iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** VI **Month of publication:** June 2022

DOI: https://doi.org/10.22214/ijraset.2022.44225

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 10 Issue VI June 2022- Available at www.ijraset.com

Prediction of Price forCars Using Machine Learning

Rithvik Raj Mekala[1], Gouru Laxmi Sevitha[2], Tushar Anumula[3], Kusuma Latha[4]
[1, 2, 3, 4]*Department of Electronics and Computer Engineering Sreenidhi Institute of Science and Technology Yamnampet, Ghatkesar, Hyderabad, India*

Abstract: Every day, many Indian customers sell their cars to other buyers, referred to as 2nd/3rd owners, etc., after they have used them. Many websites, such as cars24.com, cardekho.com, and OLX.com, provide these buyers with a place to sell their old cars, but what should thecar's price be? Machine learning algorithms may be able to overcome thisproblem. Using a history of prior used car sale data and machine learningtechniques such as Supervised Learning, I forecasted the selling price ofthe used car using machine learning algorithms such as Random Forestand Extra Tree Regression and the powerful python package Scikit- Learn. Regardless of how large or little the dataset is, the results show that both methods are highly accurate in prediction.

I. INTRODUCTION

A new car's price is established by the manufacturer, with the government incurring some additional costs in the form of taxes. Customers who buy a new car can rest assured that their money is being carefully spent. However, as new car prices rise and purchasers become unable to afford them, used car sales are on the rise all over the world. As a result, a method for estimating the valueof a used car based on a variety of characteristics is urgently needed.The current approach incorporates a technique in which a vendor sets a price at random and the buyer has no idea how much the car is worth. The vendor has no idea how much the car is worth or how much he can get for it. We've designed an extremely effective model to address this problem. Because regression methods produce a continuous value rather than a categorized value as a put, they are used. As a result, instead of estimating a car's price range, it will bepossible to estimate the car's actual price. In addition, a user interface has been developed that takes user input and displays the pricing of an automobile based on that input.

A. Machine Learning

The investigation of training PCs to work all alone without being helped what to do is known as AI. Computer-based intelligence is a subset of ML, which has turned into the most well-known expression of the twenty-first hundred years. Man-made intelligence alludes to the activity of making PCs falsely astute with the goal that they can finish responsibilities all alone. These gadgets are incredibly exact and speedy at what they will do. We create andprepare AI Techniques by utilizing different ML approaches like Supervised Learning Unsupervised Learning, and Reinforcement Learning.

Design acknowledgment might be followed back to the starting points of Machine Learning. Furthermore, Machine learning utilizes different relapses and grouping ways to deal with train models so they can learn all alone.

The capacity to apply troublesome numerical computations to a lot of information naturally - again and again, speedier - is a moderately new peculiarity. "While many AI methods have been accessible from now onward, indefinitely quite a while, the ability to apply them to a lot of information — over and over and quicker and quicker - is a new turn of events." 1 thus, the iterative part of AI turns out to be progressively significant as models are presented with new information, which they adjust to using their capacity to gainfrom past estimations, which is likewise liable for conveyingtrustworthy, repeatable decisions and results. Coming up next are the three kinds of AI strategies:

1) Supervised learnings models are to be provided with theinput details as well as the ideal answer the copy to do then endeavor to learn rules to connect that to info information to the ideal result
2) Unsupervised Learning: Models are given datasets that have no marks or biased examples, and they should derive the hidden designs.
3) Reinforcement learning: To succeed in a specific intention the model or specialist will be the cooperate with the unique reality the specialist that upheld the demonstrations of the unique world will be compensatedor rebuffed the specialist will b ultimately figure out how to do and explore the powerful world and succeed its objective in light of the prizes and disciplines it has gotten.

©IJRASET: All Rights are Reserved

Most the individuals are uncertain about the distinctions between ai and ml and dl yet truly they all work together tofoster a mindful model that can learn all alone and complete undertakings without the requirement for human communication the three spaces are looked at in the table underneath At the point when a machine finishes jobs with the assistance of a bunch of laid out decides that address difficulties (calculations), we name this "shrewd" conduct "Computerized reasoning." Machine learning is a subset ofAI, in any case, AI isn't equivalent to AI. It empowers machines to learn all alone and make the right forecasts inlight of the data given. AI is a subset of profound learning. DL calculations are generally enlivened by the data handling designs tracked down in the human cerebrum, known as Neural Networks. An assortment of these can be tracked down in DL's Neural Networks. Profound learning strategies are regularly used to prepare machines to executecomparative errands.

