2025 International Conference on Circuit Power and Computing Technologies (ICCPCT)

Social Media Aided E-Commerce Dynamic Pricing System using GPT-3.5 and XGB Regressor

M Abernakumari
Department of AI & DS
V.S.B College of Engineering Technical Campus, Coimbatore-642109, India
Abernakumari87@gmail.com

P.Venkadesh
Department of AI & DS
V.S.B College of Engineering Technical Campus, Coimbatore-642109, India
pvenkadesh2002@gmail.com

S.V.Divya
Department of CSE
V.S.B College of Engineering Technical Campus, Coimbatore-642109, India
divyasvvsb@gmail.com

C. Irfan
Department of AI & DS
V.S.B College of Engineering Technical Campus, Coimbatore-642109, India
itskickirfan@gmail.com

A F Godwin
Department of AI & DS
V.S.B College of Engineering Technical Campus, Coimbatore-642109, India
godwinspidy@gmail.com

S Roshan Kumar
Department of AI & DS
V.S.B College of Engineering Technical Campus, Coimbatore-642109, India
roshankumar9126419@gmail.com

2025 8th International Conference on Circuit, Power & Computing Technologies (ICCPCT) | 979-8-3315-4317-4/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICCPCT65132.2025.11176660

***Abstract*— At present, e-marketplaces are more competitive than ever. Most e-commerce service providers and sellers aim to personalize their services to engage their customers and make them feel valued. In the constantly changing environment, a seller has to be quick in reacting to market conditions, especially in pricing their products. Current dynamic pricing models leverage market and competition information, inventory and demand logistics to determine prices, and often lack the ability to track social trends that influence demand. However, these methods take more resources and time than machine learning models. Currently, over 70% of the products only apply dynamic pricing for one year. Our proposed method uses machine learning techniques, including ensemble learning and boosting, and the GPT-3.5 model to effectively model tweet trends and predict demand. This novel approach only uses social trends without customer data. This paper explores the technical implementation of a dynamic pricing system designed to optimize resource utilization for e-commerce. This method achieved a relative RMSE of 1.69% and a relative MAE of 1.11%.**

Keywords—Dynamic pricing, stacking ensemble, GPT-3.5, XGB Regressor, E-commerce pricing

I. Introduction

E-commerce has become an important aspect of our lifestyle. Global e-commerce sales are predicted to reach $8 billion US dollars, making it highly critical to develop and optimize performance. Retail e-commerce allows a variety of items from various categories to be available in the global marketplace. However, the heterogeneous nature of different categories of items available in the market makes it difficult to apply numerous models. While most of the tools for e-commerce, like recommendation systems, personalized systems, and supply chain management, work efficiently, the pricing of the products still feels outdated. Pricing for retail goods based on real-time events and demands is critical for achieving high revenue and customer satisfaction. Dynamic pricing in e-commerce often reflects inventory status, real-time demand, competitor price, etc. However, it does not effectively capture the trend from social media. N. Alamsyah, Saparudin, and A. P. Kurniati, in their work on optimizing event detection for airplane dynamic pricing strategy, developed social event-based pricing that works effectively to identify future demand using tweets [1]. They perform tweet labeling using a stacking ensemble but have selected keywords manually. In their case, social events directly impact price. In e-commerce, the underlying model works similarly to detect events in real time from social media and use them effectively to determine the ticket price. The key is to identify the keywords that are associated with social trends and the product. It will be difficult to identify keywords that work for all products. That is why it is best to break the products into categories such as electronics, fashion, furniture, etc., or into products like smartphones, TVs, mattresses, etc. Utilizing social media trends enables identifying trends based on events and occurrences. These are often missed by traditional time series analysis or seasonal forecasting. It helps predict future demands and customize prices based on the demand. This study focuses on complementing the traditional dynamic pricing systems in e-commerce by providing social media insights that are product-specific to deliver a pricing strategy tailored to capture both static and dynamic details of e-commerce sales. This work utilizes only keyword trends and is devoid of any customer data, building an ethical pricing system unique in this regard.

The rest of the paper is as follows: Section II reviews the relevant works; Section III describes the proposed methodology for pricing; In Section IV we discuss our results; Finally, Section V summarizes the work and discusses future development areas.

