

# Prediction of Different Types of Mobile Phone Prices based on Machine Learning Models

Yiwei Chen \*

Department of faculty of science Zhejiang University of technology, Zhejiang, China

\* Corresponding author: 202203160205@zjut.edu.cn

**Abstract.** In contemporary society, mobile communication technology is advancing rapidly. People's lives have become more convenient, significantly driving the development of human society. As mobile communication introduces many new and practical technologies, people are increasingly inseparable from it. Mobile phones, as a form of mobile communication, have gradually integrated into people's lives and become a crucial and irreplaceable product. With the continuous popularization of smartphones and the increasing richness of mobile content, choosing a cost-effective phone has become a crucial decision for many. This article is based on hardware data and prices from a dataset comprising 2000 already-released smartphones. Machine learning techniques, specifically the K-Nearest Neighbors (KNN) model and linear regression model, are employed to predict the prices of these smartphones. The study reveals that both the K-Nearest Neighbors (KNN) model and the Linear Regression model can be utilized for predicting smartphone prices. However, the KNN model exhibits a slightly higher accuracy. Specifically, when K is set to 16, the accuracy of the KNN model reaches 93.3%. The Linear Regression model achieves an accuracy of 91.3%.

**Keywords:** KNN model, linear regression model, mobile phone price prediction.

## 1. Introduction

The birth of the Internet has set off a far-reaching era change on a global scale, announcing that mankind has entered the Internet era. The development of the Internet has gradually changed people's way of thinking and improved people's life experience. Internet applications have been reflected in all aspects of human life, such as online education, online entertainment, cross-border e-commerce and so on [1]. The rapid development of the Internet has promoted the development of mobile communication rapid development of mobile communication technology has improved the rate of information transmission between people, promoted the great development of human society, and spawned various industries related to the network. These industries give people great convenience and make people inseparable from mobile communication devices [2]. With the continuous development and progress of the times, mobile phones have undergone constant updates and iterations. This has also led to a proliferation of various types of mobile phones. However, choosing a cost-effective phone among many mobile phone options can be a problem and individuals need to efficiently find the right equipment for them while saving time. There are numerous smartphone review videos on social media nowadays. However, watching and comparing all these videos can be quite time-consuming. Today, the primary channels through which people gather information about and choose smartphones are mostly derived from these smartphone reviews. If there is a model that can analyze the cost performance of smartphones, then people can choose phones more efficiently and easily. Machine learning can effectively predict whether a phone is worth buying [3]. This study, based on various hardware information and prices of 2000 already released types of smartphones, attempts to use machine learning methods to predict smartphone prices as accurately as possible. This study will utilize the KNN model and linear regression model to predict smartphone prices. The various hardware indicators of the smartphones will be treated as independent variables, while the price of the smartphones will be treated as the dependent variable. The model will be trained using 70% of the data and tested for accuracy using the remaining 30% of the data.

## 2. Predicts the Price of Different Types of Mobile Phones based on the KNN Models

### 2.1. Data Profile

This dataset for smartphone price prediction is sourced from Kaggle. This data set includes 21 indicators for 2000 types of mobile phones, and the statistical range includes battery power, internal memory, clock speed, no of cores, RAM, talk time, dual sim, 3G, 4G, WiFi, blue tooth, mobile depth, mobile weight, touch screen, screen height, screen width, front camera pixels, rear camera pixels, pixels height, pixel weight and price range. Then, calculate the mean, standard deviation, maximum, and minimum values for these data separately. In this study, variables other than the price of mobile phones are made to X. Let the variable variable "price range" be Y. We then use the train\_test\_split function to divide the dataset into a training set and a test set 70% for the training set and 30% for the test set. In the process of analyzing the data, this study found that there is a clear relationship between some data and price. For example, RAM has a clear linear relationship with price range. Figure 1 shows the relationship between RAM and price range. We can see that the chart mainly focuses on four points. That is (1, 500) (2, 1500) (3, 2500) (4, 3500). There is a roughly linear correlation between RAM and price range; the higher the RAM, the higher the price range.