Similarly, as we use our cerebrums to perceive designs and arrangeThe last result. The result is quite often the model's conclusive different kinds of information. Instances of brain networksincorporate RNNs (Recurrent Neural Networks), CNNs(Convolutional Neural Networks), and others." 2 The course of acalculation gained from a preparation dataset with the help of aboss is known as directed learning. It interprets the contribution tothe result and gives occasions for information yield matches. Themodel predicts the result utilizing haphazardly picked testinformation values from the first dataset because it as of nowcontains test information. A regularly managed learningcalculation is as per the following:

$Y=f(x)$

Where Y represents the normal result and f(x) represents the anticipated capacity. A planning capacity that gives the worth x inthe contribution of a class. This decides the normal result, Y. Thiscapacity is made by the AI model during preparation, and itattaches input credits to an anticipated result Y. 3 Supervised learning is quite possibly the most notable system, and tackling two sorts of true problems can be utilized:

This kind of undertaking arranges each of the variables that make up the result. Segment information incorporates data like conjugal status, orientation, and age. The most frequently involved model for dealing with this sort of issue is the Support Vector Machine (SVM). There is an assortment of calculations for tending to classification issues, contingent upon the rules. Direct Classifiers, Support Vector Machines, Decision Trees, K-Nearest Neighbor, and Random Forest are only a couple of the strategies that can be utilized.

B. Relapse Problems

The result factors in relapse issues are constantly designed to be agenuine number. Utilizing a direct format is normal. Coming up next are some normal rectilinear relapse conditions:

$\hat{y}= w[0] * x[0] +w[1]*x[1]+\ldots +w[i]*x[i]+b$

The plane will be utilized in two unique models. At long last, for models with multiple elements, a hyperplane will be utilized. Coming up next are instances of relapse calculations: Regression procedures incorporate Linear Regression, Logistic Regression, Polynomial Regression, ExtraTree Regression, and Random Forest.

Troupe and Bagging: The expression "gathering" alludes to the demonstration of interfacing various discrete pieces into a solitary working model or item. Various models (regularly alluded to as "powerless students") are instructed to deal with a similar issue and afterward incorporated to further develop results; the basic thoughtis that when frail models are accurately coupled, we will get more exact or potentially strong models.

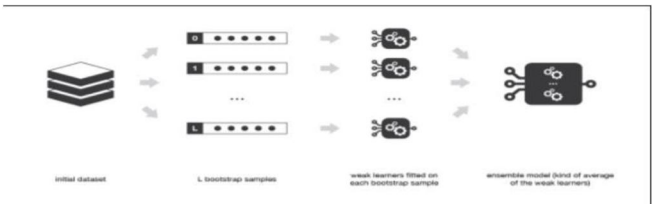


Figure 1: Bagging Technique

"Bootstrap Aggregation, generally known as sacking, is a strategy for lessening the fluctuation of high-difference calculations choice trees, order, and relapse trees" (CART). The main boundary that can be modified while sacking choice trees is the number of tests, which might be found by running the technique with various example sizes until the result exactness doesn't increment. This approach enjoys the benefit of never overfitting the preparation information, regardless of whether a critical part of the dataset consumesmost of the day to execute."

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue VI June 2022- Available at www.ijraset.com

C. Irregular Forest Algorithm

The Random Forest procedure utilizes an Ensemble-Bagging technique to fabricate various choice trees during the preparation stage. Given most of the results, the irregular backwoods picks the best other option (trees). Arbitrary Forest is equipped for consolidating bothrelapse and arrangement-regulated learning issues.

Irregular Forest is the most ideal decision for far-offsensings, for example, ETM gadgets are used to gather pictures of the world's surface since it gives higher exactness and less preparation time.

Since it performs better in extreme settings, RandomForest is utilized for Multiclass Object Detection. Agame control center utilizes this procedure to followand imitate substantial development.

The Random Forest calculation is educated to perceive body parts and afterward gains from that experience. The client's hands, feet, face, eyes, nose, and other body parts are then recognized.