II. Literature Review

Yamuna et al. examine dynamic pricing in online retail using a framework based on machine learning. They developed various models to predict optimal prices by evaluating user behavior and product characteristics, underlining the flexibility of AI in markets [2]. Likewise, Nowak and Pawłowska-Nowak compared various machine learning models for dynamic pricing in e-commerce, highlighting

979-8-3315-4317-4/25/$31.00 ©2025 IEEE

enhanced profitability achieved through precise price predictions [3]. These models work on retail, selecting the best pricing strategies and focusing on increased revenue generation, which creates several problems like price rocketing.

Padmanaaban et al. introduce an AI-driven dynamic pricing model aimed at optimizing revenue for e-commerce platforms by customer segmentation [4]. Their system combines real-time data with intelligent algorithms to adjust prices dynamically, demonstrating its scalability across various industries. This model shows existing models giving low results for less pricey products.

J. Xu, Y.-C. Hsu and W. Biscarri build on discoveries by applying dynamic pricing in finance, focusing on securities lending portfolios and utilizing historical trading data along with algorithmic tactics [5]. This shows contextual bandits outperforming rule-based and ML-based models.

The next four studies focus on existing RL-based strategies. M. Apte, K. Kale, P. Datar, and P. Deshmukh present a Q-learning framework for retail pricing, demonstrating enhanced revenue outcomes through continuous learning from transaction data [6]. This creates ethical problems due to the use of customer data and is not available in every area.

In energy management, C. Tang, Y. Qin, F. Wu, and Z. Tang introduce a pricing model for power grids that uses deep reinforcement learning to adjust prices in response to changing demand [7]. Their approach highlights the capabilities of deep reinforcement learning in areas sensitive to infrastructure changes. D. Watari, I. Taniguchi, and T. Onoye incorporate a model-free deep reinforcement learning technique to improve the duck curve and handle peak time issues, improving the peak to average ratio by 23% [8]. Similarly, an LSTM-supported actor-critic RL model using real-time pricing data to predict price and demand for energy in near real-time has been introduced by Ismail and Baysal [9]. These RL-based models require a lot of resources and are not suitable for small businesses. These solutions are solely made for tackling a single problem, making it not suitable for general adaptations.

The next two papers use LSTM-based strategies. Y. Liu, C. Yang, K. Huang, and W. Liu introduce a novel pricing strategy suitable for multivariate selection and fusion able to handle noisy financial data using a hybrid CNN-LSTM network, avoiding complex network structure for forecasting prices [10]. Gadde et al. performed a comparative analysis on various RL techniques, identifying critical challenges such as data requirements, ethical concerns and computational cost [11].

The rest of the review papers focus on ethical and systematic perspectives on pricing. An analysis by X. Wu, J. Qin, W. Qu, Y. Zeng, and X. Yang on China's high-speed railways strategy for determining ticket prices, setting price ceilings and seat allocation produced improved revenue compared to fixed prices [12]. A case study performed by M. Basal, E. Saraç, and K. Özer examined AI-based dynamic pricing framework across industries, emphasizing the need for adaption for smaller companies also states that AI's influence on demand is to be studied [13]. These studies show the adaptation of dynamic pricing by various industries.

A literature review conducted by Chenavaz and Dimitrov on the applications of AI in dynamic pricing highlights the current focus on topics such as agent-based consumption, crude-oil and price [14]. M. S. Gazi, M. R. Hasan, N. Gurung, and A. Mitra, in their recent research on AI-based pricing in the US market, developed a transparent and fair dynamic pricing model using ML models that maximize profit by learning from model features [15]. These studies emphasize the need for a fair and ethical pricing algorithm that doesn't solely use customer data to determine product prices.

This shows that existing models require high computational resources, customer data and mostly domain specific, making it difficult to apply to small businesses. The proposed model utilizes social media data and is computationally feasible, making it adaptable for small-scale businesses.

III. Research Methodology

The social media analysis for e-commerce product pricing involves identifying relevant keywords that affect the sales of that particular product. Selecting keywords manually is not suitable for e-commerce due to its heterogeneous nature. Various tweet datasets from Kaggle are used to create a dataset consisting of tweets related to TVs, television, electronics and screens from January 2023 to June 2024, particular to India, using general keywords related to the products. The price of TVs over the period is added to the dataset based on the date (source: pricebefore.com).

