OPEN ACCESS
International Journal of Innovations in Science & Technology

ML-Driven Valuation of Used Smartphones

Ubaid-ur-Rehman, Shamshad Lakho, Vishal Kumar, Abdul Basit, Shoaib Ali, Manav
Département of Computer Science Quaid-e-Awam University of Engineering Science & Technology, Nawabshah, Sindh, Pakistan
***Correspondence**: ubaid17151@gmail.com, shamshad.lakho@quest.edu.pk, vishal22370@gmail.com, basitabbasia69@gmail.com, jamshoaibali11@gmail.com, manav.kumar22455@gmail.com

Citation | Rehman. U. U, Lakho. S, Kumar. V, Basit. A, Ali. S, Manav "ML Driven Valuation of Used Smartphones", IJIST, Vol. 07 Issue. 10 pp 01-09, November 2025
Received | October 25, 2025 **Revised |** November 15, 2025 **Accepted |** November 17, 2025 **Published |** November 20, 2025.

In the current digital age, the rapid pace of smartphone upgrades complicates the accurate estimation of resale values in the second-hand market. The primary objective of this research is to develop an efficient machine learning–based system for forecasting the resale prices of pre-owned smartphones and to evaluate the effectiveness of multiple regression models. A manually collected dataset comprising more than 900 smartphone records was compiled from online marketplaces and local mobile shops to ensure realistic and up-to-date pricing information. Regression techniques, including Linear Regression, Decision Tree, Random Forest, Gradient Boosting, and XGBoost, were implemented and analyzed using standard performance metrics such as R^2, mean absolute error (MAE), and root mean square error (RMSE).. The comparative analysis identifies the most suitable model for accurate price prediction, with ensemble-based methods demonstrating superior performance over traditional approaches. The proposed methodology provides a reliable and data-driven framework that enhances pricing transparency, reduces reliance on subjective retailer assessments, and supports informed decision-making for both buyers and sellers in the used smartphone resale market.

Keywords— Smartphone Price Prediction, Regression Models, Machine Learning Used in the Smartphone Resale Market, Price Prediction, Predictive Modeling.

Introduction:

In today's digital world, mobile phones are becoming increasingly powerful in terms of functionality. They have become an essential part of people's daily lives. Almost every individual, regardless of age or profession, owns a mobile phone for communication, entertainment, business, and education. With the rapid growth of new technology, older models quickly become outdated. As a result, many people sell or exchange their old mobile phones to purchase newer ones. However, mobile phone prices in the second-hand market are affected by multiple variables such as the phone's condition, technical specifications, brand, and market demand [1].

Traditionally, the resale price of mobile phones is determined based on the experience of shopkeepers and prevailing market conditions. Although these methods provide some guidance, they offer limited accuracy because they fail to consider the relationships between specifications, condition, brand, and market behavior. Additionally, such approaches cannot quickly adapt to changes in mobile technology trends, consumer preferences, or the frequent release of new models. This leads to inconsistent and inaccurate price predictions. Therefore, a data-driven predictive system is required to estimate the resale price of used mobile phones, which improves price accuracy and transparency while minimizing reliance on third parties. [2].

In this context, machine learning presents a reliable and robust solution for forecasting the resale price of used mobile phones. By analyzing past sales data, these models can learn complex relationships among factors such as device specifications, condition, and market demand. Furthermore, their ongoing learning process provides progressively more precise and accurate predictions [2].

Literature Review:

To make a strong foundation for our project, we examined multiple research papers to obtain knowledge about how similar projects use these models in their studies.

[1] presented a ML–based approach that utilizes Linear Regression, Random Forest, and Multilayer Perceptron techniques to predict the resale prices of used electronic devices. Their findings show that Random Forest achieved the highest accuracy with lower error. with low error, showing how effectively Machine Learning captures the complex links between device features and resale values [1].

