Price Recommendation For E-Commerce Using MLTechniques

Mohit Tiwari[1] *Department of Computer Science and Engineering, Sathyabama Institute of Science and Technology ,* Semmancheri,Chennai 600119, TamilNadu,India,
mohittiwari9051@gmail.com[1]

Aryan Kumar[2], *Department of Computer Science and Engineering ,Sathyabama Institute of Science and Technology ,* Semmancheri,Chennai 600119, TamilNadu,India,
aryanrajpu2u@gmail.com[2]

Dr.B.Bharathi[3], *Dean of Academics, Department of Computer Science and Engineering,Sathyabama Institute of Science and Technology ,* Semmancheri,Chennai 600119, TamilNadu,India,
bharathi.cse@sathyabama.ac.in[3]

2024 Intelligent Systems and Machine Learning Conference (ISML) | 979-8-3503-4387-8/24/$31.00 ©2024 IEEE | DOI: 10.1109/ISML60050.2024.11007352

***Abstract*- In online commerce, pricing decisions play a critical role in determining the success and profitability of businesses, Leveraging advanced machine learning algorithms and robust data analytics, the system analyzes diverse parameters including historical sales data, market trends, competitor pricing strategies, and customer behavior patterns. By integrating these multifaceted insights, the system generates dynamic and data-driven price recommendations. These recommendations empower online sellers to optimize pricing strategies, strike a balance between competitiveness and profitability, and cater to evolving market demands. The project aims to create an algorithm for automatic price suggestions in E-commerce, leveraging user-inputted text descriptions, product category details, brand information, and item conditions. The objective is to enhance the user experience by providing pricing guidance to both sellers and buyers on an online platform. Using a supervised learning approach, the project focuses on predicting optimal product prices, employing regression models evaluated using Root Mean Squared Error (RMSE). Additionally, considering the potential impact of user-related features like location, gender, and login information, the study explores their influence on pricing accuracy. By Preprocess multilingual text data, facilitating better language recognition and categorization for more accurate pricing suggestions. The project's outcomes aim to improve transactional efficiency within the E-commerce domain by providing users with informed pricing insights before making purchasing or selling decisions.**

Keywords- Document, Tokens, Corpus, Term Frequency (TF), Inverse DocumentFrequency (IDF), ML techniques

I. INTRODUCTION

Online sellers face a myriad of challenges in determining optimal pricing strategies that attract customers while ensuring profitability. To focus on developing a sophisticated price recommendation system designed to assist online sellers in making informed decisions. These factors include the quality of the information offered, the intended audience, the degree of market competition, and the pricing schemes employed by competing goods and services[1]. Since the report's exact details are not given to provide some generic pricing advice:

1. Market research: Look for comparable papers or services in specialized market and compare their prices. Think about the report's depth, quality, and uniqueness about the competitors.

2. Value Proposition: Identify the advantages and special selling aspects of the report. Emphasize the details, observations, or tactics that make it unique. This will support charging more.

3. Consider the target audience's purchasing power and financial situation. Setting a lower price point can be necessary if they are price-sensitive. A greater price can be appropriate, though, if the research serves a niche or premium market.

4. Cost of Production: Determine how much time, effort, and money were used to produce the report. Don't forget to take into account any study, data analysis, or specialist knowledge that went into its creation.

5. Price Plan: Pick a price plan that supports company's objectives. The option of using market-based pricing, which matches or significantly undercuts competitors' prices, cost-plus pricing, or value-based pricing, which bases the price on the perceived value of the product.

6. Test and Adjust: To get feedback and gauge market reaction, think about launching with a low price or providing discounts. Keep an eye on sales and client comments, and change the price approach as necessary.