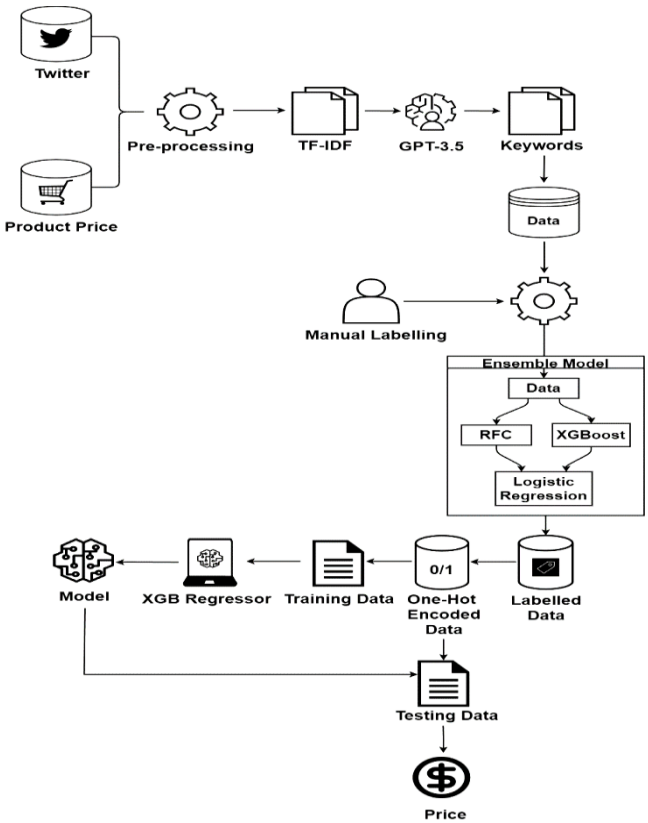

Fig. 1. The proposed model for dynamic pricing using GPT-3.5, stacking ensemble and ensemble model

The proposed model involves keyword selection using GPT-3.5, tweet labeling using a stacking ensemble model and

979-8-3315-4317-4/25/$31.00 ©2025 IEEE

product pricing using XGB Regressor. The architecture is shown in Fig. 1.

Twitter-specific text processing analyzed by D. Ramachandran and R. Parvathi is referred [16]. Using regular expression '#', '@' and 'http/s' are removed. Emojis are replaced with short codes and removed. The text is tokenized to compare with English stopwords from NLTK and returned without any stopwords. Then the tokens are grouped based on need by N-grams technique. The data is then converted into a TF-IDF matrix to assign a statistical value for each word based on the uniqueness of the word in the collection using simple arithmetic operations. P. He, J. Huang and M. L performed text extraction using GPT and produced improved results [17]. The highly significant 50–60 words are selected. A scenario is created as a product analyst analyzing keywords' effectiveness on a product, ranking it from most to least relevant in Indian market conditions, along with a small description of the top 10 keywords. The result is then further prompted to highlight the nature of the usage of the keyword in social media. On the top keywords, Pearson correlation is applied for validation against price history and the most relevant keywords are selected. The flow is shown in Fig. 2.