"An irregular woodland is comprised of an enormous number of individual choice trees that cooperate as a troupe. Each tree in the irregular timberland delivers a classexpectation, and the classification with the most votes turnsinto the forecast of our model."

The Random Forest Algorithm can likewise sort out whateach element on the conjecture is worth. The irregular woods approach is likewise easy to understand. In RandomForest, each tree is haphazardly chosen from a subset of highlights. There is less connection among trees and greater variety because of the great degree of change.

D. Hyper parameters in Random Forest

The hyperparameters add to upgrading the expectation model's exactness and speed. The boundaries of the Sk learn library are as per the following:

1) N assessors: "To pick the number of trees the calculation builds before working out the most extreme democratic or expectation midpoints." A higher tree count further develops speed and makes conjectures more steady as a rule, yet it includes some significant downfalls.
2) max highlights: When parting a hub, the irregular backwoods evaluates the number of elements.
3) This variable decides the base number of leaves expected to isolate an inner hub.
4) N occupations: This segment makes sense of the number of processors that can be utilized to run the model.
5) "Subsequently, the model's result is repeatable." "The model will constantly yield similar outcomes when the irregular state is set to a predefined esteem and the model is given similar hyperparameters and preparing information."
6) The OOB score is utilized as a cross-approval strategy in the Random Forest Algorithm.
7) Coming up next are a portion of the advantages of the Random Forest Algorithm:
8) Overfitting is stayed away from, and it is diminished to prepare time.
9) It has an elevated degree of exactness and can effectively deal with enormous data sets
10) Gauge missing qualities with high precision in any event, when there is a great deal of missing information

II. LITERATURE SURVEY

The research has already recorded numerous attempts toconstruct this technology.

Car Price Prediction using Machine Learning Techniques was developed by Enis Gegic, Becker Isakovic, Dino Keco, Zerina Masetic, and Jasmin Kevric [1]. Materials and Procedures Datais obtained from a local web portal for selling and buying cars, autopijaca. ba [9], during the winter season, the time interval has a significant impact on the pricing of cars in Bosnia and Herzegovina. Brand, model, automobile condition, fuel, year of manufacture, power in kilowatts, transmission type, mileage, color, city, state, and several doors were all recorded for each car.

Dhawal Kotak, Praful Rane, Deep Pandya [2] Prediction of the price of a used car was developed. In the system, there are two basic phases:

1) *Training Phase:* Using the data in the data set, the system is trained to fit a model (line/curve) depending on the algorithm selected.
2) *Testing Phase:* The system is given inputs and its functionality is checked. The precision is tested. As a result, the data that is used to train or test the model mustbe appropriate. Because the system is intended to detect and estimate the price of a used car, distinct algorithms must be utilized for the two jobs. Different algorithms were compared for their accuracy before being picked forfurther use. The well-suited one for the task was chosen

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

Car Price Prediction Using Machine Learning was developed by Ketan Agrahari, Ayush Chaubey, Mamoor Khan, and Manas Srivastava [3]. The major purpose of this strategy is to provide users with a precise estimate of how much they will have to pay for a specific automobile. The model may provide the consumer with a list of options for numerous autos based on the specifics of the vehicle they desire.

The system aids in presenting the customer with enough information to reach a decision. The used car market is growing at an exponential rate, and car dealers may benefit by advertising inaccurate pricing to take advantage of the demand.

"Car's Selling Price Prediction using Random Forest Machine Learning Algorithm," offered Abhishek Pandey, VanshikaRastogi, and Sanika Singh [4].

Bagging and Ensemble: In its most basic form, Ensemble refers to the word "Assemble," which means "to unite numerous diverse elements into one working model/object."

The ensemble technique works similarly, in which numerous models (commonly referred to as "weak learners") are trained to solve a similar problem and then aggregated to achieve better results.

Predicting the Price of Used Cars was created by Sameerchan d Pudaruth [5]. K-Nearest Neighbours (KNN): K-nearest neighbor (IBk in Weka) is a machine learning technique that compares new (unknown) data to all previous records to find the best match. Despite its apparent simplicity, pre-processing the data necessitates a significant amount of effort; otherwise, we risk getting off track. Only three factors were taken into account: the make, year, and cylinder volume.