[3] The authors predicted mobile phone prices based on attributes including RAM, storage, battery life, camera, and display using Support Vector Machine (SVM) and Decision Tree models. The dataset, which has 11,500 records with 178 attributes, was gathered via Kaggle. To increase the effectiveness of the model, feature selection and normalization were carried out. With an AUC score of 0.993, The Decision Tree model performed better than SVM, according to the results, showing higher accuracy. The Decision Tree model was found to be more successful in projecting smartphone pricing based on specifications [3].

[4]. The author used Decision Tree Regression and Random Forest Regression models for estimating the prices of smartphones. Data was gathered using web-scraping tools from Flipkart, such as Beautiful Soup or Selenium. Data cleaning and analysis were performed using various Python libraries. It resulted in the Random Forest Regression method performing better with $R^2 = 0.89$ as compared to the Decision Tree, which has an R^2 of 0.82. In this case study, it was concluded that Random Forest is more powerful and effective in estimating the price of smartphones based on specifications and their features [4].

[5] investigated price prediction for automotive steel materials with a wide range of machine learning regression techniques, including ANN and tree-based regressor models: Decision Tree, Random Forest, and Gradient Boosting. The study enhanced consistency in strategic purchasing decisions by merging DCV with IPT.The study enhanced consistency in strategic purchasing decisions. The findings of these experiments showed that the ANN

regression model had a higher level of prediction accuracy, indicating its potential for building intelligent and adaptive purchasing mechanisms [5]

[6] conducted a comparison of four regression-based machine learning models, which are Linear Regression, Random Forest Regressor, XGB Regressor, and Support Vector Machine Regressor, for mobile phone price prediction. By using a Kaggle dataset, the study shows that the XGB Regressor gives the highest accuracy ($R^2 = 0.95$). The finding showed that regression methods, especially XGBoost, provide highly reliable and accurate predictions for estimating mobile phone prices [6].

To anticipate Boston home prices, [7] examined Linear Regression, ANN, Random Forest, and SVR. Models were assessed using MSE, R2, and MAE following preprocessing (cleaning, feature selection, outlier treatment, and normalization). ANN outperformed SVR and Random Forest, with $R^2 = 0.86$, MSE = 0.0046, and MAE = 0.047. The least accurate method was linear regression. According to the study's findings, Random Forest and SVR are good substitutes for ANN in capturing intricate non-linear patterns [7].

SVM and rigid classifiers were compared by [8] to forecast the cost of used cell phones. SVM attained 98% training accuracy but only 85% testing accuracy following preprocessing (feature selection, scaling, and managing missing values), while Rigid consistently maintained 91–92% accuracy. According to the study's findings, both models work well, but Rigid produces less overfitting and more reliable outcomes [8].

The reviewed studies emphasize the increasing success of machine learning approaches for price prediction across multiple areas. Past research has explored both regression and classification models, including Linear Regression, Random Forest, XGBoost, ANN, SVM, LSTM, and LDA, to predict the prices of smartphones, electronic devices, agricultural products, and industrial materials. Most of the results show that advanced and ensemble-based models, especially Random Forest, XGBoost, ANN, and SVM, regularly perform better than traditional statistical methods by observing complex relationships among different features. These studies give a strong basis for the present research, which aims to use regression-based machine learning models to accurately forecast the resale prices of used smartphones.

The results of this study were compared with prior smartphone price prediction studies. Previous research by [1] and [4] shows that ensemble-based ML models, such as Random Forest, obtain higher accuracy than Standard regression techniques. Similarly, [6] reported better performance using XGBoost for mobile price estimation.

Our research differs from [1] in terms of data collection, features, and methodology. They used different data collection approaches, whereas we used a market survey. We also added new features such as resale price, device condition, and battery health. Additionally, they used the WEKA tool, while our study employs Scikit-learn with advanced ML algorithms such as XGBoost and GradientBoosting.

Predicting the prices of used mobile phones is challenging due to the rapid release of new models. Buyers often fail to accurately estimate the real price, which frequently leads to overpricing or undervaluation. Key factors such as brand, model, and RAM, etc., further complicate accurate pricing. Machine learning techniques can be used to provide reliable Price Prediction. It helps users make informed and transparent decisions in buying or selling mobile phones.