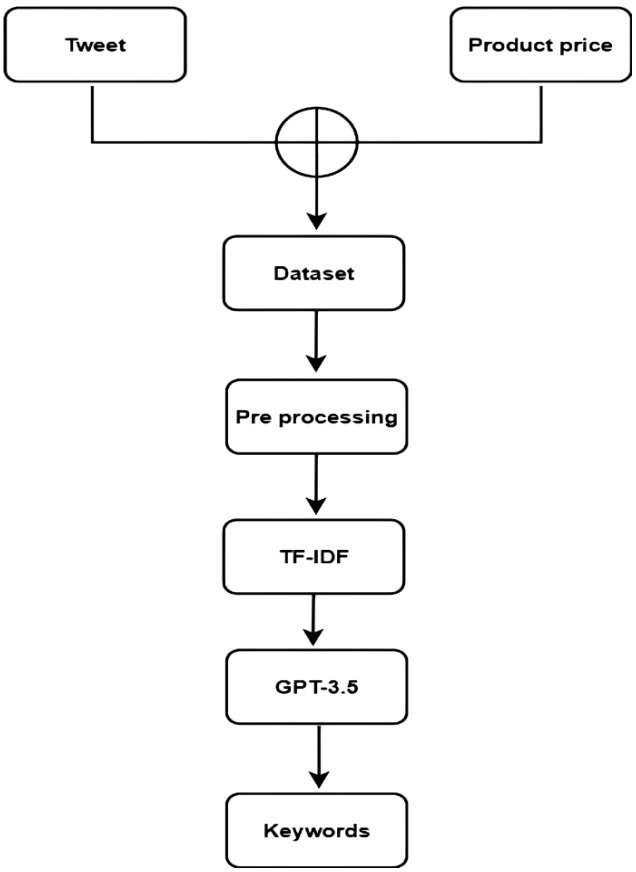

Fig. 2. Data integration and keyword selection using GPT-3.5

A. Keyword selection using GPT

The algorithm for Keyword Selection is shown below

Algorithm 1 Proposed Keyword Selection

Input: historical_tweets, price_history
Output: keywords

1. Data pre-processing(historical_tweets)
2. Removing stop-words, URLs, and emojis using NLTK
3. Perform tokenization
4. Perform N-grams
5. return historical_tweets
6. Merge historical_tweets with price_history as data
7. Keywords extraction(data)
8. Apply TF-IDF transformation
9. Sort the data in descending
10. return top 50 keywords as top_keys
11. Keywordselection(top_keys)
12. Perform relevancy analysison top_keys using GPT-3.5
13. Record relevancy score and frequency of words
14. Apply Pearson Correlation

$$\frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\left(\sqrt{\Sigma(x_i - \bar{x})^2}\right)\left(\sqrt{\Sigma(y_i - \bar{y})^2}\right)}$$

x = word frequency/sentiment per day
y = price per day

15. return top 7-10 keywords as keywords

The model identified various keywords that are general and some specific keywords that contribute to the price trend of TV. The data is collected from Indian region tweets so the result primarily consists of keywords of Indian social media trend. The terms that are selected from the dataset include 'ipl', 'festival', 'bigg boss', 'great Indian sale', 'big billion days', 'price hike', other general terms like 'tech deals', 'smart deals', etc., are omitted. After identifying the base keywords, the dataset is manually labeled.

B. Tweet labeling using stacking ensemble

The text is then converted into numerical feature representation by using TF-IDF technique.This results in a feature matrix that holds higher value for unique and rare words and lower value for common words. The vector representation of the text is suitable for text labeling. The data was labeled into 'ipl', 'festival', 'bigg boss', 'great indian sale', 'big billion days', 'price hike', and 'others'. The text is then split into test-train data frames.This data is then fed to a stacking ensemble model. It uses Random Forest Classifier(RFC) and XGBoost(XGB) as base classifiers and Logistic Regressor(LR) as meta classifier. The RFC model is selected due to its compatibility with working with high-dimensional TF-IDF data. It uses 100 estimators and a maximum depth of 10. It also works well against overfitting problems. But it is not suitable to track non-linear patterns. XGB works well with text classification and handling high-dimensional data. It uses 150 estimators which help in boosting it and a learning rate of 0.05. It is prone to overfit, but it captures non-linear relationships in the data. Thus, using RFC and XGB as base models gives an effective model that compliments the weaknesses of the individual models and delivers a strong model. The meta-classifier LG is used to avoid overfitting and learns weight for base models on itself. It uses liblinear solver. It performs especially well for

979-8-3315-4317-4/25/$31.00 ©2025 IEEE

multi-label classification. The data is trained and then tested against the test data to label the text as depicted in Fig. 3.

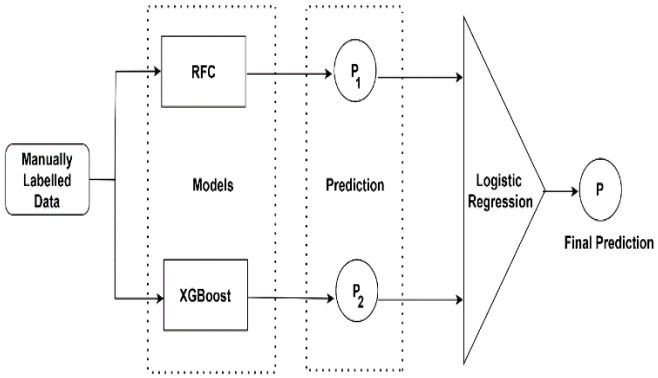

Fig. 3. Text Labeling using RFC-XGBoost hybrid stacking ensemble model

C. Determining price using XGBRegressor

The algorithm for price distribution is shown below.

