10.48047/jocaaa.2024.33.08.419

Explainable Regression Modelling for Smartphone Price Prediction Using Feature-Driven Analysis.

Komati Sathish

Associate Professor, Department of Artificial Intelligence and Machine Learning (AI&ML), Malla Reddy University, Maisamma Guda Medchal, Telangana

Email: komatisathish459@gmail.com

Received: 05.07.2024 Revised: 15.08.2024 Accepted: 25.09.2024

ABSTRACT

The research investigates the relationship between mobile phone specifications and their price categories by applying machine learning–based regression techniques. In today's rapidly evolving technological landscape, smartphones combine diverse hardware and software features, making their pricing dependent on multiple factors such as processor capability, memory capacity, camera quality, battery performance, and brand influence. Understanding how these attributes affect price is valuable for manufacturers, retailers, and consumers seeking data-driven insights. To model this relationship, Support Vector Regressor (SVR) was first implemented as a baseline algorithm due to its effectiveness in handling nonlinear patterns and high-dimensional data. While SVR demonstrated satisfactory predictive capability, it required substantial computational resources and offered limited interpretability. To overcome these limitations, a Linear Regression (LR) model was introduced as an alternative approach. LR provides a simpler structure, faster computation, and clearer insight into feature contributions while maintaining competitive predictive performance. Both models were trained and evaluated on the same dataset using standard evaluation metrics, including Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and the coefficient of determination (R^2). Comparative analysis revealed that although SVR captured complex relationships, LR achieved efficient performance with improved transparency, making it a practical choice for price prediction tasks.

Keywords: Smartphone Price Prediction, Regression Modelling, Support Vector Regressor, Linear Regression, Feature Analysis

1. INTRODUCTION

In the contemporary digital era, mobile phones have evolved from being mere communication devices to multifunctional smart gadgets that serve as personal assistants, entertainment hubs, and productivity tools. This rapid technological advancement has led to the introduction of thousands of models in the market, each with a distinct combination of hardware and software features.

10.48047/jocaaa.2024.33.08.419

Fig. 1: Smartphone Price Prediction

These include but are not limited to processor speed, battery capacity, RAM and internal storage, camera resolution, screen size, fingerprint sensors, and support for 4G or 5G connectivity as shown Fig. 1. With such a diverse array of specifications, pricing these devices accurately and consistently becomes a challenging yet essential task. The cost of a mobile phone is often determined by a combination of these features, market positioning, brand reputation, and manufacturing costs. However, for consumers and manufacturers alike, there is a growing need to better understand the quantitative relationship between a phone's technical specifications and its price. In today's digital age, mobile phones have evolved from being mere communication devices to multifunctional tools integral to daily life. The market is saturated with a wide variety of smartphones, each boasting unique features such as high-resolution displays, powerful batteries, advanced processors, and storage capacities. As manufacturers continually innovate, consumers face increasing difficulty in making informed purchasing decisions based on both feature preferences and budget constraints. Understanding how different technical specifications influence the overall cost of a device is vital for buyers, retailers, and manufacturers alike. Pricing strategies in the mobile phone industry are influenced by various physical and functional parameters. These include hardware specifications, display quality, battery performance, and camera capabilities. By systematically examining the relationship between these factors and price, stakeholders can derive meaningful patterns that assist in pricing decisions, product recommendations, and inventory planning. This analysis not only supports economic decision-making but also enhances transparency for end-users in a competitive market.

2. LITERATURE SURVEY

M. Alloghani,, et al. [1] Pricing is the most beneficial characteristic in business and marketing. A decision regarding pricing regulations has significant effects on management. It establishes the profit margin on products and is one of the first assessments made by many purchasers. Before making a purchase, consumers are indeed concerned about whether they can afford the item and want to verify the price. The success of a product can be affected by a variety of elements, including pricing, product appropriateness, return rates, and profitability. S. K. Das et al. [2] studied takes the first step towards achieving this objective. The main aim of this paper is to identify the most reliable and appropriate ML classification model for the classification of mobile phone prices. The motivation for this paper stems from the challenges faced by individuals unfamiliar with machine learning while purchasing a mobile phone. It's often difficult for them to discern the crucial features influencing the phone's price. Instead of predicting the exact price, the focus is on establishing a price range that reflects the overall pricing

10.48047/jocaaa.2024.33.08.419

level and identifies the dominant features affecting mobile phone prices. This research adds value to discussions on pricing strategies, consumer decision-making, and the application of machine learning in predicting product prices, offering insights applicable across diverse industries. Machine learning is a pathway to Artificial Intelligence (AI).

