STM JOURNALS

Journal of
Artificial Intelligence
Research & Advances

ISSN: 2395-6720
Volume 11, Issue 2, 2024
May–August
DOI Journal: 10.37591/JoAIRA

https://journals.stmjournals.com/joaira

Review JoAIRA

Prediction of Mobile Phone Price Using Machine Learning Classifiers

Gosia Sandal Parveen[1,*], Aqeel Khalique[2], Rahbre Islam[3], Imran Hussain[2]

Abstract

One cannot imagine one's life without mobile phones; in today's digital era, mobile phones have become a necessity for everyone to fulfil their various demands like messaging, communication, entertainment, productivity, research, shopping and many more. In a thriving market of mobile phones where new smartphones are launched every year with new advanced features and various designs, determining the expense of a mobile can be a trouble-some tasks for consumers. In this study, we propose an approach to predict mobile phone prices using machine learning. The dataset, sourced from kaggle.com, encompasses various attributes such as RAM, ROM, storage capacity, display size, battery life, and camera quality among others. While several machine learning algorithms exist including linear regression, decision trees, and Naive Bayes, our methodology focuses on Support Vector Machine (SVM), Decision Tree, Random Forest, and K-Nearest Neighbors (KNN) for training the model and assessing its accuracy, F1 score, precision, and recall. Through rigorous experimentation we decipher the most effective algorithms of Machine Learning and feature sets, to predict the price of mobile phone accurately. Such a model can provide valuable help rights and assistance to the consumers while buying new mobile phones and the model also helps the manufacturers to adapt to changing market conditions based on consumer behavior, thus making more profit and market share.

Keyword: Prediction model, machine learning, Support Vector Machine (SVM), Decision Tree, Random Forest, and K-Nearest Neighbors (KNN), RAM, ROM

INTRODUCTION

There is a wide range of options available in the mobile industry, thus the price of the mobile plays a crucial role in attracting more customers and contributing to higher sales. The mobile phone industry offers consumers a vast array of features and functionalities with variety of price range, so determining what will be the cost of the mobile phone having desired features and functionalities of a consumer is a challenging task. This study explores the application of the advanced predictive algorithms of machine learning to analyze dataset from Kaggle.com containing diverse features like RAM, 4G, length of screen, storage capacity, display size, battery life, camera quality and many more, to generate informed predictions with precision and accuracy. We aim to enhance market efficiency and consumer satisfaction in the mobile phone industry and to bridge the gap between consumer expectations and market realities.

The research community uses these algorithms as some of the most influential data mining tools, and model accuracy is used to assess how well they function.

***Author for Correspondence**
Gosia Sandal Parveen
E-mail: gosiasandal@gmail.com

[1]Student, Department of Computer Science & Engineering, School of Engineering Sciences & Technology, Jamia Hamdard, New Delhi, India
[2]Assistant Professor, Department of Computer Science & Engineering, School of Engineering Sciences & Technology, Jamia Hamdard, New Delhi, India
[3]Research Scholar, Department of Computer Science & Engineering, School of Engineering Sciences & Technology, Jamia Hamdard, New Delhi, India

Received Date: May 24, 2024
Accepted Date: June 22, 2024
Published Date: July 10, 2024

Citation: Gosia Sandal Parveen, Aqeel Khalique, Rahbre Islam, Imran Hussain. Prediction of Mobile Phone Price Using Machine Learning Classifiers. Journal of Artificial Intelligence Research & Advances. 2024; 11(2): 101–108p.

TECHNICAL LEARNING

The study of how computers can learn like people, develop new abilities and knowledge, and recognize information is known as machine learning. Compared to human learning, it has higher learning rates, better information acquisition, and improved performance. This technology holds the promise of significantly enhancing computer capabilities and influencing human society [1].

