

Machine Learning Driven Mobile Price Prediction Using Feature Selection and Parameter Optimization

Sarika N. Kamble^{1,*}, Rutuja Kanse², Rushil Jangam³

Abstract

Machine learning calculations are utilized in many fields like money, training, industry, medication, and online business. Machine learning calculations show execution contrasts relying upon the dataset and handling steps. Picking the right calculation, preprocessing and post-handling techniques have incredible significance in accomplishing great outcomes. The Random Forest classifier, K-nearest neighbor classifier, and support vector machine methods are evaluated to forecast mobile phone price categories. The “prediction” dataset which is taken from kaggle.com is utilized to assess strategies. First, the dataset is verified and cleaned. From that point onward, scaling is applied to datasets to get more important information for machine learning calculations. Then, highlight choice strategies that lessen the computational expense by decreasing the quantity of data sources are performed to get significant elements. At long last, the boundaries of order calculations are tuned to further develop the framework exactness. After Feature selection accuracy of models got more with selected features. It gives fulfilling precision with a base number of highlights. It is likewise seen that the support vector machine and gradient boosting have the most accurate results than the other four models.

Keywords: Random forest classifier, gradient boosting, support vector machine, feature selection, parameter optimization, decision tree

INTRODUCTION

The worldwide mobile phone market has experienced exponential expansion in recent years, marked by a proliferation of models that meet the varying demands and tastes of consumers. Because many options are available, choosing a mobile phone that not only matches their needs but also their budget may be a difficult challenge for consumers. Given this difficulty, the use of machine learning techniques to estimate mobile prices has emerged as a viable approach that can provide useful information to stakeholders in the sector as well as consumers. The goal of this project is to create and assess machine

learning models that can correctly forecast mobile phone pricing based on a collection of pertinent variables. Advanced algorithms and techniques such as feature selection and parameter optimization have been utilized. When it comes to buying a mobile phone, customers in today's fast-paced world have several options. With many different models from different manufacturers, each with an abundance of features and specifications, it can be difficult for customers to choose the model that best suits their demands and budget. This project intends to provide users with useful insights into how mobile phone qualities and pricing relate to each other by creating precise machine learning models for mobile price prediction. Equipped with this understanding, buyers may make more

*Author for Correspondence

Sarika N. Kamble
E-mail: sarika.kamble_skncoe@sinhgad.edu

¹Assistant Professor, Department of Electronics & Telecommunication Engineering, Smt. Kashibai Navale College of Engineering, Savitribai Phule Pune University, Pune, Maharashtra, India

²⁻³Student, Department of Electronics & Telecommunication Engineering, Smt. Kashibai Navale College of Engineering, Savitribai Phule Pune University, Pune, Maharashtra, India

Received Date: July 07, 2024
Accepted Date: July 26, 2024
Published Date: September 11, 2024

Citation: Sarika N. Kamble, Rutuja Kanse, Rushil Jangam. Machine Learning Driven Mobile Price Prediction Using Feature Selection and Parameter Optimization. Current Trends in Information Technology. 2024; 14(3): 18–25p.

intelligent choices, guaranteeing that they have a smartphone that fits their needs and provides the most value to their investment. The changing pricing dynamics of the mobile phone industry make it unique; demand for products, competition, and developments in technology all influence price changes. Nonetheless, consumers may find it difficult to determine the real worth of a mobile phone because of secrecy regarding the pricing tactics used by manufacturers and sellers. This initiative enhances market transparency through price predictions using machine learning techniques. This project intends to inspire consumers to make better-educated purchasing decisions and to confidently traverse the mobile phone market by providing them with accurate price projections according to a reliable analysis.

It is imperative to use machine learning for mobile price prediction, together with feature selection and parameter optimization because of its ability to empower customers, improve market transparency, optimize spending, and provide industry stakeholders with insightful information. This project tackles important issues for industry players and customers alike by creating reliable and accurate machine learning models specifically for the mobile phone market. Ultimately, this leads to a competitive, informed, and efficient market.

