Turkish Online Journal of Qualitative Inquiry (TOJQI)
Volume 13, Issue 1, January 2024: 681-688

Online Mobile Price Prediction using Machine Learning

Mr.Veer Sudheer Goud, Associate Professor, Department of Artificial Intelligence and Data Science, St.Martin's Engineering College , Hyderabad, India.

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.

Keywords- Machine Learning, Linear Regression, KNN

I. INTRODUCTION

The most influential marketing and commercial feature is price. The first query posed by the customer is regarding the pricing of the items. First and foremost, all consumers are concerned, wondering "if he can buy something with the conditions provided or not." As a result, the research's primary goal is to estimate pricing at home. This paper is simply the beginning of the journey to the above-mentioned objective. Artificial intelligence, or the ability of a computer to answer questions intelligently, is currently a very broad subject of engineering. Machine learning gives us access to the most up-to-date artificial intelligence technologies, including classification, regression, supervised and unsupervised learning, and much more. MATLAB, Python, Cygwin, WEKA, and other machine learning technologies are available. Decision trees, Nave Bayes, and a variety of other classifiers are among the options.

To choose just the best characteristics and reduce the dataset, many types of algorithms are required. The computational complexity of the issue will be reduced as a result. Because this is an optimization issue, a variety of optimization techniques are frequently employed to lower the dataset's dimensionality. Mobile is currently one of the most popular apps for sales and transactions. Every day, new mobile phones with new versions and additional apps are introduced. Every day, hundreds of thousands of cell phones are sold and purchased. As a result, the prediction of the mobile pricing class is a case study for the given issue type, namely, identifying the best product. The same method may be used to determine the true cost of any item, including cars, motorbikes, generators, motors, food, medication, and so on. Mobile Processor, for example, is one of the most essential programmes for calculating mobile costs. The time of batteries is also very important in

today's hectic human existence. Size and thickness of the mobile device are other essential factors to consider when making a selection. Internal memory, camera pixels, and video consistency must all be recalled. Internet browsing is also one of the most important technical constraints of the twenty-first century. Also, the list of various features is determined by the size of the mobile device. As a result, we'll utilise all of the aforementioned characteristics to decide if the smartphone will be very-costly, economical, pricey, or very-costly. The following is a diagram of the paper's structure. The examination of past work is the next section. Technique and Experimental Procedure are covered in the third section. The results are described in Section 4 of the report. In section 5, a comparative analysis is carried out. After the section 6 paper is completed. The work's outcomes are discussed in section 7. Finally, in the eighth part, some recommendations for further research are made.

II. RELATED WORK

The use of prior data to estimate the pricing of available and new launch products is an intriguing study background for machine-learning researchers. Sameerchand-Pudaruth[1] estimates the prices of used automobiles in Mauritius. He used a variety of approaches to forecast prices, including multiple linear regressions, k-nearest neighbours (KNN), Decision Tree, and Nave Bayes. Sameerchand-Pudaruth obtained equivalent results using all of these approaches. During study, it was discovered that the majority of standard algorithms, such as Decision Tree and Nave Bayes, are incapable of processing, categorising, and forecasting numeric data. There were only 97 instances of his work (47 Toyota+38 Nissan+12 Honda). Because of the small number of cases used, the forecast accuracies were quite low [1]. Shonda Kuiper [2] was also employed in the same profession. Kuiper utilised a multivariate regression model to predict the pricing of 2005 General Motors vehicles. He gathered the information from www.pakwheels.com, an internet resource. The "Introduction of acceptable approaches for variable selection," which assisted in determining which variables are more suitable and relevant for model inclusion, is the main component of this study effort. This (His research) enables students and potential researchers in a variety of areas to examine the conditions under which investigations should be conducted and to identify when appropriate approaches should be used[2]. Mariana Listiani[3], another researcher, uses the notion of supporting vector machine (SVM) for the same study. Listiani utilised the aforementioned method to anticipate leased car costs. When there is a big data set, it was discovered that the SVM approach is considerably better and more dependable for price prediction than other methods such as multiple linear regressions.

