Uluslararası İleri Doğa Bilimleri ve Mühendislik Araştırmaları Dergisi
Sayı 7, S. 458-462, 4, 2023

Araştırma Makalesi

IJANSER
https://as-proceeding.com/index.php/ijanser
ISSN: 2980-0811

International Journal of Advanced Natural Sciences and Engineering Researches
Volume 7, pp. 458-462, 4, 2023

Research Article

Mobile Phone Price Classification Using Machine Learning

Seda İpek AKSOY ERCAN*, Murat ŞİMŞEK [2]

[1]*Computer Engineering Department, Ostim Technical University, Türkiye*
[2] *Artificial Intelligence Engineering Department, Ostim Technical University, Türkiye*

siya.kdy@gmail.com

(Received: 12 May 2023, Accepted: 30 May 2023)

(DOI: 10.59287/ijanser.791)

(1st International Conference on Contemporary Academic Research ICCAR 2023, May 17-19, 2023)

ATIF/REFERENCE: Ercan, S. I. A. & Şimşek, M. (2023). Mobile Phone Price Classification Using Machine Learning. *International Journal of Advanced Natural Sciences and Engineering Researches*, 7(4), 458-462.

Abstract **–** Smartphones are becoming more and more important for people day by day. With the development of technology, telephones are used in many areas from daily life to business life. It does not only fulfil the function of calling someone. It also makes it possible to connect to the internet and read mail while away from the computer. For this reason, the features of the mobile phone are an important factor when purchasing a mobile phone. People who actively use the phone pay more attention to feature selection. When buying a cell phone, price-performance comparison is made. Phone features are considered as performance. There is no exact method for determining the price. Price estimation can be done with the use of machine learning algorithms. Many studies have been conducted on this subject in the literature. There are still ongoing studies on which machine learning algorithm is the most appropriate. In order to contribute to this issue, a price prediction study is conducted on different machine learning algorithms. Using a dataset of phone prices and features from Kaggle, 4 different models with 20 features are tackled. As a result, the highest value is obtained from Support Vector Machine with an accuracy value of 0.9616.

Keywords – Phone Price Prediction, Support Vector Machine, Logistic Regression, K-Nearest Neighbors, Decision Tree

I. INTRODUCTION

Over the years, the mobile phone market has grown due to increased communication on social networks and changes in working and payment conditions. In light of this continued increase, people select their phones by weighing a variety of features. According to numerous aspects, including battery, color, ram, and wifi capabilities, users tend to favor products that are worth their price. The user's primary concern is choosing a product that is worth their money [1]. As a result, it is essential to create a price projection in the background and account for numerous factors in it. The features that the user is interested in must be used to classify the pricing ranges of telephones as low cost, medium cost, high cost, and extremely high cost [2].

In this work, machine learning techniques for estimating phone prices based on attributes are examined. Using the data set containing data from Kaggle, prices are categorised into 4 groups (low cost, medium cost, high cost and very high cost). For the purpose of predicting phone prices, four different machine learning methods are examined. The findings are shown in the results section.

In this study, 20 features of the phones are taken into consideration for the phone prices determined according to the features of the phones. These features are mobile depth in cm, weight of mobile phone, number of cores of processor, primary camera mega pixels, pixel resolution height, pixel resolution width, random access memory in megabytes, screen height of mobile in cm, screen width of mobile in cm, longest time that a single battery charge will last, 3G, touch screen, wifi.

A prediction model is created using various machine learning algorithms to determine the price according to the features. The models are compared and analyzed to see which one gave the highest accuracy score. Among our results that we analyze using 20 different features for phone prices, the highest accuracy value is reached with Support Vector Machine with 0.9616.

In this study, 4 different machine learning algorithms are analyzed for phone price prediction using a dataset of 2000 phones on Kaggle (https://www.kaggle.com/code/fatemehrafiei/cell-phone-price-prediction-eda/input). It is shown that machine learning algorithms are capable of making highly accurate predictions of mobile phone prices.

A brief review of the literature is given in the remaining sections. In the results section, the results of the study are presented.

