International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

Classification Of Mobile Phone Price Dataset using Machine Learning

Mrs. Sugur Swathi (Asst.professor)
Dept of Computer Science and Engineering
Sphoorthy Engineering College
Hyderabad, India
swathi.tekumal@gmail.com

Dr.SubbaRao Kolavennu
Dept of Computer Science and Engineering
Sphoorthy Engineering College
Hyderabad, India
profrao99@gmail.com

Raja Rajeswari Mulagandla
Dept of Computer Science and Engineering
Sphoorthy Engineering College
Hyderabad, India
mrajarajeswari56@gmail.com

Devika Chowdary Sajja
Dept of Computer Science and Engineering
Sphoorthy Engineering College
Hyderabad, India
19n81a0519devika@gmail.com

Sruthi Ale
Dept of Computer Science and Engineering
Sphoorthy Engineering College
Hyderabad, India
19n81a0512sruthi@gmail.com

Abstract—**In this project, we propose to estimate the price of a mobile phone using an available dataset. Different algorithms are used to reduce complexity and identify the key selection features in order to provide the best comparison within the dataset after it is collected from the current market. The best pricing range with the maximum specifications can be found with this tool. The price range of a mobile phone is categorized as high cost or low cost depending upon on the features of the mobile phone.Many individuals fail to link the features of mobile phones with the price of mobile phones. In this case, by using machine learning algorithms like Support Vector Machine, Decision Tree, K Nearest Neighbors, and Naïve Bayes to train the mobile phone price dataset before making predictions of the price range.The appropriate algorithms is used to predict smartphone prices based on accuracy, precision, recall and F1 score. Among all the 21 features in the dataset, only top 10 features namely RAM, pixel height, battery power, pixel width, mobile weight, internal memory, screen width, talk time, front camera and screen height are selected which are used to train the model. The predicting of price range will give support to customers to choose smart phone wisely in the future. The result shows that, among all 4 classifiers which is the best algorithm is used as a model to predict the price range of the mobile phone.**

Keywords—Machine learning, Classification, Price prediction, Data Collection.

I. INTRODUCTION

The most influential marketing feature is price. As the development of technology, mobile phones have become an indispensable part of people's lives. In this modern world, science and technology have created a far more incredible world than ever before.Brand loyalty is an important determinant, [1] represents the market share of different brands of mobile phone markets in China recently, from which we can deduce that the market share is directly proportional to the consumer choices: the first few ranks is Huawei with market share 28.1%, 13.3% of Oppo and Apple (11.3%).The fluctuating prices of mobile phones are now attracting the public's attention as the price is a crucial factor affecting consumer choices when buying smartphones. Machine learning is the technique we used to predict mobile phone prices. Every day, new mobile phones with new versions and additional features are introduced, and hundreds of thousands of cell phones are sold and purchased. Factors such as brand, internal storage, Wi-Fi, battery, camera, and 4G availability currently influence consumer decisions regarding mobile phone purchases.However, people often do not connect these factors to the price of cell phones. In this case, this article aims to solve this problem by using machine learning algorithms such as Support Vector Machine, Decision Tree (DT), KNN, and Naive Bayes to train the mobile phone dataset and make price predictions. We used appropriate algorithms to predict smartphone prices based on accuracy, precision, recall, and F1 score. This approach not only helps customers make better choices when selecting mobile phones but also provides advice to mobile phone companies on how to price their products right with a wide range of features. Mobile phones are currently one of the most popular applications for sales and transactions.

Ⅱ. PROBLEM STATEMENT

In the highly competitive mobile phone market, it's not wise to make assumptions. Many people struggle to estimate the price range of mobile phones and find it challenging to determine the price of a mobile phone with specific features. As a result, they waste a lot of time searching for mobile phones with their desired features.The study aims to predict mobile phone price levels using machine learning techniques when given the smartphone's features. This helps people understand the functions in addition to the prices of mobile phones. The dataset was obtained from Kaggle's website, and after preprocessing the data and selecting the four most relevant features (ram, battery power, px_width, and px_height), four machine learning algorithms (SVM, DT, KNN, and NB) were used to fit the training dataset of mobile prices and predict the price level.Moreover, the

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

introduction of the feature selection process led to an overall improvement, laying a solid foundation to help people minimize costs when selecting mobile phones at different price levels. From a business perspective, the classification model would be useful in understanding how to set prices based on different smartphone factors. On the other hand, customers can use the results to choose mobile phones at different prices based on their needs.