SVM is also excellent at managing high-dimensional data and minimising both under-fitting and over-fitting issues, according to the study. Essential characteristics for SVM Listiani were identified via a genetic algorithm. In terms of variance and mean standard deviation, however, the approach failed to show why SVM is superior than basic multiple regression [3]. The Limsombunchai research [4] revealed that neural networks (NN) are effective at estimating property prices. When compared to the hedonic technique, his method was more dependable. Except for the fact that the model is first trained in NN and then assessed for prediction, all techniques function the same. NN produced greater R-sq and lower root mean square error (RMSE) using both approaches, but the hedonic offered lower values. The study was restricted since real housing values were unavailable, and the analysis work relied solely on estimates[4]. K Noor and Sadaqat J[5] have experimented with a variety of methods for estimating car prices. The researchers attained optimum accuracy by using several linear regressions. This article offers a method in which pricing is anticipated based on variables such as car type, make, area, edition, colour, mileage, alloy rims, and power steering [5].

III. SYSTEM STUDY

EXISTING SYSTEM

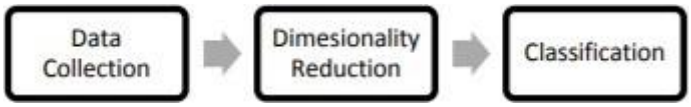
Mobile now a days is one of the most selling and purchasing device. Every day new mobiles with new version and more features are launched. Hundreds and thousands of mobile are sold and purchased on daily basis. So here the mobile price class prediction is a case study for the given type of problem i.e. 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.

PROPOSED SYSTEM

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.

IV. METHEDOLOGY

WEKA is used to carry out the experiment (Waikato Environment for Knowledge Analysis). The following are the major steps in machine learning:



Data collection

www.GSMArena.com has compiled a list of ten smartphone functions. [6] i.e. Categorization (whether the given mobile is made by Apple, Samsung, Lenovo, NOKIA etc). Whether or not a memory card slot is present, it is considered a feature. All of these features have genuine numbers with the exceptions of display size (inches), weight (g), thickness (mm), internal memory size (GB), camera pixels (MP), video quality, RAM capacity (GB), and battery (mAh).

Features	Minimum	Maximum	Mean	StdDiv
Display size(inches)	2.8	12.9	6.0	1.7
Weight(gm)	100.0	677.0	205.9	110.7
Thickness(mm)	6.0	12.8	8.2	1.1
Internal memory(GB)	0.5	256.0	39.4	33.2
Features	**Minimum**	**Maximum**	**Mean**	**StdDiv**
Camera(MP)	2.0	23.0	12.7	5.4
Video quality	240.0	2160.0	1437.7	571.1
RAM(GB)	0.5	6.0	2.9	1.4
Battery(mAh)	300.0	8827.0	3366.4	1481.7

Table 1:Dataset values

Dimensionality reduction

The technique of decreasing the number of random variables (Features) under consideration by getting a set of main variables is known as dimensionality reduction[7]. The more characteristics there are, the more difficult it is to envision the training set and subsequently work on it. Most of these characteristics are sometimes linked and therefore redundant. Dimensionality reduction techniques are useful in this situation[7]. There are two types of dimension reduction algorithms: feature selection and feature extraction.

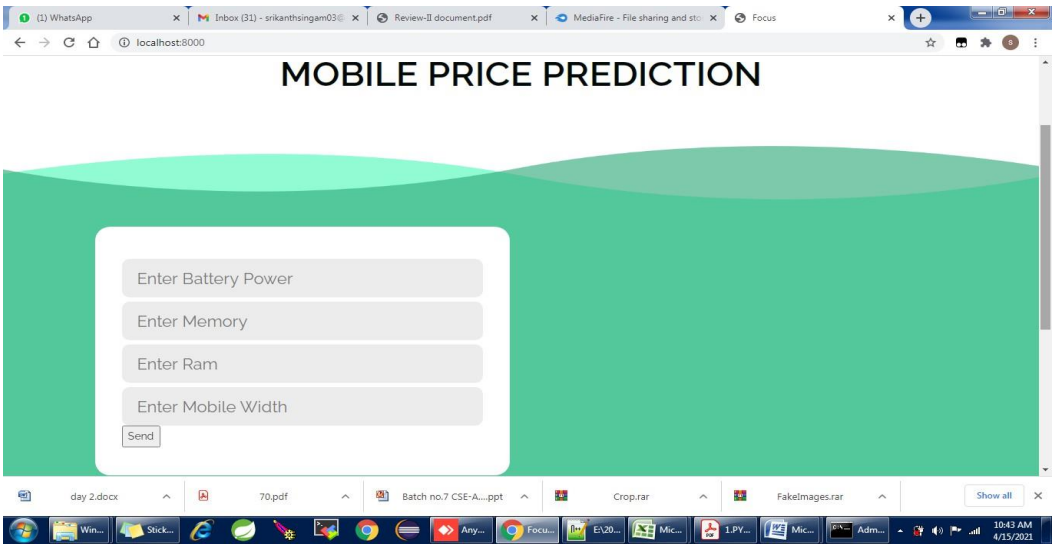
Classification

Let's go on to the final stage, classification. As previously stated, a separate test set is utilised to assess the classifier and determine accuracy. Any classification is correct if the number of correctly identified class samples (true positives), correctly identified samples that are not members of the class (true negatives), and samples that were either incorrectly assigned to the class (false positives) or not identified as class samples (false negatives) can be calculated[10]. The percentage of accurately categorised cases is called accuracy. Mathematically

