Journal of Research in Artificial Intelligence for Systems and Applications (RAISA)
2025, Vol. 01, No. 02, E-ISSN : 3090-3157

Original Research

Advanced Feature-Driven Smartphone Price Prediction Using Machine Learning

Numan Shafi[1], Farzeen Atif [2], Attroba Atif[3], Robass Atif[4]

[1,2,3,4]University of Engineering & Technology, Lahore, Lahore, Pakistan

Abstract

A rapid growth occurred in the smart phone industry owing to the launch of numerous fresh devices across various price categories. It is advantageous to both the device makers and the users to be aware of the relative price of a smartphone which is in relation to its specifications. This research focuses on artificial intelligence for the purpose of making predictions on the price ranges of the devices with some specific attributes such as: Fingerprint Sensor, storage space, 5G Support, Water Resistance and Number of Sims, Face Lock. The last one is rather significant, as in some regions of the world devices fitted with one or two SIM cards or more are likely to be used most because of the different wide range of connectivity they offer. This methodology includes cleansing, transformation of data and applying suitable techniques from the machine learning. The developed model is precise and accurate and the output information of the model suffices a good number of questions that assist companies on how to set prices and also shape their clients' buying behaviors.

Keywords

Smart phone, Predictions, Artificial Intelligence, Face Lock, Machine learning.

1. Introduction

The smartphone market is changing quickly as a result of continuous technological change along with growing competition among the manufacturers. To reach as many potential customers as possible, the said companies are ready to implement the adequate pricing models with a support of new technologies like Artificial Intelligence (AI), Machine Learning (ML), and Deep learning (DL). Their role is crucial in processing large amounts of data, spotting opportunities, and forecasting the cost structure according to the different parameters of a smartphone. Manufacturers would be better in pricing and placing the product if AI and ML were used, while consumers would get helpful tips that made the decision-making process easier. Still, and despite high penetration of these advanced technologies, consumers have problems with the choice of a smartphone. In many instances, it is the number of specifications like Fingerprint Sensor, storage space, 5G Support, Water Resistance and Number of Sims, Face Lock that knocks the buyers out of the contest and prevents them from making a rational choice. This makes it important to use smart tools with AI to help people make good choices when buying things.

One of the major problems in many past works can be said to be the fact that the supplied datasets to the prediction models are frequently incomplete or out-of- date in regards to certain crucial aspects needed for accurate pricing predictions. For instance, studies like those conducted by Kalaivani et al. [1] and Subhiksha et al. [2] relied on outdated datasets, which omitted essential features now integral to smartphones, such as 5G connectivity, Finger Print Sensor, Face Lock, and Water Resistance options. As a result, these studies achieved limited success, with Subhiksha et al. [2] reporting only an 81% accuracy. Similarly, Chen [3] applied classification techniques

*Corresponding author: Robass Atif

Email addresses:
numanshafi1@gmail.com (Numan Shafi), farzeenabaid01@gmail.com (Farzeen Atif), attroba789@gmail.com (Attroba Atif), mohammadrobass@gmail.com (Robass Atif)

Received: 19 01 2024; **Accepted:** 25 Juli 2025; **Published:** 30 Juli 2025

for price prediction, but his model did not consider modern attributes, further limiting the relevance of the study.

Knowing the limitations such as in the previous sub- section, this study seeks to address these limitations by providing a wider and up to date set of features, including but not limited to 5G, Finger Print Sensor, face recognition, water proof features and the likes. The additional features also increase the specific relevance of the prediction models set and enable one to draw conclusions about the trends of the market at the time the models are utilized. This tool assists both the producers and the consumers to make better informed decisions. In such an industry where technology changes fast and consumers' tastes vary, Such decisions are more rational.

2. RELATED WORK

Because machine learning can examine complicated information and create very accurate predictive models, it has proven to be an incredibly useful technique for predicting smartphone prices. The potential of various ways to improve the accuracy and effectiveness of these models has been investigated and shown in a number of research. For example, Kalaivani et al. [1] focused on using older datasets for price prediction, bringing to light the challenges associated with outdated data and the need to adapt to rapidly changing market trends. Subhiksha et al. [2] achieved a notable accuracy of 81%in their predictive model but acknowledged that further improvements could be realized by incorporating updated, diverse, and representative datasets.