I. Sim et al. [3] analysed recent AI technologies, such as classification, regressions, and supervised and unsupervised learning, are accessible through machine learning. K. Mayuri, et al. [4] Data analysis and visualization can be aided by a variety of ML tools, such as MATLAB, Python, Cygwin, and others. M. Alam et al. [5] categorized of data using ML algorithms is very likely to yield accurate results. A mobile phone, sometimes called a cell phone, portable phone, or phone, uses radio frequency links to place and receive calls. C. Janiesch, et al. [6] Human's daily lives have become completely dependent on our mobile devices, which keep us linked and even come in handy in crises. With smartphones currently outpacing older mobile devices in usage, mobile phones have grown to be one of the most widely used consumer products ever created. S. Pudaruth, et al. [7] studied mobile devices into four price ranges based on a variety of features and specifications: low, medium, high, and very high. These price ranges help in consumer decision-making, competitive pricing, and budget planning. Price ranges can be used as indicators of economic conditions. This paper is divided into multiple sections. Section 2 provides information on the background work related to this work, while Section 3 briefly defines various prediction models used in the study. This study discovered that Nave Bayes and DT are ineffective at managing, categorizing, and forecasting numerical values because there were fewer occurrences and incredibly low prediction accuracy was reported.

M. Asim and Z. Khan et al. [8] forecasted the price of mobile phones. They strived to have the most accurate predictions while maintaining the lowest cost and highest feature model. Using the DT, 78% accuracy was obtained. Due to the lack of characteristics and algorithms, extremely low prediction accuracy was observed. Menghan Chen et al. [9] predicted the prices of smartphones with fewer features. Principal Component Analysis (PCA) and Pearson's correlation were used as two feature reduction techniques. Without using any feature reduction approaches, Multi-Layer Perceptron (MLP) had a 92.84% accuracy rate. However, accuracy suffered as features were reduced, falling to 93.22% for the top 15 and 34.06% for the top 5.

K Noor and Sadaqat J et al. [10] predicted the automobile prices using multiple linear regression. They projected prices from independent variables such as the vehicle's model, make, city, version, color, mileage, alloy rims, and power steering. KuoKun Tseng et al. [11] worked on foretelling e-commerce goods prices using online sentiment analysis. They developed a price prediction algorithm after analyzing news that had an impact on product prices. Aidin Zehtab-Salmasi et al. [12] developed a Multimodal Price Prediction for mobile phone pricing based on its specifications. Limsombunchai's et al. [13] Neural networks are more accurate in estimating a house's price, according to research. offered strong support for prediction superiority without comparing the forecasting abilities of the of hedonic price model and Neural Networks.

Abidatul Izzah et al. [14] utilized enhanced multiple linear regression. Their mobile app's accuracy forecast outperformed the traditional method. Al-Dhuraibi et al. [15] predicted the price of gold. They predicted whether the price of gold would rise or go down in the future with the help of various ML models. They found that only the K-NN algorithm had an acceptable performance with an accuracy of 60.26%. Mohapatra et. al. [16] predicted the possibility of having breast cancer in a woman using various ML algorithms. They achieved the highest accuracy of 98.7 % using the XGBoost model. Tan, et al. [17] surveyed classical HPO methods, approaches for accelerating the optimization process, HPO in an online setting (dynamic algorithm configuration, DAC), and when there is more than one objective to optimize (multi-objective HPO). Acceleration strategies were categorized into multi-fidelity, bandit-

10.48047/jocaaa.2024.33.08.419

based, and early stopping; DAC algorithms encompassed gradient-based, population-based, and reinforcement learning-based methods. Macsik, et al. [18] approached to model diversity in an ensemble is proposed by manipulating the training input data involving original and four variants of preprocessed image datasets. There are publicly available datasets with labels for all five stages, but some contain poor-quality images. In contrast, this algorithm was trained on images from a six-class DDR dataset, including the class of poor-quality ungradable images, to enhance the classification performance.