Types of ML

1. Supervised learning uses labelled outputs that may be used in classification or regression to estimate an unknown mapping using known input and output samples.
2. Unsupervised learning performs operations like probability density function estimation and clustering using only input data.
3. Semi-supervised learning, which is employed in text or picture retrieval systems, integrates supervised and unsupervised learning by using a partially labelled dataset to infer the unlabelled portion [2].
4. A branch of machine learning called reinforcement learning carries out a task in response to rewards. The aim of this learning process is to create a medium that can interact with its surroundings in order to improve its decision-making skills and learn from the input it receives [3].

The following are some ML algorithms that are used by the researchers to develop a model.

Clustering

The basic objective of clustering, which is considered the most important unsupervised learning task, is to find patterns in an unlabelled dataset. It is employed to cluster together objects according to their shared characteristics. A cluster is a collection of things of similar sorts because it guarantees that the objects living in it are more similar to one another [4].

Naïve Bayes (NB)

Popular machine learning classification technique Naive Bayes (NB) determines an instance's class based on features that describe it and information from like cases. The Bayes theorem, which determines the likelihood of an event based on the data at hand, forms the basis of it. The mathematical expression for the Bayes theorem, which forms the basis of the Nave Bayes classification, is:

$$P\left(\frac{A}{B}\right) = \frac{P\left(\frac{B}{A}\right) * P(A)}{P(B)} \quad (1)$$

As mentioned previously, Eq. (1) addresses the conditional probability of occurrence 'A', represented by P(A/B), given the knowledge of information about 'B'. The Nave Bayes classification uses information gathered from other features to determine the likelihood of each feature. This contributes to the achievement of feature independence, which holds that characteristics are unrelated until they mutually raise the probability of an event [5].

Deep Learning

Due to its superior performance and accuracy compared to conventional machine learning algorithms, deep learning (DL) is widely utilized in the medical industry for image collection and interpretation. Consequently, it is increasingly prevalent in imaging and computer vision applications [6].

DL models use a greater number of layers to complement the structure of traditional ML models, allowing abstraction, extracting deeper characteristics, and combining fundamental concepts to comprehend complicated concepts. Due to this, DL models outperform conventional ML by making more accurate predictions from the data [7].

In recent years, deep learning has had a major influence on medical research, mostly in the areas of object identification, picture segmentation, and classification. Traditional learning systems mostly rely

on manually designed features to carry out the learning process. For example, hand-crafted structural features must be extracted in order to detect lung cancer tumours. To meet the unpredictability in patient-to-patient data with varying characteristics, however, innovative ML, or in this instance DL, approaches outperform classical learning techniques because of the latter's ability to sift through a range of attributes and do large-scale data analysis [7].

Neural Network

Most deep learning models are built on top of neural networks (NNs). Their neurons are made up of weight, bias, and activation functions. The weight and bias of the neurons are used to compute the input values linearly. Next, using an activation function like the sigmoid function, the output is subjected to element-wise non-linearity in Eq. (2), where 'w' and 'b' represent the weight and bias, respectively. Control is then transmitted to a fully linked layer network, like a multi-layered feed-forward propagation network, once these modifications take place in the network's initial layers of neurons [8]. Neural networks are capable of solving functions, regardless of their simplicity or complexity. Ensembles of related neural networks may be used to lower the prediction error on never-before-seen data points [9].

$$a=\sigma(w^T x+b) \tag{2}$$

Backpropagation uses gradient descent to modify neural network weights. The model uses the feedforward technique to find patterns in the data while it is being trained. It computes the gradient of the error function and uses it to reduce error between the computed and desired outputs when predicting the classification or regression representation [10].

RELATED WORK

There have been multiple studies on this topic of predicting mobile phone price with the help of machine learning algorithms. Different studies employ different algorithms like Support Vector Machine (SVM), Decision Tree, K-Nearest Neighbors (KNN), Random Forest, linear regression, Naive Bayes, in their studies, to analyze features of mobiles and forecast their prices based on it. Here are some of them in Table 1.

Table 1. Related work.