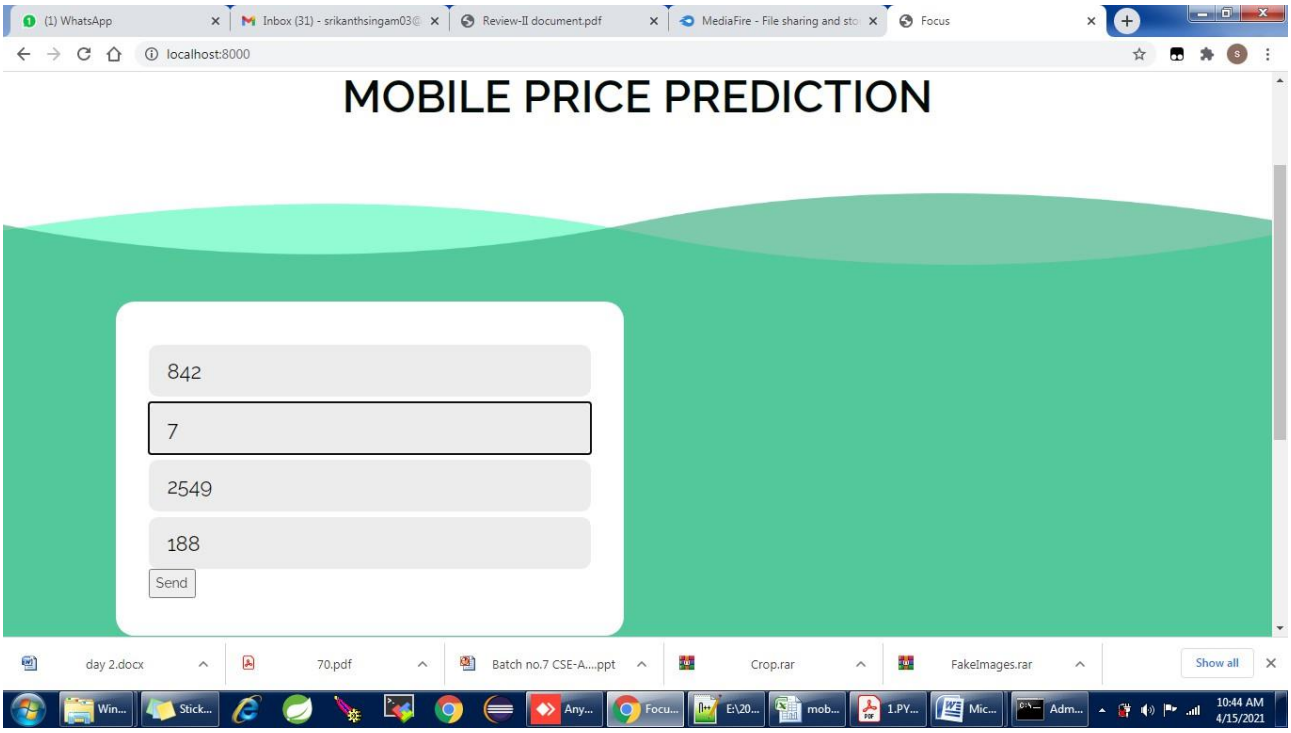
$$Accuracy = \frac{Correctly\_Classified\_Samples}{Total\_Samples} * 100$$