Chen [3] proposed an innovative classification-based model that utilized feature reduction techniques, sig- nificantly boosting prediction accuracy while reducing computational overhead. Kalmaz and Akin [4] demonstrated advanced techniques for refining price estimation, achieving high levels of precision by applying sophisticated computational methodologies. Dutta et al. [5] conducted a Detailed analysis of various super- vised learning algorithms, effectively highlighting their strength in classifying smartphone price categories with considerable reliability.

Table 1. The Comparing mode

Scientist	Finger Print	5G Support	Water Resistance	No of Sims	Face Lock
Kalaivani et al. [1]	Yes	No	No	No	No
Subhiksha et al. [2]	No	No	No	No	No
Chen [3]	No	No	No	No	No
Kalmaz & Akin [4]	No	No	No	No	Yes
Dutta et al. [5]	No	No	No	No	No
Kumuda et al. [7]	Yes	No	No	No	No
Pipalia & Bhadja [9]	No	No	No	No	No
Miao & Niu [8]	No	No	No	No	N0
Your Model	No	No	No	No	No

Kumuda et al. [7] showcased the advantages of hybrid machine learning techniques, emphasizing the potential for combining multiple methods to achieve superior predictive accuracy. Pipalia and Bhadja [9] provided a detailed analysis of supervised learning algorithms, offering critical insights into performance optimization and adaptability in price prediction scenarios. Moreover, Miao and Niu [8] underlined the importance of feature selection, demonstrating how it can enhance prediction outcomes, reduce computational costs, and streamline the model-building process.

The provided table I compares the features used in various studies related to smartphone price prediction. It lists the scientist names along with the key features they utilized, such as RAM, storage space, camera specifications, battery size, and prediction accuracy. Each study is evaluated based on whether it incorporates these features, with the table indicating 'Yes' or 'No' for each feature in the corresponding column. The model you proposed includes all these features, which is in contrast to the other studies that use fewer features or lack certain features altogether.

3. METHODOLOGY

The methodology comprises the following steps:

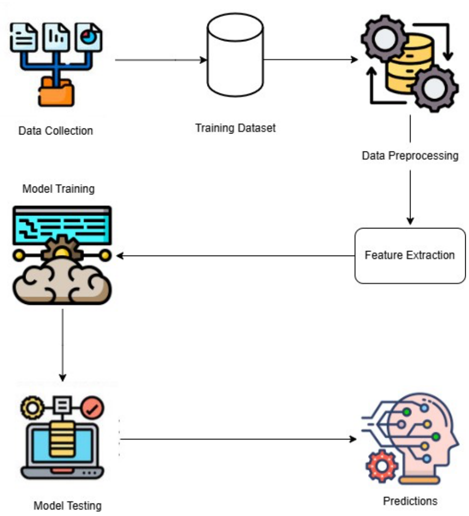

Figure 1. Flow Chart of Model

3. 1 Input Features

Table II presents a comprehensive overview of the features used in the dataset. The attributes capture the critical aspects of the smartphones, such as Name, which corresponds to the particular model, and Rating, which is based on the users' reviews. Additional details related to Performance and technical specifications are captured by attributes such as Spec_score, Processor and Pro- cessor name, whereas elements such as Ram, Battery, and Display portray the capabilities of the device. Other attributes, which include Camera, External_Memory and Inbuilt_memory, further emphasize the storage and the imaging capabilities. The dataset also addresses some fundamental features for users such as Android_version, Price and Company to serve as an overall perspective for each smartphone. To further accentuate their appeal, advanced functionalities like, Fast_charging and Screen_resolution are also included, making this dataset ready for profound examination and modeling.