Setting the appropriate prices is essential for online vendors to draw clients, increase profits, and maintain their position as leaders in the online market. Finding the ideal pricing points, however, might be difficult. Price recommendations become important in this situation. Online retailers can use price recommendations to gain important knowledge and direction when deciding on their pricing strategies. Price advice tools give vendors a thorough view of the competitive landscape by combining market data, competition analysis, and customer behavior patterns[2]. With the aid of these tools, vendors may determine the best price range for their goods or services while taking into account a variety of aspects like costs, value proposition, target market, and demand. Online vendors may find a balance between profitability and competitiveness thanks to effective price recommendations. They help vendors set prices that represent the perceived worth of their goods and services while also covering their production and operating expenses. Price recommendations also assist sellers in positioning their goods in the market effectively, whether they want to appeal to buyers who are price-sensitive or highlight the superior qualities and advantages of their goods.

It's crucial to remember that price recommendations are guidance based on data-driven insights rather than rigid restrictions. Always take into account of the own business needs when modifying the advice to fit the overall business plan. In Fig.1 given the general architecture flow. In the ever-changing internet market, maintaining the pricing plan requires regular monitoring, review, and adjustments by utilizing price recommendations.

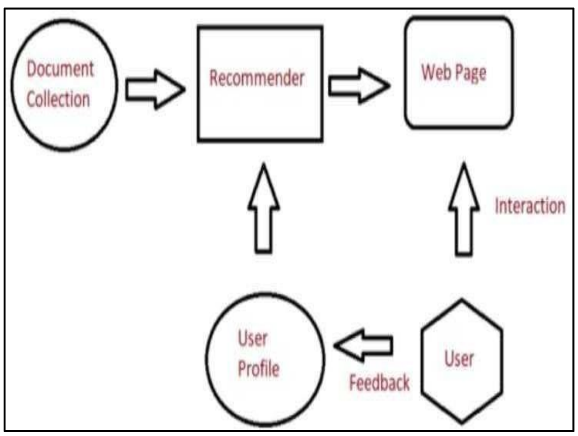

Fig 1. (Layout Model)

II. RELATED WORK

Smart Strategies: How Machine Learning Reshapes Online Selling Dynamics

The utilization of machine learning in E-commerce has evolved significantly, tracing its origins to early recommendation systems like collaborative filtering and content-based filtering. These systems laid the foundation for personalized product recommendations, helping users discover items aligned with their preferences. Over time, advancements in algorithms, data availability, and computational power have revolutionized the E-commerce landscape. The evolution of machine learning in E-commerce involves a blend of research and practical applications. Initially, models primarily focused on product recommendations, employing techniques like matrix factorization, nearest-neighbor approaches, and basic classification algorithms. As data volumes surged and computational capabilities improved, more sophisticated algorithms emerged. Deep learning models, such as neural networks, began to dominate, allowing for more complex feature extraction and representation learning from unstructured data like images and text. Moreover, research in natural language processing (NLP) significantly impacted E-commerce. Sentiment analysis, named entity recognition, and semantic understanding enhanced search relevance, improved product descriptions, and facilitated better customer interactions. Furthermore, reinforcement learning gained attention in pricing strategies and dynamic pricing optimization, enabling E-commerce platforms to adjust prices in real time based on market conditions, demand, and competitor pricing. Research efforts in E-commerce and machine learning continue to expand. Academia and industry collaborate to explore novel applications, such as recommendation systems leveraging graph neural networks, interpretable machine learning models for a better understanding of user behavior, and AI-driven chatbots for enhanced customer service. Additionally, the integration of machine learning with other technologies, like computer vision for visual search and augmented reality, reshapes the online shopping experience[3]. The evolving landscape of machine learning in E-commerce underscores a convergence of diverse research areas. It's an amalgamation of advancements in recommendation systems, natural language processing, reinforcement learning, and more, all aimed at enhancing customer experiences, optimizing operations, and driving business growth in the competitive realm of online commerce. The integration of machine learning in E-commerce began to take shape in the early 2000s, marked by the emergence of recommendation systems as one of the foundational applications. This period witnessed the inception of collaborative filtering and content-based filtering techniques, laying the groundwork for personalized product recommendations in online retail. Companies like Amazon and Netflix were pioneers in implementing these systems to suggest items or movies based on users' historical preferences and behavior. As data availability expanded and computational capabilities advanced, machine learning algorithms in E-commerce evolved[4]. More sophisticated approaches, including matrix factorization, nearest neighbor methods, and basic classification algorithms, began to enhance recommendation systems' accuracy and scalability. Furthermore, advancements in natural language processing (NLP) and computer vision technologies further enriched E-commerce applications. NLP techniques improved search relevance, sentiment analysis, and semantic understanding of product descriptions, while computer vision facilitated visual search and image-based recommendations. The early stages of machine learning in E- commerce set the stage for personalized user experiences, optimizing product discovery, and fundamentally altering the way online platforms interacted with consumers[5]. This period laid the groundwork for subsequent advancements and innovations in the integration of machine learning algorithms across various facets of online retail. ML algorithms optimize logistics, transportation routes, and warehouse management, reducing costs and enhancing delivery efficiency.

