DOI : https://doi.org/10.56452/7-2-549

ISSN: 0974-5823 Vol. 7 No. 2 February, 2022

International Journal of Mechanical Engineering

The Application of Machine Learning Algorithms to a Dataset of Mobile Phone Prices for Classification

Anil Kukreti
Faculty, School of Computing, Graphic Era Hill University, Dehradun, Uttarakhand India 248002,

Abstract: A mobile phone is now one of the most ubiquitous consumer goods. Every year, thousands of new mobile phone models are introduced, each with its own set of features, specifications, and designs. The fundamental issue that needs answering before a mobile phone can be successfully marketed and released to the public is, of course, how much will it cost. The marketability and longevity of a product are now largely dependent on its price. The marketing of a mobile device and its competitiveness in the market are both affected by its pricing. In addition to meeting customer demands for features and aesthetics, having access to sufficient funds is crucial for success in the marketplace. Customers often verify whether or not their desired features and the anticipated cost are feasible. Before launching the cellphone, it is crucial to have a good idea of the market and the prices of similar products. In this Forecast, we take a dataset from the current market and apply several algorithms to it in order to simplify it, highlight the most important elements for making a decision, and get the most accurate comparison possible within the dataset.The highest quality at the lowest possible price may be found with this tool.

Keywords: Machine Learning, Data Collection, Forward Selection, Backward Selection

Introduction

When deciding whether or not to purchase a thing, the buyer will constantly be thinking, "what is the worth, and is it good to buy within this range?" When introducing a new product to the market, many things must be taken into account. This is particularly true in the mobile industry, where numerous features and specifications, such as memory, must be taken into account. There are various features and specifications to consider when purchasing a mobile device. There are various limitations imposed by the need for the product's pricing to be reasonable and within consumers' financial means. When choosing and comparing mobile phones, price and features are of primary importance. The best attributes and dataset for comparison are chosen using a variety of tools and Classifiers. Dataset collection is difficult because of the large number of new mobile devices introduced annually. In this way, specific features are used to simplify the dataset and provide an approximate cost at which to bring the product to market. The precise outcomes of the predicted pricing and other attributes depend on a wide range of factors. Buyers, marketers, and app creators may all benefit from the mobile dataset's rich historical data by making more informed decisions based on hard facts. In order to evaluate the dataset, the KNN Model technique is implemented in this project to calculate the distances between K models. K model is used to calculate accuracy with the help of the training model. The KNN model is used to the price prediction model. When it comes to making a purchase decision, consumers generally base their choice on the price of a product. The price of a product may make or break it in today's unpredictable market. Prior to releasing a product, it is crucial for a business to choose the best pricing point. Setting the market price for a product requires careful consideration, and a tool that delivers an estimate of the price after factoring in the characteristics it provides might be helpful. The customer may input their desired product characteristics and get a pricing estimate from the tool. Depending on the data at hand and the goal of the project at hand, the appropriate machine learning method may be selected. Python, MATLAB, Java, WEKA, Cygwin, Octave, etc. are only some of the languages and tools that may be used for machine learning. Naive Bayes, K-NN, etc., are only a few examples of popular algorithms. To improve a model's accuracy and reduce its computing time, feature selection methods may be used to choose and choose the most useful parameters with which to train it. Predicting a product's price may be accomplished using any of these techniques, provided sufficient training data is made available. A mobile is a must-have item these days. It's the technology market's quickest-changing and quickest-moving product. There is a constant influx of new, improved mobile phones into the market. Daily sales of mobile phones are in the thousands. A mobile firm has to charge competitive pricing in order to survive in an ever-changing industry. The process of pricing anything begins with an estimate based on the item's qualities. To that end, this study sets out to create a machine learning model that can predict the cost of a smartphone in terms of its specifications. A prospective customer may use the model to get an idea of the cost of a mobile phone by specifying the features they need. Most items with identical independent variable characteristics may have their prices estimated using the same method as used to establish a prediction model. Many factors, like the phone's CPU, battery life, camera quality, display size, thickness, etc., affect how much it costs. Phones may be sorted into different price points and quality levels based on these characteristics. This work employs supervised ML techniques since the dataset utilised has a clear classification for pricing.

Literature Review

Md. Hafizur Rahman et.al.,(2021) The prediction of future stock prices is an essential part of the area of study that has focused on the volatility of stock prices. Estimating the future price of a stock and, by extension, its worth, is a common use of predictive modelling. In this research we investigate using machine learning models with Long Short-Term Memory (LSTM) for predicting stock prices. It is possible to make very precise forecasts about stock prices if one uses a suitable model. This study focuses on stock market forecasting using LSTM models, despite the fact that there is a large body of literature on the fundamental analysis of stock prices, which emphasises the identification and understanding of stock price patterns. Our analysis relied on data from the 30 most successful corporations in the Dow Jones Sustainability Index (DSE30). Stock price forecasts were run via two LSTM models and compared. These models were trained using these firms' stock data from January 2019 through January 2021. The main goal was to find the model with the best prediction accuracy using the LSTM architecture.

