Predicting The Price Of Used Electronic Devices Using Machine Learning Techniques

Muhammad Hasnain[1], Abdul Sajid[2], M. Arshad Awan[4]
Department of Computer Science
Allama Iqbal Open University Islamabad Pakistan
E-mail: mabbasi103@gmail.com, asajids84@gmail.com, arshad.awan@aiou.edu.pk

Amani Ayeb
Department of Computer Science
National Institute of applied sciences and technology INSAT, Tunisia
E-mail: amani.ayeb@insat.ucar.tn

Received: 11th April 2023; Accepted: 15th June 2023; Published: 1st July 2023

Abstract: The trade of used electronics devices plays vital role towards the growth of economy of the nation. In recent years, the commercial field has adopted many technological transformations, except for the used electronics. Moreover, the amount of market data is also increasing at an exponential rate. Hence, in this paper we proposed a predictive model using three Machine Learning (ML) algorithms to forecast the price of used smartphones, which are the most widely used electronic devices. In addition, the data set was collected over the last five years using web scraping techniques for modeling with machine learning. The performance of proposed model is evaluated in terms of Absolute Percentage Difference (APD) and Root Means Square Error (RMSE). Finally, the results show that the Random Forest model predicted the used smartphone price with lower prediction error and better generalization ability compared to Linear Regression and Multi-Layer Perceptron.

Key Words: Prediction, Supervised Learning, Machine Learning, Multi-Layer Perceptron, Linear Regression.

1. INTRODUCTION

Predicting the price of used electronic devices i.e. smartphones, tablets, LEDs, laptops, Scanner, Printer, Digital Camera, etc is an important and interesting problem. Due to pandemics i.e. Covid-19, which badly influenced economic conditions throughout the world. On the other hand, with the rapid growth of technology, it is being expected that the sale of used/secondhand imported devices and used electronic items will increase. Whereas, buyers apply traditional techniques to assess the value of these devices. Also, the purchaser focuses on a considerable number of distinctive qualities and properties to examine the reliability. Hence, to gauge the accurate prediction of the price of these items using modern technology i.e. Machine Learning (ML) is the need of today. Moreover, it can also help buyers to make a informed buying decisions and allow vendors to maximize their profits. However, determining the right price of used appliance can be difficult. Factors such as the age of the device, its history of use, its physical condition and the accessories included can all affect the price of a device. Although many research studies are conducted on the price prediction of electronic devices, these studies focused on the price estimation of new devices to compete in the market for launching a new product. In developed countries or even in developing countries people are becoming more brand-conscious and also prefer to have a new version of the product as it comes to the market. Nevertheless, a large number of people in the developing world cannot afford the new electronic devices, so they rely on the old electronic devices to fulfill their needs and wishes. It is observed that existing literature is not focusing on the analysis of the price estimation of used electronic items whereas, the choice of parameters is also diverse in the sale and purchase of used electronic devices. On the other hand, the price predicting accuracy also varies for different domains and factors. Therefore, it is worthy to incorporate modern technologies like ML to perform price prediction of used e-devices in general and especially smartphones.

This paper we proposed a machine learning based approach to predict the selling price. Our model considers various characteristics of the device and its usage history to predict its price. These characteristics are extracted from a dataset that contains information about different cell phones, such as brand, model, storage capacity, screen size, as well as the number of previous owners. We will also use other relevant characteristics such as the condition of the device, its usage time, and the presence of accessories. The experimental results show that our prediction model has a high-level of accuracy in predicting the selling price of used mobile devices. The results are compared with previous studies and our model is shown to be more accurate than existing models. The proposed approach has the potential to significantly improve the price of used devices, making the market more efficient for buyers and sellers.

For example, in the study of [1], cell phone price prediction using classification techniques was performed. The original dataset consisted of four price classes and 20 feature sets. Then, an analysis of variance measure and mutual information score was applied to the dataset for attribute elimination. Therefore, they compared Random Forest Classifier (RFC), Logistic Regression Classifier (LRC), Decision Tree Classifier (DTC), Linear Discriminant Analysis (LDA), K-Nearest Neighbor Classifier (KNNC), and Support Vector Regression (SVR) to evaluate the performance of the methods used to predict cell phone prices.

A similar study was conducted by [2], and a framework for cell phone price prediction via ML was proposed for four cell phone price ranges. The data was collected from an online website, and then different feature selection techniques were applied to identify, to remove less important and redundant features. The labeled dataset consists of eight features. In this paper, two classifiers, decision tree and Naive Bayes, were implemented.

