Mobile Price prediction using Machine Learning Techniques

B.Balakumar[1], P.Raviraj[2], V.Gowsalya[3]

[1]*Assistant Professor, Centre for Information Technology and Engineering, Manonmaniam Sundaranar University, Tirunelveli, India.*[1]*balakumarmsu@gmail.com*

[2]*Professor, Department of CSE, GSSS Institute of Engineering and Technology for Women,KRS Road,Metagalli,Mysore, Karnataka-570016.*
[2]*drpraviraj@gmail.com*

[3]*PG Scholar, Centre for Information Technology and Engineering, Manonmaniam Sundaranar University, Tirunelveli, India,*
[3]*vgowsalya6076@gmail.com*

ABSTRACT

To predict "If the mobile with given features will be Economical or Expensive" is the main motive of this research work. Real Dataset is collected from website https://www.kaggle.com Different feature selection algorithms are used to identify and remove less important and redundant features and have minimum computational complexity. Different classifiers are used to achieve as higher accuracy as possible. Results are compared in terms of highest accuracy achieved and minimum features selected. Conclusion is made on the base of best feature selection algorithm and best classifier for the given dataset. This work can be used in any type of marketing and business to find optimal product (with minimum cost and maximum features). To predict the accuracy of the mobile price range.

General Terms

Machine Learning

Keywords- Machine Learning, Linear Regression, KNN

1. INTRODUCTION

Price is the most effective attribute of marketing and business. The very first question of costumer is about the price of items. All the costumers are first worried and thinks "If he would be able to purchase something with given specifications or not".

Artificial Intelligence-which makes machine capable to answer the questions intelligently-now a days is very vast engineering field. Machine learning provides us best techniques for artificial intelligence like classification, regression, supervised learning and unsupervised learning and many more. Different tools are available for machine learning tasks like MATLAB, Python, Cygwin, WEKA etc. We can use any of classifiers like, Linear Regression, KNN and many more. Different type of feature selection algorithms are available to select only best features and minimize dataset. This will reduce computational complexity of the problem. As this is optimization problem so many optimization techniques are also used to reduce dimensionality of the dataset

Mobile now a days is one of the most selling and purchasing device. Every day new mobiles with new version and more features are launched.

Hundreds and thousands of mobile are sold and purchased on daily basis. So here the mobile price class prediction is a case study for the given type of problem i.e. finding optimal product. The same work can be done to estimate real price of all products like cars, bikes, generators, motors, food items, medicine etc.

Many features are very important to be considered to estimate price of mobile. For example 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 we will use many of above mentioned features to classify whether the mobile would be very economical, economical, and expensive or very_ expensive.

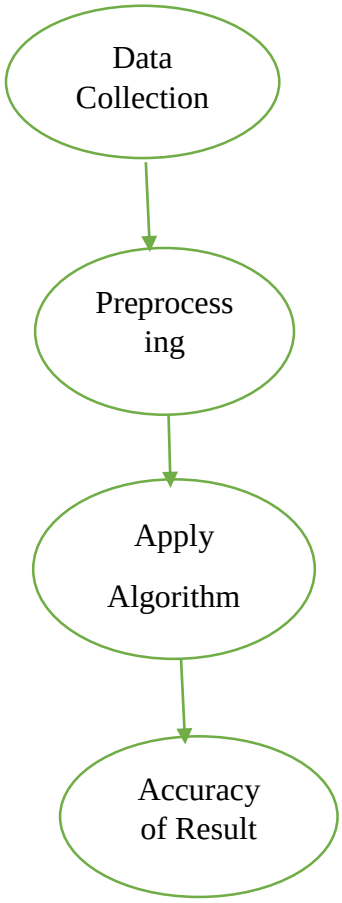
2. PREVIOUS WORK

Using previous data to predict price of available and new launching product is an interesting research background for machine learning researchers. Sameerchand-Pudaruth [1] predict the prices of second hand cars in Mauritius. He implemented many techniques like multiple linear regression, k-nearest neighbors (KNN), Decision Tree, and Naïve Bayes to predict the prices. Sameerchand-Pudaruth got Comparable results from all these techniques. During research it was found that most popular algorithms

Shonda Kuiper [2] has also worked in the same field. Kuiper used multivariate regression model to predict price of 2005 General Motor cars. He collected the data from available online source www.pakwheels.com. The main part of this research work is "Introduction of suitable variable selection techniques, which helped to find that which variables are more suitable and relevant for inclusion in model. This (His research) helps students and future researchers in many fields to understand the conditions under which studies should be conducted and gives them the knowledge to discern when appropriate techniques should be used[2].

