Smart Phone Price Prediction using Machine Learning

Meenakshi Sudhish, Prisly Mary Jacob, Vinayak Pulkuzhiyil

Abstract

The report presents a study on prediction of the price of a smart phone using machine learning algorithms, given its input. We also aim to identify what other machine learning algorithms can be used to get a better prediction. Regression models like Linear Regression, Ridge Regression, KNN(k-nearest neighbors algorithm), Decision Tree regression, and Classification models like Decision Tree classifier, Random Forest Classifier and SVM(Support Vector Machine) were experimented with.

1 Introduction

The increasing usage and importance of smartphones in our day-to-day lives necessitates accurate price prediction models that can assist stakeholders(manufacturers, retailers, and consumers) in making informed decisions. This report presents a study on Smart Phone Price Prediction using various Machine Learning techniques. The primary aim of this research is to develop a model to predict the price of a smartphone given the specifications.

Predictive Machine Learning works by taking in data as input to develop and train a prediction model and the trained model is used to predict the outcome of future data instances. The performance of our model is evaluated using accuracy, and (R^2) score as metrics. Supervised Machine Learning algorithms make use of data that contains a pre-defined class label, which is the price of a mobile in our case.

Exploratory data analysis is performed on the dataset obtained from Kaggle to gain insights into the dataset, uncover trends, and identify the significant features that influence smartphone prices.

Python is used due to its readily accessible ML libraries. Various algorithms were used to train the model to try and find the algorithm that is able to predict the mobile price class most accurately.

2 Related works

Several studies have explored the field of smartphone price prediction using machine learning techniques. These works have focused on various aspects, including feature selection, algorithm selection, and evaluation metrics.

S. Subhiksha conducted a study on Prediction of Phone Prices Using Machine Learning Techniques. Random forest classifier, support vector machine and logistic regression were used primarily. Among the three classifiers chosen, logistic regression and support vector machine had the highest accuracy of 81%. Further, logistic regression was used to predict the prices of the phone. [1]

A Varun Kiran and Dr. Jebakumar R published a study and the model trained using LDA(Linear Discriminant Analysis) was found to predict mobile price classes most accurately (95%). The accuracy of the models can be improved by doing some data preprocessing steps like normalization and standardization [2]

Muhammad Asim, Zafar Khan, published about Mobile Price Class prediction using Machine Learning Techniques to predict if the mobile with given features will be economical or expensive. Data was collected from the website www.GSMArena.com. Their work concluded comparable results of both Feature selection algorithms and classifier. Comparing the results maximum accuracy achieved is 78%, when WrapperattributEval algorithm is used for feature selection and Decision tree as a classifier. [3]

3 Evaluations

3.1 Regression Algorithm

3.1.1 Linear Regression

The linear regression model achieved a Coefficient of Determination (R^2) value of 0.92 suggests that approximately 92% of the variation in smartphone prices can be explained by the features used in the model.

3.1.2 Ridge Regression

Simple ridge regression gives (R^2) value of 0.92. Applying stacking ensemble method also gives the same result, but applying Random Forest Regressor gives an improved result of 0.94.

3.1.3 KNN

The KNN model outperformed the previous models, achieving a Coefficient of Determination value (R^2) of 0.96, suggesting that around 96% of the price variation can be accounted for by the features used in the KNN model.

3.1.4 Decision Tree Regression

The Decision Tree Regression model achieved a Coefficient of Determination (R^2) of 0.88. This indicates that the model provides a reasonably good fit to the data, explaining approximately 88% of the price variation. However, the performance of this model is relatively lower compared to the other models employed. But upon applying random forest regressor, the resultis an improved value of 0.94.

3.2 Classification Algorithms

3.2.1 Logistic Regression

Logistic Regression achieved an accuracy of 0.955. This model demonstrates the ability to classify smartphones into the correct price range with an accuracy of 95.5%.

3.2.2 Decision Tree Classifier

The Decision Tree Classifier achieved an accuracy of 0.8275, indicating the proportion of correctly classified price ranges, and upon optimizing parameters, we get an improved accuracy of 0.845.