The specific objectives of this research are as follows:

To collect and refine the dataset for accurate mobile price prediction.

To apply and compare various Machine Learning Models for forecasting price.

To evaluate the performance of various Machine learning models.

Novelty of the Study:

Unlike earlier studies that mostly depend on already available online datasets, this research presents a newly developed and manually collected dataset from online marketplaces and local mobile markets, providing more realistic and up-to-date pricing information.

The research further evaluates different machine learning models on this novel dataset to determine the most accurate and reliable method for price prediction.

Methodology:

This method is applied to create an all-encompassing ML-based system for forecasting the actual market price of pre-owned smartphones. The system utilized past pricing data and smartphone characteristics like brand, model, RAM, storage, condition, and original price to forecast precise pricing. This system is trained using multiple regression models and also seeks to produce equitable and reliable price estimates. The approach depicted in Figure 1 is created to be scalable, enabling it to manage larger datasets. Additionally, this system improves pricing clarity by reducing human error in judgment and providing data-driven price assessments for buyers and sellers. This method enhances decision-making in the used smartphone market through equitable and automated price forecasts.

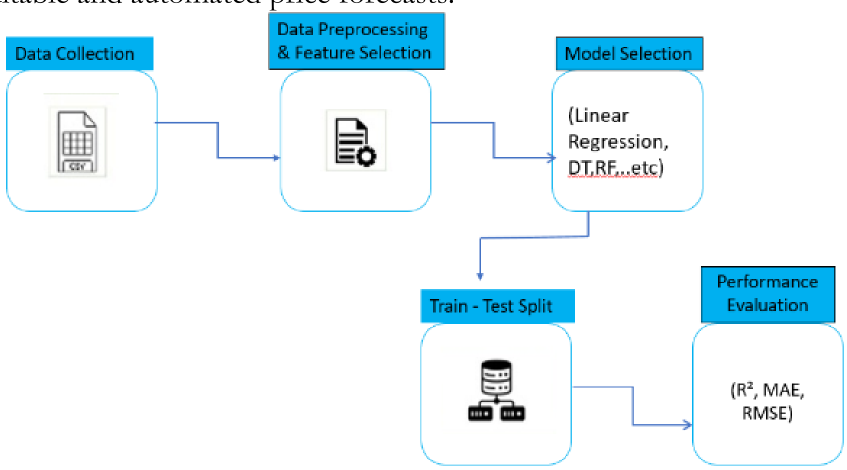

Figure 1. Proposed Methodology Workflow

Figure 1 shows the complete workflow of the proposed methodology. The process starts with data collection and data preprocessing, followed by the selection of machine learning regression models and train–test split. Finally, model performance is evaluated using standard regression metrics, including R^2, mean absolute error (MAE), and root mean square error (RMSE).

Data Collection:

Data has been sourced from multiple channels, including online marketplaces, mobile retail websites, and local market surveys. The dataset includes key features listed in Table 1, such as brand, model, RAM, storage, and selling price.

Table 1. Description of Dataset Features Used for Smartphone Resale Price Prediction

Features	Description
Brand	Indicates the manufacturer of the smartphone and reflects market reputation and demand.
Model	Represents the specific version of the device, which affects performance and resale value.
Ram_gb	Shows the memory capacity that influences the phone's speed and multitasking ability.
Storage_gb	Defines internal storage space, which impacts usability and price.
Condition	Describes the physical and functional state of the smartphone (e.g., excellent, good, fair).

Battery Health	Measures battery efficiency and charging capacity over time.
Original Price	The launch price of the smartphone when it was new.
Resale Price	The target variable represents the estimated market value in the used condition.

Data Preprocessing:

The collected dataset was cleaned to remove missing, duplicate, and irregular values.. Feature encoding and normalization were applied to prepare the data for training.

Model Selection:

Various machine learning regression models, including Linear Regression, Decision Tree Regression, Random Forest Regression, Gradient Boosting Regression, and XGBoost Regression, were selected to analyze and predict resale prices.