Mustafa Çetın et.al.,(2021) Several sectors, including business, education, manufacturing, healthcare, and e-commerce, are increasingly using machine learning (ML) algorithms. The effectiveness of ML algorithms is highly reliant on the quality of the dataset and the processing techniques used. Selecting the right algorithm, preprocessing, and postprocessing methods is essential for achieving accurate results. This study compares the performance of six different classification approaches for predicting mobile phone price categories: the Random Forest Classifier, the Logistic

Regression Classifier, the Decision Tree Classifier, the Linear Discriminant Analysis, the K-Nearest Neighbour Classifier, and the SVC. We use the "Mobile Price Classification" dataset gathered from Kaggle.com to do our analysis. We start by making sure there are no gaps in the information we have. Scaling the dataset improves the input data's use for ML algorithms. As part of our efforts to reduce the number of required inputs and the associated computing cost while maintaining the integrity of the most important qualities, we also use feature selection techniques. Last but not least, we fine-tune the parameters of categorization algorithms to improve the system's accuracy. Our findings suggest that the ANOVA f-test feature selection technique is the best option for this dataset, as it achieves satisfactory accuracy with a minimum number of features. Furthermore, the SVC classifier has the highest test accuracy compared to other models.

Haomeng Liu et.al.,(2020) The recycled mobile phone industry has expanded greatly in recent years, drawing further focus on the device's autonomous price dilemma. Due to the existence of nonlinear mapping connections and fuzzy notions, the pricing issue for recycled mobile phones is complicated, making an accurate evaluation of the worth of these phones challenging. This research presents a new pricing model using Fuzzy Neural Networks with Momentum in Updating the Parameters (FNN-MUP) to improve precision in pricing. In the first step, we use principal component analysis to identify the most important factors influencing the cost of mobile phone recycling. We then develop a pricing model using fuzzy neural networks to relate the primary features to the cost of the mobile device. The parameters of the fuzzy neural network are then optimised and updated using the momentum approach. We compare the suggested approach to six other pricing strategies by analysing 1,200 data points culled from actual sales of discarded mobile phones. According to the findings, the suggested model outperforms the competing models in terms of both average relative error and prediction accuracy.

Prediction

There are many different methods and algorithms that may be used in machine learning to predict the price of a mobile phone.

Gather Information: To begin, amass a database including information on mobile phones, including their ages, conditions, prices, brands, models, RAM, storage capacities, cameras, etc..

Data Preprocessing: After gathering the data, it must be preprocessed by being cleaned, transformed, and normalised so that it can be used by the machine learning algorithm.

Feature Engineering: Data preparation is the first step in engineering novel characteristics that may enhance a model's functionality. Screen size, pixel density, and cost-to-features comparisons are just some of the metrics that may be determined.

Model Selection: Select a machine learning method that works well with the available data and solves the issue at hand. Linear regression, decision trees, random forests, and neural networks are only a few examples of often used methods for such regression situations.

Train the ModelSeparate the data into a test set and a training set. Train the model with one set of data, and then test it on another to see how well it did.

