Universitat Politecnica De Catalunya (UPC)

Mobile Price Prediction using Support Vector Machine (SVM)

ML PROJECT - BDMA

Asha Seif (asha.seif@estudiantat.upc.edu) & Hang Yu (hang.yu1@estudiantat.upc.edu)

INTRODUCTION

The most difficult challenge that is faced by business companies is selecting the best price for their new products. To solve this problem Machine Learning (ML) algorithms are developed to predict the price of the new product, based on features of existing similar products.

The aim of this project is to predict the price range of mobile phones, based on features such as RAM, internal memory, and other features. This problem is not about predicting the actual price of the phone (i.e not a regression) but is about predicting the price range which can be either low (0), medium(1), high(2), or very high (3).

Five ML algorithms such as Random Forest, Support Vector Machine (SVM), Naïve Bayes, Extreme Gradient Boost (xgboost), and Decision Tree will be tested and evaluated to find the best estimator for solving this problem.

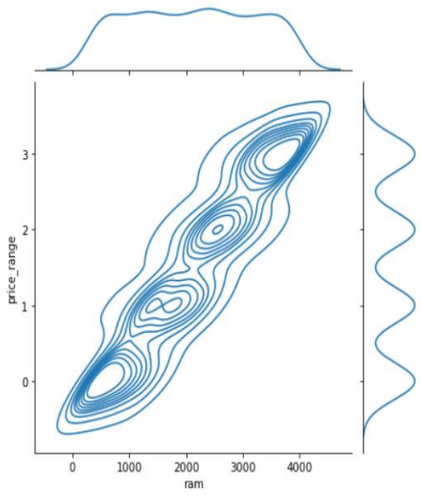
DATA

The data used was collected from Kaggle [1] which consists of 2000 samples with 21 features such as RAM, battery power, size of internal memory, and other features.

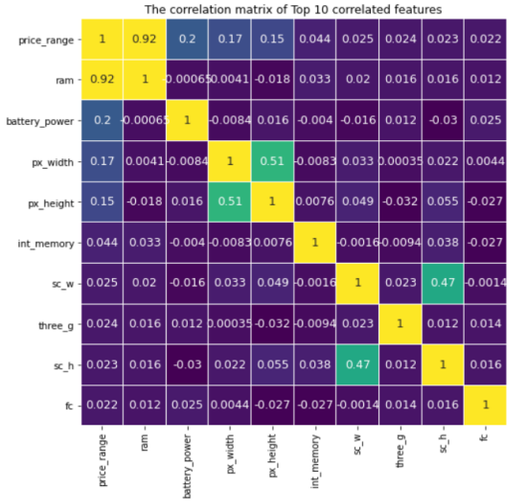
EXPLORATORY DATA ANALYSIS (EDA)

- The price of the mobile phone is highly dependent on features like RAM, battery power, pixel resolution height, and width.
- The phones which support 4G and 3G share the same price range