III. IMPLEMENTATION

The research design for Classification of mobile phone price using machine learning involves several key steps, including using machine learning algorithms, data collection and pre-processing, feature selection and implementation,and model evaluation.The following sections provide a detailed overview.

A. Background of Machine Learning Algorithms1.Support Vector Machine

In the era of big data, Support Vector Machine (SVM) is widely used as a machine learning classification and regression method.There are some applications of SVM, for example, Cristina uses SVM to predict liquid-crystalline property [2], Laurinda adopts SVM to classify prostate cancer [3] and Parisa employs SVM for the food classification like single food portion, a non-mixed and mixed plate of food [4].SVM works best when the training data is linearly separable, and it tries to find the partition hyperplane with hard margin maximization since such hyperplane produces the most robust classification results. In this case, the classifier is known as a linearly separable support vector machine.Moreover, when the training data is roughly linearly separable, the classifier can also be learned through soft margin maximization, which is called a linear support vector machine. However, when the data cannot be separated linearly, kernel methods and soft interval maximization can be used to acquire non-linear support vector machines.

2.**Decision Tree**

Decision Trees is a supervised machine learning method used for classification. The goal of decision trees is to build a prediction model and infer the values of target variables from data features by learning direct decision rules. A tree can be viewed as a piece-wise constant estimation. [5] explains the application of the decision tree to predict the type of ECG beats while Abraham and Mark [6] find the way to acquire the most.Additionally, due to the simplicity of the algorithm and the effectiveness of visualization, decision trees are used in various fields such as medicine and data mining analysis.

3.K-Nearest Neighbors

KNN is one of the most commonly used techniques that relies on the nearest neighbors of each query point. The steps involved in the KNN classifier algorithm are easy to follow.In terms of the usage of this model, in [7], researchers perform the classification of ECG signal which is helpful for identifying heart diseases, obtaining the highest accuracy of 97.5%.The first step is to assign a value to k, which determines the number of nearest neighbors we should choose. Next, we calculate the distance between the testing objects and every object in the set of training objects, and pick the closest training object with respect to the test object. Then, we select the class with the largest number of matched objects. Finally, we repeat the process until the same class is obtained.

4.Naïve Bayes Classification

Naïve Bayes algorithm is a supervised learning algorithm, which is based on Bayes theorem and used for solving classification problems.It is mainly used in text classification that includes a high-dimensional training dataset. [8] The way of Naïve Bayes works is to find the maximum conditional probability of label L and features F, which is denoted by Formula (1). $P(L|F) = P(F|L) \times P(L)$ $P(F)$ Naïve Bayes Classifier is one of the simple and most effective Classification algorithms which helps in building the fast machine learning models that can make quick predictions.

Dataset

The dataset is taken from Kaggle.com namedand used to evaluate dataset. The initial crucial stage is to gather data after defining the business problem. It is essential to comprehend the sources of data. The data gathered during this phase is in its raw form, as it may come from various sources and systems, and hence, it is not organized. The dataset obtained from Kaggle is categorized as binary data. The dataset has 2000 samples with 21 features. The features presented in the dataset.

Battery_power , bluetooth , clock speed, dual_sim, fc four_g, int_memory, m_dep, mobile _wt, n_cores, pc, px_height, px_ width, ram, sc_h, sc_w, talk_time, three_g,touch_screen, wifi, price_range.

B. Data Acquisition and Pre-processing:

The first step in this research design is to collect and pre-process a dataset from publicly available sources. This step is critical to ensure that the machine learning models can effectively identify fraudulent claims based on relevant features. The data is pre-processed by cleaning and normalizing it, and removing any duplicate or irrelevant information. Feature selection is also performed to select the most relevant features for the machine learning models. The target column selected for this analysis is the price range column, which is classified into four ranges: very high price level (3), high price level (2), median price level (1), and low price level (0), as shown in Figure 1.

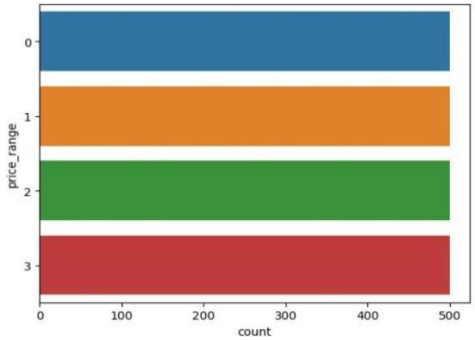

Fig 1. Bar graph for Price level

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

C. Feature Selection and Implementation:

The process of feature selection has existed because it can reduce overfitting and improve accuracy. The specific purpose of feature selection is to solve the problem of dimensionality by selecting only the most relevant features from the original feature set. This allows the learning model to obtain higher performance results, with a predictable model and lower execution time. Among the 21 original features, only 10 features were selected in the feature selection step. These 10 features are battery_power, int_memory, mobile_wt, pc, px_height, px_width, RAM, sc_h, sc_w, and talk_time.The more specific purpose of feature selection is mentioned in [9-11], which describes that in order to solve the problem of dimensionality, feature selection becomes an efficient solution.