Model Training and Testing:

The dataset was split into training and testing sets. Model training was performed using the Python scikit-learn framework.

Evaluation:

The model was evaluated using metrics such as R^2, mean absolute error (MAE), and mean squared error (MSE) to determine the most accurate model. R^2 measures how much a regression model shows the variance in the dependent variable and predicts, with 1 = perfect, 0 = mean-level, and <0 = worse. SSR is the sum of squared errors, and SST is the total variance.

$$R^2 = 1 - \left(\frac{SSres}{SStot}\right)$$

Where:

$$SSres = \Sigma\,(yi - \hat{y}i)^2$$
$$SStot = \Sigma\,(yi - \bar{y})^2$$

MAE measures the average absolute difference between predicted and actual values, showing prediction accuracy in the same units as the target.

$$MAE = (1 / n)\,\Sigma\,|\,yi - \hat{y}i\,|$$

RMSE tells us how far, on average, the predicted values are from the actual values. It gives more weight to bigger errors, so big mistakes affect the score more, helping us see how accurate the model really is.

$$\mathbf{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y})^2}$$

Tools and Techniques:

The machine learning workflow is implemented by using the following tools:

Python was used as the primary high-level programming language to implement and execute the complete machine learning workflow.

Pandas & NumPy libraries are utilized for data loading, data cleaning, numerical computations, and efficient data handling and analysis.

Scikit-Learn provides essential tools for building, training, and evaluating machine learning models.

XGBoost is applied as an advanced gradient boosting algorithm to enhance prediction accuracy by combining multiple weak learners.

A decision tree is employed as a supervised machine learning algorithm that splits data into branches to make predictions.

Random Forest is an ensemble learning technique that improves model accuracy by aggregating the results of multiple decision trees.

Gradient Boosting is used as an ensemble method that builds models sequentially to minimize prediction errors.

Linear Regression is implemented as a basic predictive algorithm for estimating continuous numerical values based on input features.

Matplotlib & Seaborn are used for data visualization and graphical representation of model performance and results.

Tkinter, a Python-based GUI library, was used to develop an interactive and user-friendly application interface.

Results & Discussion:

The predictive performance of all models is summarized in Table 2 and Figure 4 using the coefficient of determination (R^2), mean absolute error (MAE), and root mean square error (RMSE). R^2 measures the proportion of variance explained by the model, with values closer to 1.0 indicating a superior fit. MAE and RMSE quantify average prediction errors in price units, where lower values signify higher accuracy.

Gradient Boost Regression achieved the highest precision (R2 = 0.9952, MAE = 480.44, RMSE = 1039.15), followed closely by XGBoosting Regression (R2= 0.9950, MAE = 504.05). These results, along with the strong performance of Random Forest Regression (R^2 = 0.9857), demonstrate the superior ability of ensemble-based methods to capture complex, nonlinear relationships—such as those among original price, storage, and battery health—that dictate smartphone resale value. In contrast, simpler models like Linear Regression and Decision Trees produced significantly higher error rates, highlighting their limitations in this domain.

Table 2. Performance Metrics (R^2, MAE, RMSE) of Regression Models for Smartphone Price Prediction

Models	R2	MAE	RMSE
Linear Regression	0.8978	2819.57	4805.37
Decision Tree	0.9549	1287.01	3193.04
Random Forest Regression	0.9857	785.31	1798.18
XGBoosting Regression	0.9950	504.05	1064.90
Gradient Boost Regression	0.9952	480.44	1039.15