Table 2. ***Input Features with Advanced Smartphone Attributes***

Attribute	Description	Type
Name	Smartphone model name.	Categorical
Rating	User rating of the smartphone.	Numerical
Spec score	Specification score of the device.	Numerical
No of sim	Number of SIM slots.	Numerical
Ram	Random access memory (GB).	Numerical
Battery	Battery capacity (mAh).	Numerical
Display	Screen size (inches).	Numerical
Camera	Rear camera resolution (MP).	Numerical
Front camera	Front camera resolution (MP).	Numerical
External Memory	External memory support.	Boolean
Android version	Android OS version.	Numerical/ Categorical
Price	Smartphone price.	Numerical
Company	Manufacturer or brand.	Categorical
Inbuilt memory	Internal storage (GB).	Numerical
Fast charging	Fast charging support.	Boolean
Screen resolution	Screen resolution.	Categorical
Processor	Processor type.	Categorical
Processor name	Processor model.	Categorical
Refresh rate	Display refresh rate (Hz).	Numerical
Fingerprint sensor	Presence of a fingerprint sensor.	Boolean
Face unlock	Face unlock feature support.	Boolean
5G support	Support for 5G connectivity.	Boolean
Water resistance	Water resistance rating (e.g., IP68).	Categorical
NFC	Near Field Communication (NFC) support.	Boolean
Dual speakers	Availability of stereo speakers.	Boolean

3. 2 Data Preprocessing

The clever techniques for data preprocessing improve the quality and use of the data with the aim of achieving stronger modeling. Below are the few important actions that were taken, backed up by visuals.

Encoding and Transformations Categorical variables, such as No of_sim, were encoded using dummy variables, expanding them into binary features for better model compatibility. Columns like Battery, Ram, and Display were cleaned and converted to appropriate numerical formats (e.g., mAh, GB, and inches).

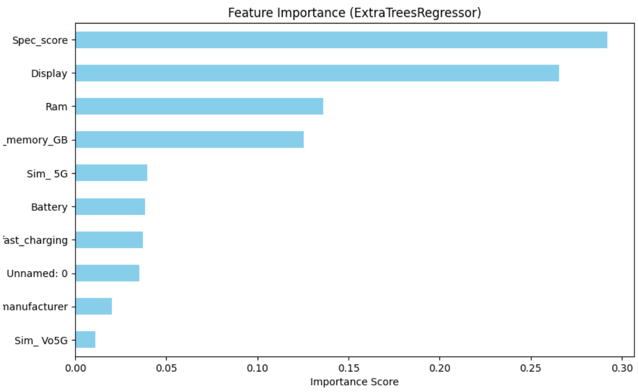

Figure 2. **Feature Importance (Extra Trees Regressor)**

Figure 2 illustrates the importance of various features in the prediction task. The most influential features, such as Ram, Processor name manufacturer, and Battery, were identified as having the greatest impact on price prediction. These features exhibited higher importance scores, highlighting their critical role in determining the target variable and contributing significantly to the model's performance.

Standardization and Feature Selection Company names and categorical values (e.g., Processor_name_manufacturer) were standardized for consistency. Text values were encoded numerically using Label Encoder. Features irrelevant to the target variable (e.g., External_Memory, Screen_resolution) were dropped.

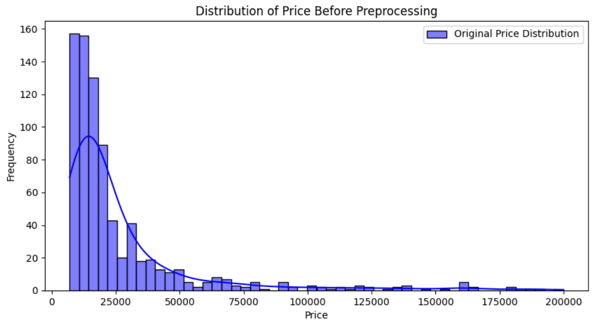

Figure 3. **Price Distribution Before Preprocessing**

As shown in Figure 3, the price distribution before preprocessing reveals a positive skew, with a higher frequency of lower prices observed between 25,000 and 150,000. This implies that more reasonably priced goods or services are more common within the dataset since the bulk of the data points are concentrated in the lower price range. The frequency of occurrences dramatically drops as the price rises over this range, emphasizing that more expensive things are comparatively uncommon. Before using specific machine learning models or analysis tools, such a distribution could indicate the presence of outliers or the necessity of data transformation techniques, like log transformation, to normalize the distribution. Understanding the nature of the data requires this insight, which can also help with preprocessing choices like how to handle outliers or skewed distributions in later stages of analysis.

Train-Test Split and Feature Importance The dataset was split into training (80%) and testing (20%) subsets, ensuring reliable evaluation.