Algorithm 2 Proposed Product Pricing

Input: labelled_data
Output: product_price

1. Data preparation(labelled_data)
2. Perform one-hot encoding
3. return as data
4. Prepare the data as test_x, train_x, test_y, train_y
5. Model development(test_x, train_x, test_y, train_y)
6. Initialize XGBoost Regressor
7. Parameters:
8. Estimators=200-220, depth=6
9. Train the model against train_x, train_y
10. Test the model against test_x for test_y
11. Analyse the results(RMSE)
12. return XGBoost Regressor as model
13. Individual price determination(model)
14. Get input (choices - 0 or 1) for keywords
15. Apply the developed model to determine the product_price

A new dataset is created using the labeled data. The label column is converted into a label list of unique label values. Using multilabel-binarizer the list is converted into encoded labels, and then it is converted into a dataframe. The values are then grouped by date. The data is then split as test-train data. Francisco Louzada, Kleython José Coriolano Cavalcanti de Lacerda, Paulo Henrique Ferreira and Naomy Duarte Gomes produced good results with XGBoost for categorical data [18]. XGB-Regressor is applied to the training data as it is more suitable for numerical value prediction using one hot encoding of a multi-labelled dataset. For hyperparameters, it uses 200 estimators, 5 as the minimum child weight, a low learning rate of 0.05, 0.8 sub-sample to avoid overfitting, objective is set to squared error.

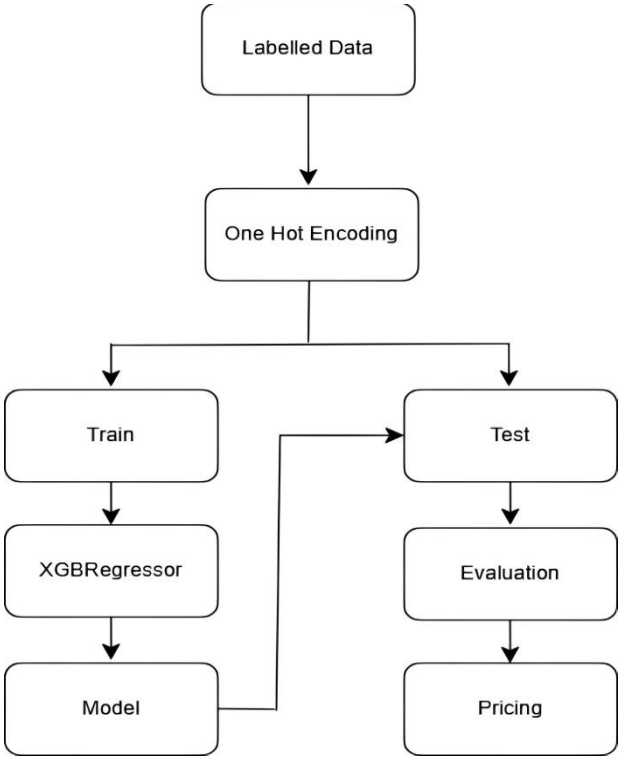

Fig. 4. One-hot encoding and product pricing using XGB-Regressor

This helps the model to learn from the dataset effectively while avoiding overfitting. The model is tested against the training data for determining the price as shown in Fig.4.

IV. Results And Discussion

A. GPT-3.5 Keyword Selection

The GPT-3.5 for identifying keywords worked efficiently; however, it also identified some keywords which did not directly affect product prices. So, determining the keywords comes down to market understanding from there. To analyze the working efficiency of GPT-3.5 model, some other text processing models like TF-IDF, Universal Sentence Encoder and Glove are applied on the same data to select the top 5 words and the correlation of the words with price is given as a column chart in Fig. 5. It follows an ordinal range of 1-5 where 1 gives the lowest correlation and 5 gives the highest correlation, positive or negative.

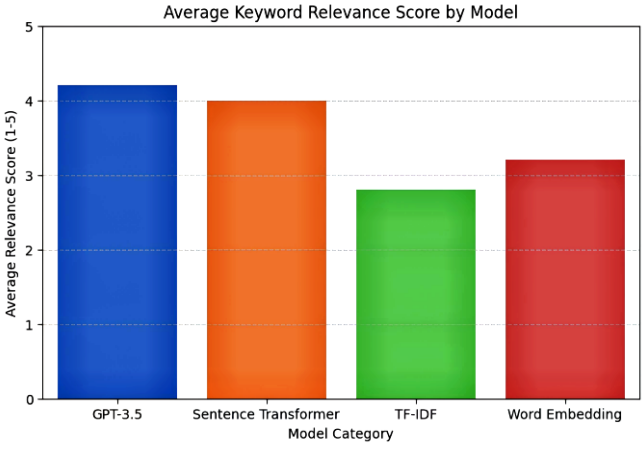

Fig. 5. Keyword relevancy: GPT-3.5 Vs Other Text Processing methods

To improve reliability of keyword selection certain changes could be made such as experimenting with minimum and

979-8-3315-4317-4/25/$31.00 ©2025 IEEE

maximum document frequencies for TF-IDF and improving the prompt suitable as needed based on response.

