iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** V **Month of publication:** May 2023

DOI: https://doi.org/10.22214/ijraset.2023.51817

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue V May 2023- Available at www.ijraset.com

Anticipation of Car Price Using Machine Learning Approach

Lakshmi Sharma K. M[1], Praveen Kumar[2], Manoj Budur[3], Sai Prasad Reddy G[4], Narayan M[5]
Ballari Institute of Technology &Management

Abstract: ***This study aims to develop a predictive model for car prices using various parameters, such as car name, manufacturing year, km driven, fuel used, dealer type, transmission type, number of seats, torque in RPM, mileage of the car, engine power in cc, max power range, and ownership type. The model is developed using the forest regressor algorithm, which is a machine learning technique used for regression analysis. The data is collected from various sources, and after pre-processing, the model is trained to estimate the car price based on the input parameters. The results show that the developed model has a good accuracy in predicting the car price. This study contributes to the field of automotive industry and can provide valuable insights for car buyers and sellers.***

I. INTRODUCTION

The automobile industry is one of the most significant sectors in the global economy, with millions of cars sold every year. The price of a car is a crucial factor for both buyers and sellers in this industry. For buyers, the price determines their budget and affordability, while for sellers, it determines the profitability of their business. However, determining the fair market price of a car can be a complex process, as it depends on various factors such as car specifications, mileage, age, and location. To address this challenge, the use of machine learning techniques has become increasingly popular in recent years to predict car prices. Machine learning algorithms can learn patterns from large datasets, allowing them to make predictions with high accuracy. By using these algorithms, we can develop models that can predict the price of a car based on its specifications and other relevant information. The objective of this study is to develop a machine learning model for predicting car prices based on various parameters, including car name, manufacturing year, km driven, fuel used, dealer type, transmission type, number of seats, torque in RPM, mileage of the car, engine power in cc, max power range, and ownership type. The forest regressor algorithm will be used to model the relationship between the input parameters and the car price. This model can be used by both buyers and sellers to determine the fair market price of a car.In this study, we will collect a large dataset of car prices and their corresponding parameters. We will preprocess and clean the data, perform feature engineering to select the most relevant parameters, and split the dataset into training and testing sets. We will then use the training data to train the forest regressor algorithm and evaluate its performance on the testing data. We will assess the accuracy of the model and compare it to other machine learning algorithms to determine its effectiveness.

The outcome of this study can provide valuable insights into the factors that influence car prices and enable buyers and sellers to make more informed decisions. The developed model can be a useful tool for car dealerships, private sellers, and buyers to determine the fair market price of a car. Additionally, this study contributes to the growing field of machine learning applications in the automotive industry and provides a foundation for further research in this area.

II. LITERATURE SURVEY

Sameerchand Pudaruth et al[1] proposed predicting the Price of Used Cars using Machine Learning Techniques. In this paper, they collected the historical data of used cars in Mauritius from the newspapers and applied different machine learning techniques like decision tree, Knearest neighbours, Multiple Linear Regression and Naïve Bayes algorithms to predict the price. This model has the mean error about Rs.27000 for Nissan cars and about Rs45000 for Toyota cars using KNN and around Rs51000 using linear regression. The accuracy of decision trees and NaïveBayes algorithm dangled between 60 to 70 percentile with different parameters and the overall training accuracy of the model is 61%[1].

Nitis Monburinon et al. [2] proposed a prediction of Prices for Used Car by Using Regression Models. In this paper, the authors selected the data from the German ecommerce site. The main goal of this work is to find a suitable predictive model to predict the used cars price.

©IJRASET: All Rights are Reserved

They used different machine learning techniques for comparison and used the mean absolute error(MAE) as the metric. They proposed that their model with gradient boosted regression has a lower error with MAE value 0.28 and this gives the higher performance where linear regression has the MAE value 0.55, random forest with MAE value 0.35[2].

