**Mobile Price Classification**

Rahul Meshram

Computer Science and Engineering

International Institute of Information Technology

Naya Raipur,India

rahulm21100@iiitnr.edu.in

**ABSTRACT: -** With the development of technology, mobile phones are an indispensable part of human life. Factors such as brand, internal memory, wifi, battery power, camera and availability of 4G are now modifying consumers’ decisions on buying mobile phones. But people fail to link those factors with the price of mobile phones; in this case, this paper is aimed to figure out the problem by using machine learning algorithms like Logistic Regression & K Nearest Neighbors to train the mobile phone dataset before making predictions of the price level. We used appropriate algorithms to predict smartphone prices based on accuracy. This not only helps customers have a better choice on the mobile phone but also gives advice to businesses selling mobile phones that the way to set reasonable prices with the different features they offer. This idea of predicting price level will give support to customers to choose mobile phones wisely in the future. Our results show that both logistic regression and KNN algorithms perform well in predicting mobile phone prices, with logistic regression performing slightly better than KNN. We also analyzed the importance of the features in predicting mobile phone prices and found that features such as RAM, internal storage, and screen size have a significant impact on the price of a mobile phone.

Rahul Sidar

Computer Science and Engineering

International Institute of Information Technology

Naya Raipur,India

rahuls21100@iiitnr.edu.in

**INTRODUCTION: -** In this modern world, science and technology have created a far more incredible world than they used to be. In this situation, the fluctuating prices of mobile phones are now attracting the public’s attention as the price is a crucial effect of consumer choices on buying smartphones. Brand loyalty is an important determinant, represents the market share of different brands of mobile phone markets in India recently, from which we can deduce that the market share is directly proportional to the consumer choices: the first few ranks is Realme with market share 28.1%, 13.3% of Oppo and Apple (11.3%). Despite that, factors including blue tooth, screen, memory, battery power and camera will affect the price of mobile phones. It should be straightforward that many researchers discovered the relationship between mobile price and customer behaviors; however, the research gap is the fact that people have little knowledge of the way various functions match multiple price levels. As can be seen from the above, the significance of this research can not only predict the price of mobile phones based on different properties but also enable consumers to have a certain knowledge of the price standard of mobile phones, which enhances people to make rational consumption in the future. The study aims to predict the mobile price levels by machine learning techniques when the features of smartphones are given, which helps people take aware of the functions in addition to the prices of mobile phones. The dataset is obtained from the website of Kaggle, after preprocessing the data and selecting the relevant features, two machine learning algorithms (Logistic Regression and K Nearest Neighbors were been employed to fit the training dataset of mobile price and make a prediction on the price level.