Ref. No.	Year	Algo	Summary	Accuracy
[11]	2023	KNN, SVM, Decision Tree	Top 10 features are selected from dataset to train the model.	94%
[12]	2022	KNN	Simplify the dataset and provide an approx. cost	68%
[13]	2021	LR, SVM, XGBoost	To estimate the price of a new smartphone	97%
[14]	2018	Naive Bayes, Decision Tree	If the mobile with given features will be Economical or Expensive.	78%
[15]	2023	Bagging classifier, Naive Bayes, CatBoost	The dataset used here is small and the research focuses on comparison of multiple models.	78–80%
[16]	2024	KNN, SVM, Random Forest, Decision Tree	In this work, a variety of models were trained using mobile features, and a considerable prediction of the range of mobile prices was made.	88.8%
[17]	2023	KNN, SVM, Random Forest, Decision Tree	The performance of the mobile phone varies depending on the features used. As the performance improves, so does the price range.	97%
[18]	2021	KNN	KNN algorithm is used to find the distances between K models and model for testing the dataset. It uses Feature selection to give precise features to be selected and get maximum accuracy results.	

METHODOLOGY

The machine learning algorithms used in this study are listed below:

Support Vector Machine (SVM)

This algorithm, which falls under supervised machine learning, serves purposes in both classification and regression tasks [19]. The goal of the Support Vector Machine (SVM) algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future.

Decision Tree (DT)

This classifier utilizes a tree structure, where internal nodes depict dataset features, branches embody decision criteria, and leaf nodes signify outcomes [20]. It serves as a graphical model to explore potential solutions to a problem or decision based on specified conditions.

Random Forest (RF)

Random Forest stands out as a robust and adaptable supervised machine learning technique. By assembling numerous decision trees into a "forest [21]", it excels in tasks involving classification and regression. For instance, it can effectively determine if an email falls under the categories of "spam" or "not spam".

K-Nearest Neighbor (KNN)

The K-Nearest Neighbor (K-NN) algorithm represents a straightforward Supervised Learning technique [22]. It operates under the assumption that new data points resemble existing cases and are therefore categorized with the most similar ones.

The following Figure 1 shows the methodology used in this study:

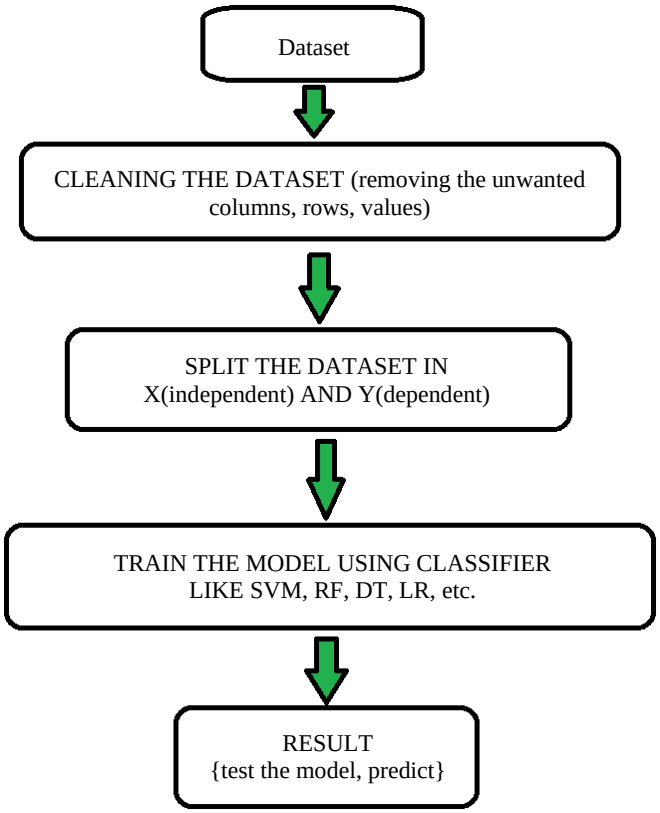

Figure 1. Workflow Diagram.