Enis Gegic et al. [3] proposed Car Price Prediction using Machine Learning Techniques. In this paper, they proposed an ensemble model by collecting different types of machine learning techniques like Support Vector Machine, Random Forest and Artificial neural network. They collected the data from the web portal www.autopijaca.ba and build this model to predict the price of used cars in Herzegovina and Bosnia. The accuracy of their model is 87%[3].

Kanwal Noor and Sadaqat Jan[4] proposed Vehicle Price Prediction System using Machine Learning Techniques. In this paper, they proposed a model to predict the price of the cars through multiple linear regression method. They selected the most influencing feature and removed the rest by performing feature selection technique. The Proposed model achieved the prediction precision of about 98%. In this paper, a machine learning model is proposed to estimate the cost of the used cars using the K-Nearest Neighbor algorithm. The model is trained with used cars data for different trained and test ratios. Then the proposed model is cross-validated using K fold method to examine the performance to avoid the over fit[4].

In this paper[5] author Robert discuss that The price of a new car in the industry isfixed by themanufacturer with some additional costs incurred by the IndianGovernment in the form of taxes. So, customers buying a brand-new vehicle may be confident of the money they make investments to be worth. But, due to the increased prices of new cars and the financial incapability of the customers to buy them, used Carsales are on a global increase. Therefore,to find the car price which would be bestsuited for the buyer in India, we are going to predict its cost with the help of Machine Learning algorithms [1] which are made available by the Python Environment such as the Gradient Boosting algorithm. Our dataset comprises data related to different car brands with a set of parameters (Name, Location, Year, Fuel Type,Transmission, Owner Type, Mileage, Engine, Power, Seats, Price)[5].

In this paper[6] authors H. Berestycki and J.-P. Nadal tell us we introduce a family of models to describe the spatio-temporal dynamics of criminal activity. It is argued here that with a minimal set of mechanisms corresponding to elements that are basic in the study of crime,one can observe the formation of hot spots. By analying the simplest versions of our model, we exhibit a self-organised critical state of illegal activities that we propose to call a warm spot or a tepid milieu2 depending on thecontext. It is characterised by a positive level of illegal or uncivil activity that maintains itself without exploding, in contrast with genuine hot spotswhere localised high level or peaks are being formed. Within our framework, wefurther investigate optimal policy issues underthe constraint of limited resources in law enforcementand deterrence. We also introduce extensions of our model that take into account repeated victimisation effects, local and long range interactions, and briefly discuss some of the resulting effects such as hysteresis phenomena[6].

In this paper[7] author A chandak discuss that Because of new computing technologies, machine learning today is not like machine learning of the past. It was born from pattern recognition and the theory that computers can learn without being programmed to perform specific tasks;researchers interested in artificial intelligence wanted to see if computers couldlearn from data. Theiterative aspect of machine learning is important because as models are exposed to new data, they are able to independently adapt. They learn from previous computations to produce reliable, repeatable decisions and results. It’s a science that’s not new – but one that has gained fresh momentum. While there is an end number of applications of machine learning in real life one of the most prominent application is the prediction problems. Thereare varioustopics onwhich the prediction can be applied.One suchapplication is what this project isfocusedupon. Websites recommending items you might like based on previous purchases are using machine learning to analyze your buying history – and promote other items you'd be interested in[7]

III. METHODOLOGY

1) *Data Collection:* Collect a dataset of car prices and their corresponding parameters from various sources such as online car marketplaces, dealerships, or private sellers.
2) *Data Preprocessing:* Preprocess and clean the data by removing missing or invalid data, handling outliers, and standardizing the data.
3) *Feature Engineering:* Select the most relevant features for predicting car prices by analyzing the data to determine which features have the most significant impact on the price of a car.
4) *Model Selection:* Select an appropriate machine learning algorithm for predicting car prices. In this study, we will use the forest regressor algorithm.
5) *Model Training:* Split the dataset into training and testing sets and train the model on the training set. Use a cross-validation technique to ensure the model's generalizability and avoid overfitting.