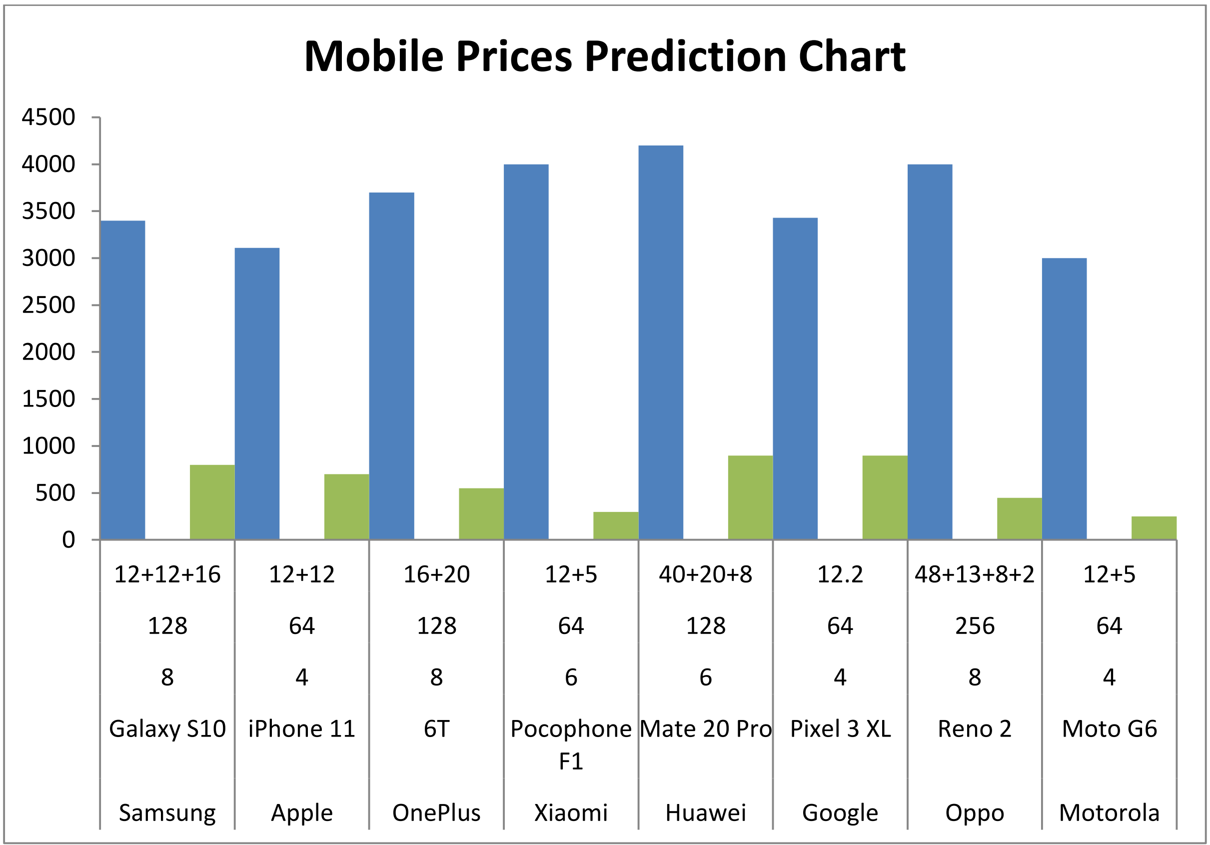
Evaluate the Model: Use measures like mean absolute error (MAE), mean squared error (MSE), and coefficient of determination (R-squared) to assess the model's efficacy.

DOI : https://doi.org/10.56452/7-2-549

Predict Mobile Prices: After the model has been trained and assessed, it may be used to provide pricing predictions for upcoming mobile devices.

Brand	Model	RAM (GB)	Storage (GB)	Camera (MP)	Battery Capacity (mAh)	Screen Size (inches)	Price ($)
Samsung	Galaxy S10	8	128	12+12+16	3400	6.1	799
Apple	iPhone 11	4	64	12+12	3110	6.1	699
OnePlus	6T	8	128	16+20	3700	6.41	549
Xiaomi	Pocophone F1	6	64	12+5	4000	6.18	299
Huawei	Mate 20 Pro	6	128	40+20+8	4200	6.39	899
Google	Pixel 3 XL	4	64	12.2	3430	6.3	899
Oppo	Reno 2	8	256	48+13+8+2	4000	6.55	449
Motorola	Moto G6	4	64	12+5	3000	5.7	249

Brand, model, RAM, storage, camera, battery capacity, screen size, and price are all shown in this table for each mobile device. Our ultimate goal is to use the other characteristics to make predictions about the pricing. Using this information, we may train a machine learning model to anticipate the cost of upcoming mobile devices on the basis of their specifications.



DOI : https://doi.org/10.56452/7-2-549

Random Forest Classifier

Random forest algorithm can be used for both regression and classification tasks. For classification tasks, each tree in the forest predicts a class and the class with the highest number of votes becomes the predicted class for the random forest. For regression tasks, the average of the predicted values by each tree in the forest becomes the final predicted value for the random forest.

One of the advantages of random forest is its ability to handle missing data and maintain accuracy even with large datasets. It also avoids overfitting and reduces the variance by aggregating multiple trees. Random forest is widely used in various fields, including finance, healthcare, marketing, and image recognition.