RESULTS

In this section, we present multiple diagrams which shows what features affect the price of mobile phone, Confusion Matrix of Support Vector Machine (SVM) Model, Random Forest Model, K-Nearest Neighbors (KNN) Model for Mobile Price Prediction; it shows how much our predictions are accurate to the real values and also comparison between the accuracy of different Machine learning Algorithms is shown in Figures 2–6.

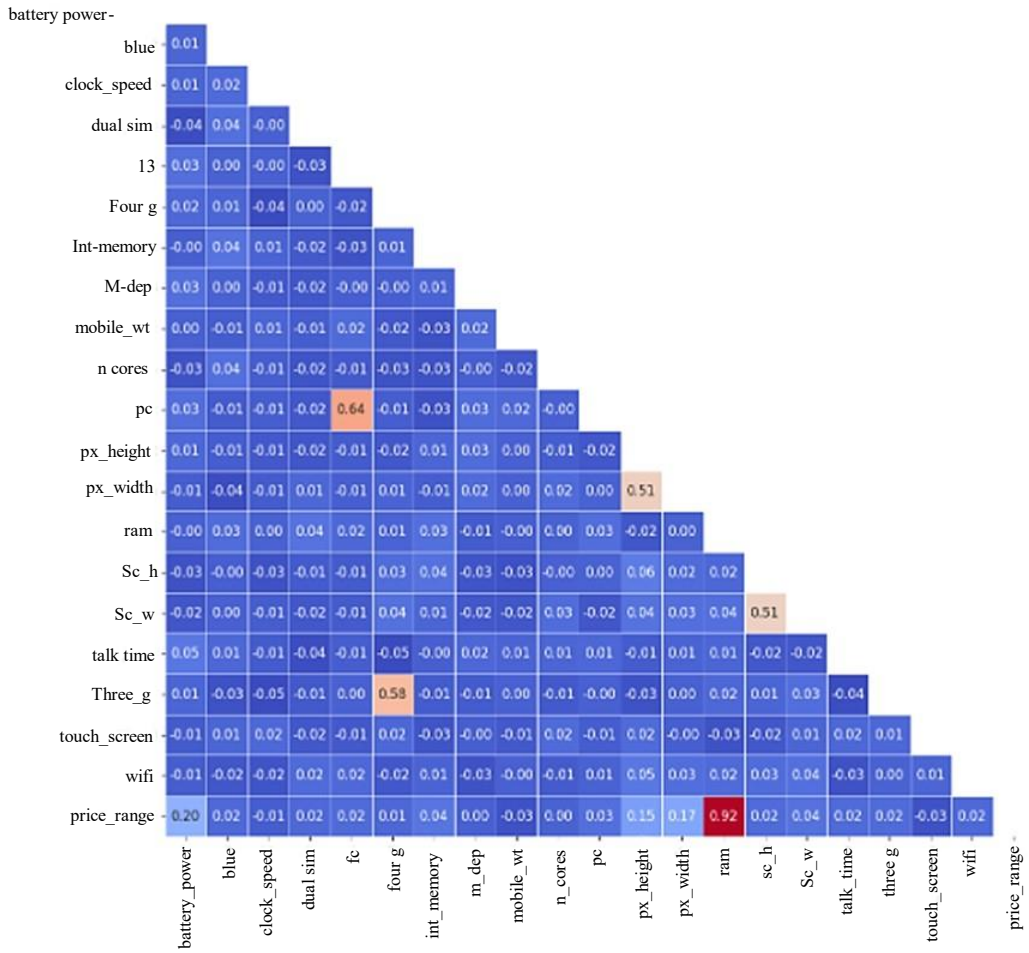

Figure 2. Correlation Heatmap.