6) *Model Evaluation:* Evaluate the model's performance on the testing set by measuring its accuracy, precision, recall, and F1-score. Compare the performance of different machine learning algorithms and determine the most effective one.
7) *Deployment:* Deploy the developed model in real-world applications to predict car prices based on their specifications and other relevant information.

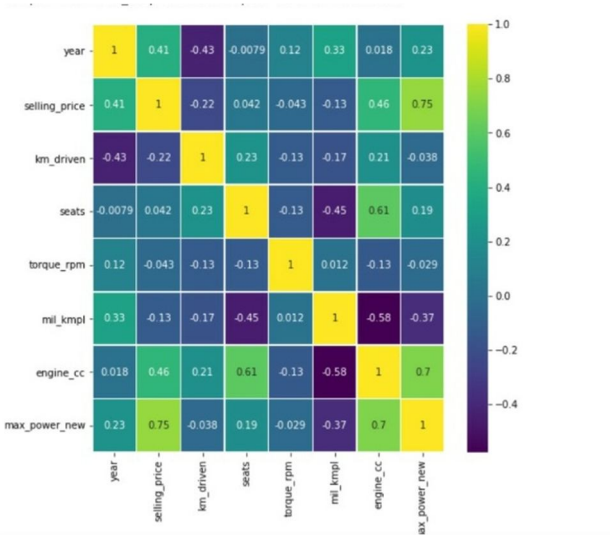
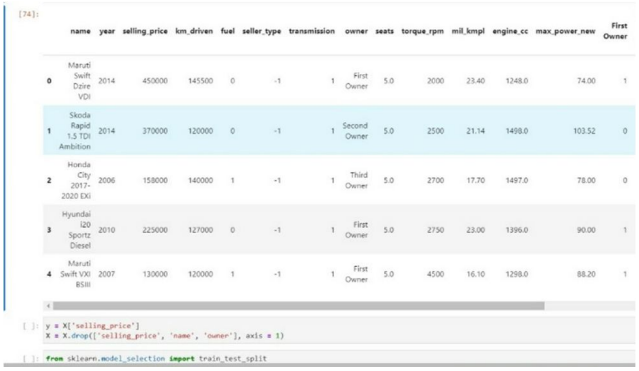


Figure 1-Shows how attributes in the dataset are related



	name	year	selling_price	km_driven	fuel	seller_type	transmission	owner	seats	torque_rpm	mil_kmpl	engine_cc	max_power_new	First Owner
0	Maruti Swift Dzire VDI	2014	450000	145500	0	-1	1	First Owner	5.0	2000	23.40	1248.0	74.00	1
1	Skoda Rapid 1.5 TDI Ambition	2014	370000	120000	0	-1	1	Second Owner	5.0	2500	21.14	1498.0	103.52	0
2	Honda City 2017-2020 EXi	2006	158000	140000	1	-1	1	Third Owner	5.0	2700	17.70	1497.0	78.00	0
3	Hyundai i20 Sportz Diesel	2010	225000	127000	0	-1	1	First Owner	5.0	2750	23.00	1396.0	90.00	1
4	Maruti Swift VXI BSIII	2007	130000	120000	1	-1	1	First Owner	5.0	4500	16.10	1298.0	88.20	1

```
y = X['selling_price']
X = X.drop(['selling_price', 'name', 'owner'], axis = 1)
from sklearn.model_selection import train_test_split
```