V. RESULT

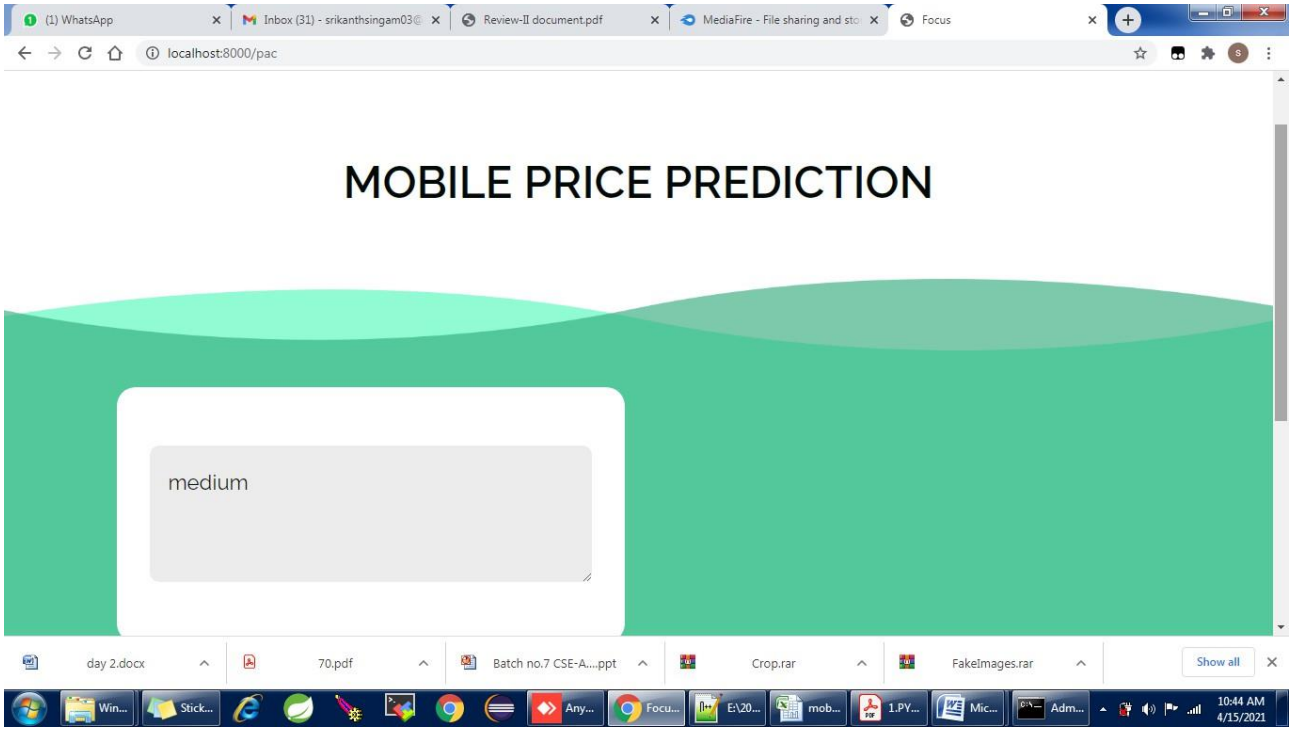
Input form:



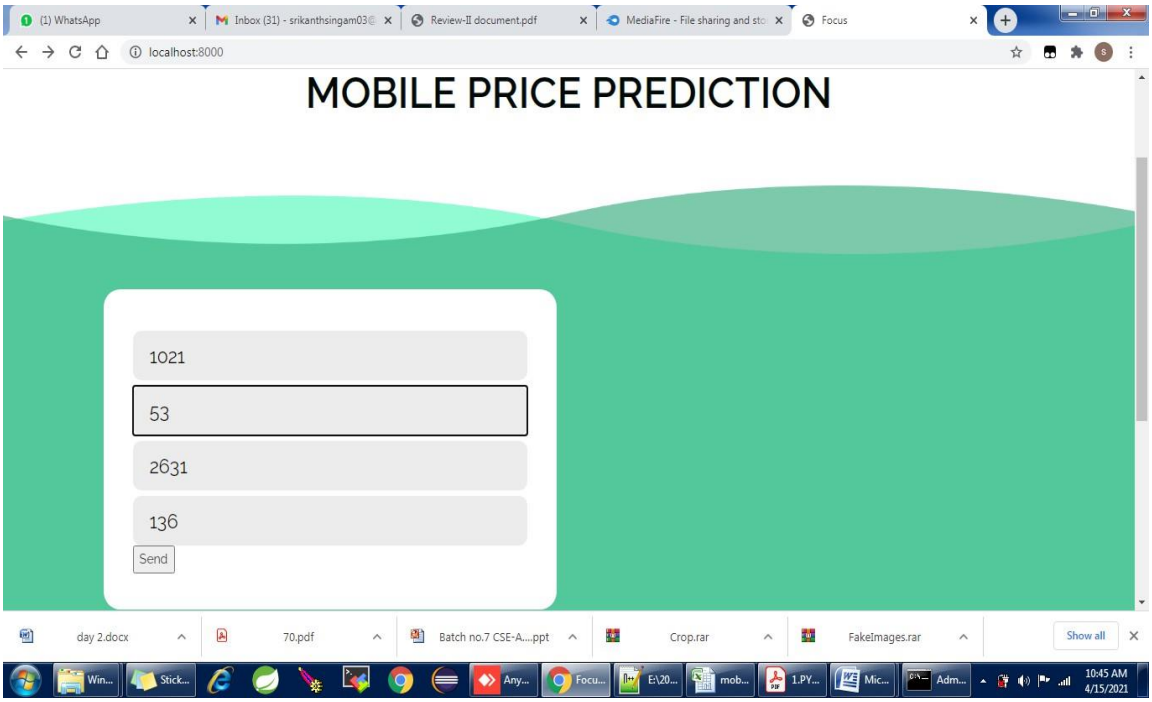
Entering values:



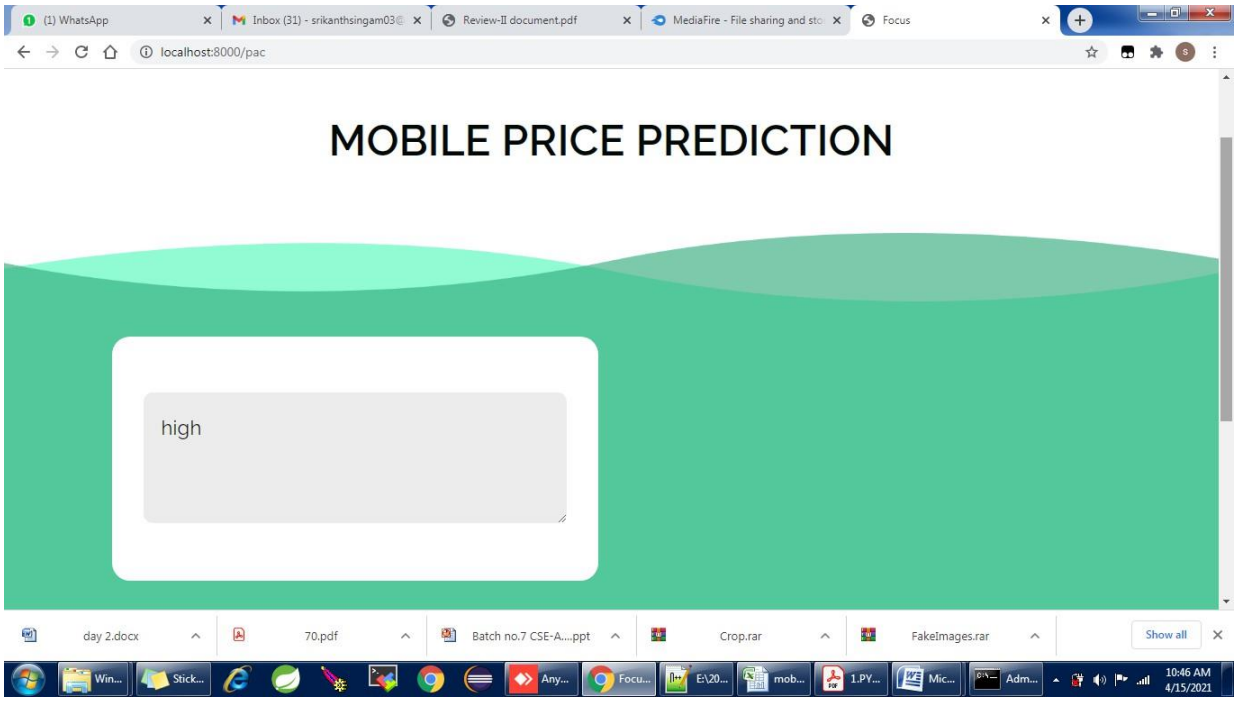
Mechine predictor Results :