Automatic Language Recognition

Automatic Language Recognition (ALR) and text classification are vital components facilitated by techniques like numerical statistics used in information retrieval and text mining to evaluate the importance of a word in a document relative to a collection of documents or a corpus. ALR involves identifying*, and categorizing the language of text data, essential when dealing with user descriptions in various languages. TF-IDF, a technique used in text analysis, assigns weights to words based on their frequency in a document relative to a corpus, aiding in language recognition by highlighting distinctive terms across languages. Text classification, a broader concept, encompasses tasks like sentiment analysis, topic modeling, and language identification[6]. Text classification might involve categorizing product descriptions into specific classes or categories (e.g., electronics, clothing, books) based on their content. The patterns from labeled data to classify or categorize unseen text accurately. For language recognition within the E-commerce context, TF-IDF can help preprocess and represent text data by capturing language-specific features, allowing subsequent classification models to effectively differentiate between languages[7]. By employing text classification techniques, the system can intelligently process and categorize product descriptions, enabling better organization, search relevance, and user experience within the E-commerce platform. Overall, these methods of ALR, TF-IDF, and text classification play pivotal roles in understanding, organizing, and optimizing the handling of multilingual textual data within the project's framework, enhancing the overall functionality and usability of the E-commerce platform. Additionally, text classification techniques like TF-IDF assist in preprocessing textual information by assigning weights to words based on their relevance within individual descriptions, aiding in language recognition and subsequent categorization. Machine learning models trained on TF-IDF representations can effectively distinguish between languages or classify product descriptions into specific categories, enabling accurate suggestions for pricing[8].

Analysis of market basket

In the context of suggesting product prices based on user-inputted text descriptions within the E-commerce domain, adapting MBA techniques to textual data presents both opportunities and challenges[9]. Traditionally, MBA analyzes transactional data to identify patterns in customers' purchasing behavior. However, The adaptation involves exploring text-based associations within user descriptions, aiming to extract implicit connections among various product attributes. By scrutinizing the co-occurrences of product categories, brand mentions, and item conditions within these textual descriptions, MBA seeks to unveil underlying relationships and preferences that could influence pricing decisions. One of the primary advantages of employing an MBA in this context is its potential to unravel implicit associations within unstructured textual data. By identifying frequent combinations of attributes in user descriptions, an MBA aids in understanding potential affinities between specific brands and product categories or correlations between item conditions and pricing preferences[10]. These inferred relationships offer valuable insights into user preferences, potentially guiding pricing strategies and enhancing the overall customer experience by tailoring pricing recommendations to align with inferred user preferences[11]. However, significant challenges arise when applying an MBA solely to textual inputs. The absence of explicit transactional records limits the depth and accuracy of association mining. Quantifying the strength of associations purely from text introduces complexities in representing nuanced relationships, potentially limiting the completeness and accuracy of insights derived from an MBA. Moreover, textual data's inherent complexity and variability pose challenges in accurately capturing and quantifying associations, thereby impacting the reliability of insights generated solely from text-based analysis[12].