3.2.3 Random Forest

The Random Forest model exhibited improved performance with an accuracy of 0.8575. Optimizing hyperparameters slightly increased accuracy to 86.5%.

3.2.4 SVM

The SVM (Support Vector Machine) model achieved an accuracy of 0.91, showcasing its effectiveness in accurately classifying mobile phones into price categories. SVM models are known for their ability to handle complex decision boundaries and perform well in classification tasks.

4 Future Works

1. Enhanced Price Prediction Accuracy:

   To further improve the accuracy of price prediction, future research can focus on exploring advanced and sophisticated machine learning algorithms and techniques. Capturing complex relationships between smartphone features and prices may offer higher precision and more accurate price estimations.

2. Incorporation of New Features:

   As the smartphone market continuously evolves with the introduction of new specifications and features, future works can include a broader range of attributes to enhance the predictive power of the models. Models should be able to capture the changing dynamics of the market and provide more accurate price predictions.

3. Incorporation of External Data Sources:

   Integrating external data sources like market trends, economic indicators, consumer sentiment, and competitor pricing information can help leverage a more comprehensive set of factors to make more accurate price predictions.

4. Evaluation of Alternative Evaluation Metrics:

   While this study evaluated the models based on common metrics like Coefficient of Determination (R^2), and accuracy, future research can explore the application of alternative evaluation metrics, which will help gain additional insights into the performance of the models and offer a more comprehensive evaluation of their predictive capabilities.

5 Results and Discussion

5.1 Regression Algorithms

5.1.1 Linear Regression Model

Coefficient of determination (R^2) = 0.92

5.1.2 Ridge Regression Model

Coefficient of determination (R^2) = 0.94

5.1.3 KNN Model

Coefficient of determination (R^2) = 0.96

5.1.4 Decision Tree Regression Model

Coefficient of determination (R^2) = 0.94

5.2 Classification Algorithms

5.2.1 Logistic Regression Model

Accuracy = 0.955

5.2.2 Decision Tree Classifier Model

Accuracy = 0.845

5.2.3 Random Forest Classifier Model

Accuracy = 0.865

5.2.4 SVM Model

Accuracy = 0.91

The regression algorithms, including Linear Regression, Ridge Regression, KNN, and Decision Tree Regression, were evaluated. The Linear Regression model achieved a coefficient of determination (R^2) of 0.92, indicating a reasonable fit to the data. The Ridge Regression model improved upon this with a coefficient of determination of 0.94, incorporating regularization techniques to enhance price estimation. The KNN model performed exceptionally well, achieving a coefficient of determination of 0.96. This model exhibited a strong correlation between the features and prices, enabling accurate price predictions. The Decision Tree Regression model also showed promise, achieving a coefficient of determination of 0.94, utilizing a tree-like structure to segment the data and make price predictions.

In terms of classification, Logistic Regression, Decision Tree Classifier, Random Forest Classifier, and SVM models were used to categorize the mobile phones into price ranges. The Logistic Regression model achieved an accuracy of 0.955, indicating its effectiveness in accurately classifying phones. The Decision Tree Classifier model achieved an accuracy

of 0.845, while the Random Forest Classifier model achieved an accuracy of 0.865. These models demonstrated their ability to classify phones into price ranges based on their features. The SVM model achieved an accuracy of 0.91, showcasing its capability to accurately classify phones into different price categories. .

6 Conclusion

Overall, the results highlight the effectiveness of machine learning algorithms in predicting mobile phone prices and classifying them into price ranges. The regression models provided reasonable price estimation, with the KNN model exhibiting the highest accuracy. The classification models demonstrated their ability to categorize phones accurately, with the Logistic Regression model achieving the highest accuracy. These findings emphasize the potential of machine learning techniques in assisting with mobile phone price prediction and classification tasks, offering valuable insights for both manufacturers and consumers in the ever-evolving smartphone market.

References

(1) Subhiksha, S.; Thota, S.; Sangeetha, J. In *Data Engineering and Communication Technology: Proceedings of 3rd ICDECT-2K19*, 2020, pp 781–789.

