid* 10(2):2326–2329
18. Ren M, Liu J, Feng S, et al (2020) Complementary product pricing and service cooperation strategy in a dual-channel supply chain. *Discrete Dynamics in Nature and Society* 2020
19. Shahrel MZ, Mutalib S, Abdul-Rahman S (2021) Pricecop-price monitor and prediction using linear regression and lsvm-abc methods for e-commerce platform. *Int J Inf Eng Electron Bus* 13(1)
20. Spuritha M, Kashyap CS, Nambiar TR et al (2021) Quotidian sales forecasting using machine learning. In: 2021 International conference on innovative computing. Intelligent communication and smart electrical systems (ICES). IEEE, pp 1–6
21. Sun X, Liu M, Sima Z (2020) A novel cryptocurrency price trend forecasting model based on lightgbm. *Finance Res Lett* 32(101):084
22. Wei Y, Li F (2015) Impact of heterogeneous consumers on pricing decisions under dual-channel competition. *Math Probl Eng* 2015
23. Wiyati R, Dwi Priyohadi N, Pancaningrum E, et al (2019) Multifaceted scope of supply chain: evidence from Indonesia. *International Journal of Innovation, Creativity and Change* www.ijicc.net 9(5)
24. Yang M, Xia E (2021) A systematic literature review on pricing strategies in the sharing economy. *Sustainability* 13(17):9762
25. Yu Z (2021) Data analysis and soybean price intelligent prediction model based on lstm neural network. In: 2021 IEEE conference on telecommunications, optics and computer science (TOCS). IEEE, pp 799–801

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.