Ram vs Price

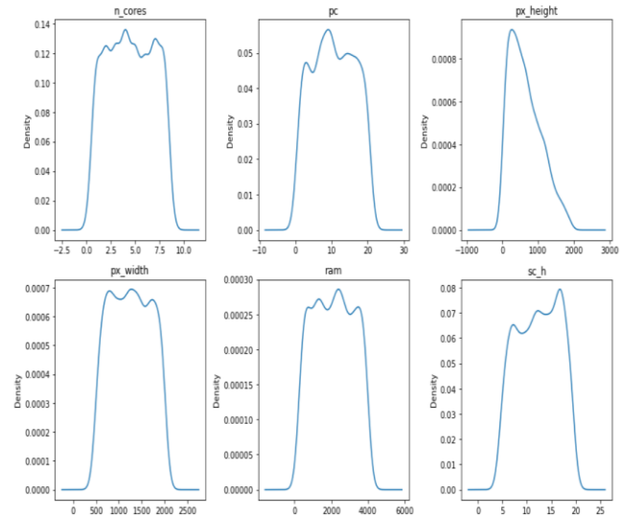


Top 10 Most Correlated Features



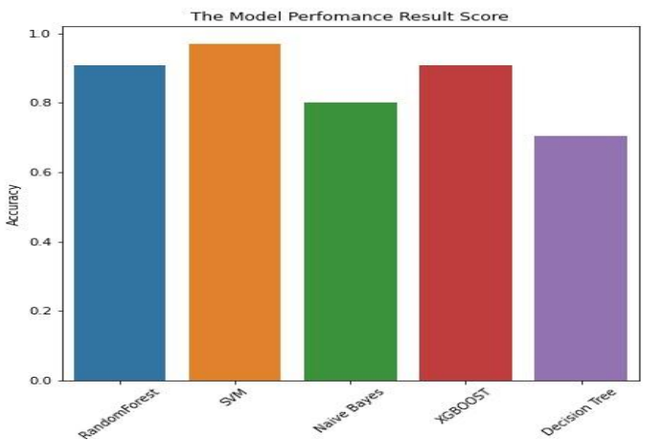
DATA DISTRIBUTION AND PREPROCESSING

- Most of the features are well distributed
- Incorrect data were computed using 1KNN
- Outliers were computed using Median Imputation



MODEL SELECTION & VALIDATION

The Five algorithms were tested to find the best estimator by using **5-fold Cross-Validation** and **GridSearch** (to find the best parameters for each algorithm).



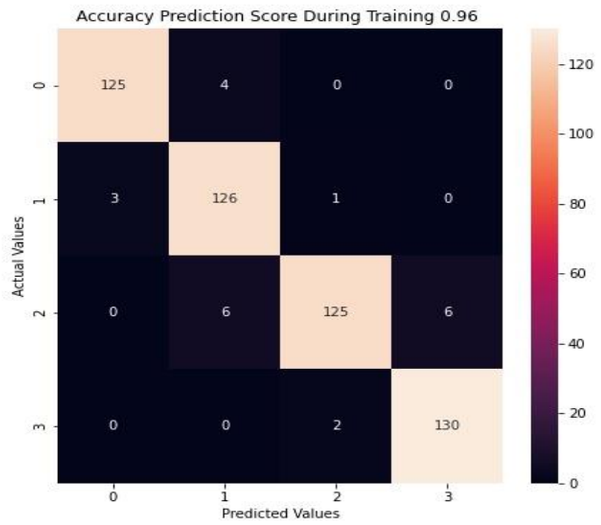
Results of The Scores of Each Algorithm

RandomForest (0.906 ≈91%), SVM (0.970≈97%), Naïve Bayes (0.801 ≈80%), XGBOOST(0.906≈91%), Decision Tree (0.703 ≈70%)

FINAL MODEL TRAINING AND ANALYSIS

The best estimator of SVM with parameters (SVC(C=1, gamma=0.1, kernel='linear')) was used as a final model.

During the training process, SVM acquires 96% prediction accuracy with approximately 4% mean error.



CONCLUSION & FUTURE WORK

Prediction Results & Conclusion

- SVM got the highest estimator score of approximately 97% compared to other algorithms.
- In Final testing, 400 samples(unseen data) were tested, 7 were incorrectly classified, and the remaining 393 were correctly classified. Hence SVM attains a prediction accuracy of 98% and an error of 2%.
- The price range of mobile phones is much dependent on RAM, battery power, pixel resolution height, and width.
- The results did not show how the price is affected by internal memory which is the main feature most people consider when buying a mobile phone.

Future Work

- More machine learning techniques such as Artificial Neural Networks can be used to predict the price
- Data can be improved to see how the internal memory affects the price of the phone as used in relation to normal life.

REFERENCES

[1] A.Sharma. (2018). *Mobile Price Classification*. Retrieved from Kaggle: https://www.kaggle.com/datasets/iabhishekofficial/mobile-price-classification?select=train.csv

[2] J.Brownlee. (2020, July 31). *How to configure k-fold-cross-validation*. Retrieved from Machine Learning Mastery: https://machinelearningmastery.com/how-to-configure-k-fold-cross-validation/

[3] J.Brownlee. (2020, April 8). *4 Types of Classification in Machine Learning*. Retrieved from Machine Learning Mastery: https://machinelearningmastery.com/types-of-classification-in-machine-learning/