Sadaqat Kanwal Noor Jan [6] used Machine Learning Techniques to create a Vehicle Price Prediction System. The goal of this study is to create a good regression model that can accurately predict car prices. To accomplish so, we'll need some past used automobile data, for which we'll use pricing and other basic attributes. The cost of the car is the dependent variable, while the other characteristics are the independent variables.

Predicting the Price of Used Cars Using Machine Learning Techniques was created by Sameerchan d Pudaruth [7]. The decision tree was constructed using just Nissan and Toyota vehicles. Because most popular decision tree algorithms cannot handle numeric outputs, the prices were divided into six nominal groups [13]. There are several gaps in the ranges that have been specified due to the lack of automobiles inside these ranges, yet fresh data that fits within these zones is probably feasible. These big gaps helped define the classborders.

III. METHODS USED

A. *Preprocessing of Data*

"Information Preprocessing is the main stage and will be the essential advance before preparing for each model utilizing a technique."There are a few designated spots (steps) given.

1) Import Libraries: For information handling and examination, I utilized Pandas, Numpy, Matplotlib, and Seaborn, as well as Matplotlib and Seaborn for better visuals and graphical measurements.
2) Import the Dataset: I got this dataset from Kaggle and afterward utilized Panda's library to download it.
3) Managing Missing Data in the Dataset: There were no missing qualities in this dataset when I explored it.
4) Parceling the Dataset into a Training and a Testing Set: For preparing our AI model, I utilized serious areas of strength for Python learning module, sci-unit learn, or sklearn, to part this dataset into Test and Train datasets. To create testing information, it utilizes its model choice method, which includes picking irregular qualities from a given dataset for model expectation or directed learning. Highlight Scaling: I use no component scaling approaches since the information is all in a standard organization.

B. *Information Training and Modeling*

To prepare and foster a model, we should initially indicate the reliant and free factors. I originally used to find the connection between's the result factors, and afterward, I separated my factors into two tomahawks, which we call x and y, with the x-hub holding every one of the autonomous factors and the they-hub containing the reliant variable, which in our model is the Used Cars selling cost.

The ideal hyperparameters for our model expectation are found utilizing the sklearn. model determination bundle and its train test split capacity, and this dataset is additionally appropriated in the train-test dataset utilizing Randomized SearchCV tuning.

IV. PROPOSED SYSTEM

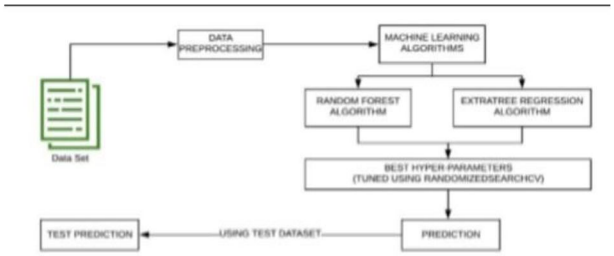


Figure 2: Flow chart proposed model

The proposed model consolidates two AI calculations: the Random Forest Method and the Extra Tree Regression strategy. This model loads the dataset at first so it very well might be explored further. This model was made with the assistance of a Kaggle Dataset. In the wake of performing information pre- handling steps on this dataset, for example, overseeing missing qualities and hot encoding of unmitigated factors, we start preparing the model for the circulated dataset into two 1. Preparing Dataset and 2. Test Dataset. This test information was picked aimlessly from the first dataset. For result expectation, we utilized the Random Forest Algorithm and the Extra Tree Regression Algorithm, as well as Randomized Search CV and Hyper-Parameters. Hyperparameters. When the model predicts an outcome, I'll test it utilizing the test dataset made with the scikit- Learn module and measure its exactness.

A. Model Prediction and Cross-Approval.

"A measurable logical technique for testing how a disclosure sums up to an alternate informational index is cross- approval." The most widely recognized reason for cross- approval is to ensure that the conjecture is right and that the model is functioning admirably." 11 After cross-approval and assessment of any remaining model exhibition and representation rules, the accompanying outcome is gotten.