An Extra Trees Regressor model revealed the importance of features in predicting the target (Price).

3. 3 Feature Engineering

To establish a relationship between price ranges and the relevant features, key determinants were analyzed. Correlation analysis and determination of feature importance were performed to improve the efficiency of the models.

A number of advanced algorithms were tried with following aim of predicting price for smartphone devices. As a point of departure classifier, logistic regression was employed, which is simple to interpret and effective, but it does not seem to be able to explain intricate patterns within the data.

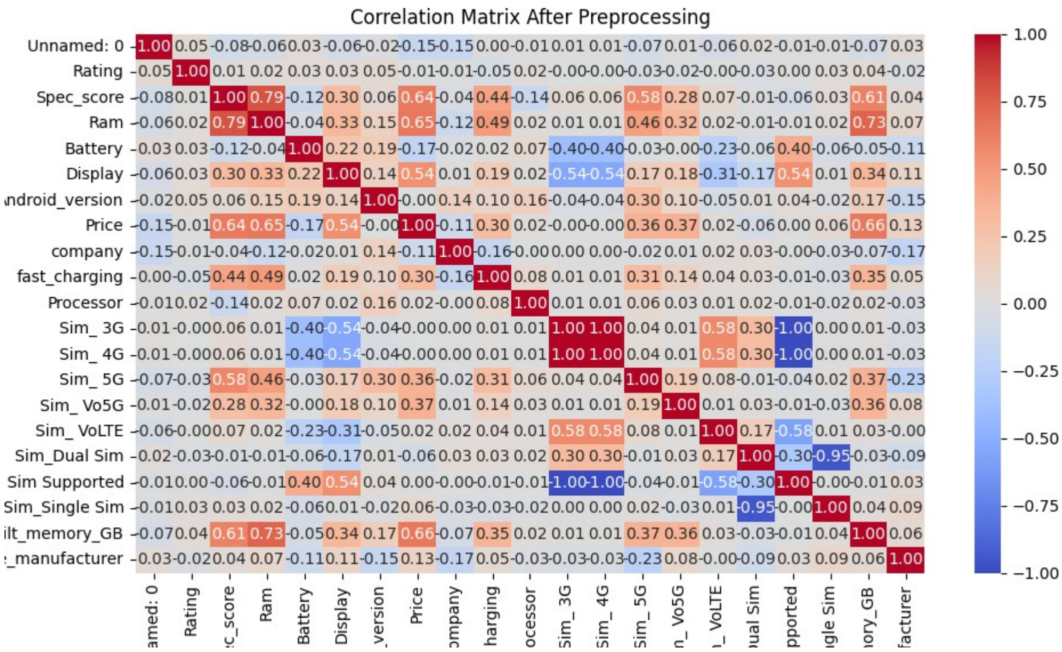

Figure 4. **Correlation Heatmap After Preprocessing**

The decision trees were able to provide decent models that could manage non-linear relationships although they would be susceptible to over-learning of smaller datasets. Random forests are ensemble approaches that facilitate accuracy and generalization by voting across a number of decision trees.

Gradient Boosting is an effective boosting approach that achieved better results on high dimensional datasets by integrating weak learners.

Extra trees regressor is another ensemble method that mitigates overfitting and decrease variance by capturing complex features through averaging of predictions of random decision trees.

A custom ensemble model fused multiple techniques' advantages, hence best performed on the dataset maximizing accuracy and generalization.

The Ensemble Model was chosen as the final model due to its superior performance, achieving a highest accuracy. This result underscores its strength and capability to handle complex and high-dimensional data. The Extra Trees Regressor also performed well, achieving an accuracy of 89.45%, and was instrumental in forming the ensemble.

Visualization techniques, including scatterplots, histograms, and boxplots, were employed to analyze and compare the actual and predicted prices. These plots highlighted the model's precision and areas for potential refinement. In conclusion, the Ensemble Model proved to be the most reliable and effective solution for predicting smartphone prices in this study.

3. 4 Model Comparison and Selection

In order to establish the best model for predicting smartphone prices, various machine learning algorithms were applied and compared against each other. The models were robust in that they were able to extend their knowledge to new and unseen data such that they would perform well in real life.