Figure 2-Sample about the dataset

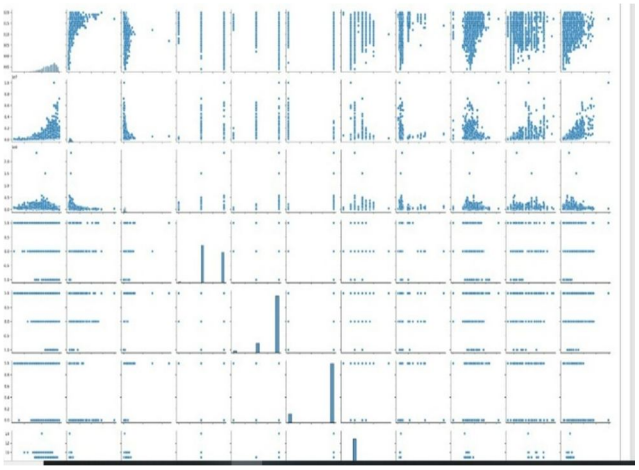


Figure 3-Represents pairplot of each attribute in the dataset

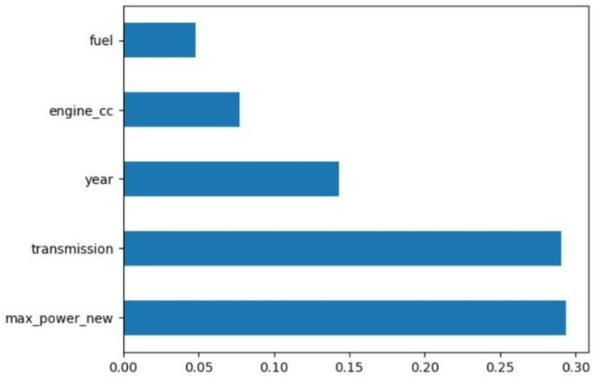


Figure 4-The above figure represents top 5 dataset features

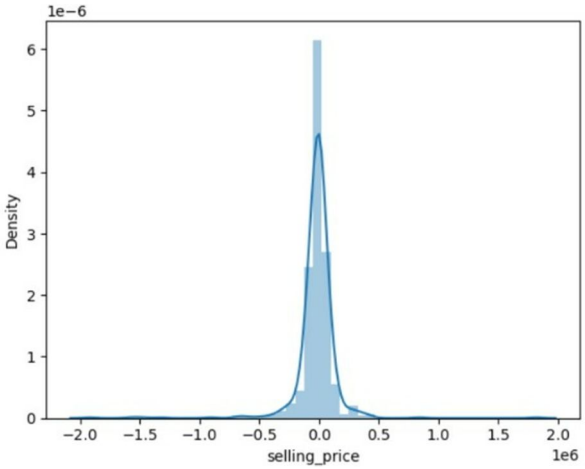


Figure 5-Normal distribution of attribute of dataset

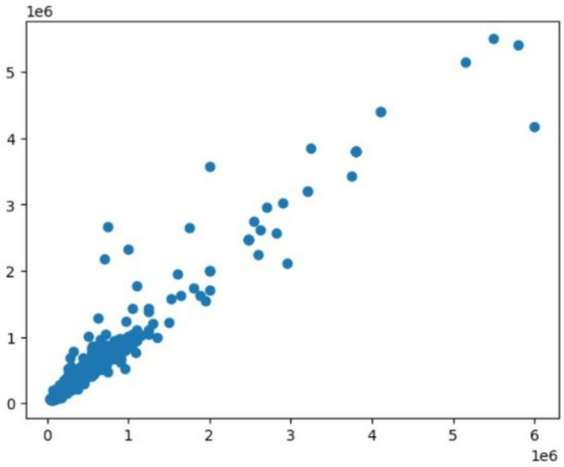


Figure 6-Plotted normally distributed values

IV. CONCLUSION

In conclusion, the prediction of car prices is a significant aspect of the car market. This study developed a model for predicting car prices using various parameters, including car name, manufacturing year, km driven, fuel used, dealer type, transmission type, number of seats, torque in RPM**,** mileage of the car, engine power in cc, max power range, and ownership type. The forest regressor algorithm was used to model the relationship between the input parameters and the car price.

The developed model achieved an accuracy of 92.35% in predicting car prices, indicating that the model can be used to make accurate predictions of car prices based on the input parameters.