One more study presented Laptops price prediction model by applying the supervised ML techniques [3] adopted multiple linear regressions as the ML prediction method which offered 81% prediction precision. According to the study, using multiple linear regressions, there were multiple independent variables and proposed a system where they consider price as the dependent variable which is predicted on the bases of Laptops' features set i.e. i) Model, ii) RAM, iii) ROM (HDD/SSD), iv) GPU, v) CPU, vi) IPS Display, and vii) Touch Screen.

In the study of [9] introduced another price prediction model of crude oil and incorporates long-term and short-term memory using the deep learning approach of ML, based on prior knowledge. According to the study, the proposed model also offers a data transfer according to previous knowledge and inherits the characteristics of long and short-term memory. The basic purpose of the proposed model is to train using deep learning algorithms with long short-term memory factors. According to the study, the result of the proposed predictive model improved the abilities of the model either with data transfer or without data transfer also compared it by several evaluation methods. The results of the study suggested that data transfer attributes can greatly improve the prediction accuracy of long and short-term memory.

According to the study of [10] the ARIMA models have been explored in detail using the time series prediction method. As per the study, they present an extensive process of building a stock price predictive model based on ARIMA algorithms. They obtained and collected published stock market data from two stock exchanges one is located in the USA i.e. New York Stock Exchange (NYSE) while the other one is in Nigeria i.e. Nigeria Stock Exchange (NSE). The data of both stock exchange markets are used for price forecasts in the proposed study. The results of the study revealed that the selected model i.e. ARIMA has a strong potential for short-term prediction and it can compete favorably with existing techniques for gauging stock market price estimates.

Finally, another paper by [4] proposed and developed a predicting model of the price of mobile using an Artificial Neural Network (ANN). Various factors like power and life of battery, Central Processing Unit (CPU) clock speed, Dual SIM, and Front Camera megapixels were considered as feature set. Accordingly, another study highlighted and emphasized [14] for the use of emerging technologies to predict the infectious disease using Machine Learning classification techniques. In the following introduction, data collection is described. Then, the exploratory analysis of the data set is presented. Next, the experimental results obtained are compared to each model to evaluate the prediction model's performance. Accordingly, existing literature is not focusing on the analysis of the price estimation of used electronic items whereas, the choice of parameters is also diverse in the sale and purchase of used electronic devices. On the other hand, the price predicting accuracy also varies for different domains and factors. Therefore, it is worthy to incorporate modern technologies like ML to perform price prediction of used e-devices in general and especially smartphones.

2. DATA COLLECTION

The first step in ML is data collection. In this paper, the data used on smartphone pricing was collected by web scraping from e-commerce and online websites i.e. www.whatmobile.com [12] and www.olx.com.pk [11]. As, no dataset of the used smartphones is found from online data repository i.e. Kaggle, UCI, etc. In fact, web scraping is a technique to extract data from websites using a scraping robot, which automates the whole process. This method allows quick access to large-scale web data. Usually, the price of a used item is lower as compared to a brand new item. As per the market survey, the price of used items depends on various factors like its condition, make, not repaired, having an original charger, within the warranty period, no scratches, etc. In this regard, a review is conducted for the assessment of the price of used mobile phone. Hence, both technical and non-technical features of the used smartphone have been selected for the preparation of dataset for applying ML regression technique. Subsequently, it is supplied to a data mining most useful tool i.e. WEKA [13] for the evaluation, analysis, and finding best accuracy of the ML classifier algorithms. The WEKA, a data mining tool described above is used for experiments. It is a collection of ML algorithms and various other components that help from data loading to classification/regression model building. The 10-folds cross-validation (CV) method is used for experiments as it is a standard evaluation method to perform a systematic way of running repeated percentage splits on the dataset. Also, it further divides a dataset into ten pieces known as "folds" also it can be increased or decreased, then it holds out each piece in turn by turn for testing and training on the remaining (9 x folds) together. Our dataset consists of 379 examples and 24 features, the most important of which are presented in Table 1.

Table 1: Features Dataset

Variable	Description
Brand	It is a categorical variable and indicates the name of the company that manufactures the device, such as Apple, Samsung, Huawei…
Year	It is a numerical variable. It refers to the year in which the device was released.
Weight_G	It is a numerical variable, indicating the weight of the device in grams
RAM	It is a numerical variable, indicating the amount of random-access memory (RAM) that the device has usually measured in gigabytes (GB).
Reparable	It is a binary variable, indicating whether the device can be repaired or not.
Approx_Price_ Pkr	It is a numerical variable, indicating the approximate price of the device in Pakistani Rupees when the manufacturer released it.
Orignal_Charger	It is a binary variable, indicating whether the device comes with an original charger or not.
Warranty	It is a binary variable, indicating whether the device is covered by a warranty from the manufacturer or not.
Internal_Mem_Gb	It is a numerical variable, indicating the amount of internal storage memory that the device has, usually measured in GB.

