American Journal of Electronics & Communication

Subhomoy Chattopadhyay et. al., American Journal of Electronics & Communication, Vol. II (2), 17-18

Classification of Mobile Price Range with Different Machine Learning Algorithms and Optimized Hyperparameters

Subhomoy Chattopadhyay
Computer Science and Engineering Department
PES University
Bangalore, India
chattopadhyay.subhomoy@gmail.com

Kishore S
Computer Science and Engineering Department
PES University
Bangalore, India
kishoresshankar@gmail.com

Abstract— This paper discusses the classification of mobile phones into different price ranges using basic features of mobile. Several supervised machine learning algorithms like Logistic Regression, KNN, Naïve Bayes, Random Forest, AdaBoost, XGBoost, Gradient Boosting with different optimization techniques like Grid-Search and Randomized-Search have been applied to get the best classification model. The most important features which influences mobile price have been extracted using feature selection methods.

Keywords— Machine Learning, Supervised Learning, Feature Engineering, Classification problem, XGBoost, Random Forest.

I. INTRODUCTION (*HEADING 1*)

The price of a product is the most important attribute of marketing that product. One of those products where price matters a lot is a smartphone because it comes with a lot of features so that a company thinks a lot about how to price this mobile which can justify the features and also cover the marketing and manufacturing costs of the mobile. In this research paper, the task of mobile price classification has been executed with Machine Learning using Python.

Mobile phones are the best-selling electronic devices as people keep updating their cell phones whenever they find new features in a new device. Thousands of mobiles are sold daily, in such a situation it is a very difficult task for someone who is planning to set up their own mobile phone business to decide what the price of the mobile should be.

Many features are very important to be considered to estimate the price of the mobile like processor of the mobile. Battery timing is also very important in today's busy schedule of human being. Size and thickness of the mobile are also important decision factors. Internal memory, Camera pixels, and video quality must be under consideration. Internet browsing is also one of the most important constraints in this technological era of 21st century. And so is the list of many features based upon those, mobile price is decided. So there will be many of the above mentioned features to classify whether the mobile would be very economical, economical, expensive or very expensive.

II. DATASET AND FEATURES

A. *Pre-processing and Exploratory Data Analysis*

price_range	1.000000
ram	0.917046
battery_power	0.200723
px_width	0.165818
px_height	0.148858
int_memory	0.044435

Fig.1: Top 5 features that are highly co-related with price range

B. *Feature Engineering and splitting the data into train-test*

After performing the steps of data cleaning and exploratory data analysis, feature engineering takes a major role.

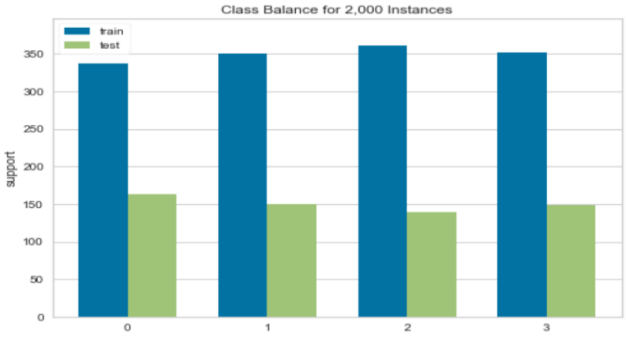

Fig.2: Feature unbalancing check

III. EXPERIMENTING WITH DIFFERENT MODELS

Fig.3: Train and test accuracy of different models

A. *Analysis of different models*

The training and testing accuracy have been shown for different models in the above data-frame. It is pretty evient that most of the models are over-fitting. So, the models have been tried to tune so that its behavior stays same for various kind of datasets.