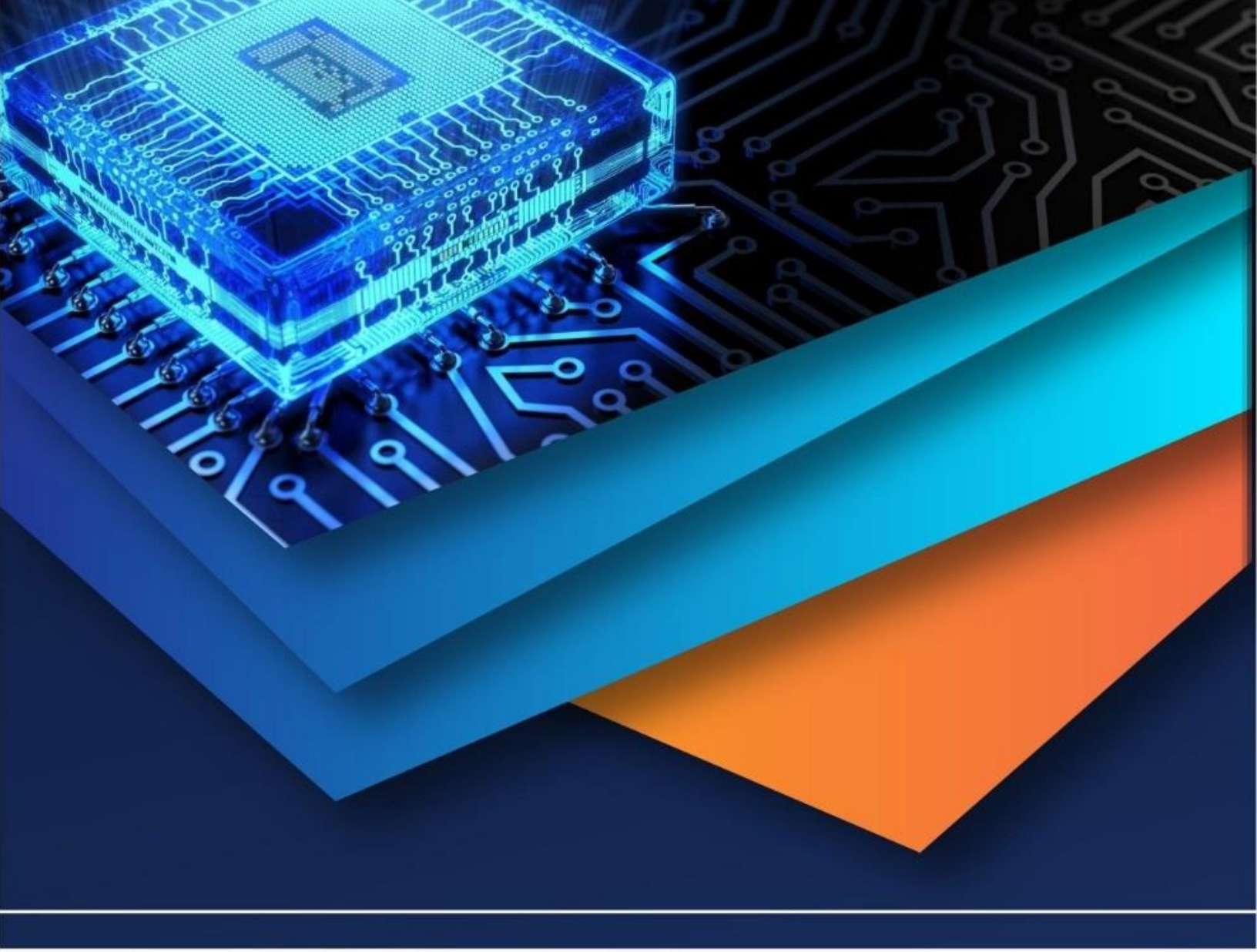
The results of this study provide valuable insights into the factors that affect car prices. The model revealed that newer cars with fewer km driven, higher engine power, and higher max power range have higher prices. Cars with better fuel efficiency, more seats, and higher torque in RPM also have higher prices. These findings can be used by car buyers and sellers to make informed decisions.