RELATED WORK

The paper titled “*Mobile Phone Price Class Prediction Using Different Classification Algorithms with Feature Selection and Parameter Optimization*” (2021) [1] describes how machine learning algorithms are used in various fields such as finance, education, industry, medicine, and e-commerce, as explained by Istanbul Technical University. The dataset was obtained from Kaggle.com, which was used to evaluate the methods. Then, they checked the dataset values. Choosing the correct model is of great importance for achieving good results. To obtain more relevant data, the authors applied a scaling method to obtain additional relevant data. They then applied feature selection methods, which provided greater accuracy in the model. Finally, the parameters of the classification algorithms are determined to obtain meaningful features.

The paper titled “*Prediction of Mobile Model Price Using Machine Learning Techniques*” (2021) [2] by Kumuda, Vishal Karur, and Karthick Balaje discussed various mobile specifications and new designs. The authors classified mobile prices based on these features. In this prediction, the dataset was collected from the existing market, and different algorithms were applied to reduce the complexity, identify the major selection features, and obtain the best comparison within the data. This tool was designed to determine the best price with maximum specifications.

The paper titled “*Classification of Mobile Phone Price Dataset Using Machine Learning Algorithms*” (2022) [3] authored by researchers from the Department of Mathematics at the University of London, examines mobile phone attributes. This study addresses the challenge of predicting price levels by employing machine learning algorithms such as support vector machine (SVM), decision tree, K-nearest neighbors (KNN), and Naïve Bayes to train the dataset. The performance of these algorithms is evaluated based on metrics such as accuracy, precision, recall, and F1 score. The results show that among the four classifiers, SVM delivers the best performance with 94.8% accuracy, 97.3 F1 score (without feature selection), 95.5% accuracy, and 97.7 F1 score (with feature selection).

Ensemble-based feature selection and classification of gene expression using a support vector machine, K-nearest neighbor, decision tree (2019) [4] focused on three main classification algorithms: SVM, KNN, and decision tree. These algorithms enhance performance and increase confidence in the results. The ensemble was performed using the majority-voting method, which outperformed the base classifier. To increase the efficiency of the classifier, a signal-to-noise ratio algorithm is used to obtain the most relevant genes.

In the study titled “*Prediction of Mobile Phone Price Class using Supervised Machine Learning Techniques*”, the author developed a model to predict the price class of mobile phones based on their

specifications and seeks to identify the most accurate machine learning algorithm for this task. The mobile price class dataset sourced from the Kaggle data science community was used. Various classification algorithms were employed to train the model to determine the algorithm that could most accurately predict the mobile price class. Metrics, such as the accuracy score and confusion matrix, were used to evaluate the trained model and identify the most suitable algorithm. Predictive machine learning uses input data to develop and train a model, which is subsequently used to predict future data instances. Supervised machine learning algorithms utilize data with predefined class labels, which are attributes to be predicted [5].

The study titled “*Mobile Phone Price Prediction Based on Supervised Learning Algorithms*” highlights that mobile phones have become essential for many people, with 91% of the global population owning a mobile phone and 83.3% specifically owning a smartphone. Over 170 mobile phone brands from 38 countries compete in both national and international markets. Market acceptance of mobile phones depends on various factors, including price, features, shape, and brand. In such a competitive market, price plays a crucial role in determining mobile phone success. Some studies have predicted mobile phone price ranges using machine learning algorithms, such as neural networks, KNNs, deep neural networks, random forests, logistic regression, decision trees, linear discriminant analysis, and support vector machines. These studies focused on predicting the price classes of mobile phones rather than actual prices [6].