B. Stacking Ensemble Labeling

The stacking ensemble model for labeling data into related keywords achieved an accuracy of 0.99. It labels about 2300 tweets effectively labeling text into multiple keywords or labels depicted in Fig. 6.

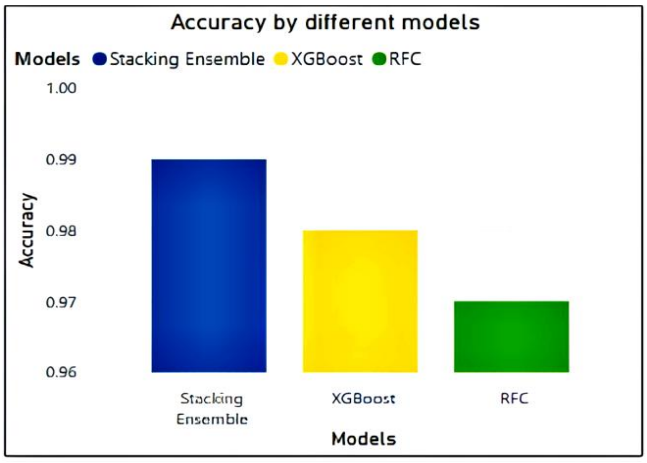

Fig. 6. Stacking ensemble model accuracy Vs RFC Vs XGBoost

The model is tested against the test data in 20/80 test train data to generalize results and the performance is tabulated in table.1

Table 1. Stacking Ensemble Model Performance against test data

Label	Precision	Recall	F1-Score
IPL	0.99	1	0.99
Festival	0.99	1	0.98
Bigg Boss	0.99	1	0.99
Great Indian Sale	0.99	1	0.99
Big Billion Days	0.99	1	0.99
Price Hike	0.99	1	0.99
Others	0.95	0.94	0.97
Average(Macro)	0.99	0.99	0.99

C. XGB Regressor Pricing

The average price of SONY Bravia 108 cm (43 inches) Full HD LED Smart Google TV is approximately 41293. The price is determined in two ranges: from Jan 2023–Dec June 2024 used for training and testing while July 2024-Dec 2024 is used as unseen data. RMSE, MAE and MAPE are used as validation metrics. Cross-validation and residual analysis are further performed to validate the model. The price determined by the ensemble model based on different event availability on the unseen data gave a mean RMSE value of 698.52 and a mean MAE of 457.36, resulting in a relative RMSE of 1.69% and a relative MAE of 1.11%. Low RMSE and MAE closer to RMSE show there aren't many large deviations. It resulted in a 1.17% MAPE. The cross-validation score for RMSE is marked in Fig. 7.

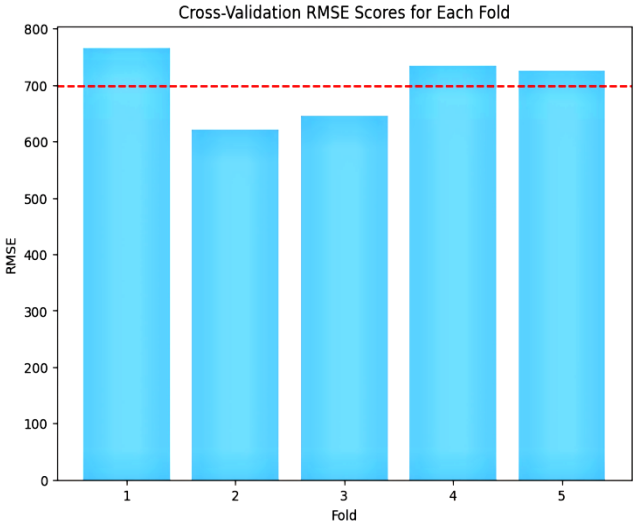

Fig. 7. Cross Validation of XGB Regressor model for price determination; the mean RMSE is marked by red dotted line

The error distribution in determination of price for products under various events is shown in Fig. 8. This shows the degree of variation in pricing is less and most results come closer to MAE value.

