Smart Mobile Phone Price using Machine Learning Techniques

A Nitish and M Ganga Vasudha and MV Sai Keerthana

Alliance University - Central Campus, Chandapura-Anekal, Main Road - Bengaluru, Karnataka - 562106, India.

# ABSTRACT

# The goal of this research is to create a model that accurately predicts the price of a mobile phone when provided with its specifications. To achieve this, research will leverage predictive analytics, specifically through the use of machine learning. By utilizing machine learning algorithms, the research aims to develop and train a prediction model using input data. The trained model will then be employed to forecast the prices of future mobile phone instances. Supervised machine learning algorithms will be utilized, which rely on data that includes a pre-defined class label, in this case, the mobile phone price. To train the prediction model, the Mobile Price Class dataset from the Kaggle data science community-website: (<https://www.kaggle.com/iabhishekofficial/mobile-price-classification>).

# This dataset categorizes mobile phones into different price ranges. Python was chosen as the programming language due to its extensive ML libraries. Several classification algorithms were implemented to train the model, with the objective of identifying the algorithm that offers the most accurate predictions of mobile phone prices. Performance metrics such as accuracy score and confusion matrix were utilized to evaluate the trained models and determine the most suitable algorithm among those used.

# INTRODUCTION

The pricing of a product holds immense significance in its marketing, often serving as the ultimate determining factor for consumer purchase decisions. In a constantly evolving and unpredictable market, pricing can make or break the success of a product. Hence, it is crucial for companies to establish an optimal price before launching a product. A valuable tool in this regard is the ability to estimate prices based on product features, enabling informed pricing decisions. This estimation tool proves beneficial not only for companies but also for consumers, who can obtain estimated prices based on their desired product features.

Machine learning algorithms offer a wide range of capabilities that can be tailored to specific data and task objectives. Various tools and programming languages, such as Python, MATLAB, Java, WEKA, Cygwin, and Octave, are available to facilitate machine learning tasks. Commonly used algorithms include Naïve Bayes, K-NN, among others. Feature selection algorithms play a crucial role in training models by identifying and extracting the most relevant parameters. This process optimizes accuracy and reduces computational time. The choice of the specific method for predicting product prices depends on the available data and the specific requirements of the model training process.

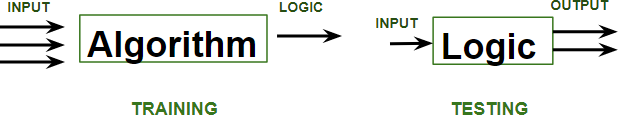
In today's world, cellphones have become an essential accessory for individuals. They are at the forefront of technology's constant evolution and innovation. New mobile models with updated features are introduced to the market at a rapid pace, and thousands of units are sold each day. In such a fast-paced and competitive market, mobile companies must set optimal prices to effectively compete with their rivals.

The initial step in determining a suitable price for a mobile phone is to estimate it based on the device's features. The objective of this research is to develop a machine learning (ML) model that accurately predicts the price of a mobile phone by analyzing its specifications. This model can also be beneficial to potential buyers who wish to estimate the price of a mobile phone based on their desired features.

The same approach used to create the prediction model can be applied to develop price estimation models for other products with similar independent variables. Mobile phone prices are influenced by various features, such as the processor, battery capacity, camera quality, display size, and thickness. These features can be utilized to classify phones into different categories, such as entry-level, mid-range, flagship, or premium. To accomplish this, supervised ML algorithms are employed, leveraging a dataset that includes definitive class labels for different price ranges.

# RESEARCH METHODOLOGY

The research was performed using the Python kernel in Google Colab, specifically focusing on the workflow diagram for supervised machine learning tasks is as follows:



(Ref: <https://www.superannotate.com/blog/supervised-learning-and-other-machine-learning-tasks> )

The dataset is split into two sections: the training set and the test set. The purpose of the training set is to train the model, while the test set is used to assess its performance. Through analyzing the features of mobile phones, the computer strives to comprehend the underlying logic behind their pricing. This understanding is then applied to make accurate predictions for future instances.

# UNDERSTANDING THE DATASET

The Mobile Price Class dataset that used in our project, obtained from the Kaggle data science community website (https://www.kaggle.com/iabhishekofficial/mobile-price-classification), was utilized to train the prediction model. This dataset categorizes mobile phones into different price ranges.

The dataset consists of a total of 21 attributes, which include 20 features and a class label representing the price range. The features encompass various aspects such as battery capacity, RAM, weight, camera pixels, and more. The class label indicates the price range and is categorized into four ordinal values: 0, 1, 2, and 3. These values correspond to increasing degrees of price, where higher values indicate higher price ranges. Specifically, these values can be interpreted as economical, mid-range, flagship, and premium categories.