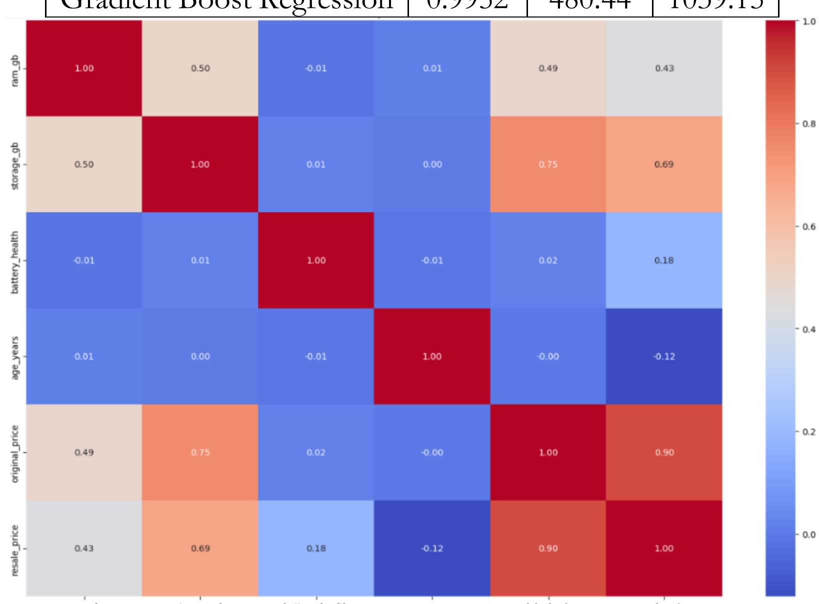

Figure 2. Correlation Heatmap Showing Relationships Between Smartphone Features and Resale Price