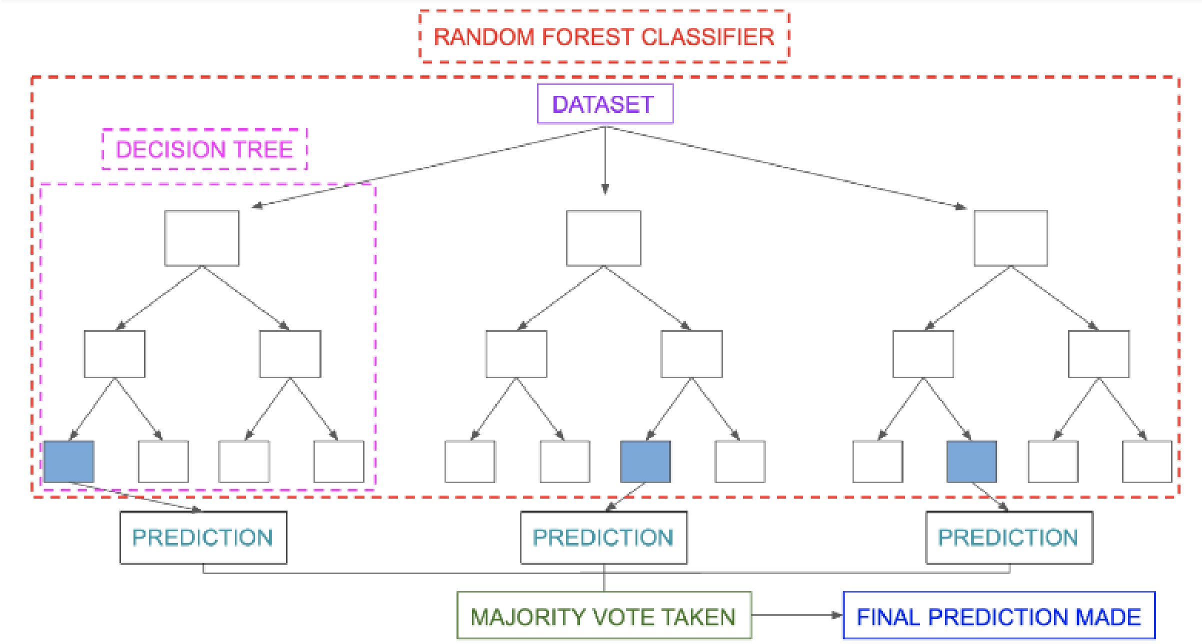


Figure 2.Random Forest Classifier Prediction

ImageSource: https://miro.medium.com/max/5752/1*5dq_1hnqkboZTcKFfwbO9A.png

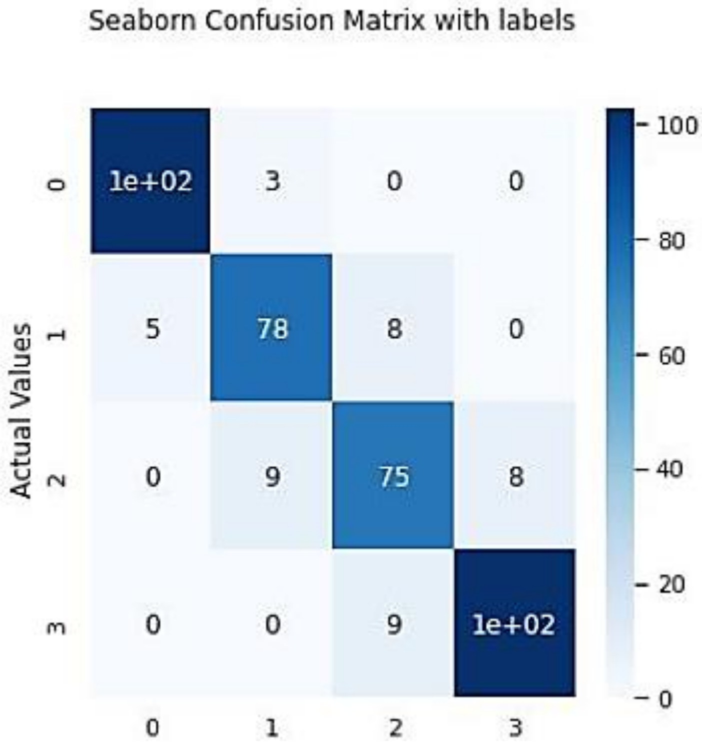


Figure 3 Regression Model for mobile price prediction

Conclusion

Our study demonstrates the potential of using machine learning models for predicting mobile phone prices based on various features. The Random Forest Regressor model provided the best accuracy, and screen size, battery capacity, internal memory, and RAM were identified as the most important features for predicting mobile phone prices. This information can provide valuable insights for manufacturers and retailers when deciding on the pricing of their products. Our study also highlights the importance of data preprocessing, feature engineering, and model selection in achieving accurate predictions