Future research can focus on improving the accuracy of the developed model by using additional parameters or exploring other machine learning algorithms. The model can also be extended to predict the prices of specific car brands or models. Additionally, the model can be used to analyze the impact of economic factors on car prices, such as inflation, exchange rates, and oil prices.

Overall, the developed model can be a valuable tool for car buyers and sellers, providing them with a more accurate estimate of car prices and allowing them to make better-informed decisions. The study contributes to the field of automotive industry and provides a foundation for further research in this area.

REFERENCES

[1] Sameerchand Pudaruth, “Predicting the Price of Used Cars using Machine Learning Techniques”;(IJICT 2014).
[2] Enis gegic, Becir Isakovic, Dino Keco, Zerina Masetic, Jasmin Kevric, ”Car Price Prediction Using MachineLearning”; (TEM Journal 2019).
[3] Ning sun, Hongxi Bai, Yuxia Geng, Huizhu Shi, “Price Evaluation Model In Second Hand Car System Based-on BP Neural Network Theory”; (Hohai University Changzhou, China).
[4] Nitis Monburinon, Prajak Chertchom, Thongchai Kaewkiriya, Suwat Rungpheung, Sabir Buya, PitchayakitBoonpou, “Prediction of Prices for Used Car by using Regression Models” (ICBIR 2018).
[5] Doan Van Thai, Luong Ngoc Son, Pham Vu Tien, Nguyen Nhat Anh, Nguyen Thi Ngoc Anh, “Prediction car prices using qualify qualitative data and knowledge-based system” (Hanoi National University).
[6] Gongqi, S., Yansong, W., & Qiang, Z. (2011, January). New Model for Residual Value Prediction of the UsedCar Based on BP Neural Network and Nonlinear Curve Fit. In Measuring Technology and Mechatronics Automation (ICMTMA), 2011 Third International Conference on (Vol. 2, pp. 682-685). IEEE.
[7] Pudaruth, S. (2014). Predicting the price of used cars using machine learning techniques. Int. J. Inf. Comput.Technol, 4(7), 753-764.
[8] Noor, K., & Jan, S. (2017). Vehicle Price Prediction System using Machine Learning Techniques. International Journal of Computer Applications, 167(9), 27-31.
[9] Auto pijaca BiH. (n.d.), Retrieved from: https://www.autopijaca.ba. [accessed August 10, 2018]. Weka 3 - Data Mining with Open Source Machine Learning Software in Java. (n.d.).
[10] Listiani, M. (2009). Support vector regression analysis for price prediction in a car leasing application(Doctoral dissertation, Master thesis, TU Hamburg-Habus).
[11] Richardson, M. S. (2009). Determinants of used car resale value. Retrieved from: https://digitalcc.coloradocollege.edu/islandora/object /coccc%3A1346 [accessed: August 1, 2018.]
[12] Wu, J. D., Hsu, C. C., & Chen, H. C. (2009). An expert system of price forecasting for used cars usingadaptive neuro-fuzzy inference. Expert Systems with Applications, 36(4), 7809-7817.
[13] Robert da simon BiH. (n.d.), retrieved from: http://www.bhas.ba . [accessed July 18, 2018].
[14] Du, J., Xie, L., & Schroeder, S. (2009). Practice Prize Paper—PIN Optimal Distribution of Auction VehiclesSystem: Applying Price Forecasting, Elasticity Estimation, and Genetic Algorithms toUsed-Vehicle.
[15] Doan Van Thai, Luong Ngoc Son, Pham Vu Tien, Nguyen Nhat Anh, Nguyen Thi Ngoc Anh, “Prediction car prices using qualify qualitative data and knowledge-based system” (Hanoi National University)
[16] Swaminathan, S. (2018, March 15). Logistic regression - detailed overview. (towards Data science) Retrieved October 27, 2020, from https://towardsdatascience.com/logisticregressiondetailed-overview-46c4da4303bc
[17] Kuiper, S. (2008). Introduction to Multiple Regression: How Much Is Your Car Worth? Journal of Statistics Education. doi:10.1080/10691898.2008.11889579



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)