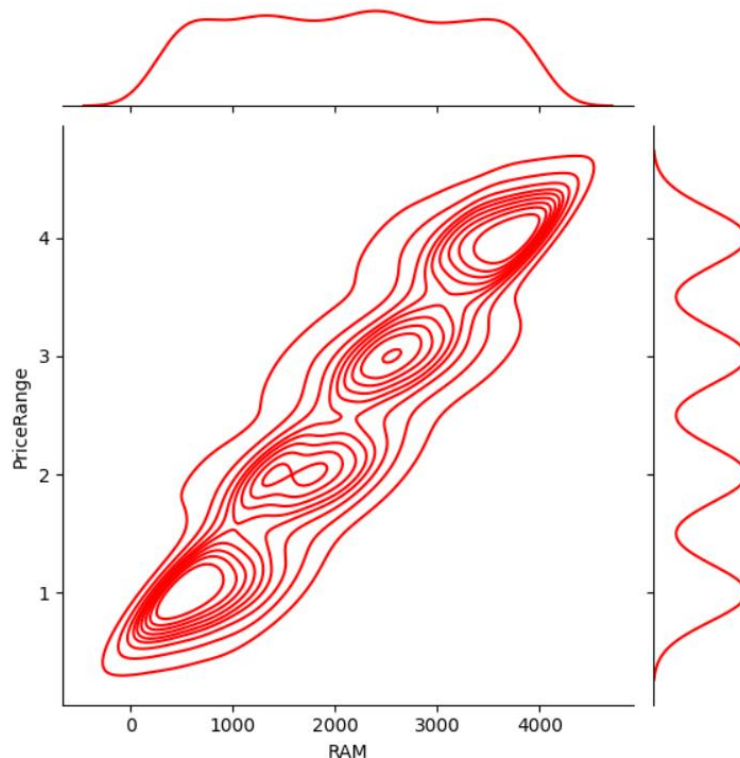


Figure 1. The relationship between RAM and price range

### 2.2. Model Introduction

The KNN model, whose full name is the K-Nearest Neighbor model, is one of the simplest machine learning algorithms. It is a nonlinear classifier that can be used in the fields of classification and regression and is widely used in the fields of finance, image recognition, data mining and so on [3]. This model aggregates all the data and then inputs the data to be predicted. The model calculates the K nearest data points to the given data, and the prediction for this data is determined by the most common outcome among these K data points.

### 2.3. Interpretation of Result

The first step is to build a KNN model. In this study, the K-Nearest Neighbors classifier from the scikit-learn library was used to create and train this model. Then program K to equal 10. When K is set to 10, the KNN model's predictions are relatively accurate, with an accuracy of 92.3%. Next, this paper will study the impact of different K values on the accuracy of the model. First, create an empty list to count different error rates for different K values. Then use for loops to calculate the different error rates of K from 1 to 19. Then plot a line graph where the x-axis represents the number of neighbors (1 to 19), and the y-axis represents the error rate. The graph is depicted with a blue dashed line, and data points are marked with red markers. As can be seen from Figure 2, the error is minimal when the K value is 16,17 and 19. So, when K is equal to 10 is not the most accurate time. When the K value is 16, 17, and 19, the accuracy of the KNN model reaches 93.3%.