PROPOSED SYSTEM

The main goal of this project is to create a reliable machine learning model that can reliably predict mobile phone pricing. Consumers are confronted with a bewildering array of mobile devices in today's fast-paced technological landscape, with each offering a unique set of capabilities at different price points. Accurately estimating a mobile phone's price can help manufacturers and customers make wise choices. This study uses cutting-edge machine learning approaches to accomplish this goal, with a particular emphasis on feature selection and parameter optimization to improve prediction performance. A key component of model building is featuring selection, which involves removing superfluous or irrelevant features while selecting the most pertinent ones that have a substantial impact on price prediction [7]. This reduces complexity, enhances readability, and prevents overfitting. In addition, parameter optimization seeks to optimize the hyperparameters of the machine learning methods of the model, which may include support vector machines, decision trees, and random forests. We can optimize the predictive accuracy and generalization skills of the model by methodically modifying these parameters to guarantee trustworthy price forecasts for a variety of mobile devices. The proposed task starts with data collection, which involves gathering a wide dataset of mobile phone characteristics and associated costs from reliable sources. The algorithms used for machine learning were trained, validated, and tested using the dataset. The most informative features for price prediction are then identified using a thorough feature engineering and selection process. Techniques, such as principal component analysis and correlation analysis, can be utilized for this purpose. In parallel, a range of machine learning methods will be investigated and their predictive power for mobile phones will be assessed. To maximize the predictive power of these algorithms, they undergo a rigorous parameter-tuning process that makes use of strategies such as grid search and random search. Comprehensive assessment measures, including the mean absolute error, root mean square error, and R-squared value for the training and testing datasets, were used to evaluate the effectiveness of the proposed model. To guarantee the model's usefulness in actual situations, its interpretability and robustness will also be examined [8].

Architecture of System

The front-end interface of the system architecture includes a web component that facilitates user interaction. Users can interact with the given features of the backend model and obtain information about mobile prices. Backend services include machine learning to respond to customer inquiries and a prediction engine that uses machine learning techniques to provide mobile price predictions. The backend provides the entire dataset of operations and programming used for the interface. Several

important components are included in the backend. Data preparation is the first step, which includes cleaning and preparing the unprocessed mobile phone dataset for analysis [9]. This involves scaling numerical features, encoding categorical variables, and handling missing values. Subsequently, the pre-processed data were used for model training, employing various machine learning methods, such as decision trees, random forests, and support vector machines, to develop the prediction model. The hyperparameters of the machine learning algorithms are then adjusted for improved prediction accuracy using parameter optimization approaches such as grid search and random search. Finally, based on user input, an Application Programming Interface (API) is created to provide predictions, enabling easy integration with online interfaces or mobile applications for real-world applications, as shown in Figure 1.

METHODOLOGY

The “mobile price prediction” system was created using a thorough design process that ensures efficiency and accuracy as shown in Table 1. Systematic data gathering with features, including price range, battery power, RAM, processor cores, SIM status, and other mobile information, is part of the first step. The prediction engine uses machine learning algorithms and statistical models to generate precise price range predictions, and the assembled dataset forms its foundation. The following are the various modules that constitute the proposed system.

Dataset Description

In the development of “mobile price prediction” a prediction dataset is employed to improve accuracy and efficiency. This dataset provides various mobile features for performing operations. It has features, including price range, battery power, RAM, cores, sim status (whether it is dual sim or not), mobile weight, and depth. In this dataset, the “price range” serves as the target variable, representing the primary focus of the predictive analysis. The goal is to leverage machine learning methodologies to accurately forecast the price range of mobile phones by utilizing pertinent features extracted from the dataset.

Dataset Processing

A variety of techniques are used to process data, which is necessary for the system to work.

Data Cleaning

This step provides a technique for cleaning data, checking missing values, detecting outliers if needed, and standardizing them. Techniques, such as imputation or removal, are employed to address missing data points and ensure the completeness of the dataset.