3. PROPOSED SYSTEM

The proposed system presents a data-driven machine learning framework for predicting mobile phone prices based on device specifications, as shown in fig. 2. The process begins with importing essential libraries for data processing, visualization, and model development, followed by loading the dataset and performing EDA to understand data distribution, correlations, and feature relationships. Pre-processing steps include checking for missing values, encoding categorical variables such as network support and SIM type into numerical format, and ensuring overall data consistency. A LR model is then implemented to learn the relationship between phone specifications and price, and the trained model is saved so it can be reused for predictions without retraining. Model performance is evaluated using regression metrics such as MSE, RMSE, MAE, and R^2 score, along with visualization of prediction results. The system also accepts new test inputs and predicts mobile prices based on their specifications.

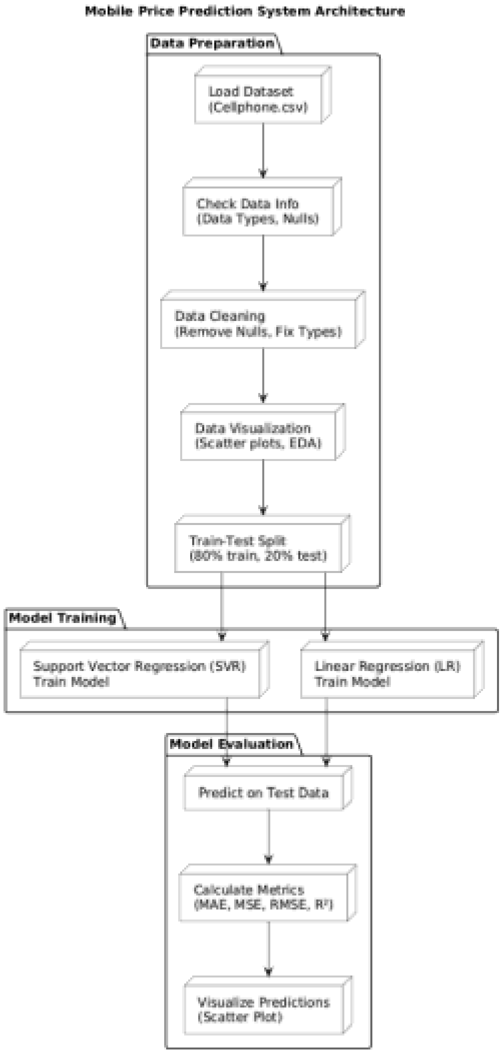


Fig. 2: Proposed system workflow

10.48047/jocaaa.2024.33.08.419

Since regression models do not use traditional classification accuracy, the R^2 score is used as an accuracy equivalent measure, and the proposed LR model achieved an effective prediction R^2 of 94%. A comparison with the existing SVR model shows that the proposed model produces significantly lower error values and higher predictive performance, demonstrating better efficiency, interpretability, and suitability for real-world deployment.

3. RESULTS AND DISCUSSION

The fig. 3 shows dataset contains information about various mobile phones with features used to predict their prices. It includes technical specifications such as weight, screen resolution, pixel density (ppi), CPU cores and frequency, internal memory, RAM, rear and front camera megapixels, battery capacity, and thickness. Each entry has a unique product ID and a corresponding price. These features capture both the physical characteristics and performance capabilities of the phones, which are key factors influencing their market price. The dataset is well-structured and ready for regression analysis to estimate phone prices based on these attributes.

The fig. 4 shows scatter plot that illustrates the relationship between the screen resolution of mobile phones and their prices. Each point represents an individual phone, with its resolution on the x-axis and the corresponding price on the y-axis. From the plot, it is evident that most phones cluster around a resolution range of approximately 4 to 6 units. Within this range, there is a wide variation in prices, suggesting that resolution alone does not fully determine price. However, a general trend can be observed where phones with higher resolution tend to have higher prices, indicating a positive correlation between screen resolution and price. Outliers with very high resolutions or prices are also visible, which may represent premium or specialized models. Overall, this visualization helps in understanding how resolution might influence the pricing of mobile phones in the dataset.