NFC	It is a binary variable, indicating whether the device is equipped with Near Field Communication technology or not.
Sell_Price	It is a Class Label that is a numeric attribute-dependent Feature set of the dataset. It's just denoting prices of each used mobile device demanded by the seller.

3. EXPLORATORY ANALYSIS

The dataset was collected from different online marketplaces and contains information about the brand, model, year of release, age of the device, RAM, network technology, network brand, GPRS, EDGE, weight in grams, SIM, WLAN, GPS, NFC, radio, approximate price, warranty, reparability, scratches, original charger, and selling price of the device. In the exploratory analysis, a correlation matrix to understand the relationship between the features and the target variable is presented in Figure 1. The Pearson correlation coefficient measures the linear relationship between two variables. Only those features with a correlation coefficient above a certain threshold are considered relevant for the prediction model. The results of the correlation matrix analysis showed that features such as internal memory, RAM, year, Approx_Price_Pkr, and NFC had the strongest correlation with the sales price. These features are very informative and should be selected for the model. In addition, this reduces the size of the data set for easy interpretation and visualization.

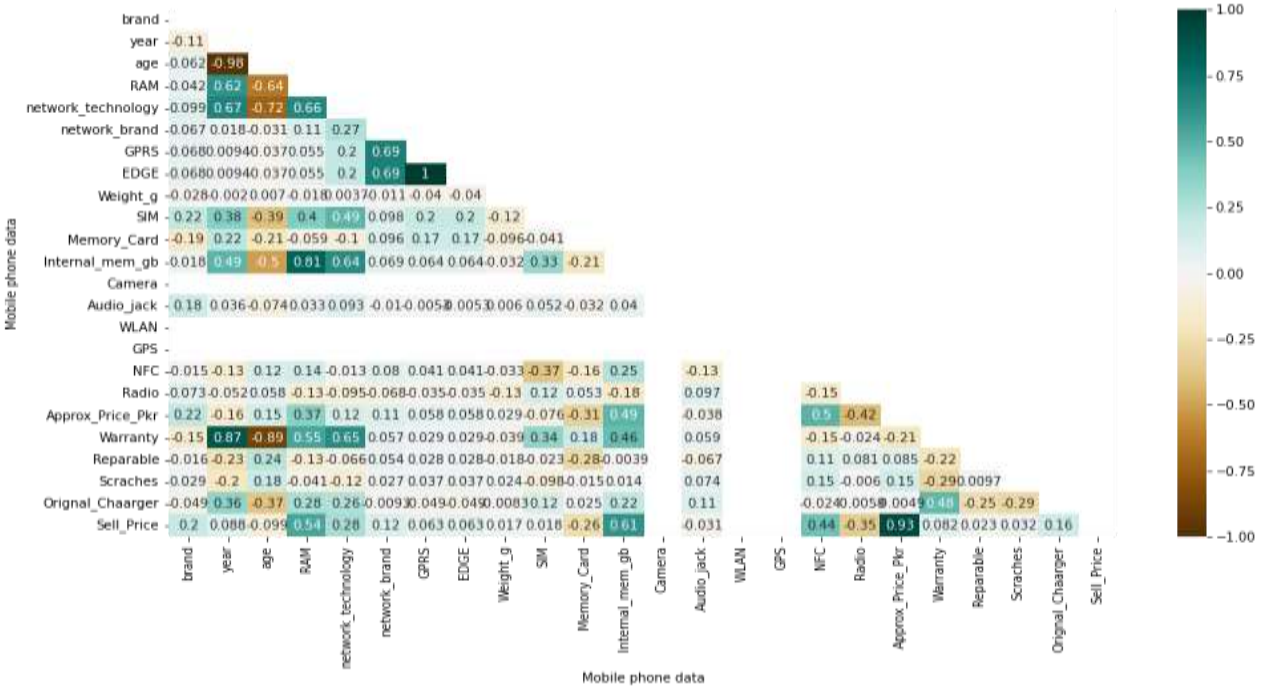

Fig.1. Triangle Correlation

4. MACHINE LEARNING ALGORITHMS

Machine learning is a class of algorithms that helps systems leverage data to efficiently perform a particular function without being explicitly programmed, based on relationships and patterns. Then, the machine learning model will use these patterns to make predictions on new data. In this research, we focus on supervised machine learning to predict, in particular,

the selling price. Using the labeled dataset, a predictive model is built to determine a numerical variable, the selling price of a used cell phone, based on the output and input data.