The group's Output Models Are Classified As:

Logistic Regression: A basic and uncomplicated model that serves as the source of reference in other comparisons. While it serves its purpose well in fitting linear relationships, it is quite ineffective in identifying intricate relationships.

Decision Trees Models: These models have the ability to deal with nonlinear dependencies and make interpretations that make sense and are easy to explain. However, they often over-fit smaller data sets which reduces their generalizing capabilities.

Random Forest: A hybrid method that combines several decision trees so that the average of their outputs is produced to enhance accuracy. It improves generalization and reduces the amount of overfitting that occurs when a single decision tree is used.

Gradient Boosting: A very rich model that creates a strong model predictor by applying many weak learners sequentially. It works extremely well with high dimensional datasets enabling it to outperform other methods. Extra Trees Regressor: This model is basically a random forest but constructs the decision trees in a different manner by randomly picking the features for each tree.

This model fine-tunes complex structures and over-all gives a reasonable bias variance trade off.

A final Ensemble Model was created by combining the strengths of the aforementioned algorithms. This ensem- ble approach improved overall performance, achieving an highest accuracy. The ensemble model demonstrated robust generalization capabilities and excelled in predict- ing smartphone prices, outperforming individual models such as Random Forest and Extra Trees Regressor.

4. Results and Discussion

Table 3 summarizes the performance of the evaluated algorithms.

Table 3. **Accuracy of Different Machine Learning Algorithms**

Algorithm	Accuracy	Precision	Recall	F1-Score
Logistic Regression	77.45%	0.76	0.75	0.76
Decision Trees	84.25%	0.83	0.84	0.83
Random Forest	89.45%	0.88	0.89	0.88
Ensemble Model	94.47%	0.90	0.91	0.91

The results indicate that the ensemble model out performed the other algorithms with a highest accuracy, demonstrating its ability to effectively capture the complexities in the smartphone price prediction task.

4. 1 Ensemble Model Results

The Extra Trees Regressor achieved an accuracy of 89%, indicating strong predictive performance. As shown in Figure 5, the relationship between actual and predicted prices is visualized using a scatter plot. In com- parison, the ensemble model achieved a higher accuracy, illustrating the benefit of combining multiple models to enhance performance

4. 2 Discussion

Overall, the scatter plot confirms existence of strong linear relationship between actual prices and predicted ones as most of the plotted points are closely around the line that runs through the origin on the one-diagram, which represents the ideal prediction, (y = x).

However, there are predictions for that range of prices as well, but they tend to be forecasts that are slightly below reality. This could be due to not enough of such data points that fall within those price ranges being included in the training set, resulting in higher price predictions not being precise.

The ensemble model provided an accuracy of 94.47% which proves that the model was able to learn the intricacies that existed within the dataset and generalize well to unseen data. This result shows that the ensemble approach is a more effective model than the individual models Extra Trees Regressor, Random Forest and Gradient Boosting in regard to predictions of smartphone prices based on the chosen parameters.

Figure 5. **Actual vs Predicted Prices for Ensemble Model**

5. Conclusion

This research implements the various techniques of machine learning in predicting the prices of smartphones' based on, their specifications, models, and brands. The best performance among the methods tested was achieved with the Ensemble Model yielding an accuracy of approximately 94%, indicating the advantages of multiple models combining to provide a more accurate estimate than one model. Results show that the prediction of prices can be effectively improved by more sophisticated machine learning algorithms like Gradient Boosting and Ensemble methods.

Acknowledgements (14 Pt)

All sections of this paper were conceptualized and written by the author, Robass Atif. The study was carried out under the supervision of Dr. Numan Shafi, who provided valuable guidance on the use of Machine Learning algorithms. The paper was reviewed and ap- proved by Attroba Atif and Farzeen Atif, whose insights contributed to its refinement. A heartfelt appreciation goes to everyone who assisted in enhancing this research. The author confirms that no funding was received for this research.

References

[1] K. S. Kalaivani, N. Priyadharshini, S. Nivedhashri, and R. Nandhini,'Predicting the price range of mobile phones using machine learning techniques,' AIP Conference Proceedings, vol. 2387, no. 1, 2021, AIP Publishing.