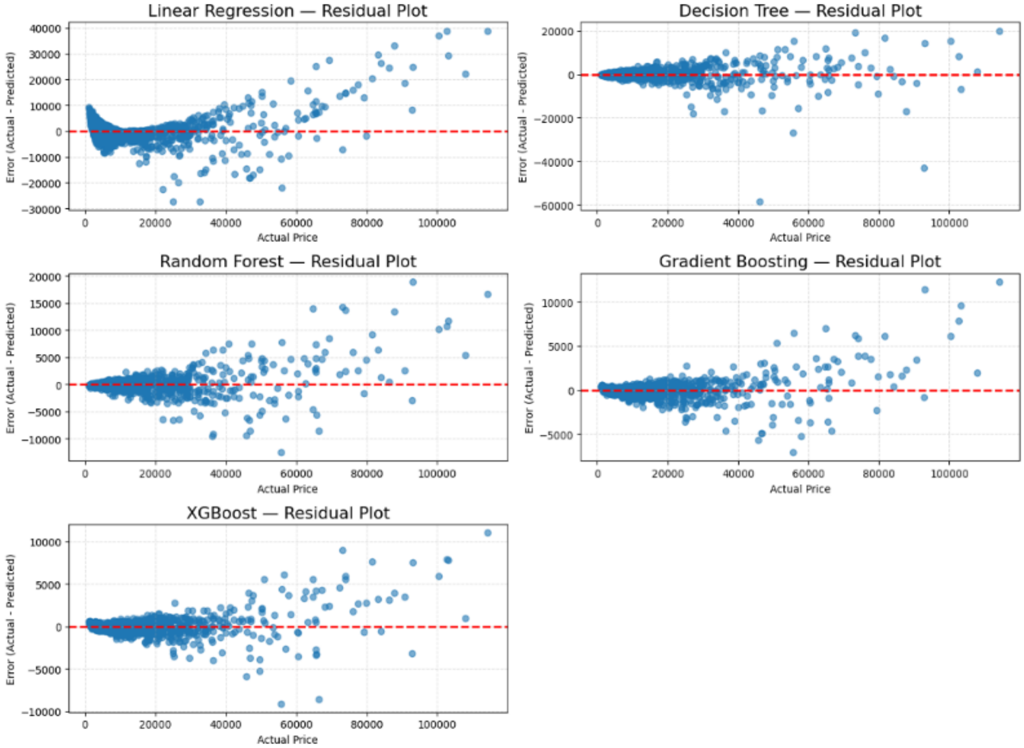

Figure 3. Residual Plots of Machine Learning Models

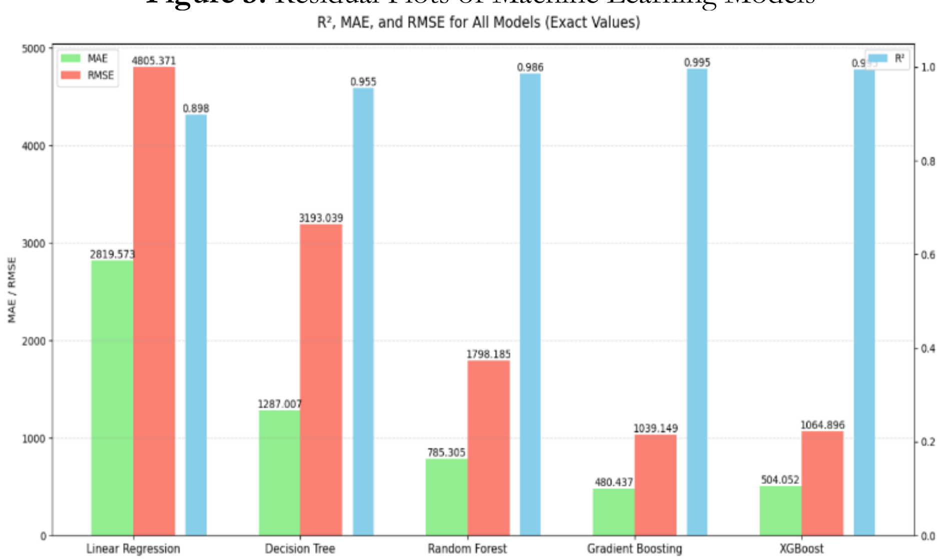

Figure 4. Model Evaluation Metrics Bar Chart

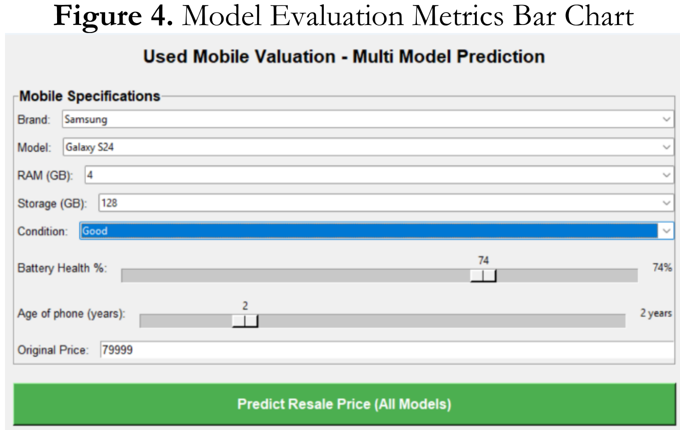

Figure 5. Interface of Project

The results of this study indicate that Machine learning models, specifically tree-based ensemble methods such as Gradient Boosting and XGBoost, performed exceptionally well for predicting smartphone resale prices. as visualized by the performance metrics in Figure 4 and the tight grouping of errors around the zero line in the residual plots shown in Figure 3, these advanced algorithms perform much better than traditional methods.

Figure 2 presents the correlation heatmap illustrating the relationship between smartphone features and resale price. The resale price shows a strong positive correlation with the original price ($r = 0.90$), indicating that devices with higher launch prices tend to retain greater resale value. Storage capacity also demonstrates a moderate positive correlation ($r = 0.69$), followed by RAM ($r = 0.43$), highlighting consumer preference for higher technical specifications.

Battery health exhibits a weaker yet positive relationship ($r = 0.18$), indicating that while it influences resale value, its impact is secondary to that of price and storage. Conversely, device age exhibits a negative correlation ($r = -0.12$), confirming that resale prices decline as smartphones age. These findings validate the inclusion of hardware specifications and depreciation factors for resale price prediction.

Figure 3 illustrates the residual plots for all models, where residuals are calculated as the difference between actual and predicted prices. Ideally, residuals should be randomly distributed around the zero-error line.

The Linear Regression model exhibited systematic patterns and wide error dispersion, particularly for higher-priced devices, indicating underfitting and violation of linear assumptions. The Decision Tree model improved performance but still exhibited large variance, especially at extreme price ranges.