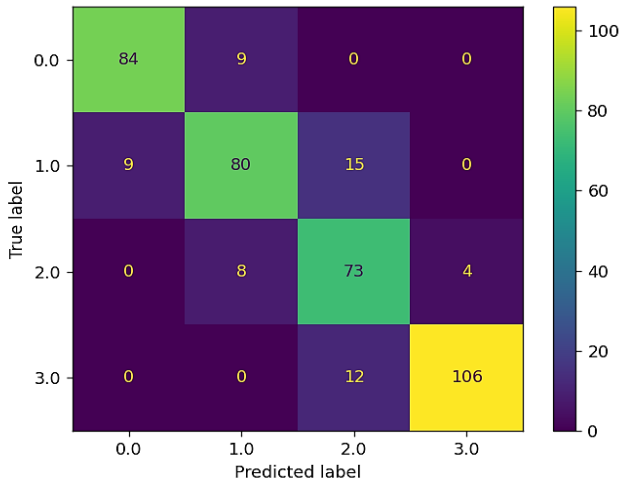

Figure 3. Confusion Matrix of SVM Model.

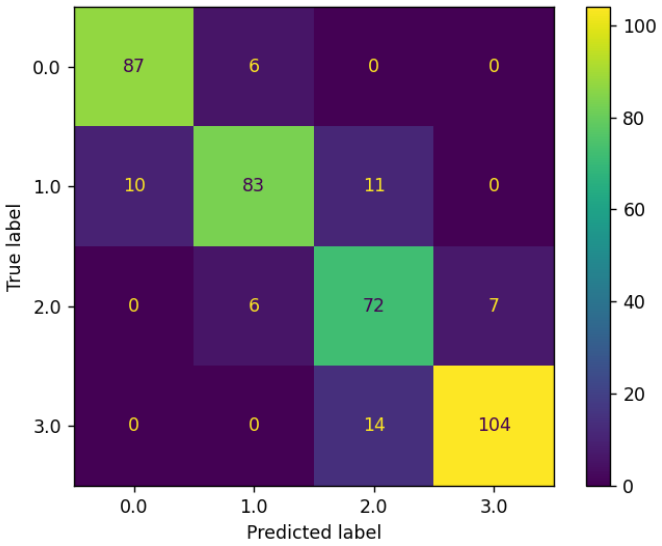

Figure 4. Confusion matrix of random forest model.

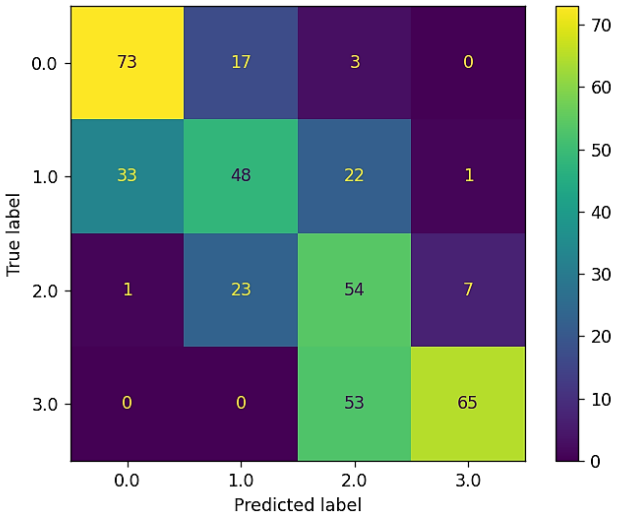

Figure 5. Confusion matrix of k-nn model.