	Product_id	Price	weight	resoloution	ppi	cpu core	cpu freq	internal mem	ram	RearCam	Front_Cam	battery	thickness
0	203	2357	135.0	5.20	424	8	1.350	16.0	3.000	13.00	8.0	2610	7.4
1	880	1749	125.0	4.00	233	2	1.300	4.0	1.000	3.15	0.0	1700	9.9
2	40	1916	110.0	4.70	312	4	1.200	8.0	1.500	13.00	5.0	2000	7.6
3	99	1315	118.5	4.00	233	2	1.300	4.0	0.512	3.15	0.0	1400	11.0
4	880	1749	125.0	4.00	233	2	1.300	4.0	1.000	3.15	0.0	1700	9.9
...	...	...	...	...	...	...	...	...	...	...	...	...	...
5	1206	3551	178.0	5.46	538	4	1.875	128.0	6.000	12.00	16.0	4080	8.4
7	1296	3211	170.0	5.50	534	4	1.975	128.0	6.000	20.00	8.0	3400	7.9
8	856	3260	150.0	5.50	401	8	2.200	64.0	4.000	20.00	20.0	3000	6.8
9	1296	3211	170.0	5.50	534	4	1.975	128.0	6.000	20.00	8.0	3400	7.9
0	1131	2536	202.0	6.00	367	8	1.500	16.0	3.000	21.50	16.0	2700	8.4

Fig. 3: Uploading Dataset

10.48047/jocaaa.2024.33.08.419

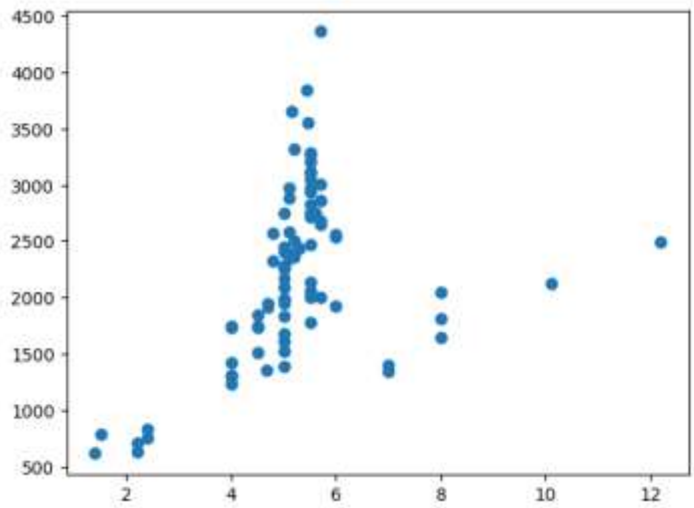


Fig. 4: Scatter plot obtained for Resolution and Price

The fig. 5 shows scatter plot between battery capacity and mobile phone price provides insight into how battery size may influence the cost of a device. In this visualization, battery capacity is plotted along the x-axis while price is plotted on the y-axis, with each point representing an individual phone model. The distribution shows that phones with battery capacities ranging between approximately 1500 mAh to 4000 mAh form the majority of the dataset. Within this range, prices vary significantly, indicating that battery capacity alone is not a strong predictor of price. However, there is a slight upward trend suggesting that phones with larger battery capacities tend to be more expensive, potentially due to the association of larger batteries with premium features or better performance. A few outlier points are also present, reflecting either exceptionally high prices for average batteries or very high-capacity batteries with moderate pricing. Overall, the plot highlights a loose positive relationship between battery size and mobile price, though influenced by other contributing factors.