II. Background Survey

In the literature, many people have conducted phone price prediction studies with machine learning algorithms. However, there are studies on different data sets and different machine learning algorithms. The size of the data in the data sets and the proportion of the tested data can change the accuracy. This section presents the results of some studies in the literature.

For the purpose of predicting the pricing of mobile phones, Koc and Cetin [3] compared the Random Forest Classifier, Logistic Regression Classifier, Decision Tree Classifier, Linear Discriminant Analysis, K-Nearest Neighbors Classifier, and Support Vector Machine techniques. According to the results obtained, it is seen that the ANOVA f-test feature selection method is more appropriate for this data set.

Kıran and Jebakumar [4], who set out to develop a model to predict mobile phone price according to mobile phone features and to find the machine learning algorithm that predicts the price most accurately, made an evaluation using test data sampled with 20% of the data set. Decision Tree, Linear Discriminant Analysis, Naive Bayes Classifier, K- Nearest Neighbors Classifier and Random Forest Classifier are compared. The result is obtained with 95% accuracy using Linear Discriminant Analysis is found to be the most accurate prediction algorithm.

A study by Hu [5] uses Support Vector Machine, Decision Tree, K-Nearest Neighbors and Naive Bayes for mobile phone price level prediction. Support Vector Machine has the highest accuracy, recall and F1 score of 97.7%. Naive Bayes has the worst success (81.2%).

Kalaivani et al. [6] use 3 machine learning models as Support Vector Machine, Random Forest Classification and Logistic Regression for mobile phone price prediction. The first ten characteristics are chosen and the model is trained using chi-square based feature selection. The accuracy of the Support Vector Machine model is the highest with 95% before feature selection. After feature selection, the highest accuracy is again achieved with Support Vector Machine (97%).

Güvenc, Cetin and Kocak [7] compare K-Nearest Neighbors and Deep Neural Network models for a dataset of basic features for mobile phone price prediction. According to the results of the study, Deep Neural Network achieved higher success with 94% accuracy rate.

Many different machine learning models have been tested. In some studies, the number of features are low and in some studies, external evaluations are also made.

This study aims to contribute to the literature on phone price prediction through 4 different algorithms.

III. Materials and Method

Detailed information on 4 different machine learning models used in the study is provided.

A. *Logistic Regression Classifier*

Logistic Regression allows to predict the probability of the dependent variable from a given set of independent variables. There are probability values between 0 and 1 [6].

B. *Support Vector Machine*

This classifier is called a linear support vector machine if the training data can be separated linearly, albeit coarsely. If the data are not linearly

separated, it is preferable to use kernel methods and soft range maximization to obtain nonlinear support vector machines [5]. The data is separated into two classes and placed on a plane known as the marginal plane. Points from two classes that are adjacent to the line are called support vectors. The goal of Support Vector Machine is to choose the marginal plane with the greatest separation between two data points [6].

C. *Decision Tree*

There are decision and leaf nodes according to the goal and independent variables in the decision tree algorithm. Because it is a categorization technique that produces a tree-like structure, it is known as a decision tree. The data set entries are processed into a tree to be used for classification, after which the classification procedure is carried out. Algorithms can follow different paths in the selection of root, node and branching criteria [2].

D. *K-Nearest Neighbors*

K-Nearest Neighbors is a well-known classification technique that bases predictions on finding the closest neighbors in classes that share a lot of characteristics. Since the dataset is scanned one by one to find the nearest neighbors, the performance of the algorithm decreases. It is also known as the lazy learning method. It works slowly in large volumes of data [4].

Describe in detail the materials and methods used when conducting the study. The citations you make from different sources must be given and referenced in references.

IV. RESULTS

In this study, an accuracy score comparison is made between Logistic Regression, Support Vector Machine, Decision Tree and K-Nearest Neighbors machine learning algorithms. Colab is used as the development environment in the model. It is written in Python programming language.

4 different machine learning methods are used to predict phone prices. Logistic Regression, Support Vector Machine, Decision Tree and K-Nearest Neighbors have accuracy rates of 91%, 96%, 82%, and 41% respectively. The best result among the models is achieved with the Support Vector Machine model.

Fig 1 Fig *1* Correlation Heatmapshows the correlation heatmap, illustrating the correlations between 20 features.