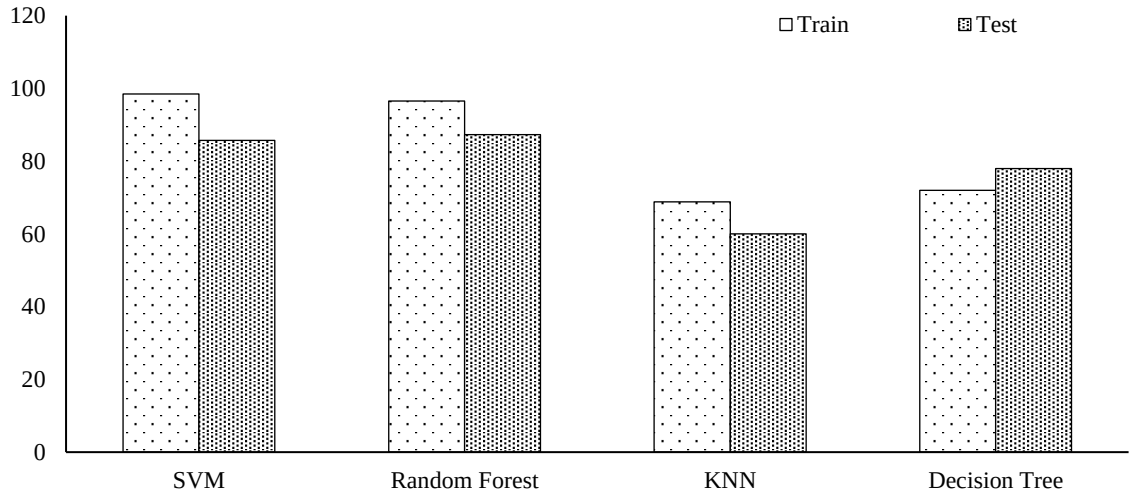

Figure 6. Accuracy % of prediction model using different classifiers.

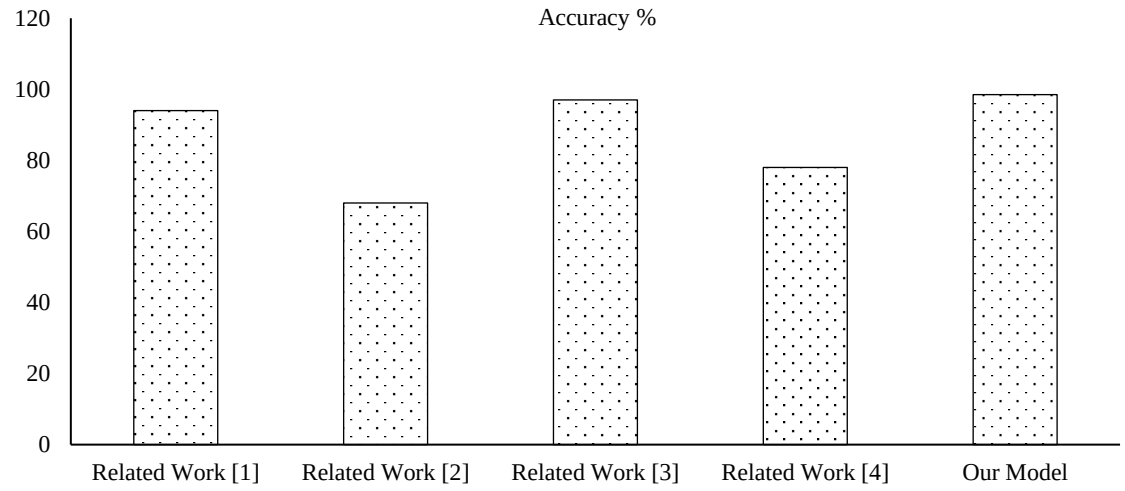

Figure 7. Comparison of Accuracy % of related work and our model.

In our predicted model, we are getting 98% accuracy using Support Vector Machine (SVM) classifier, which is highest among others. As evident in related work, our accuracy is more than the accuracies in related work. Figure 7 shows the comparison of accuracies of related work and our predicted model.

CONCLUSION

Through this study we came to know that the utilization of machine learning algorithms on a clean dataset containing variety of features for predicting mobile phone prices represents a significant development in mobile industry by enhancing market efficiency and consumer satisfaction; also, this study demonstrates the virtue of machine learning models. All the machine learning models used in this study like K-Nearest Neighbors (KNN), Decision Tree, and Random Forest predicted the price of mobile phone, but the model trained using Support Vector Machine (SVM) model was found to be most accurate, like 98.5%, in predicting the price of mobile phone. It is important to highlight that RAM emerged as the primary determinant in predicting mobile prices. As the mobile phone industry continues to evolve, such machine learning-based price prediction models promise to be a valuable tool for maximizing profitability, empowering the manufacturers and meeting consumer preferences in the future.