In contrast, ensemble models such as Random Forest, Gradient Boosting, and XGBoost demonstrated tightly clustered residuals around the zero line, confirming reduced bias and improved generalization. The Gradient Boosting model shows the most compact residual distribution, supporting its superior predictive capability.

This ML method offers a distinct advantage over conventional shopkeeper price assessments by taking into account various elements. This system provides a straightforward and reliable predictive framework, depicted in Figure 5, that reduces human bias. The use of real-world and locally collected data verifies that the forecasts are connected to changing market trends. In summary, this method improves clarity in the used smartphone market, allowing buyers and sellers to make informed decisions based on reliable and verifiable data.

Conclusion:

This study focuses on developing a machine learning–based framework to estimate the resale value of used smartphones by creating a reliable dataset and comparing several regression models. The novelty of this research lies in the manual collection of data from online marketplaces and local mobile markets, along with the use of additional features such as battery health and device condition to better reflect real market conditions. Experimental results indicate that ensemble methods, particularly Gradient Boosting and XGBoost, achieved the highest prediction accuracy. The developed system reduces dependence on retailers and encourages transparent, data-driven decision-making, enabling both buyers and sellers to determine fair resale prices in the used smartphone market.

Future Work:

Future studies could build upon this research in various ways:

Gathering large datasets from various cities and online sources to improve the model's forecasting abilities.

Analyzing advanced neural network techniques and ensemble algorithms to boost overall performance.

Additional attributes such as warranty, ratings, and repair history can be included.

A website or mobile app can be created for user convenience.
For forecasting future resale trend a new time-series analysis can be applied based on historical price movements.

References:

[1] "Predicting The Price Of Used Electronic Devices Using Machine Learning Techniques." Accessed: Mar. 09, 2026. [Online]. Available: https://www.researchgate.net/publication/377526585_Predicting_The_Price_Of_Used_Electronic_Devices_Using_Machine_Learning_Techniques

[2] Seda İpek, Aksoy Ercan, "Mobile Phone Price Classification Using Machine Learning," *Int. J. Adv. Nat. Sci. Eng. Res.*, vol. 7, no. 4, 2023, [Online]. Available: https://as-proceeding.com/index.php/ijanser/article/view/791

[3] Ms.Mubeen Begum, SHAIK ABUBAKER SIDDIQ, "Using Machine Learning to Forecast Mobile Phone Prices," *Int. J. Eng. Res. Sci. Technol.*, vol. 21, no. 2, 2025, [Online]. Available: https://ijerst.org/index.php/ijerst/article/view/848

[4] Richard Honey, "Smartphone Price Prediction Using Machine Learning Techniques," *Int. J. Innov. Res. Eng.*, vol. 4, no. 2, pp. 603–607, 2023, [Online]. Available: https://www.theijire.com/archiver/archives/smartphone_price_prediction_using_machine_learning_techniques.pdf

[5] Rabiah Al-Qudah, Moayad Aloqaily, "Data-Driven purchasing strategies: Price prediction models and strategy development," *Expert Syst. Appl.*, vol. 266, p. 125986, 2025, [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417424028537

[6] Qipeng Liang, "Mobile phone price prediction: A comparative study among four models," *Appl. Comput. Eng.*, vol. 48, no. 1, 2024, [Online]. Available: https://www.researchgate.net/publication/379076746_Mobile_phone_price_prediction_A_comparative_study_among_four_models

[7] Luke Mangala Soegianto, Alvrian Timotius Hinandra, "Comparison of Model Performance on Housing Business Using Linear Regression, Random Forest Regressor, SVR, and Neural Network," *Procedia Comput. Sci.*, vol. 245, pp. 1139–1145, 2024, doi: https://doi.org/10.1016/j.procs.2024.10.343.

[8] "Intelligent Assessment of Secondhand Mobile Phone Prices by Machine Learning Techniques | Journal of Computing & Biomedical Informatics." Accessed: Mar. 09, 2026. [Online]. Available: https://jcbi.org/index.php/Main/article/view/351