[2] S. Subhiksha, S. Thota, and J. Sangeetha, ”Prediction of phone prices using machine learning techniques,” Data Engineering and Communica- tion Technology: Proceedings of 3rd ICDECT-2K19, pp. 781–789, 2020, Springer.

[3] Chen, M., "Mobile Phone Price Prediction with Feature Reduction," Highlights in Science, Engineering and Technology, vol. 34, pp. 155- 162, 2023.

[4] Kalmaz, A., & Akin, O., "Estimation of Mobile Phone Prices with Machine Learning," in 2022 International Conference on Engineering and Emerging Technologies (ICEET), pp. 1-7, IEEE, Oct. 2022.

[5] Dutta, A., Mallick, P. K., Mohanty, N., & Srichandan, S., "Supervised Learning Algorithms for Mobile Price Classification," in Cognitive Informatics and Soft Computing: Proceeding of CISC 2021, pp. 653- 666, Springer Nature Singapore, 2022.

[6] Global smartphone sales to end users since 2007, Statista. Available: https://www.statista.com/statistics/263437/global-smartphone-sales-to- end-users-since-2007/.

[7] Kumuda, S., Vishal Karur, and Karthick Balaje S. E., "Prediction of Mobile Model Price using Machine Learning Techniques," International Journal of Engineering and Advanced Technology, vol. 11, no. 1, pp. 273-275, 2021.

[8] J. Miao and L. Niu, "A survey on feature selection," Procedia Computer Science, vol. 91, pp. 919-926, 2016.

[9] K. Pipalia and R. Bhadja, "Performance Evaluation of Different Su- pervised Learning Algorithms for Mobile Price Classification," In- ternational Journal for Research in Applied Science & Engineering Technology, vol. 8, no. 6, pp. 1841-1848, 2020.

[10] S. Dreiseitl and L. Ohno-Machado, “Logistic regression and artificial neural network classification models: a methodology review,” Journal of Biomedical Informatics, vol. 35, no. 5-6, pp. 352-359, 2002.

[11] Mobile PriceClassification Dataset, Kaggle. Available: https://www.kaggle.com/datasets/iabhishekofficial/mobile-price- classification.

[12] A. Famili, W. M. Shen, R. Weber, and E. Simoudis, “Data preprocessing and intelligent data analysis,” Intelligent Data Analysis, vol. 1, no. 1, pp. 3-23, 1997.

[13] P. J. M. Ali, R. H. Faraj, and E. Koya, “Data normalization and standardization: a technical report,” Machine Learning Technical Report, vol. 1, no. 1, pp. 1-6, 2014.

[14] F. Murtagh, “Multilayer perceptrons for classification and regression,” Neurocomputing, vol. 2, no. 5-6, pp. 183-197, 1991.

[15] T. Menzies, E. Kocaguneli, B. Turhan, L. Minku, and F. Peters, “Sharing data and models in software engineering,” Morgan Kaufmann, 2014.

[16] V. Nair and G. E. Hinton, “Rectified linear units improve restricted Boltzmann machines,” in International Conference on Machine Learn- ing, 2010.

[17] P. Schober, C. Boer, and L. A. Schwarte, “Correlation coefficients: appropriate use and interpretation,” Anesthesia & Analgesia, vol. 126, no. 5, pp. 1763-1768, 2018.

[18] S. Wold, K. Esbensen, and P. Geladi, “Principal component analysis,” Chemometrics and Intelligent Laboratory Systems, vol. 2, no. 1-3, pp. 37-52, 1987.

[19] Asim, M., & Khan, Z. (2018). Mobile price class prediction using machine learning techniques. International Journal of Computer Ap- plications, vol. 179, no. 29, pp. 6-11.

[20] Sakib, A. H., Shakir, A. K., Sutradhar, S., Saleh, M. A., Akram, W., & Biplop, K. B. M. B. (2022, January). A hybrid model for predicting Mobile Price Range using machine learning techniques. In 2022 The 8th International Conference on Computing and Data Engineering, pp. 86-91.

[21] Gu¨venc¸, E., C¸ etin, G., & Koc¸ak, H. (2021). Comparison of KNN and DNN classifier’s performance in predicting mobile phone price ranges. Advances in Artificial Intelligence Research, vol. 1, no. 1, pp. 19-28.