Ref.no	**Algorithm**	**Data Preprocessing**	**Accuracy**
1.	Corpus	Yes	87.1%
2.	Count Vectorizer	Yes	88%
3.	Tfidf Vectorizer	Yes	80%
4.	LabelBinarizer	Yes	83%
5.	Cross Validation	Sparse Matrix	97%
6.	BGM	RMSLE Techniques	96.50%
7.	Ridge Regression	RMSLE Techniques	98.87%

Table 1 - Summary For Comparison

III. METHODS

The project's core aims revolved around refining the pricing strategy for resale products on an e-commerce platform. The focus was on ensuring consistency in product prices, considering all product attributes for accurate price predictions. A standardized pricing structure aimed to leverage more referral opportunities while guiding customers toward purchasing new resale items. Ultimately, these strategies aimed to bolster sales for the online store offering resale goods.

Data Collection and Description: The data contained comprehensive information: 7 columns and 1,482,535 rows of product data. Preprocessing steps encompassed handling missing values, converting categorical values, and employing Natural Language Processing (NLP) techniques to process textual data.

Developing a methodology for price recommendation in e-commerce, integrating LightGBM (Gradient Boosting) and Ridge Regression, involves a comprehensive approach to leverage the strengths of both algorithms for optimal results. The first step in this methodology is data preprocessing, where raw data is cleaned, normalized, and transformed to ensure its suitability for model training. Feature engineering plays a crucial role, as it involves selecting and creating relevant features that contribute to pricing decisions. Categorical variables may be one-hot encoded, and continuous variables may be scaled to standardize their values.

Next, the data is split into training and validation sets to assess the model's performance. LightGBM, a gradient-boosting framework, is employed for its efficiency and speed in handling large datasets. This algorithm is particularly effective in capturing complex relationships within the data and provides accurate predictions. During the training phase, hyperparameter tuning is crucial to optimize the model's performance. Grid search or random search can be employed to find the best combination of hyperparameters.

After the initial model is trained, Ridge Regression is introduced as a complementary technique. Ridge Regression helps prevent overfitting by adding a regularization term to the cost function, which penalizes large coefficients. This is especially valuable when dealing with high-dimensional data, as it encourages the model to generalize well to new, unseen data. The Ridge Regression model is trained on the same dataset, and its predictions are combined with those of the LightGBM model.

The ensemble of the LightGBM and Ridge Regression models is created by combining their predictions through weighted averaging. The weights assigned to each model can be determined through cross-validation or based on domain expertise. The combined model leverages the strengths of both algorithms, resulting in a more robust and accurate price recommendation system.

Once the ensemble model is established, it undergoes thorough evaluation using metrics such as Root Mean Squared Error (RMSE) on the validation set. This step ensures that the model performs well on unseen data and provides reliable price recommendations. Further refinement can be achieved by iteratively adjusting hyperparameters and features based on the evaluation results. The final step involves deploying the model to the e-commerce platform for real-time price recommendations. Continuous monitoring and updating of the model are essential to adapt to changing market dynamics and consumer behavior. Regular retraining of the model with new data ensures that it remains accurate and effective over time.

A binning column was added due to the skewed price by taking into account the price's interquartile range, which places values between 3 and 10 in q1. Q2 and so on will have values that are larger than 2 and less than 17. Additionally, the q4's range is rather wide, which causes the data in Figure 5 to be skewed. The Distribution of price in binning is given in Fig.2.

	count	mean	std	min	25%	50%	75%	max
price_bin								
q1	375615.0	7.715192	2.077888	3.0	6.0	8.0	10.0	10.0
q2	378177.0	13.842940	1.794584	10.5	12.0	14.0	15.0	17.0
q3	359743.0	22.555694	3.337832	17.5	20.0	22.0	25.0	29.0
q4	368123.0	63.527701	63.508250	29.5	35.0	45.0	66.0	2000.0

Fig 2. Distribution of price in binning

There were a total of 4809 unique values for a brand name feature. The two brands with the highest sales were Pink and Nike. As seen in the figure, the brand name was absent for the majority of the products. In Fig .3 Different Brand Analysis is given.