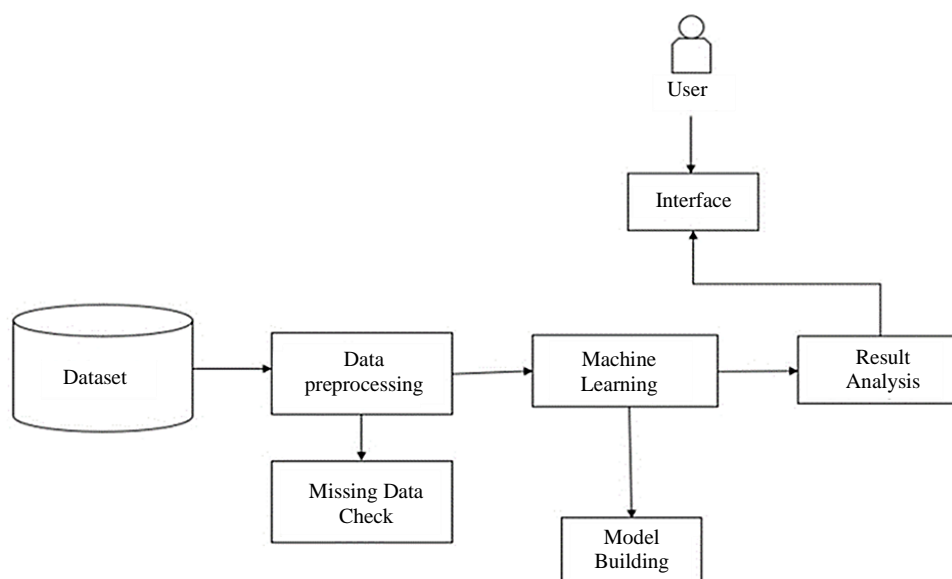


Figure 1. Block diagram of mobile price prediction.

Outliers were identified through statistical methods or domain knowledge, and appropriate treatment strategies such as trimming or transformation were applied to mitigate their influence on the analysis. Numerical features are scaled to a common scale using methods such as normalization or standardization to promote consistency and comparability across variables.

Feature Selection

Selected features with potential predictive power for the mobile phone price range, considering factors such as specifications, brand, and market trends, are extracted. Engineered new features through transformations, interactions, or domain-specific knowledge to enhance the predictive capability of the model. Only those features that are important and can increase the efficiency of the model were selected.

Here, the correlation between the target variables and each feature is calculated as:

$$\text{Correlation} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}}$$

Encoding Categorical Variables

Applied techniques, such as one-hot encoding or label encoding, convert categorical variables into numerical representations, facilitating their incorporation into machine learning algorithms. Using this encoding technique, string data that would not have been supported by a computer were encoded. This process involves assigning a unique integer to each category.

Data Splitting

The dataset was partitioned into training, validation, and testing sets to enable model training, evaluation, and validation and to provide an unbiased performance assessment. The X-category data were divided into training and test data. Y categorical data, which is a target variable, are also divided into training and test data. Data splitting is based on time. The dataset was split based on time, where earlier data were used for training and validation and later data were used for testing. This is common in time series analyses.

Model Selection

Model selection is the process of determining which machine learning algorithm or model structure is best suited for a particular task based on complexity, interpretability, performance measures, and other pertinent considerations. In this system, five to six machine learning algorithms were employed to train the model.

K-nearest Neighbor Classifier

K-nearest neighbor classifier (KNN) is a non-parametric algorithm used for both classification and regression tasks. It predicts the label or value of a new data point based on the majority class or the average of its nearest neighbors in the training dataset. For the predicted price range, \hat{y} is determined as follows.

$$\hat{y} = \text{mode}(\text{neighbor})$$

Y neighbor represents a set of price ranges closest to the input data point.

Random Forest Classifier

Renowned for its robustness and versatility, the Random Forest model operates by constructing a large number of decision trees during the training phase. Together, the individual trees in the ensemble contribute to the final forecast by separately processing the input variables, which include characteristics like the price range, battery power, and RAM. The main advantage of Random Forest is its capacity to combine the predictions of several trees to reduce overfitting and improve prediction accuracy. In our particular application, the model predicts which mobile device will have a higher price

using features such as price range, battery power, and RAM. The predictive formula for the Random Forest classifier (RFC) is provided as follows:

$$\hat{y} = \text{Mode}(y_1, y_2, \dots, y_n)$$

Support Vector Machine

Support vector machine is a supervised machine learning model employed for both classification and regression tasks. It identifies the optimal hyperplane that separates data points into distinct classes with the maximum margin. An SVM can handle high-dimensional data and can effectively address nonlinear relationships using kernel functions. It is robust to overfit and works well for both linearly separable and non-separable datasets, making it widely used in various domains. We devised a technique for computing the mobile price range based on different machine learning models and data processing.

RESULT AND ANALYSIS

In this section, we compare the effectiveness of many machine learning models developed for mobile pricing prediction. A test dataset was used to assess each model's accuracy, and the results are compiled in the following chart [10]. By contrasting their accuracies, we can learn more about how well the various models forecast mobile pricing. This research helps in making better decisions and mobile pricing strategies by determining which model is most suited for implementation in practical situations, as shown in Figure 2.