Despite price being traditionally treated as a numeric problem, in this case, the machine learning task falls under classification rather than regression. This is because the class label consists of discrete values rather than continuous ones. The advantage of classification algorithms, such as Naïve Bayes and Decision Tree, is their compatibility with discrete data, which makes them suitable for this problem.

The dataset contains 2000 records in total.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

The df.info() function is commonly used in pandas, a popular data manipulation and analysis library in Python. When applied to a DataFrame object named df, it provides a summary of the dataset's information.

A screenshot of a computer

Description automatically generated

The code tdf = df[df['sc\_w'] != 0] filters the DataFrame df to select rows where the 'sc\_w' column is not equal to 0. The resulting DataFrame tdf contains the filtered data. tdf.shape retrieves the dimensions (number of rows and columns) of the DataFrame tdf.

# TRAINING THE PREDICTION MODEL

The first step in creating a model is to extract the required features for training from the dataset and assigning the parameter that is to be the class label.



In this code snippet, the first 20 attributes are being extracted to serve as the training parameters and the final attribute (price\_range) is used as the class label.

A close-up of text

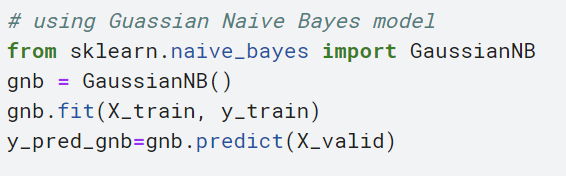
Description automatically generated

# Prediction Model

1.Random Forest

# 

2.Naive Bayes



3.KNN

A screenshot of a computer code

Description automatically generated

4.SVM

A screen shot of a computer code

Description automatically generated

# RESULTS AND DISCUSSION

A random forest system is built on a variety of decision trees. Every decision tree is made up of nodes that represent decisions, leaf nodes, and a root node. The leaf node of each tree represents the decision tree’s final result. The final product is chosen using a majority-voting procedure. In this situation, the output picked by the majority of the decision trees becomes the random forest system’s ultimate output. Let us now implement the random forest algorithm.

A screenshot of a computer screen

Description automatically generated

Gaussian Naive Bayes is a Naive Bayes variation that allows continuous data and follows the Gaussian normal distribution. The Bayes theorem is the foundation of a family of supervised machine learning classification algorithms known as naive Bayes. It is a basic categorization approach with a lot of power. When the dimensionality of the inputs is high, they are useful. The Naive Bayes Classifier may also be used to solve complex classification issues.

A screenshot of a screen

Description automatically generated

The K Nearest Neighbor method is a type of supervised learning technique that is used for classification and regression. It’s a flexible approach that may also be used to fill in missing values and resample datasets. K Nearest Neighbor examines K Nearest Neighbors (Data points) to forecast the class or continuous value for a new Datapoint, as the name indicates.

A screenshot of a computer screen

Description automatically generated

The SVM algorithm’s purpose is to find the optimum line or decision boundary for categorising n-dimensional space so that we may simply place fresh data points in the proper category in the future. A hyperplane is the optimal choice boundary.

A screenshot of a computer screen

Description automatically generated

# CONCLUSION

The SVM model demonstrated the highest accuracy of 96% in accurately predicting mobile price classes. To further improve the model's performance, data pre-processing techniques like normalization and standardization can be employed. Additionally, the application of feature selection and extraction algorithms can help eliminate unsuitable and redundant features, leading to more refined results. This research methodology can be extended to predict prices for various other products such as cars, bikes, and houses, utilizing historical data that incorporates relevant features such as cost and specifications. Such an approach would provide valuable insights for organizations and consumers, enabling them to make well-informed pricing decisions.

# REFERENCES

# D. Banerjee and S. Dutta, "Predicting the housing price direction using machine learning techniques", 2017 IEEE International Conference on Power Control Signals and Instrumentation Engineering (ICPCSI), pp. 2998-3000, 2017.

# Kanwal Noor and Sadaqat Jan, “Vehicle Price Prediction System using Machine Learning Techniques”, International Journal of Computer Applications (0975 – 8887) Volume 167 – No.9, June 2017.

# R. Gareta, L.M. Romeo and A. Gil, "Forecasting of electricity prices with neural networks", Energy Conversion and Management, vol. 47, pp. 1770-1778, 2006.

# https://[www.kaggle.com/iabhishekofficial/mobile-price-](http://www.kaggle.com/iabhishekofficial/mobile-price-) classification

# https://[www.gsmarena.com/](http://www.gsmarena.com/)

# Sameerchand Pudaruth . “Predicting the Price of Used Cars using Machine Learning Techniques”, International Journal of Information & Computation Technology. ISSN 0974-2239 Volume 4, Number 7 (2014), pp. 753-

# 764.