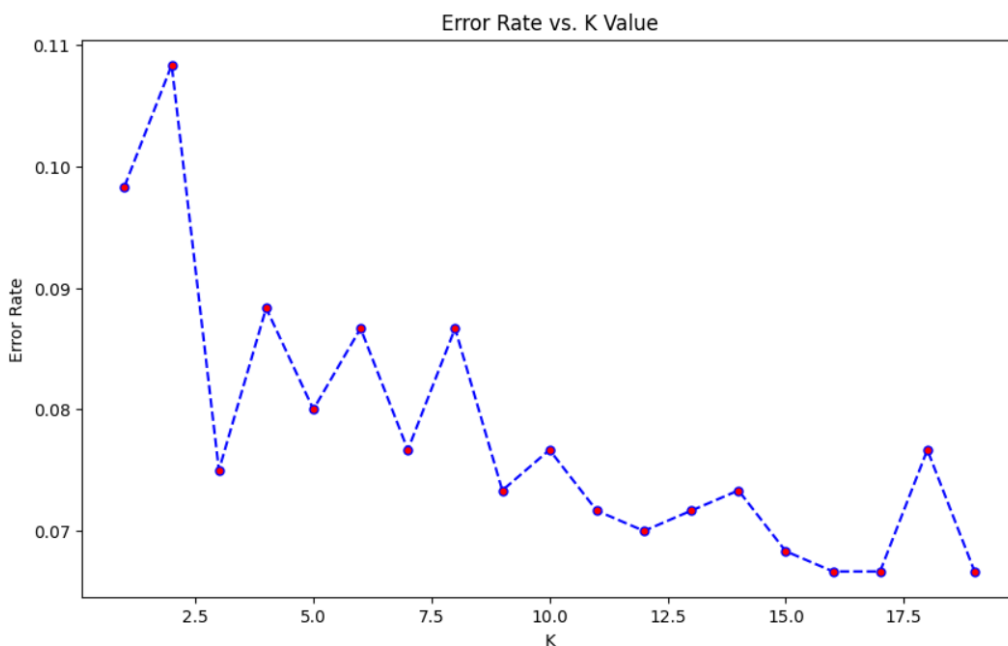


Figure 2. Error rate corresponding to different K values

## 3. Analysis on the Influencing Factors of Mobile Phone Price based on the Linear Regression Model

### 3.1. Model Introduction

The linear regression model is a method for fitting a straight line to the independent and dependent variables. By adjusting the intercept and slope, the goal is to make the line more accurately predict the prices of mobile phones. To make the prediction model of mobile phones more accurate, this paper decides to delete some nonlinear independent variables from the database. The independent variable selected in this paper is battery power, internal memory, clock speed, no of cores, RAM, talk time, mobile depth, mobile weight, screen height, screen width, front camera pixels, rear camera pixels, pixels height, and pixel weight.

### 3.2. Variable settings

In order to treat the variables more effectively, this paper has done the following treatment on the variables. Let's call battery power X variable 1, internal memory X variable 2, and so on. The details are shown in Table 1.

**Table 1.** Name of variable

battery power	Variable 1
internal memory	Variable 2
clock speed	Variable 3
no of cores	Variable 4
RAM	Variable 5
talk time	Variable 6
mobile depth	Variable 7
mobile weight	Variable 8
screen height	Variable 9
screen width	Variable 10
front camera pixels	Variable 11
rear camera pixels	Variable 12
pixels height	Variable 13
pixel weight	Variable 14

### 3.3. Interpretation of Result

These data are imported into Excel, which can be used to perform linear regression and calculate its regression coefficient [5]. And then this formula will be derived.

$$\begin{aligned}
 Y = & -0.58402924 + 0.000511183X_1 + 0.000868599X_2 \\
 & - 0.012443849X_3 + 0.001894775X_4 + 0.000946712X_5 \\
 & + 0.00040328X_6 - 0.008525803X_7 - 0.000874794X_8 \\
 & + 0.001150058X_9 - 0.000289723X_{10} + 0.000971987X_{11} \\
 & + 0.000124085X_{12} + 0.000274972X_{13} + 0.000279327X_{14}
 \end{aligned}$$

It can be seen from the formula that clock speed and mobile depth have a high linear correlation with price range. Since Excel is not convenient to calculate the accuracy of the model, this paper decides to use Python to calculate. This paper uses the Linear Regression model from the scikit-learn library. Then the model is trained using the training set, and the accuracy of the model is calculated. The linear regression accuracy of 91.3%.

## 4. Conclusion

This study found that both the KNN model and linear regression model have high prediction accuracy. The accuracy of the KNN model is 93.3%. The accuracy of the linear regression model is 91.3%. When studying the KNN model, the prediction accuracy of the model corresponding to different K values is different. Therefore, the accuracy of the KNN model corresponding to different K values is calculated in this paper, to find a KNN model with higher accuracy. In this paper, the value of K is 16, and then the KNN model reaches a high accuracy. In the study of the linear regression model, the formula of linear regression is calculated by Excel. From the formula obtained, it can be seen that clock speed has the greatest influence on this model. Then python can be used in calculating the accuracy of the linear regression model. It is found that the prediction accuracy of the linear regression model is also quite high. In general, the prediction accuracy of the KNN model is higher, and this model can be used for prediction. This paper hypothesizes that the weak predictive ability of linear regression models may be because some variables do not show a particularly obvious linear relationship. In this paper, only two machine learning models are studied, and more models can be studied in the future to increase the accuracy of predicting mobile phone prices.

## References

- [1] Ke Qi. Study on the influence of Internet development level on labor productivity. Guangdong University of Finance and Economics, 2020.
- [2] Peng Yuchao, Sun Minghong, Zhao Yiran. Research on the contribution of mobile communication technology to economic development and its transmission ways -- taking 3G popularization as an example. Gongxin financial technology, 021, (03): 99 - 126.
- [3] Lin Wenlong. Research on mobile phone network sales based on machine learning. Guizhou University of Finance and Economics, 2022.
- [4] Fu Jiahua. Improved SVM stock trend forecasting method based on KNN and its application. Henan University, 2023.
- [5] Wang Zhe, Wu Zhenjun, Wu Xu. Application of multiple linear regression analysis in predicting cold strength of coke. Metal world, 2023, (06): 81 - 84.