We also compared all the machine learning algorithms in Figure 3. The user is provided with an interface to select the desired feature requirements. Various options display features that are selected for prediction.

Table 1. Comparison of machine learning SVM, RF, K-NN.

Algorithm	Accuracy	Precision	Recall	F1 score
Support vector machine	0.75	0.763	0.75	0.749
Random Forest	0.805	0.811	0.802	0.805
K-nearest neighbor	0.760	0.776	0.767	0.770

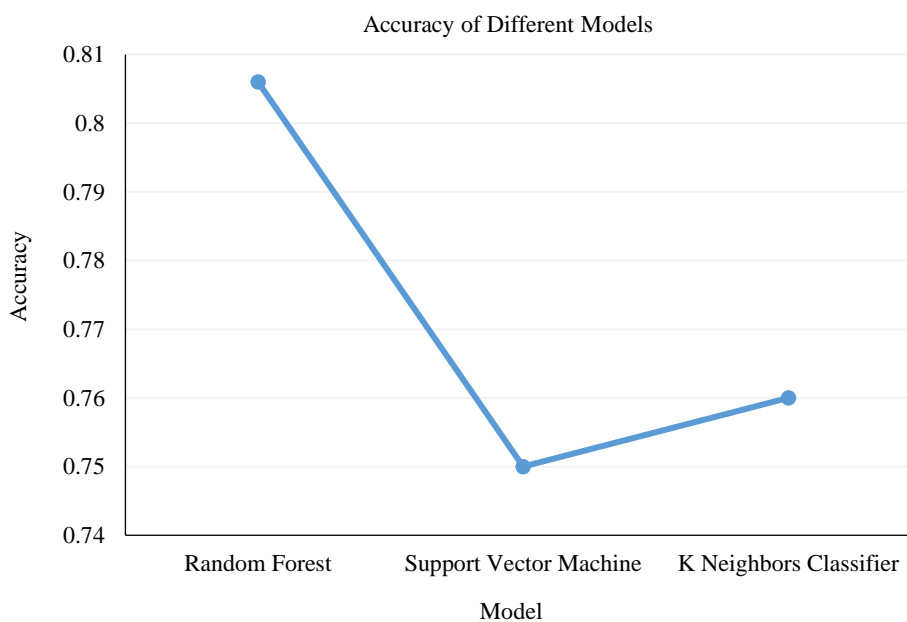


Figure 2. Comparison of accuracy among machine learning algorithms.

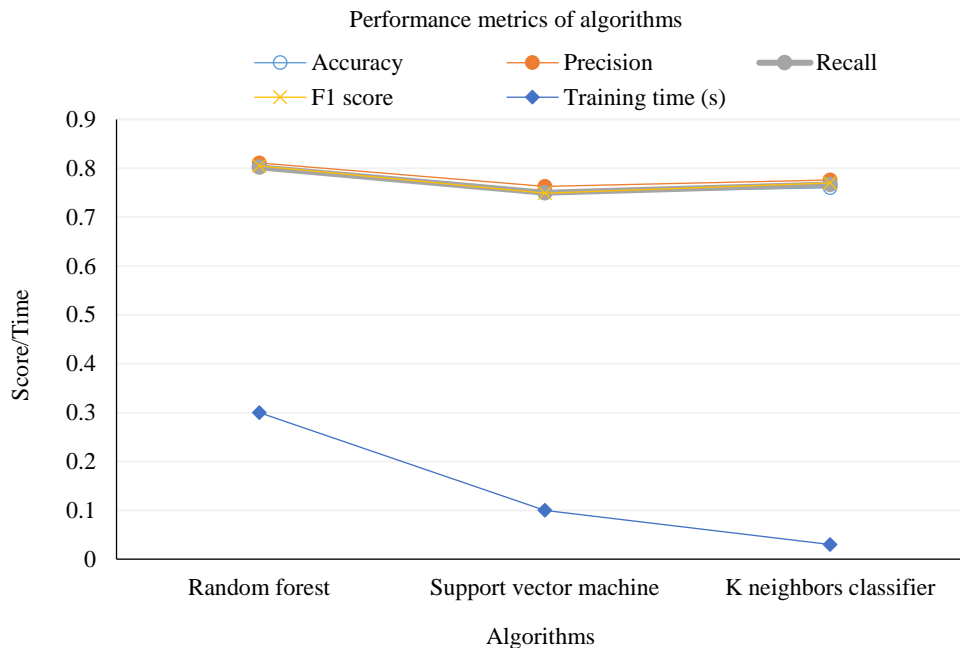


Figure 3. Comparison of machine learning algorithms.