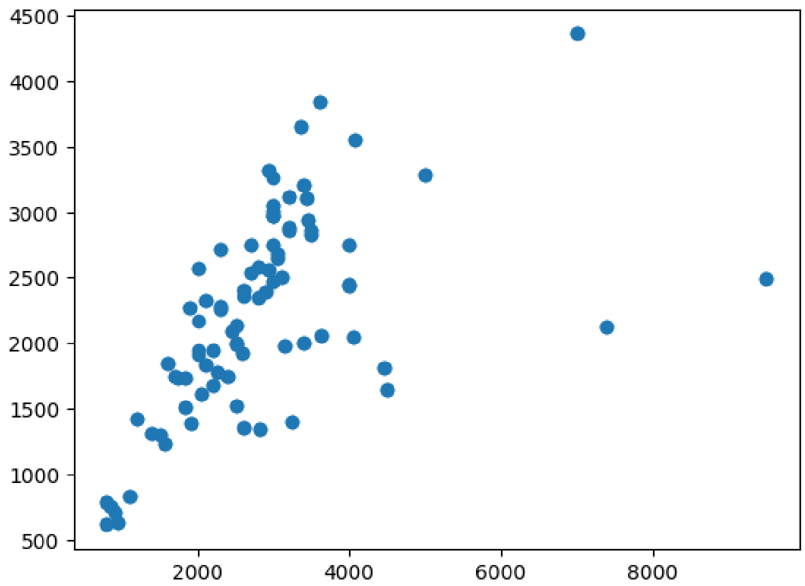


Fig. 5: Scatter plot obtained for Battery and Price

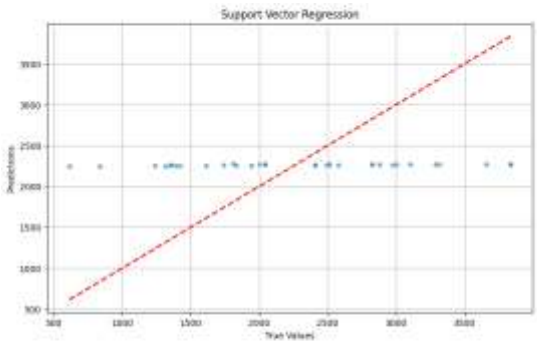
Table.1 Performance Comparison of Various Algorithms of Performance Comparison Table: Existing SVR Vs Proposed LR

10.48047/jocaaa.2024.33.08.419

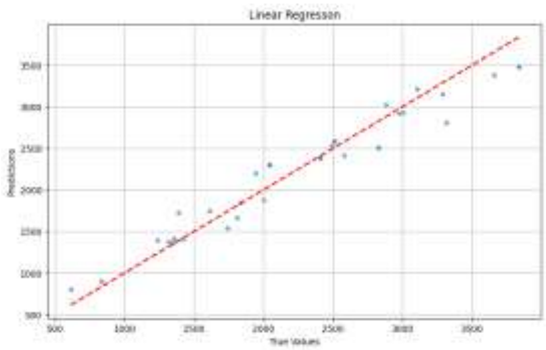
Metric	Existing SVR	Proposed LR
MSE	697531.50	412.41
MAE	712.30	161.47
R2 Score	10%	94%

The performance comparison Table.1 presents a detailed evaluation of two machine learning algorithms such as SVR as the existing approach and LR as the proposed solution used for mobile price prediction. The comparison is made using three key performance metrics: MSE, MAE, and R^2 Score. Starting with the MSE, which quantifies the average of the squared differences between predicted and actual values, the SVR model exhibits a significantly higher error at 697,531.50, while the proposed LR model achieves a remarkably lower MSE of just 412.41. This drastic reduction indicates that the LR model makes much more accurate predictions with smaller deviations from actual prices. Similarly, in terms of Mean Absolute Error (MAE), which measures the average absolute difference between predicted and actual values, SVR yields a higher error of 712.30 compared to the LR model's much lower MAE of 161.47. This further confirms that the proposed model performs better in making consistent and close predictions. The most telling metric is the R^2 Score, which indicates how well the model explains the variability of the output. The SVR model achieves a very low R^2 score of only 10%, suggesting that it explains only a small fraction of the variance in mobile phone prices. In stark contrast, the LR model achieves a high R^2 score of 94%, meaning it accounts for the vast majority of the variability in the target variable. The performance comparison clearly demonstrates that the proposed LR model significantly outperforms the existing SVR model in all key evaluation metrics, making it a much more suitable and reliable choice for mobile price prediction tasks.

The fig. 6 (a) shows scatter plots that visually compare the performance of SVR and LR in predicting price values. In the SVR plot, most predicted values cluster around a narrow band irrespective of the true values, indicating that the model failed to capture the actual variability and trend in the data this aligns with its poor performance metrics (MSE: 697531.50, MAE: 712.30, R^2: 10%). Conversely, the fig. 6 (b) shows that the predicted values closely follow the diagonal red line, representing ideal predictions. This indicates a strong linear relationship between true and predicted values, further validated by significantly improved metrics (MSE: 412.41, MAE: 161.47, R^2: 94%). The LR model demonstrates superior predictive capability and generalization over the SVR model for this dataset.