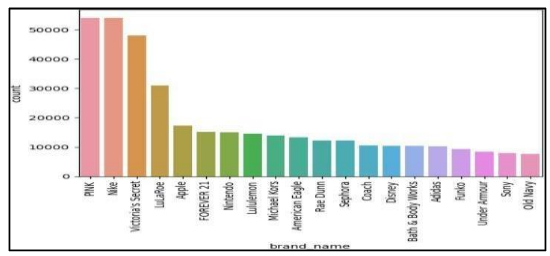

Fig 3 Brand Analysis

To create tokens, all of the data from the item description and name was first converted using natural language processing.

Recommendation

The process begins with the continuous collection of user interactions and relevant contextual data as users navigate through an online platform. This data includes but is not limited to clicks, searches, purchases, and dwell times on product pages. This constant influx of real-time data is crucial for maintaining the accuracy and relevance of recommendations. As users engage with the platform, the recommendation system processes and updates user profiles dynamically. These profiles evolve based on the latest interactions and preferences, ensuring that recommendations stay current. It plays a pivotal role in analyzing this real-time data. The algorithms adapt and learn from the changing patterns of user behavior, making them highly responsive to shifts in consumer preferences and market trends.

Simulation Prognosis

LGBM: LightGBM's capability to handle categorical features and its native support for efficient handling of sparse data aligns well with the nature of textual data often encountered in E-commerce. By efficiently handling high-dimensional and sparse feature spaces, LightGBM can effectively capture complex relationships and patterns present in user descriptions, enabling accurate price predictions. Its ability to handle large-scale data ensures scalability, crucial in dealing with diverse product catalogs and user-generated content in an E-commerce platform. Moreover, LightGBM's boosting technique allows it to sequentially learn from its mistakes, iteratively improving predictive performance, which is vital for refining pricing suggestions in an evolving market environment.

Ridge regression: RidgeRegression's regularization term helps mitigate the impact of multicollinearity among features, allowing for more stable estimates of coefficients and reducing model sensitivity to outliers or noise present in the data. This aspect is particularly beneficial when dealing with text-derived features where the presence of correlated variables is common. By penalizing large coefficients, Ridge Regression promotes smoother and more generalized models, thus contributing to better generalization and performance on unseen data.

Moreover, Ridge Regression's ability to handle high-dimensional data efficiently aligns with the nature of textual descriptions often encountered in E-commerce platforms. It aids in managing the sparsity and complexity of textual features, providing a robust framework to incorporate textual information into the pricing prediction model. While Ridge Regression might not capture intricate non-linear relationships as complexly as some other algorithms, its simplicity, interpretability, and regularization properties make it a valuable candidate for incorporating textual features and improving the accuracy of price predictions within the E-commerce setting.

Model Assessment
Prognosis

The weights for evaluating a suggestion model are r-squared and root mean squared error(rmse). The models underwent rigorous tuning and optimization by adjusting hyperparameters and exploring different algorithms to enhance their predictive accuracy and recommendation capabilities. Model performance was iteratively analyzed, and adjustments were made based on the assessment metrics to ensure the best possible outcomes. In conclusion, the model assessment phase encompassed a comprehensive analysis using various evaluation metrics, cross-validation techniques, and iterative refinement to gauge the prediction accuracy and recommendation effectiveness of the Price Recommendation System. In Fig.4 Design flow is given

.

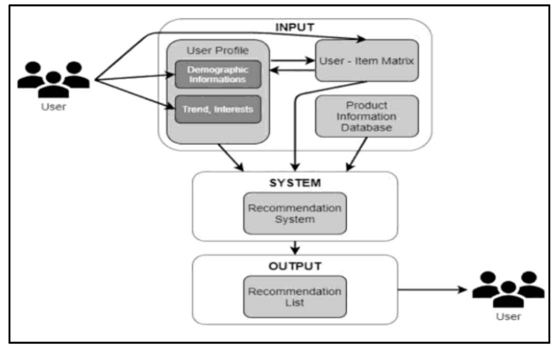

Fig 4. Design Architecture

IV. CONCLUSION

For e-commerce websites that sell secondhand goods, this article implements two things: price prediction and product recommendation. The algorithms chosen are those that are less CPU-intensive given the amount of the data. Two techniques are used to estimate the price. Ridge regression and LGBM were performed using conventional NLP methods. By about 3%, ridge regression outperformed LGBM. the most common way for textbook pre-processing and this allowed me to prepare myself for unborn work whenever textbook data again. Another concept to value further is the choice of algorithms and how important calculation is whenever dealing with large datasets.Text data is far and wide and it can get messy. Understanding the fundamentals of how to attack these problems will surely help me out in the future.