The main objective of our paper is to predict the price of used smartphones based on the recorded attributes. Based on the literature review of similar works, it has been identified that linear regression, multilayer perceptron and random forest algorithms have high prediction accuracy on numerical data. Therefore, these regression algorithms are selected. The performance of the model will be evaluated using metrics such as mean absolute error (MAE), root mean square error (RMSE), relative absolute error (RAE) and relative square error (RRSE). The 10-fold cross-validation technique was used. With this method, the data set is randomly divided into 10 parts. Nine of these parts are reserved for training and a tenth for testing. This procedure is repeated 10 times, reserving a different tenth for testing each time. This technique reduces the variance of the performance estimate and can provide a more accurate measure of model performance.

Multi-linear regression (MLR) assumes [5] that there is a linear relationship between the predictor variables and the target variable, and estimates the values of the coefficients of the predictor variables that minimize the difference between the predicted and actual values. Once the coefficients are estimated, the model can be used to predict the value of the target variable for new data. The Multilayer Perceptron (MLP) model [6] is a type of artificial neural network that is widely used in supervised learning problems, such as regression and classification. It consists of several layers of interconnected nodes, which process and transmit information between input and output layers. Random Forest (RF) [8] is an ensemble learning method for classification, regression, and other tasks that works by building a multitude of decision trees at training time and producing the class that is the mode of the classes or the average prediction of the individual trees.

5. RESULTS AND DISCUSSIONS

Since the parameter space of ML algorithms such as RF and MLP includes spaces with real and unbounded values, we can obtain optimal values of model hyper parameters by implementing grid search in Python. This is the most popular algorithm for optimizing hyper parameters. We also use 10-fold cross-validation testing techniques to ensure that the model does not over fit the training data.

Table 2: Results according to the standard metrics for each model before applying the Grid search method.

	Linear Regression	Random Forest	Multi-layer Perceptron
Mean absolute error	3047.021	934.443	4429.459
Root mean squared error	4353.790	3135.187	6854.860
Relative absolute error	0.215	0.066	0.313
Root relative squared error	0.247	0.178	0.389
Coefficient of determination	-2.641	0.992	-3.526

All models must be evaluated and compared on the basis of MAE, RMSE, RAE and RRSE to determine where exactly one model outperforms the other. In addition, we apply the square root (R2) or coefficient of determination. After observing the cross-validation results 10 times of each model, the experiment is done with the best cross-validation. In Table 2 above, the standard performance measures are calculated. In conclusion, Random Forest is a good choice for predicting prices of

used appliances because it has the best performance. In addition, it has a low tendency to over fit, which means that it generalizes well to new data, which is important for price prediction because the model needs to make accurate predictions about new appliances that have never been seen before.

6. CONCLUSION

In this work, we applied test score prediction with k-fold cross validation and hyper parameter optimization based on a grid search strategy. Three machine learning models were employed to predict the selling price of used phones; we also evaluate which parameter values are best and can be used to achieve relevant performance. A comparison of the machine learning models was performed to test their results. We compare our proposed methods with the statistical method. This comparison is done using standard metrics such as MAE, RMSE, RAE, and RRSE. We conclude that the RF model performs better than the other models in these experiments. This research can be extended to the estimation of prices of older electronic devices commonly sold and purchased, e.g., laptops, scanners, washing machines, microwave ovens, LEDs, etc.

ACKNOWLEDGMENT

First and foremost, I express my sincerest gratitude to the Almighty Allah for the blessings and guidance that have been bestowed upon me throughout my life. I would like to extend my deepest appreciation to my respected research supervisor, Dr. Arshad Awan, for his invaluable guidance and expertise, which have played a pivotal role in shaping my research endeavors. Additionally, I am immensely thankful to Abdul Sajid and Amani Ayeb for their support and contributions to my academic journey.

Furthermore, I am indebted to my beloved family for their unwavering support, encouragement, and understanding. Their constant motivation has been the driving force behind my achievements. Moreover, I would like to express my sincere thanks to Mr. Hani Khalifa, Syed Tehseen Shah sb, Syed Rifaqat Husain Shah sb, Raza Ahmed Shah sb, Niyyar Abbasi, Azrar Abbasi, Syed Talal, Syed Afzal, Tahir Javed, Shasha abbasi, Moon Abbasi and Gohar Khan sb for their exceptional support and assistance.