In the cross-validation step, the train-test split method was implemented. This involved using 1600 samples for the training data and the remaining 400 samples for the testing data. The figure [2] below shows the correlation heatmap, which illustrates the correlations between the 21 features.

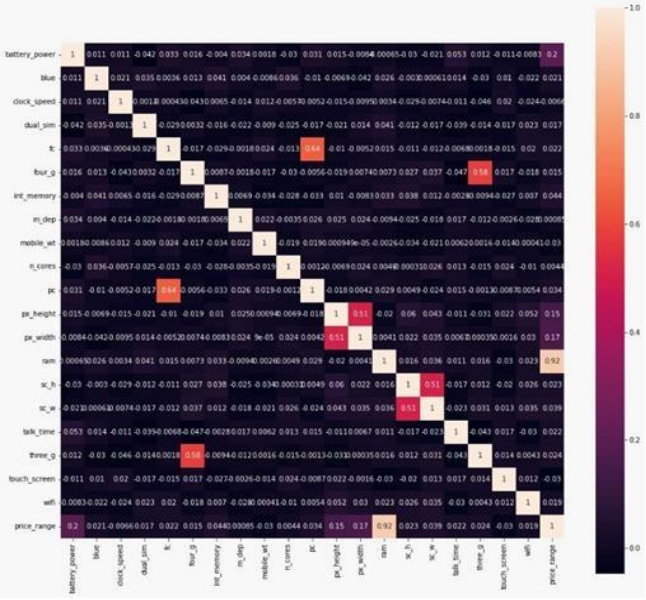

Fig 2. Correlation heatmap

IV. EXPERIMENTAL RESULTS

A. Model Evaluation

The model is evaluated using standard performance metrics, including accuracy, precision, recall, and F1 score. The samples are split into 80% training data and 20% testing data. The formulas below briefly explain some parameters in the F1 score system. Based on these parameters, the performance of the appropriate algorithm can be determined.For instance, accuracy is the ratio of correctly predicted labels to the total number of samples, which is the ratio between true positives + true negatives and true positives + false positives + false negatives + true negatives. Precision is the ratio between true positives and the sum of true positives and false positives. The parameters are briefly expressed in Table I below

TABLE I. FORMULAS OF PARAMETERS

Parameter	Formula
Accuracy	$\frac{TP+TN}{TP+FP+FN+TN}$
Precision	$\frac{TP}{TP+FP}$
Recall	$\frac{TP}{TP+FN}$
F1score	$2 \times \frac{Recall \times Precision}{Recall+Precision}$

B. Results

The confusion matrix and results are shown below. The parameter values of accuracy, recall, F1 score, precision are acquired for every algorithm in the below Table II.Theconfusion matrix is indicated by using within selected features for every algorithm as shown in below figure [3].

TABLE II. CLASSIFIER PERFORMANCE SELECTION

	SVM	DT	KNN	NB
Accuracy	94%	95%	71%	82%
Precision	94%	95%	73%	82%
Recall	94%	96%	71%	82%
F1score	94%	94%	72%	82%

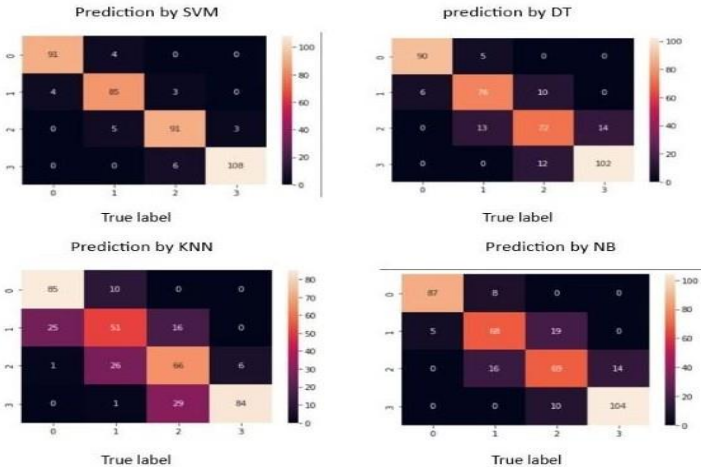

Fig 3. Confusion matrix of algorithms

The SVM classifier performed the best, with the highest accuracy, recall, and F1 score, all exceeding 93.75%,While other classification models performed poorly, with low accuracy, recall, and F1 score values. The Decision Tree