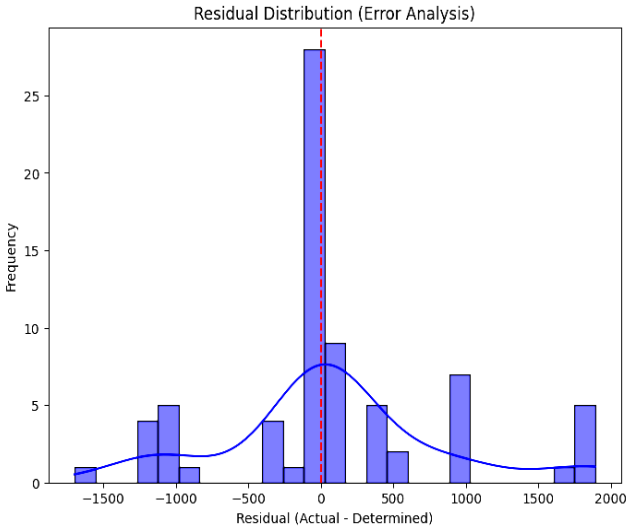

Fig. 8. Residual Distribution of XGB Regressor error in determining product price

The execution time and computational resources required for the proposed model to determine price dynamically is way less than existing models which differ in exponential scale; it is depicted in Fig. 9. It shows that the model requires way less resources than existing pricing models.

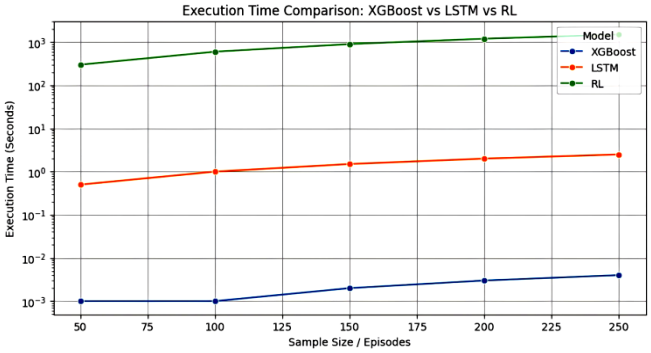

Fig. 9. Execution time comparison: XGBoost Vs LSTM Vs RL in exponential power scale(seconds)

979-8-3315-4317-4/25/$31.00 ©2025 IEEE

Scaling the data, adding some product information and tuning the hyperparameters could help improve RMSE and MAE scores.

Table 2. Performance comparison of various models and the proposed model

Model	RMSE %	MAE %	MAPE%
Proposed Model	1.69	1.1	1.17
K-means +Regression	17.5	22.5	-
XG-Boost	6.62	7.34	-
XNG-Boost	4.23	3.82	-
LSTM	1.73	1.33	1.59
ARIMA	1.78	1.37	1.63
CNN-LSTM	0.59	0.39	0.28
LSTM-RL	-	0.22	0.158

The comparison summary of pricing is given in table 2. It shows that the proposed model outperforms all basic models. Heavy models like CNN-LSTM and LSTM-RL performed better than the proposed model however they use customer data to train and run also take lot of time to train.

V. Conclusion

The keyword identification for products using pre-trained GPT-3.5 model uses less resources than training a BERT model. It identifies keywords related to the product with higher correlation than other models avoiding the need for expert selection. It is supposed to be run only once, making it a one-time expensive move. The stacking ensemble labels text data with high accuracy. It outperforms both of the base models. It works fairly similar to other stacking ensemble hybrid models. The ensemble pricing model determines price based on labelling rather than time specific analysis which allows capturing non time specific social patterns supporting pricing of most categories of products. Current dynamic pricing systems mostly stop altering price of products nearly after a year mostly due to its high computational cost. This method of determining price for products using social media driven demand can be an efficient alternative than leaving the price at an all-time high. It is a low cost fair approach that doesn't require any customer data. In future, the model should be applied to different products testing the validity for different products and categories. Integrating twitter API to handle demand in real time could the next work. This model can be integrated with existing reinforcement models to strengthen the pricing ability with social media insights.

References

[1] N. Alamsyah, Saparudin, and A. P. Kurniati, "Event detection optimization through stacking ensemble and BERT fine-tuning for dynamic pricing of airline tickets," IEEE Access, vol. 12, pp. 145254–145269, 2024, doi: 10.1109/ACCESS.2024.3466270.

[2] G. Yamuna, D. P. Dhinakaran, C. Vijai, P. S. J. Kingsly, R. Raynukaazhakarsamy, and S. R. Devi, "Machine learning-based price optimization for dynamic pricing on online retail," in Proc. 2024 9th Int. Conf. Sci. Technol. Eng. Math. (ICONSTEM), Chennai, India, 2024, pp. 1–5, doi: 10.1109/ICONSTEM60960.2024.10568763.