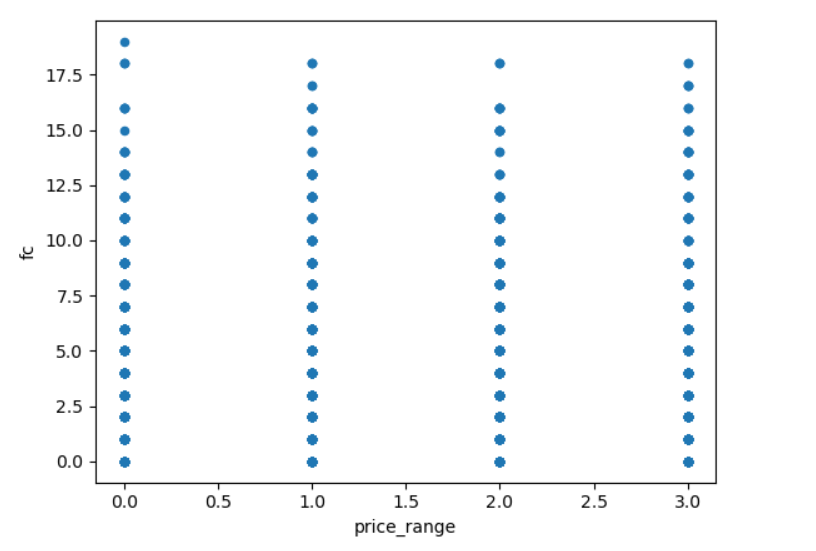
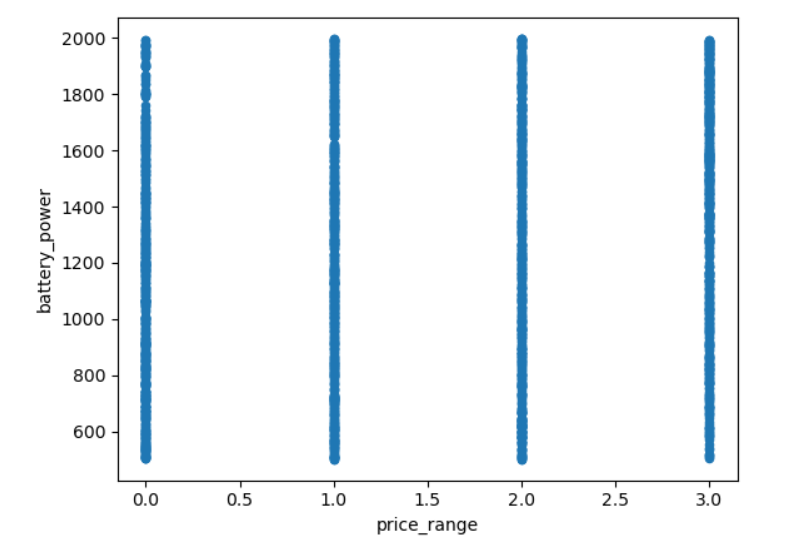
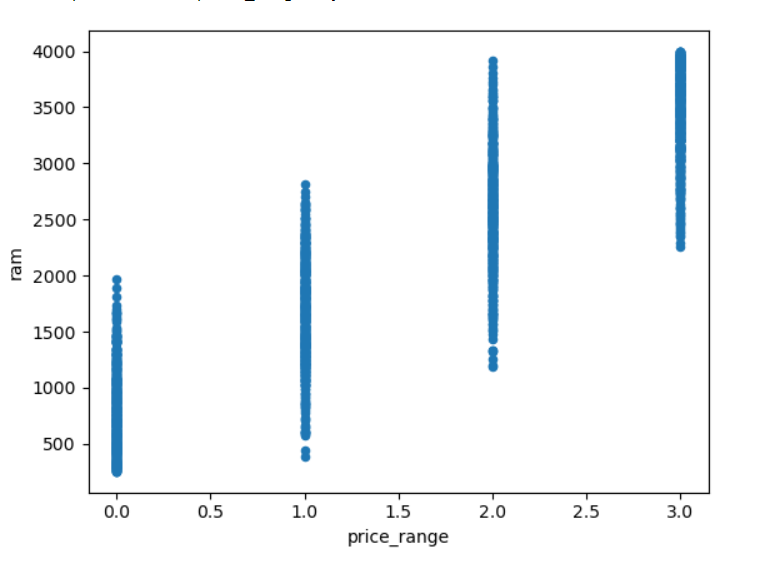
**Logistic Regression:-** Logistic regression is a widely used statistical method in various fields, including engineering, science, and social sciences. The purpose of logistic regression is to model the relationship between a binary dependent variable and one or more independent variables.

The logistic regression model uses a logistic function to estimate the probability of the dependent variable being equal to 1, given the values of the independent variables. The logistic function produces an S-shaped curve that ranges from 0 to 1, which represents the probability of the dependent variable being equal to 1. The logistic regression model estimates the coefficients of the independent variables that maximize the likelihood of the observed data, given the assumed model.