REFERENCES

1. Olivas ES, Guerrero JD, Martinez-Sober M, Magdalena-Benedito JR, Serrano L, editors. Handbook of research on machine learning applications and trends: Algorithms, methods, and techniques. IGI global; Pennsylvania, United States; 2009 Aug 31.
2. El Naqa I, Murphy MJ. What is machine learning? Cham: Springer International Publishing; 2015.
3. Mahesh B. Machine learning algorithms-a review. Int J Sci Res (IJSR). 2020 Jan; 9(1): 381–386.
4. Omran MG, Engelbrecht AP, Salman A. An overview of clustering methods. Intell Data Anal. 2007 Jan 1; 11(6): 583–605.
5. Ramalingam VV, Dandapath A, Raja MK. Heart disease prediction using machine learning techniques: a survey. Int J Eng Technol. 2018 Mar; 7(2.8): 684–7.
6. Sharma S, Chaudhary P. Machine learning and deep learning. Quantum Comput. Artif. Intell. Train. Mach. Deep Learn. Algorithms Quantum Comput. Berlin, Boston: De Gruyter; 2023 Aug 21; 71–84.
7. Suganyadevi S, Seethalakshmi V, Balasamy K. A review on deep learning in medical image analysis. Int J Multimed Inf Retr. 2022 Mar; 11(1): 19–38.
8. Dev S, Wang H, Nwosu CS, Jain N, Veeravalli B, John D. A predictive analytics approach for stroke prediction using machine learning and neural networks. Healthcare Anal. 2022 Nov 1; 2: 100032.

9. Hansen LK, Salamon P. Neural network ensembles. IEEE Trans Pattern Anal Mach Intell. 1990 Oct; 12(10): 993–1001.
10. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015 May 28; 521(7553): 436–44.
11. Swathi MS, Sajja DC, Kolavennu S, Mulagandla RR, Ale S. Classification of Mobile Phone Price Dataset using Machine Learning. Int J Adv Res Sci Technol. 2023; 13(06): 445–448.
12. Kukreti A. The Application of Machine Learning Algorithms to a Dataset of Mobile Phone Prices for Classification. Int J Mech Eng. 2022; 7(2): 4591–4597.
13. Jose J, Raj V, Seaban SV, Jose DV. Machine Learning Algorithms for Prediction of Mobile Phone Prices. In International Conference on Innovative Computing and Communication. Singapore: Springer Nature Singapore; 2023 Feb 17; 81–89.
14. Asim M, Khan Z. Mobile price class prediction using machine learning techniques. Int J Comput Appl. 2018 Mar; 179(29): 6–11.
15. Jun Wang. Mobile Phone Price Range Prediction Based on Machine Learning Algorithms. Proceedings of the 2nd International Conference on Financial Technology and Business Analysis. 2023; 46: 260–268.
16. Sunariya N, Singh A, Alam M, Gaur V. Classification of Mobile Price Using Machine Learning. In Symposium on Computing & Intelligent Systems (SCI). 2024; 55–66.
17. Singh A, Thakur N, Sharma A. A review of supervised machine learning algorithms. 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 2016, p. 1310–5.
18. Karur K, Balaje K. Prediction of mobile model price using machine learning techniques. Int J Eng Adv Technol. 2021; 11(1): 273–275.
19. IBM. (2023). What Is Support Vector Machine? [Online]. IBM. Available from: https://www.ibm.com/topics/support-vector-machine
20. Javatpoint. (2021). Decision Tree Algorithm in Machine Learning. [Online]. Available from: https://www.javatpoint.com/machine-learning-decision-tree-classification-algorithm
21. CareerFoundry. (2020). What is Random Forest? [Beginner's Guide + Examples]. [Online]. Available from: https://careerfoundry.com/en/blog/data-analytics/what-is-random-forest/
22. Javatpoint. (2021). K-Nearest Neighbor (KNN) Algorithm for Machine Learning. [Online]. Available from: https://www.javatpoint.com/k-nearest-neighbor-algorithm-for-machine-learning