[3] M. Nowak and M. Pawłowska-Nowak, "Dynamic pricing method in the e-commerce industry using machine learning," Appl. Sci., vol. 14, no. 24, p. 11668, 2024, doi: 10.3390/app142411668.

[4] P. Padmanaaban, A. Jamal, A. Garg, K. Chauhan, K. Chanda, and A. Kumar, "E-commerce management and AI-based dynamic pricing revenue optimization strategies," J. Inform. Educ. Res., 2024, doi: 10.52783/jier.v4i2.998.

[5] J. Xu, Y.-C. Hsu, and W. Biscarri, "Dynamic pricing in securities lending market: application in revenue optimization for an agent lender portfolio," in Proc. 5th ACM Int. Conf. AI in Finance, Brooklyn, NY, USA: ACM, Nov. 2024, pp. 513–520, doi: 10.1145/3677052.3698611.

[6] 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, Nov. 2024, doi: 10.48550/arXiv.2411.18261.

[7] C. Tang, Y. Qin, F. Wu, and Z. Tang, "Dynamic demand-aware power grid intelligent pricing algorithm based on deep reinforcement learning," IEEE Access, vol. 12, pp. 75809–75817, 2024, doi: 10.1109/ACCESS.2024.3406338.

[8] D. Watari, I. Taniguchi, and T. Onoye, "Duck curve aware dynamic pricing and battery scheduling strategy using reinforcement learning," IEEE Trans. Smart Grid, vol. 15, no. 1, pp. 457–471, Jun. 2023, doi: 10.1109/TSG.2023.3288355.

[9] A. Ismail and M. Baysal, "Dynamic pricing based on demand response using actor–critic agent reinforcement learning," Energies, vol. 16, no. 14, p. 5469, 2023, doi: 10.3390/en16145469.

[10] Y. Liu, C. Yang, K. Huang, and W. Liu, "A multi-factor selection and fusion method through the CNN-LSTM network for dynamic price forecasting," Mathematics, vol. 11, no. 5, p. 1132, 2023, doi: 10.3390/math11051132.

[11] N. Gadde, A. Mohapatra, S. Dey, I. Das, V. Bhatia, and G. Reddy, "Optimizing dynamic pricing through reinforcement learning: techniques, case studies, and implementation challenges," Int. J. Sci. Res. (IJSR), vol. 13, no. 11, pp. 159–165, Nov. 2024, doi: 10.21275/SR241028211931.

[12] X. Wu, J. Qin, W. Qu, Y. Zeng, and X. Yang, "Collaborative optimization of dynamic pricing and seat allocation for high-speed railways: an empirical study from China," IEEE Access, vol. 7, pp. 139409–139419, 2019, doi: 10.1109/ACCESS.2019.2943229.

[13] M. Basal, E. Saraç, and K. Özer, "Dynamic pricing strategies using artificial intelligence algorithm," Open J. Appl. Sci., vol. 14, pp. 1963–1978, 2024, doi: 10.4236/ojapps.2024.148128.

[14] R. Y. Chenavaz and S. Dimitrov, "Artificial intelligence and dynamic pricing: a systematic literature review," J. Appl. Econ., vol. 28, no. 1, 2025, doi: 10.1080/15140326.2025.2466140.

[15] M. S. Gazi, M. R. Hasan, N. Gurung, and A. Mitra, "Ethical considerations in AI-driven dynamic pricing in the USA: balancing profit maximization with consumer fairness and transparency," J. Econ. Finance Account. Stud., vol. 6, no. 2, pp. 100–111, Apr. 2024, doi: 10.32996/jefas.2024.6.2.8.

[16] D. Ramachandran and R. Parvathi, " Analysis of twitter specific preprocessing technique for tweets," Procedia Computer Science, vol. 165, pp. 245–251, 2019, doi: 10.1016/j.procs.2020.01.083.

[17] P. He, J. Huang and M. Li, "Text keyword extraction based on GPT," 2024 27th International Conference on Computer Supported Cooperative Work in Design (CSCWD), Tianjin, China, 2024, pp. 1394-1398, doi: 10.1109/CSCWD61410.2024.10580849.

[18] Francisco Louzada, Kleython José Coriolano Cavalcanti de Lacerda, Paulo Henrique Ferreira and Naomy Duarte Gomes, " Smart renting: harnessing urban data with statistical and machine learning methods for predicting property rental prices from a tenant's perspective," Stats, vol. 8, no. 1, p. 12, Jan. 2025, doi: 10.3390/stats8010012.

979-8-3315-4317-4/25/$31.00 ©2025 IEEE