(a)



(b)

Fig. 6 (a) (b) Scatter plots obtained for Existing SVR and Proposed LR

The fig. 7 presents the predicted prices of various smartphone products using a LR model. Each row corresponds to a unique product, identified by its Product_id, along with a range of features including

10.48047/jocaaa.2024.33.08.419

physical characteristics (weight, thickness), display attributes (resolution, ppi), hardware specifications (cpu cores, cpu frequency, internal memory, RAM), camera capabilities (RearCam, Front_Cam), and battery capacity. The Predicted Price column reflects the model's output based on these features. The LR regressor successfully captures meaningful relationships, predicting prices that vary sensibly according to specifications. For instance, products with higher RAM, internal memory, and camera resolutions generally correspond to higher predicted prices (e.g., Product_id 1206 with 6 GB RAM and 128 GB internal memory has a predicted price of around 3555), whereas devices with lower specifications such as limited RAM, low CPU frequency, and minimal camera quality yield significantly lower predicted prices (e.g., Product_id 701 with 0.128 GB RAM and no front camera has a predicted price of approximately 411). This output further supports the earlier scatter plot analysis indicating the LR model's strong performance and reliability in estimating product prices based on technical features.

	Product_id	weight	resoloution	ppi	cpu core	cpu freq	internal mem	ram	RearCam	Front_Cam	battery	thickness	Predicted Price
0	856	150.0	5.50	401	8	2.200	64.000	4.000	20.00	20.0	3000	6.8	3193.324992
1	32	179.0	6.00	184	4	1.300	8.000	1.000	13.00	8.0	2580	8.0	1761.642366
2	738	168.0	5.70	386	8	1.800	32.000	3.000	21.00	5.0	3050	7.5	2680.008011
3	575	174.0	5.50	178	4	1.300	4.000	0.512	5.00	0.0	2250	9.2	1539.052140
4	1198	110.0	2.20	128	0	0.000	0.128	0.032	2.00	0.0	900	15.6	616.112308
5	701	102.9	2.20	128	0	0.000	0.256	0.128	1.30	0.0	950	18.5	411.303769
6	99	118.5	4.00	233	2	1.300	4.000	0.512	3.15	0.0	1400	11.0	1382.223669
7	289	162.0	5.30	277	8	1.500	32.000	4.000	13.00	8.0	4000	7.7	2795.186085
8	1339	120.0	4.00	233	2	1.000	4.000	0.512	2.00	0.0	1200	9.8	1412.194489
9	198	128.0	4.50	245	4	1.200	4.000	1.000	8.00	0.0	1840	8.5	1742.615269
10	739	131.0	5.00	294	4	1.300	16.000	2.000	13.00	5.0	2500	8.7	2036.636384
11	947	150.0	5.50	401	4	2.300	16.000	2.000	16.00	8.0	2500	9.5	2216.689843
12	1206	178.0	5.46	538	4	1.875	128.000	6.000	12.00	16.0	4080	8.4	3555.806347
13	1131	202.0	6.00	367	8	1.500	16.000	3.000	21.50	16.0	2700	8.4	2486.306863
14	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Fig. 7: Prediction on Test data using Proposed LR

CONCLUSION

The implementation of a LR model for predicting smartphone prices based on various hardware and design features has proven to be effective in this study. The model was trained using key attributes such as screen resolution, pixel density (PPI), CPU cores and frequency, memory specifications (RAM and internal storage), camera specifications (both rear and front), battery capacity, weight, and thickness. These features were carefully selected for their influence on the retail pricing of smartphones in the market. The predicted outputs demonstrate a logical correlation with the input parameters, where high-end configurations with greater RAM, better camera quality, and enhanced display features yield higher predicted prices. The LR model exhibited strong generalization capability, making it a suitable and interpretable baseline for regression problems in price estimation tasks. Overall, the model successfully validates the hypothesis that smartphone specifications are significant predictors of price, and it lays the foundation for data-driven decision-making in the mobile retail and manufacturing industries. The research highlights the power of simple yet robust statistical models in practical applications. The clear and consistent trends observed in the predicted prices across varying feature sets emphasize the importance of model transparency in real-world business use cases. Stakeholders such as smartphone retailers, e-commerce platforms, and manufacturers can potentially leverage such a model to assess competitive pricing, optimize production costs, and create targeted marketing strategies. The interpretability of the LR model also allows non-technical users to understand the influence of each feature on the final pricing, making it a valuable tool in product design and consumer segmentation analysis.