References

1. M. H. Rahman, S. I. Nahid, I. H. Al Fahad, F. M. Nahid and M. M. Khan, "Price Prediction Using LSTM Based Machine Learning Models," *2021 IEEE 12th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, Vancouver, BC, Canada, 2021, pp. 0453-0459, doi: 10.1109/IEMCON53756.2021.9623120
2. M. Çetın and Y. Koç, "Mobile Phone Price Class Prediction Using Different Classification Algorithms with Feature Selection and Parameter Optimization," *2021 5th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)*, Ankara, Turkey, 2021, pp. 483-487, doi: 10.1109/ISMSIT52890.2021.9604550.
3. M. H. Rahman, S. I. Nahid, I. H. Al Fahad, F. M. Nahid and M. M. Khan, "Price Prediction Using LSTM Based Machine Learning Models," *2021 IEEE 12th Annual Information*

Technology, Electronics and Mobile Communication Conference (IEMCON), Vancouver, BC, Canada, 2021, pp. 0453-0459, doi: 10.1109/IEMCON53756.2021.9623120.

4. . H. Liu, J. Huang, H. Han and H. Yang, "An Improved Intelligent Pricing Model for Recycled Mobile Phones," *2020 Chinese Automation Congress (CAC)*, Shanghai, China, 2020, pp. 3724-3731, doi: 10.1109/CAC51589.2020.9327611.
5. X. Yuan, M. Sun, Q. Fang and C. Du, "DLECP: A Dynamic Learning-based Edge Cloud Placement Framework for Mobile Cloud Computing," *IEEE INFOCOM 2019 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, Paris, France, 2019, pp. 1035-1036, doi: 10.1109/INFCOMW.2019.8845174.
6. J. Ren *et al.*, "Incentivized Social-Aware Proactive Device Caching With User Preference Prediction," in *IEEE Access*, vol. 7, pp. 136148-136160, 2019, doi: 10.1109/ACCESS.2019.2942440.
7. C. Wang and Q. Gao, "High and Low Prices Prediction of Soybean Futures with LSTM Neural Network," *2018 IEEE 9th International Conference on Software Engineering and Service Science (ICSESS)*, Beijing, China, 2018, pp. 140-143, doi: 10.1109/ICSESS.2018.8663896.
8. S. Karasu, A. Altan, Z. Saraç and R. Hacioğlu, "Prediction of Bitcoin prices with machine learning methods using time series data," *2018 26th Signal Processing and Communications Applications Conference (SIU)*, Izmir, Turkey, 2018, pp. 1-4, doi: 10.1109/SIU.2018.8404760.
9. A. Izzah, Y. A. Sari, R. Widyastuti and T. A. Cinderatama, "Mobile app for stock prediction using Improved Multiple Linear Regression," *2017 International Conference on Sustainable Information Engineering and Technology (SIET)*, Malang, Indonesia, 2017, pp. 150-154, doi: 10.1109/SIET.2017.8304126.
10. Y. Udagawa, "Predicting Stock Price Trend Using Candlestick Chart Blending Technique," *2018 IEEE International Conference on Big Data (Big Data)*, Seattle, WA, USA, 2018, pp. 4162-4168, doi: 10.1109/BigData.2018.8622402.
11. B. Jeevan, E. Naresh, B. P. V. kumar and P. Kambli, "Share Price Prediction using Machine Learning Technique," *2018 3rd International Conference on Circuits, Control, Communication and Computing (I4C)*, Bangalore, India, 2018, pp. 1-4, doi: 10.1109/CIMCA.2018.8739647.
12. G. Smail and J. Weijia, "Techno-economic analysis and prediction for the deployment of 5G mobile network," *2017 20th Conference on Innovations in Clouds, Internet and Networks (ICIN)*, Paris, France, 2017, pp. 9-16, doi: 10.1109/ICIN.2017.7899243.
13. B. Jeevan, E. Naresh, B. P. V. kumar and P. Kambli, "Share Price Prediction using Machine Learning Technique," *2018 3rd International Conference on Circuits, Control, Communication and Computing (I4C)*, Bangalore, India, 2018, pp. 1-4, doi: 10.1109/CIMCA.2018.8739647.
14. J. Zhang, J. Xie and M. Yuan, "A User-Oriented Approach toward Price Prediction for IaaS Service," *2018 IEEE International Conference on Web Services (ICWS)*, San Francisco, CA, USA, 2018, pp. 343-346, doi: 10.1109/ICWS.2018.00057.
15. M. Chaitanya Lahari, D. H. Ravi and R. Bharathi, "Fuel Price Prediction Using RNN," *2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, Bangalore, India, 2018, pp. 1510-1514, doi: 10.1109/ICACCI.2018.8554642.