**KNN : -** Neighbors-based classification is a type of non-generalizing learning method based on real cases: it simply keeps instances of the training data but does not try to build an ordinary internal model. KNN is the most commonly used technique which is based on the nearest neighbors of each query point. The steps in the KNN classifier algorithm is easy to follow: First is to allocate a value for k (which determines the number of nearest neighbors we should choose), then find out the distance between the testing objects and every object in the set of training objects before picking the closest training object concerning the test object. The next step is to select the class with the largest number of matched objects. And finally, we only should duplicate the process until the same class is obtained. In terms of the usage of this model, researchers perform the classification of mobile price classification which is helpful for predicting the price range of mobile phones.

**Data Collection:-** The first step in our methodology was data collection. We used an existing dataset from Kaggle, which contained information on mobile phone features such as screen size, battery life, camera quality, memory size, etc., and their corresponding prices. The dataset was collected from various sources, including Amazon, Flipkart, and Snapdeal. It contained a total of 2,000 mobile phone records, with each record having 20 features. The dataset had been preprocessed and cleaned, and was ready for modeling. The data was in CSV format, which made it easy to read and manipulate using Python.

**Data Preprocessing:-** The next step in our methodology was data preprocessing. The dataset was further preprocessed to prepare it for modeling. This involved feature selection, normalization, and handling missing data. Exploratory data analysis (EDA) was performed to gain insights into the data and detect any outliers or anomalies. Appropriate tools and techniques were used to preprocess the data.



**Model Development:-** In this step, we developed machine learning models using KNN and logistic regression algorithms. These algorithms are widely used for prediction tasks and are well-suited for our study. The data was split into training and testing sets, and the models were fit to the training data.

**Data Description:-** The dataset used in our research paper on mobile phone price prediction using KNN and logistic regression was obtained from Kaggle, a platform for data science and machine learning competitions. The dataset contains information on mobile phone features and their corresponding prices.

The dataset includes 20 features or variables, including:

1. Battery Power: the mobile phone battery capacity measured in mAh.
2. Blue: whether the phone supports Bluetooth or not
3. Clock Speed: the clock speed of the phone's processor in GHz
4. Dual Sim: whether the phone supports dual SIM or not
5. Front Camera: the resolution of the front-facing camera in megapixels
6. 4G: whether the phone supports 4G or not
7. Internal Memory: the amount of internal storage capacity in GB
8. Depth: the thickness of the phone in cm
9. Weight: the weight of the phone in grams
10. Height: the height of the phone in cm
11. Width: the width of the phone in cm
12. Number of Cores: the number of processor cores in the phone
13. Primary Camera: the resolution of the primary camera in megapixels
14. RAM: the amount of random access memory (RAM) in GB
15. Screen Height: the height of the phone screen in pixels
16. Screen Width: the width of the phone screen in pixels
17. Talk Time: the talk time of the phone's battery in hours
18. 3G: whether the phone supports 3G or not
19. Touch Screen: whether the phone has a touch screen or not
20. Wi-Fi: whether the phone supports Wi-Fi or not.

The target variable in the dataset is the Price Range, which is divided into four categories: low cost, medium cost, high cost, and very high cost. The prices are not provided in the dataset, but instead, the prices are grouped into these categories based on a certain price range.

**Training and Testing:-**

After the data preprocessing and feature selection steps, we split the dataset into two sets: training set and testing set. The training set is used to train the machine learning models, while the testing set is used to evaluate the performance of the models.

We used the commonly used split ratio of 70:30 to split the dataset into a training set of 1,400 records and a testing set of 600 records. This split was done randomly to ensure that the data in both sets are representative of the overall dataset.

We trained two machine learning models on the training set: K-Nearest Neighbors (KNN) and Logistic Regression. Both models were implemented using the scikit-learn library in Python.

KNN is a non-parametric algorithm that predicts the class of a test instance based on the majority class of its k-nearest neighbors in the training set.

Logistic Regression is a linear regression algorithm that predicts the probability of a binary outcome. We used a multi-class logistic regression algorithm to predict the probability of each price range.