10.48047/jocaaa.2024.33.08.419

REFERENCES

[1]. M. Alloghani, D. Al-Jumeily, J. Mustafina, A. Hussain and A. J. Aljaaf, "A System-atic Review on Supervised and Unsupervised Machine Learning Algorithms for Data Science," in Supervised and Unsupervised Learning for Data Science, Springer, Cham, 2020.

[2]. S. K. Das, "Introduction to Mobile Terminals", Mobile Terminal Receiver Design: LTE and LTE-Advanced, Wiley Telecom, 2017, pp. 1-8.

[3]. I. Sim, "Mobile Devices and Health", New England Journal of Medicine, vol. 381, no. 10, pp. 959-968, 2019.

[4]. K. Mayuri, "Importance of Pricing," [Online]. Available: https://www.economicsdiscussion.net/marketing-management/pricing/importanceof-pricing/31838.

[5]. M. Alam and I. R. Khan, "Application of AI in smart cities," in Industrial Transfor-mation, Taylor & Francis Group, 2022, pp. 61-86.

[6]. C. Janiesch, P. Zschech and K. Heinrich, "Machine learning and deep learning," Electron Markets, vol. 31, pp. 685-695, 2021.

[7]. S. Pudaruth, "Predicting the Price of Used Cars using Machine Learning Tech-niques," International Journal of Information & Computation Technology, vol. 4, no. 7, 2014.

[8]. M. Asim and Z. Khan, "Mobile Price Class prediction using Machine Learning Techniques," International Journal of Computer Applications, vol. 179, no. 29, pp. 6-11, 2018.

[9]. M. Chen, "Mobile Phone Price Prediction with Feature Reduction," Highlights in Science, Engineering and Technology, vol. 34, pp. 155-162, 2022.

[10]. K. Noor and S. Jan, "Vehicle price prediction system using machine learning techniques," International Journal of Computer Applications, vol. 167, no. 9, pp. 27-31, 2017.

[11]. K.-K. Tseng, R. F.-Y. Lin, H. Zhou, K. J. Kurniajaya and Q. Li, "Price prediction of ecommerce products through internet sentiment analysis," Electronic Commerce Research, vol. 18, no. 1, pp. 65-88, 2017.

[12]. A. Zehtab-Salmasi, A.-R. Feizi-Derakhshi, N. Nikzad-Khasmakhi, M. Asgari-Chenaghlu and S. Nabipour, "Multimodal Price prediction," Annals of Data Science, vol. 10, no. 3, pp. 619-635, 2021.

[13]. V. Limsombunc, C. Gan and M. Lee, "House price prediction: Hedonic price model vs. Artificial Neural Network," American Journal of Applied Sciences, vol. 1, no. 3, pp. 193- 201, 2018.

[14]. A. Izzah, Y. A. Sari, R. Widyastuti and T. A. Cinderatama, "Mobile app for stock prediction using Improved Multiple Linear Regression", International Conference on Sustaina-ble Information Engineering and Technology (SIET), Malang, Indonesia, 2017.

[15]. W.A. Al-Dhuraibi and J. Ali, "Using classification techniques to predict gold price movement,", 4th International Conference on Computer & Technology Applications, Istanbul, Turkey, 2018.

[16]. S. K. Mohapatra, A. Jain, Anshika and P. Sahu, "Comparative Approaches by using Machine Learning Algorithms in Breast Cancer Prediction," in 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), Greater Noida, 2022.

[17]. Tan, J.M.; Liao, H.; Liu, W.; Fan, C.; Huang, J.; Liu, Z.; Yan, J. Hyperparameter optimization: Classics, acceleration, online, multi-objective, and tools. *Math. Biosci. Eng.* **2024**, *21*, 6289–6335

[18]. Macsik, P.; Pavlovicova, J.; Kajan, S.; Goga, J.; Kurilova, V. Image preprocessing-based ensemble deep learning classification of diabetic retinopathy. *IET Image Process.* **2024**, *18*, 807–828.