(2) Kiran, A. V.; Jebakumar, R. Prediction of Mobile Phone Price Class using Supervised Machine Learning Techniques. *International Journal of Innovative Science and Research Technology* **2022**, *7*, 248–251.

(3) Asim, M.; Khan, Z. Mobile price class prediction using machine learning techniques. *International Journal of Computer Applications* **2018**, *179*, 6–11.

(4) Zehtab-Salmasi, A.; Feizi-Derakhshi, A.-R.; Nikzad-Khasmakhi, N.; Asgari-Chenaghlu, M.; Nabipour, S. Multimodal price prediction. *Annals of Data Science* **2021**, 1–17.

(5) Brackmann, C.; Hütsch, M.; Wulfert, T. Identifying Application Areas for Machine Learning in the Retail Sector: A Literature Review and Interview Study. *SN Computer Science* **2023**, *4*, 426.

(6) Sarker, I. H. Machine learning: Algorithms, real-world applications and research directions. *SN computer science* **2021**, *2*, 160.

(7) Nasser, I. M.; Al-Shawwa, M. O.; Abu-Naser, S. S. Developing Artificial Neural Network for Predicting Mobile Phone Price Range. **2019**.

(8) Chen, M. Mobile Phone Price Prediction with Feature Reduction. *Highlights in Science, Engineering and Technology* **2023**, *34*, 155–162.

(9) Güvenç, E.; Çetin, G.; Koçak, H. Comparison of KNN and DNN classifiers performance in predicting mobile phone price ranges. *Advances in Artificial Intelligence Research* **2021**, *1*, 19–28.

(10) Mobile Price Classification, `https://www.kaggle.com/datasets/iabhishekofficial/mobile-price-classification`.

(11) The research of regression model in machine learning field, https://www.researchgate.net/publication/326121964_The_research_of_regression_model_in_machine_learning_field.

(12) Learn Mobile Price Prediction Through Four Classification Algorithms, https://www.analyticsvidhya.com/blog/2022/02/learn-mobile-price-prediction-through-four-classification-algorithms/.

(13) A Practical Approach to Linear Regression in Machine Learning, https://towardsdatascience.com/linear-regression-5100fe32993a.

(14) Handling Missing Values with Pandas, https://towardsdatascience.com/handling-missing-values-with-pandas-b876bf6f008f.

(15) Python — Decision Tree Regression using sklearn, https://www.geeksforgeeks.org/python-decision-tree-regression-using-sklearn/.

(16) Machine Learning Series Day 6 (Decision Tree Regressor), https://becominghuman.ai/machine-learning-series-day-6-decision-tree-regressor-82a2e2f873a?gi=8c145752cb71.

(17) Prediction of Phone Prices Using Machine Learning Techniques, https://www.researchgate.net/publication/338471736_Prediction_of_Phone_Prices_Using_Machine_Learning_Techniques.

(18) Top Evaluation Metrics for Regression Problems in Machine Learning, https://www.freecodecamp.org/news/evaluation-metrics-for-regression-problems-machine-learning/#:~:text=RMSE\%20is\%20a\%20popular\%20evaluation,the\%20effect\%20of\%20large\%20errors.

(19) A Comprehensive Guide on Hyperparameter Tuning and its Techniques, https://www.analyticsvidhya.com/blog/2022/02/a-comprehensive-guide-on-hyperparameter-tuning-and-its-techniques/.

(20) 3 Best metrics to evaluate Regression Model?, https://towardsdatascience.com/what-are-the-best-metrics-to-evaluate-your-regression-model-418ca481755b.

(21) Cross-Validation, https://docs.aws.amazon.com/machine-learning/latest/dg/cross-validation.html.

(22) How To Conduct Exploratory Data Analysis in 6 Steps, https://www.indeed.com/career-advice/career-development/how-to-conduct-exploratory-data-analysis.

(23) Quick Introduction to Boosting Algorithms in Machine Learning, https://www.analyticsvidhya.com/blog/2015/11/quick-introduction-boosting-algorithms-machine-learning/.

(24) Random Forest Feature Importance Plot, https://www.analyseup.com/learn-python-for-data-science/python-random-forest-feature-importance-plot.html.