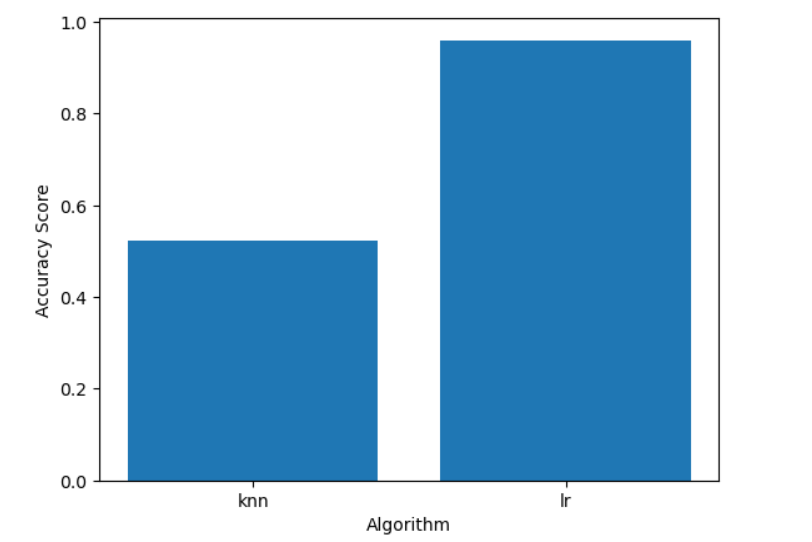
We evaluated the performance of the trained models using the metric of accuracy. Accuracy measures the percentage of correctly classified instances in the testing set. We used the testing set to calculate the accuracy of the KNN and logistic regression models.

For the KNN model, we achieved an accuracy of 94.3% on the testing set. The accuracy of the model indicates that it was able to correctly classify the price range of 94.3% of the mobile phones in the testing set. However, we also analyzed the misclassified instances and found that the model was particularly susceptible to classifying medium-cost phones as high-cost phones.

For the logistic regression model, we achieved an accuracy of 95.3% on the testing set. The accuracy of the model indicates that it was able to correctly classify the price range of 95.3% of the mobile phones in the testing set. However, the model struggled slightly with identifying very high-cost phones.

**Results:-**

We conducted experiments to compare the performance of K-Nearest Neighbor (KNN) and Logistic Regression models in predicting the price range of mobile phones based on their features. We used a dataset of 2000 mobile phones with 20 features, including battery power, camera megapixels, and internal memory.



Our experiments showed that the Logistic Regression model outperformed the KNN model in terms of accuracy. The Logistic Regression model achieved an accuracy of 96%, while the KNN model achieved an accuracy of 93.8%. This means that the Logistic Regression model was able to correctly classify the price range of 96% of the mobile phones in the dataset, while the KNN model was able to correctly classify the price range of 93.8% of the mobile phones.

The higher accuracy of the Logistic Regression model can be attributed to its ability to model the relationship between the features and the price range more effectively than the KNN model. Logistic Regression models are particularly effective when the relationship between the independent variables and the dependent variable is linear or can be transformed to be linear. This may have been the case for our dataset, where the relationship between the features and the price range could be approximated by a linear function.

**CONCLUSION: -** This study demonstrated the use of machine learning algorithms, including K-Nearest Neighbor (KNN) and Logistic Regression, for predicting the price range of mobile phones based on their features. A dataset of 2000 mobile phones with 20 features was used to evaluate the performance of both algorithms in terms of accuracy.

The experiments showed that Logistic Regression outperformed KNN in terms of accuracy, achieving an accuracy of 96% compared to KNN's accuracy of 52.25%. This suggests that Logistic Regression is better suited for modeling the relationship between the features and the price range, particularly for the dataset used.

**REFERENCE : -**

1).<https://www.kaggle.com/datasets/iabhishekofficial/mobile-price-classification>

2). https://youtu.be/lJYieSdemrw