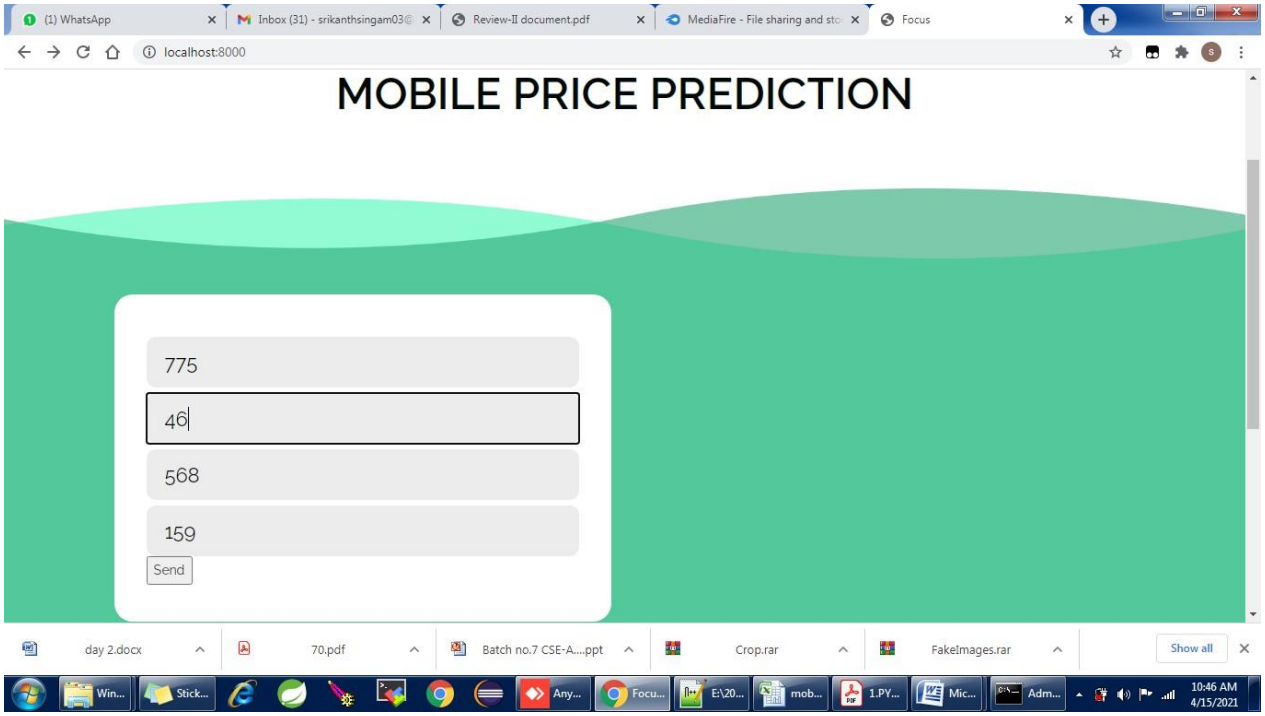
Input values:



Mechine predictor Results :



Input values:



Mechine predictor Results :

VI. 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.

Future Enhancements:

More advanced artificial intelligence algorithms can be utilised to improve accuracy and accurately anticipate product prices. Any software or mobile app that predicts the market price of a newly introduced product can be produced. More and more examples should be added to the data set to attain maximal accuracy and forecast more accurately. Additionally, picking more relevant characteristics can improve accuracy. To obtain greater accuracy, the data collection should be big, and more appropriate characteristics should be used.

VII. REFERENCES

[1] Pudaruth Sameerchand. "Predicting the price of used vehicles using machine learning techniques," International Information and Computer Technology Magazine. ISSN 0974-2239 Volume 4, Number 7 (2014), pp. 753- 764

[2] Shonda Kuiper, "Multiple Regression Introduction: How much is your car worth? "November 2008, Journal of Statistics Education

[3] Mariana Listiani , 2009. “Support Vector RegressionAnalysis for Price Prediction in a Car LeasingApplication”. 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] Kanwal Noor and Sadaqat Jan, “Vehicle Price Prediction System using Machine Learning Techniques” , International Journal of Computer Applications (0975 – 8887) Volume 167 – No.9, June 2017.

[6] Mobile data and specifications online available from https://www.gsmarena.com/ (Last Accessed on Friday, December 22, 2017, 6:14:54 PM)

[7] Introduction to dimensionality reduction, A computer science portal for Geeks.https://www.geeksforgeeks.org/dimensionalityreduction/ (Last Accessed on Monday , Jan 201822, 3 PM).

[8] Ethem Alpaydın, 2004. Introduction to Machine Learning, Third Edition. The MIT PressCambridge, Massachusetts London, England.

[9] Info Gain Attribute Eval-Weka Online available fromhttp://weka.WrapperattributEval/doc.dev/weka/attributeS election/InfoGain Attribute Eval.html (LastAccessed in Jan 2018).

[10] Thu Zar Phyu, Nyein Nyein Oo. Performance Comparison of Feature Selection Methods. MATEC Web of Conferences42, (2016).