An intensity map in Figure 3 shows how the characteristics are all associated. In the sidebar, the dim blue tone addresses a positive association between the qualities of the x and y tomahawks, while the white variety shows emphatically adversely connected factors. "Selling Price" and "Present Price" are decidedly related in this guide, and they can be a significant calculation assessing current selling costs after vehicles have been utilized. This map also reveals a negative link between "Present Price,""Fuel Type," and "Seller Type."

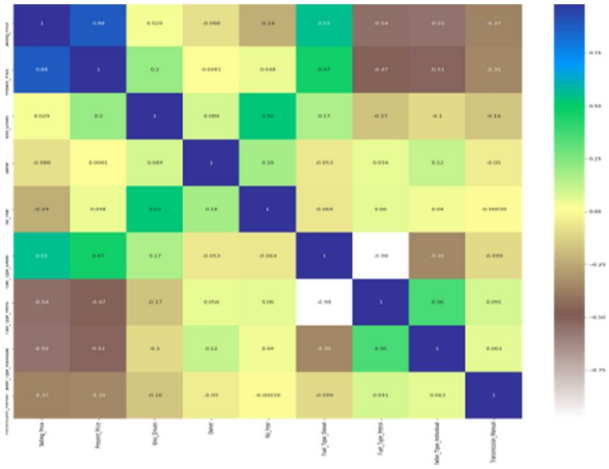


Figure 3: Heat map Showing Top Correlation Features

The distplot in figure 4 shows the model's ordinary appropriation with the test dataset, it is legitimate to demonstrate that the model. As an outcome, we might induce that the guess made by this model is very precise.

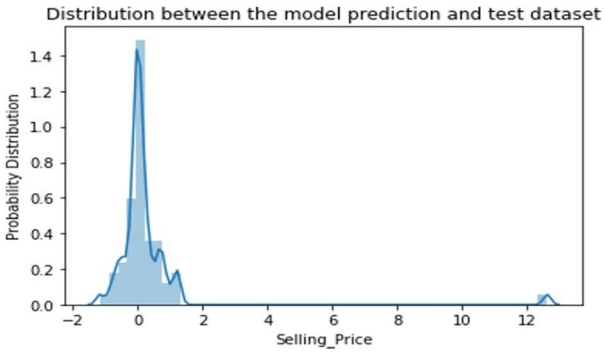


Figure 4: Displot showing the Distribution

The scatterplot in Figure 5 shows a straight circulation, showing that the model is correct and permitting us to derive that the selling cost estimatein light of the accessible dataset is right.

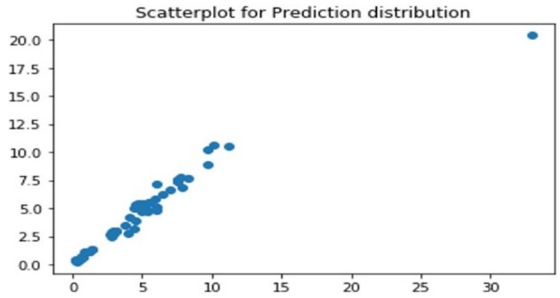


Figure 5: Scatterplot showing the distribution.

At last, we can utilize the Heroku stage to send this modelas is used to enter the information important to prepare themodel Kaggle. Kaggle is a Google organization that has an internet-based local area of information researchers and AI specialists. Clients might utilize Kaggle to look and distribute informational indexes, review and build models in an online information science climate, team up with different information researchers and AI specialists, and contend with information science challenges. Kaggle started in 2010 with AI challenges and has now extended to incorporate a public information stage, a – cloud-based information science workbench, and Artificial Intelligencecourses. Kaggle is an open-source stage that gives informational indexes to AI models to be prepared.

V. DESIGN FLOW

A. Design Flow of Car Price Prediction

Figure 6: Flowchart of car price prediction

This is the flow chart for applying machine learning to anticipate vehicle prices.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue VI June 2022- Available at www.ijraset.com

VI. RESULTS

After experimenting we have come up with some results. The numeric values for each attribute in order will be takenas input to the Random Forest Regressor trained model anddisplays the output with the accuracy of 98 percent.
This system considers features and attributes in used cars like –

1) Car brand
2) Fuel Type Model
3) Vehicle Age
4) Kilometers driven
5) Mileage
6) Engine Power
7) Max Power
8) No. of Seats