CONCLUSIONS

The machine learning-based mobile pricing prediction system is a noteworthy development in the practical use of artificial intelligence. Utilizing machine learning techniques, the system provides a reliable method for determining mobile phone costs based on features. In conclusion, the proposed machine learning based mobile pricing prediction system that has been suggested is a complex solution with wide-ranging effects. In addition to addressing the pressing need for precise pricing predictions in the mobile phone market, this study paves the way for upcoming advancements in pricing techniques in a variety of industries. These types of predictive models will become increasingly essential as technology develops to improve consumer experience and market dynamics.

Acknowledgment

We also express our gratitude to, *Prof. S.N. Kamble (professor, department of Electronics and Telecommunication Engineering, Smt. Kashibai Navale College of Engineering, Vadgaon Bk, Pune, Maharashtra, India.)* for sharing their pearls of wisdom with us during this study. We are also immensely grateful to him for his comments on an earlier version of the manuscript, although any errors were our own and should not tarnish the reputation of these esteemed people.

REFERENCES

1. Çetin M, Koç Y. Mobile phone price class prediction using different classification algorithms with feature selection and parameter optimization. 021 5th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), Ankara, Turkey, 2021, pp. 483–7. DOI: 10.1109/ISMSIT52890.2021.9604550.
2. Kumuda K, Karur V, SE KB. Prediction of mobile model price using machine learning techniques. *Int J Eng Adv Technol.* 2021;11:273–5. DOI: 10.35940/ijeat.A3219.1011121.
3. Hu N. Classification of mobile phone price dataset using machine learning algorithms. 2022 3rd International Conference on Pattern Recognition and Machine Learning (PRML), Chengdu, China, 2022, pp. 438–43. DOI: 10.1109/PRML56267.2022.9882236.
4. Nair AJ, Rasheed R, Maheeshma KM, Aiswarya LS, Kavitha KR. An ensemble-based feature selection and classification of gene expression using support vector machine, K-nearest neighbor, decision tree. 2019 International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 2019, pp. 1618–23. DOI: 10.1109/ICCES45898.2019.9002041.

-
5. Varun Kiran A, Jebakumar R. Prediction of mobile phone price class using supervised machine learning techniques. *Int J Innov Sci Res Technol*. 2022;7(1):248–51.
 6. Maesya A, Yanfi Y. Mobile phone price prediction based on supervised learning algorithms. *Int J Appl Eng Technol*. 2023;5:41–4.
 7. Aksoy Ercan Sİ, Şimşek M. Mobile phone price classification using machine learning. *Int J Adv Nat Sci Eng Res*. 2023;7:458–62. DOI: 10.59287/ijanser.791.
 8. Kumar M, Pilaian U, Varshney C. Predicting mobile phone prices with machine learning. 2023 3rd International Conference on Advancement in Electronics & Communication Engineering (AECE), Ghaziabad, India, 2023, pp. 181–5. DOI: 10.1109/AECE59614.2023.10428614.
 9. Reddy KVK, Reddy BE. SBM for mobile phone price classification. 2023 3rd International conference on Artificial Intelligence and Signal Processing (AISP), VIJAYAWADA, India, 2023, pp. 1–5. DOI: 10.1109/AISP57993.2023.10134978.
 10. Subhiksha S, Thota S, Sangeetha J. Prediction of phone prices using machine learning techniques. In: Raju KS, Senkerik R, Lanka SP, Rajagopal V, editors. *Data Engineering and Communication Technology. Advances in Intelligent Systems and Computing*, vol 1079. Singapore: Springer; 2020. p. 781–9. DOI: 10.1007/978-981-15-1097-7_65.