classifier performed well, with accuracy, recall, and F1 score exceeding 85.0%. The Naive Bayes classifier also performed reasonably well, with accuracy, recall, and F1 score exceeding 82.0%. The KNN classifier had relatively lower performance, with accuracy, recall, and F1 score exceeding only 71.5%. However, the SVM classifier performed the best, with the highest accuracy, recall, and F1 score, all exceeding 93.75%.

V. CONCLUSION

The primary goal of this study was to predict mobilePhoneprices based on features such as Ram and battery power,and we have successfully achieved this objective.In conclusion,the four machine learning techniques employed were effectivein predicting mobile phone prices using the features of RAM,battery power, pixel resolution height, and width, based ondata collected from the Kaggle website. We used SupportVector Machine, Decision Tree, K Nearest Neighbors, andNaïve Bayes classifiers to predict the price level, and theirPerformancewas evaluated based on accuracy, precision, recalland F1 score. Based on the results obtained, the SVMclassifier performed the best, while the other classification models performed poorly, compared to SVM.Furthermore, there was a general improvementafter the feature selection process was implemented becauseit provides a strong framework for people to use to cut costs when choosing mobile phones at various price points. Afterusing other types of feature selection with various machine learning algorithms, more study may be required to make cpredictions. One the one hand, the forecast would be helpful from a commercial standpoint in understanding how to set prices according to various smartphone features. On the other hand, the outcome might help buyers choose cell phones at various price points. The relationship betweensmartphonefunctionality and price ranges is now better understood as a result.

REFERENCES :

[1] Li H., Wei Y,"Analysis the Impacting of "User Experience" for ChineseMobile Phone's Brands Market Changing"Design, User ExperienceandUsability.Practice and Case Studies, pp277 -287, 2019.
[2] Cristina Butnariu, Catalin Lisa, Florin Leon and Silvia Curteanu,"Prediction of liquid-crystalline property using support vector machineclassification", Journal of Chemometrics,June 2013.
[3] Kassio M.G. Lima, Laurinda F.S. Siqueira, Camilo L.M.Morais,
"SVMfor FT-MIR prostate cancer classification: An alternative to the traditionalmethods" Journal of Chemometrics, July 2018.
[4] Parisa Pouladzadeh, Shervin Shirmohammadi,Aslan Bakirov, AhmetBulut & Abdulsalam Yassine, "Cloud-based SVM for food categorization", Multimedia Tools and Applications, June 2014.
[5] Mohebbanaaz, L.V.Rajani Kumari & Y.Padma Sai,"Classification
OfECG beats using optimized decision tree and adaptive boosted Optimizeddecision tree", Signal, Image and Video Processing, Oct 2021.
[6] Abraham Itzhak Weinberg & Mark Last,"Selecting a representative
decision tree from an ensemble of decision-tree models for fast big data classification", Journal of Big Data, Feb 2019.
[7] C. Venkatesan, P. Karthigaikumar & R. Varatharajan, "A novel LMS algorithm for ECG signal preprocessing and KNN classifier-based abnormality detection", Multimedia Tools and Applications, Mar 2018.
[8] Shi H., Liu Y., "Naïve Bayes vs. Support Vector Machine:
Resilience to Missing Data", Artificial Intelligence and Computational Intelligence, pp680-687, 2011.
[9] Thomas Rincy N, Roopam Gupta, "An efficient feature subset selection approach for machine learning", Multimedia Tools and Applications, Jan 2021.
[10] A.N.M. Bazlur Rashid, Choudhury T, "Knowledge management overview of feature selection problem in high-dimensional financial data: cooperative co-evolution and MapReduce perspectives", Problems and Perspectives in Management, vol 17, pp.340-359, Dec 2019.
[11] Rashid AB, Mohiuddin Ahmed, Leslie F. Sikos & Paul Haskell-Dowland," Cooperative co-evolution for feature selection in Big Data with random feature grouping", Journal of Big Data, Dec 2020.