- If the difference between the training accuracy and the testing accuracy comes out as very high, it is said

	Algorithm_name	Train_accuracy	Test_accuracy
0	Naive Bayes	0.818571	0.791667
1	KNN	0.711429	0.623333
2	LogisticRegression	0.981429	0.951667
3	DecisionTreeClassifier	1.000000	0.825000
4	RandomForest	1.000000	0.880000
5	Ada-Boost	0.770714	0.756667
6	Gradient Boosting	0.999286	0.890000
7	XGBoost	1.000000	0.913333
8	XGBoost(Grid_search)	0.844286	0.800000
9	RandomForest_Tuned	1.000000	0.880000
10	XGBoost_imp	1.000000	0.916667
11	XGBoost_imp(Grid_search)	0.844286	0.800000
12	RandomForest_imp	1.000000	0.913333
13	RandomForest_imp_CV	1.000000	0.913333

that the model is over-fitting.

- To avoid over-fitting the model should get exposed to almost all of the varieties of data during training time. If it gets trained in such manner it can predict correctly for those data which are not known previously.
- K-Fold cross-validation and Grid-Search CV are two well-known methods to reduce over-fitting. These methods help to tune the hyper-parameters of the

	precision	recall	f1-score	support
0	0.92	0.96	0.94	163
1	0.89	0.88	0.89	150
2	0.90	0.88	0.89	139
3	0.95	0.94	0.95	148
accuracy			0.92	600
macro avg	0.92	0.92	0.92	600
weighted avg	0.92	0.92	0.92	600

model so that it gets exposed to varieties of data and it gets to learn even more.

B. *Fine-Tuning the models*

The models have been tried to be tuned by using grid-search cv method. And it is evident that the random forest model is over-fitting even after getting tuned. The Extreme gradient boosting model is giving better result after getting tuned with grid-search cv provided hyper-parameters. So, the XG-Boost model with grid-search cv is taken as the best model even though the random forest model is giving very high accuracies. The random forest model is giving the difference between the training accuracy and the test accuracy very high.

C. *Train and test accuracy of the best model*

Fig. 4: Train and test accuracy of the best model

D. *Confusion Matrix*

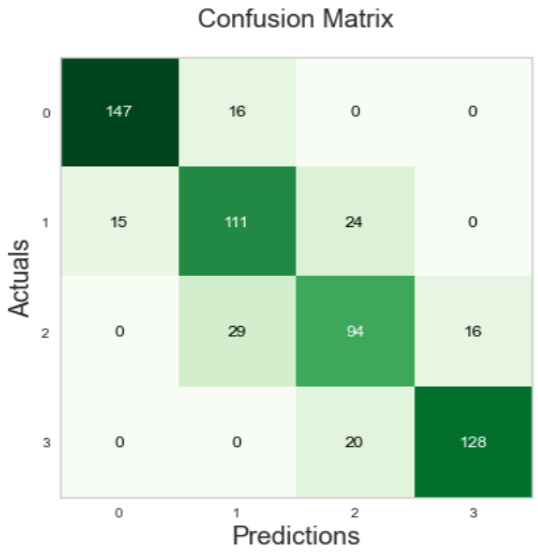

Fig. 5: Confusion Matrix of the best model

CONCLUSIONS

- RAM is the most important feature in deciding the price of the mobile phone.
- The XG-Boost model is the best model among all.

REFERENCES

[1] 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

[2] Shonda Kuiper, "Introduction to Multiple Regression: How Much Is Your Car Worth? " , Journal of Statistics Education · November 2008.

[3] Mariana Listiani , 2009. "Support Vector Regression Analysis for Price Prediction in a Car Leasing Application". Master Thesis. Hamburg University of Technology.

[4] Limsombunchai, V. 2004. "House Price Prediction: Hedonic Price Model vs. Artificial Neural Network", New Zealand Agricultural and Resource Economics Society Conference, New Zealand, pp. 25-26. 2004

[5] Mobile data and specifications online available from https://www.gsmarena.com/ (Last Accessed on Friday, December 22, 2020, 6:14:54 PM)

Identify applicable funding agency here. If none, delete this text box.