REFERENCES

[1] Bauer, J., & Jannach, D. (2018). Optimal pricing in e-commerce based on sparse and noisy data. Decision Support Systems, 106, 53-63. URL:http://www.sciencedirect.com/science/article/pii/S016792361730221X

[2] Cakir, O., & Aras, M. E. (2012). A recommendation engine by using association rules. Procedia - Social and Behavioral Sciences, 62, 452-456. World Conference on Business, Economics and Management (BEM-2012), May 4-6, 2012, Antalya, Turkey. URL:http://www.sciencedirect.com/science/article/pii/S187704281203515X

[3] Chen, C., Zhang, Q., Ma, Q., & Yu, B. (2019). LightGBM-PPI: Predicting protein-protein interactions through LightGBM with multi-information fusion. Chemometrics and Intelligent Laboratory Systems, 191, 54-64. URL:http://www.sciencedirect.com/science/article/pii/S016974391930262X

[4] Peng, L., Zhang, W., Wang, X., & Liang, S. (2019). Moderating effects of time pressure on the relationship between perceived value and purchase intention in social e-commerce sales promotion: Considering the impact of product involvement. Information & Management, 56(2), 317-328. Social Commerce and Social Media: Behaviors in the New Service Economy. URL: http://www.sciencedirect.com/science/article/pii/S0378720617305712

[5] Petersen, N. C., Rodrigues, F., & Pereira, F. C. (2019). Multi-output bus travel time prediction with convolutional LSTM neural network. Expert Systems with Applications, 120, 426-435. URL: http://www.sciencedirect.com/science/article/pii/S0957417418307486

[6] Tripathy, A., Agrawal, A., & Rath, S. K. (2015). Classification of sentimental reviews using machine learning techniques. Procedia Computer Science, 57, 821-829. 3rd International Conference on Recent Trends in Computing 2015 (ICRTC-2015). URL: http://www.sciencedirect.com/science/article/pii/S1877050915020529

[7] Tripathy, A., Agrawal, A., & Rath, S. K. (2016). Classification of sentiment reviews using n-gram machine learning approach. Expert Systems with Applications, 57, 117-126.

URL:

http://www.sciencedirect.com/science/article/pii/S095741741630118X

[8] Valle, M. A., Ruz, G. A., & Morris, R. (2018). Market basket analysis: Complementing association rules with minimum spanning trees. Expert Systems with Applications, 97, 146-162.
URL:http://www.sciencedirect.com/science/article/pii/S0957417417308503

[9] Greenstein-Messica, A., & Rokach, L. (2018). Personal price aware multi-seller recommender system: Evidence from eBay. Knowledge-Based Systems, 150, 14-26.
URL:http://www.sciencedirect.com/science/article/pii/S0950705118300893

[10] Guo, J., Gao, Z., Liu, N., & Wu, Y. (2018). Recommend products with consideration of multi-category inter-purchase time and price. Future Generation Computer Systems, 78, 451-461.
URL:
http://www.sciencedirect.com/science/article/pii/S0167739X1730256X

[11] Jiang, Y., Shang, J., Liu, Y., & May, J. (2015). Redesigning promotion strategy for e-commerce competitiveness through pricing and recommendation. International Journal of Production Economics, 167, 257-270.URL:http://www.sciencedirect.com/science/article/pii/S092552731500208X

[12] Kaur, M., & Kang, S. (2016). Market basket analysis: Identifying changing trends in market data using association rule mining. Procedia Computer Science, 85, 78-85. International Conference on Computational Modelling and Security (CMS 2016).
URL:http://www.sciencedirect.com/science/article/pii/S1877050916305208