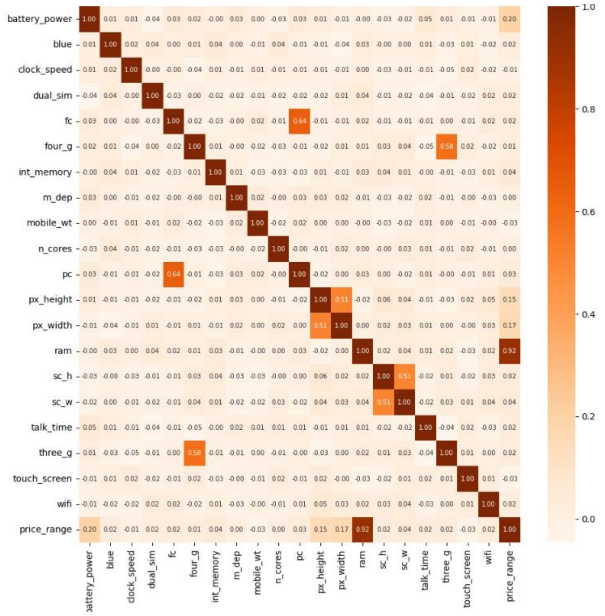

Fig 1 Correlation Heatmap

The confusion matrix and results for each model are presented in Fig. 2, Fig. 3, Fig.4 and Fig. 5.

The confusion matrix presents the distribution of estimated prices according to the outputs of the dataset. This matrix summarizes the performance of the classification algorithm in a visual way. The number of correct and incorrect predictions is tabulated. It provides performance measurement.

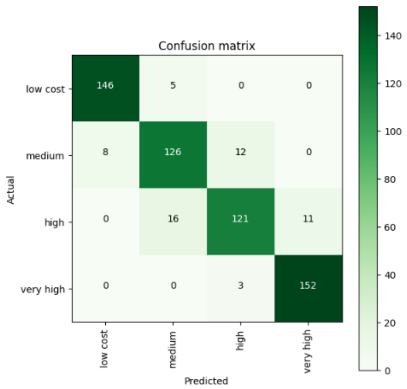

Fig. 2 Confusion Matrix for Logistic Regression

As seen in Fig. 2, in the logistic regression, the number of actual low cost data is 146, while the number of estimated low cost data is also 146. 8 data are incorrectly estimated as low cost when in reality they are medium cost.

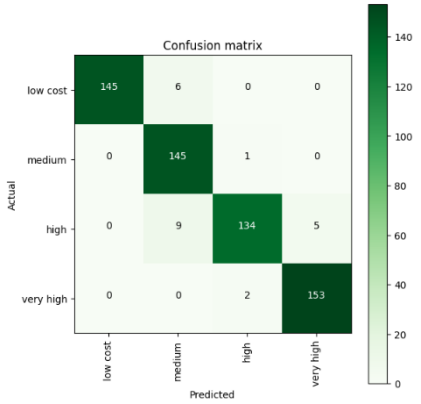

Fig. 3 Confusion Matrix for Support Vector Machine

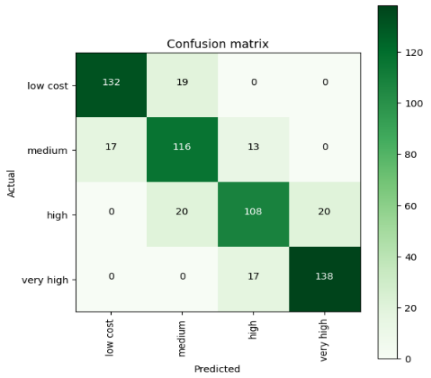

Fig.4 Confusion Matrix for Decision Tree

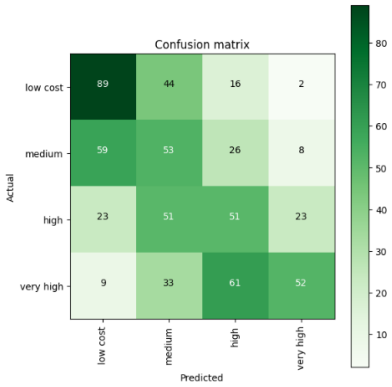

Fig. 5 Confusion Matrix for K-Nearest Neighbors

Table 1, Table 2, Table 3 and Table 4 present the performance metrics including accuracy, precision, recall and f1 score for each model.