.

REFERENCES

[1] E. Güvenç, G. Çetin, and H. Koçak, “Comparison of KNN and DNN Classifiers Performance in Predicting Mobile Phone Price Ranges”, Advances in Artificial Intelligence Research, vol. 1, no. 1, pp. 19-28, Jan. 2021

[2] Asim, M., and Khan, Z. “Mobile price class prediction using machine learning techniques”. International Journal of Computer Applications, 179(29), 6-11, 2018.

[3] Surjuse, V., Lohakare, S., Barapatre, A., & Chapke, A. (2022). Laptop Price Prediction using Machine Learning. International Journal of Computer Science and Mobile Computing, 11(1), 164-168.

[4] Nasser, I. M., Al-Shawwa, M. O., and Abu-Naser, S. S. “Developing Artificial Neural Network for Predicting Mobile Phone Price Range ”, 2019.

[5] Varun Kiran, A. and Jebakumar, R. “ Prediction of Mobile Phone Price Class using Supervised Machine Learning Techniques ”,7 (1), 248-251, 2022.

[6] Çetın, M., & Koç, Y. “ Mobile Phone Price Class Prediction Using Different Classification Algorithms with Feature Selection and Parameter Optimization”. International Symposium on Multidisciplinary Studies and Innovative Technologies, 483–487, 2021.

[7] Gegic, E., Isakovic, B., Keco, D., Masetic, Z., & Kevric, J. . “Car price prediction using machine learning techniques”. TEM Journal, 8(1), 113-118, 2019.

[8] Chandrashekhara, K. T., Thungamani, M., Gireesh Babu, C. N., & Manjunath, T. N. "Smartphone price prediction in the retail industry using machine learning techniques". Emerging Research in Electronics, Computer Science and Technology, 363-373, 2019.

[9] Cen, Z., & Wang, J. Crude oil price prediction model with long short term memory deep learning based on prior knowledge data transfer. Energy, 169, 160-171, 2019.

[10] Yu, P., & Yan, X. Stock price prediction based on deep neural networks. Neural Computing and Applications, 32(6), 1609-1628. 2020.

[11] https://www.olx.com.pk/

[12] https://www.whatmobile.com.pk/

[13] https://www.cs.waikato.ac.nz/ml/weka/

[14] A. Sajid and M. A. Awan, "Monitoring, Tracking and Diagnosing Dengue Fever using Smartphone, GPS and Machine Learning," 2021 16th International Conference on Emerging Technologies (ICET), Islamabad, Pakistan, 2021, pp. 1-6, doi: 10.1109/ICET54505.2021.9689785.

Authors' Profiles

Muhammad Hasnain, an accomplished individual, holds a Master of Science in Computer Science (MSCS) degree from the esteemed Department of Computer Science at Allama Iqbal Open University in Islamabad, Pakistan. He has gained valuable knowledge and expertise in the field. With a strong focus on artificial model designing and machine learning model design, Muhammad Hasnain has showcased his remarkable skills in developing automated systems. His innovative approach and dedication have led him to create models that possess the ability to work autonomously, marking a significant achievement in the realm of Artificial intelligence. For any inquiries or communication, you can reach out to Muhammad Hasnain via email at mabbasi103@gmail.com

Abdul Sajid received Master of Science in Computer Science (MSCS) degree from Department of Computer Science, Allama Iqbal Open University Islamabad, Main Campus, Islamabad, Pakistan. His research interests include diagnose infections disease using Machine Learning and his research work published in conference paper. Contact him at asajids84@gmail.com

Amani Ayeb holds a PhD in Industrial IT and Automation. With a strong educational background and expertise in the field, Amani is highly skilled in areas such as machine learning, data science, and industrial automation. With a passion for utilizing technology to drive innovation, Amani strives to make a positive impact in the field of IT and automation. Email : amani.ayeb@insat.ucar.tn .

Muhammad Arshad Awan received a PhD in Computer Science from Yonsei University, South Korea; currently he is working as Assistant Professor in the Department of Computer Science, Allama Iqbal Open University, Islamabad, Pakistan. His research interests include human activity recognition, context aware system for healthcare, smart home/cities based on semantic enable data/information fusion & integration, Ontology based activity recognition & assistive technologies, and Interface system that support autonomy, learning & adaptability. His work has appeared in over 18 publications. He can be reached at the email Arshad.awan@aiou.edu.pk .