Support Vector Machine (SVM) concept is used by one another researcher Mariana Listiani [3] for the same work. Listiani predicted prices of leased cars using above mentioned technique. It was found in this research that SVM technique is far better and accurate for price prediction as compared to other like multiple linear regression when a very large data set is available. The researcher also showed that SVM also handles high dimensional data better and avoids both the under-fitting and over-fitting issues. To find important features for SVM Listiani used Genetic Algorithm. However, the technique failed to show in terms of variance and mean standard deviation why SVM is better than simple multiple regression [3].

3. METHODOLOGY



3.1Data Collection

The features of mobiles are collected from https://www.kaggle.com

Category:

Id,Batterypower,bluetooth,clock_speed, dual_sim,Front Camera mega pixels,four_g,int_memory,Mobile Depth in cm,Weight of mobile phone,n_cores:Number of cores of processor, Primary Camera mega pixels,px_height:Pixel Resolution Height,px_width:Pixel Resolution Width,ram:Random Access Memory in Megabytes,sc_h:Screen Height of mobile in cm,sc_w:Screen Width of mobile in cm,talk_time:longest time that a single battery charge will last when you are,three_g,touch_screen,wifi

USE:

This kind of prediction will help companies estimate price of mobiles to give tough competion to other mobile manufacturer

Also it will be useful for Consumers to verify that they are paying best price for a mobile.

4. RESULT

4.1 Linear Regression Model

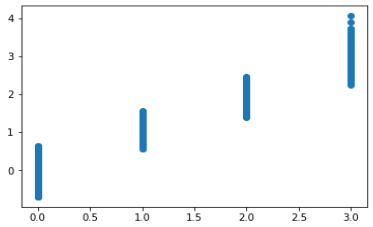
```
In [20]: from sklearn.linear_model import LinearRegression
         lm = LinearRegression()

In [21]: lm.fit(X_train,y_train)
Out[21]: LinearRegression(copy_X=True, fit_intercept=True, n_jobs=1, normal
         ize=False)

In [22]: lm.score(X_test,y_test)
Out[22]: 0.91328014881852748
```

4.2 KNN Model

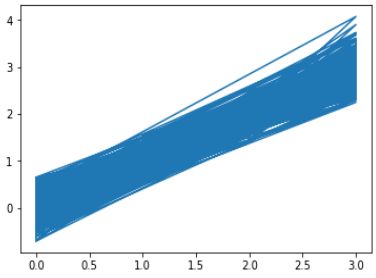
```
In [23]: from sklearn.neighbors import KNeighborsClassifier
         knn = KNeighborsClassifier(n_neighbors=10)
         knn.fit(X_train,y_train)
Out[23]: KNeighborsClassifier(algorithm='auto', leaf_size=30, metric='minko
         wski',
                    metric_params=None, n_jobs=1, n_neighbors=10, p=2,
                    weights='uniform')

In [24]: knn.score(X_test,y_test)
Out[24]: 0.92121212121212126
```

4.3 Best Accuracy: Linear Regression

```
In [38]: y_pred=lm.predict(X_test)

In [39]: plt.scatter(y_test,y_pred)

Out[39]: <matplotlib.collections.PathCollection at 0x7ff5a80c0048>
```




```
In [40]: plt.plot(y_test,y_pred)

Out[40]: [<matplotlib.lines.Line2D at 0x7ff5a80178d0>]
```



5. CONCLUSION

This work can be concluded with the comparable results of both Feature selection algorithms and classifier. This combination has achieved maximum accuracy and selected minimum but most appropriate features. It is important to note that in Forward selection by adding irrelevant or redundant features to the data set decreases the efficiency of both classifiers. While in backward selection if we remove any important feature from the data set, its efficiency decreases. The main reason of low accuracy rate is low number of instances in the data set. One more thing should also be considered while working that converting a regression problem into classification problem introduces more error

6. FUTURE WORK EXTENSION

More sophisticated artificial intelligence techniques can be used to maximized the accuracy and predict the accurate price of the products.

Software or Mobile app can be developed that will predict the market price of any new launched product.

To achieve maximum accuracy and predict more accurate, more and more instances should be added to the data set. And selecting more appropriate features can also increase the accuracy. So data set should be large and more appropriate features should be selected to achieve higher accuracy.

7. REFERENCES

[1] Sameer Chand 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. 753764

[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]https://www.kaggle.com/vikramb/mobile-price-prediction

[5] file:///C:/Users/Lenovo/Downloads/asim-2018-ijca-9165551%20(1).pdfMuhammad Asim UET Lahore Pakistan, Zafar Khan UET Lahore Pakistan

[6]https://github.com/vikram-bhati, Classification classify mobile price range

[7] Introduction to dimensionality reduction, A computer science portal for Geeks. https://www.geeksforgeeks.org/dimensionality-reduction