Table 1. Classifier Performance for Logistic Regression

	precision	recall	f1-score	support
0	0.95	0.97	0.96	151
1	0.86	0.86	0.86	146
2	0.89	0.82	0.85	148
3	0.93	0.98	0.96	155
accuracy			0.91	600
macro avg	0.91	0.91	0.91	600
weighted avg	0.91	0.91	0.91	600

According to the values obtained from the confusion matrix in Table 1, it is seen that the validation data are predicted with 91% accuracy according to the modeled f1-score values created using logistic regression. Precision and recall values close to 1 indicate that the performance of the model increases. Accuracy supports this.

Table 2. Classifier Performance for Support Vector Machine

	precision	recall	f1-score	support
0	1.00	0.96	0.98	151
1	0.91	0.99	0.95	146
2	0.98	0.91	0.94	148
3	0.97	0.99	0.98	155
accuracy			0.96	600
macro avg	0.96	0.96	0.96	600
weighted avg	0.96	0.96	0.96	600

Table 3. Classifier Performance for Decision Tree

	precision	recall	f1-score	support
0	0.89	0.87	0.88	151
1	0.75	0.79	0.77	146
2	0.78	0.73	0.76	148
3	0.87	0.89	0.88	155
accuracy			0.82	600
macro avg	0.82	0.82	0.82	600
weighted avg	0.82	0.82	0.82	600

Table 4. Classifier Performance for K-Nearest Neighbors

	precision	recall	f1-score	support
0	0.49	0.59	0.54	151
1	0.29	0.36	0.32	146
2	0.33	0.34	0.34	148
3	0.61	0.34	0.43	155
accuracy			0.41	600
macro avg	0.43	0.41	0.41	600
weighted avg	0.44	0.41	0.41	600

Results should be clear and concise. The most important features and trends in the results should be described but should not interpreted in detail.

V. CONCLUSION

In this study, an analysis is made to predict the phone price with machine learning algorithms. As a result, 4 different algorithms are used on a dataset containing 20 different features. The results are evaluated according to the accuracy score. Among the algorithms applied for phone price prediction, the highest value is obtained with an accuracy value of 0.9616 in the Support Vector Machine algorithm. The analysis of 4 different machine algorithms is intended to contribute to the literature. In future studies, more diverse machine learning algorithms

can be analyzed using large data sets with more features.

REFERENCES

[1] M. Chen, “Mobile Phone Price Prediction with Feature Reduction”, Highlights in Sicence, *Engineering and Technology*, vol. 34, pp. 155–162, Feb. 2023.

[2] S. Aydın, *Using Machine Learning Algorithms in the Classification of Prices on Mobile Phones*, International Research in Science and Mathematics, pp. 201-212, 2022.

[3] M. Çetın and Y. Koç, "Mobile Phone Price Class Prediction Using Different Classification Algorithms with Feature Selection and Parameter Optimization," *2021 5th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT*), Ankara, Turkey, 2021, pp. 483-487, doi: 10.1109/ISMSIT52890.2021.9604550.

[4] A. V. Kiran and Jebakumar R., “Prediction of Mobile Phone Price Class using Supervised Machine Learning Techniques,” *International Journal of Innovative Science and Research Technology*, vol. 7, pp. 248-251, Jan. 2022.

[5] N. Hu, "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. 438-443, doi: 10.1109/PRML56267.2022.9882236.

[6] KS Kalaivani, N. Priyadharshini, S. Nivedhashri and Nandhini, "Predicting the Price Range of Mobile Phones Using Machine Learning Techniques," *AIP Conference Proceedings*, Vol. 2387. No. 1. AIP Publishing LLC, 2021.

[7] E. Güvenç, G. Çetin and H. Koçak, "Comparison of KNN and DNN Classifiers Performance in Predicting Mobile Phone Price Ranges," *Advances in Artificial Intelligence Research*, vol. 1, no. 1, pp. 19-28, Jan. 2021.