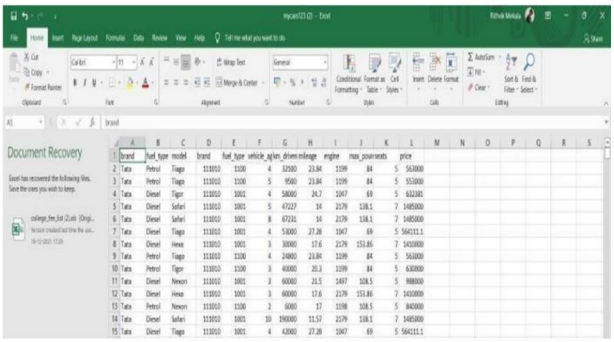


Figure 7: Features and attributes of the cars

The above figure shows our CSV file with multiple datasets and the above-mentioned attributes cum features including their prices. So using these datasets we can predict any used car values by giving all the attributes as input using our trained model system.

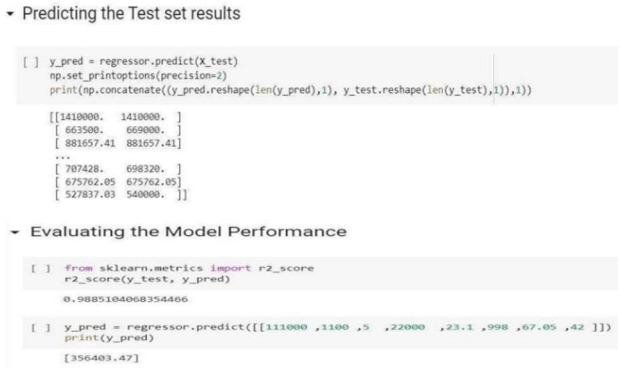


Figure 8: Final accuracy and example output

The above figure shows the accuracy of the trained model post testing the data by it. As we used a Supervised machine learning model the training accuracy is high andcomplete, so we need not calculate training accuracy but the testing accuracy as we got is 98 percent. We also can see the predicted test set examples above. We have given the attribute inputs for the desirable car price output.

VII. CONCLUSION

Utilizing AI calculations and the Kaggle dataset, we endeavored to expect the selling cost of pre-owned vehicles. This dataset was anticipated utilizing Random Forest and Extra Tress Regressor. The expectation of the model is contrasted with a test dataset made by haphazardly picking values from the first dataset, and the forecast is surveyed utilizing an assortment of techniques. We might presume that the expectation model is exceptionally precise after a far-reaching study and that Random Forest and Extra Tree Regression are among thebest calculations for relapse issues. These two methodologies are amazingly precise and quick, no matter what the size of the dataset.

©IJRASET: All Rights are Reserved

VIII. FUTURE SCOPE

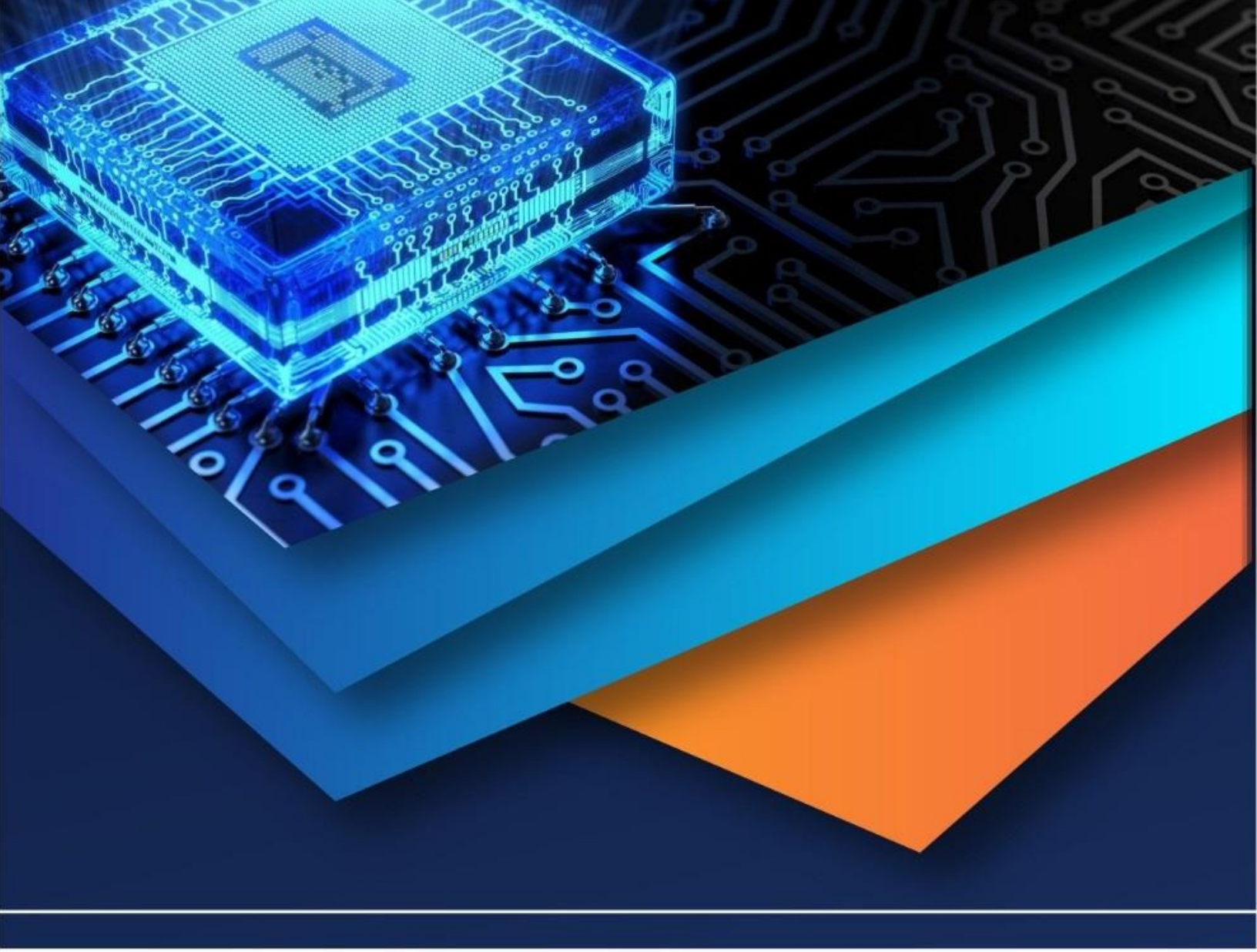
This AI model could ultimately be connected to an assortment of sites that give constant information to cost forecast. We may likewise incorporate a great deal of past vehicle cost information to help the AI model increment its precision. As a UI, we can make an Android application. We propose to utilize movable learning rates and train on groups of information rather than the whole dataset to further develop execution. Utilized vehicle deals are expanding internationally because of rising new vehicle costs and purchasers' monetary failure to buy them. Thus, there is a squeezing need for a Used Car Price Prediction framework that can precisely survey a vehicle's worth given a scope of variables. The proposed procedure will help with deciding a precise cost expectation for a trade-in vehicle.

IX. ACKNOWLEDGEMENT

We'd like to express our deepest appreciation to everyone who hashelped us create this project report. We'd like to thank K. Kusuma Latha, an assistant professor in the department of electronics and computer engineering, for guiding us through the project. We appreciate your excellent advice and encouragement during the project, as well as your efforts to ensure that we function systematically. It is a wonderful pleasure for us to be able to collaborate with him.

REFERENCES

[1] Enis genic, Becker Isakovic, Dino Keco, Zerina Masetic, Jasmin Kevric," Car Price Prediction Using Machine Learning"; (TEM Journal 2019) J. Brownlee – "Bagging and Random Forest Ensemble Algorithms for MachineLearning.
[2] Praful Rane1, Deep Pandya2, Dhawal Kotak3 "Used car price prediction ";International Research Journal of Engineering and Technology (IRJET). (2021)
[3] katamagrahari1, Ayushchobey2, Mamurkhan3, Manas srivasthava4" Car price prediction using machine learning" JUNE 2021.
[4] Abhishek pandey1, Vanshika Rastogi2, Shanika Singh " Car selling price prediction using random forest Machine learning Algorithm, MAY 2021.
[5] "Sameerchand Pudaruth, "Predicting the Price of Used Cars using Machine Learning Techniques" ;(IJICT 2014)
[6] SAS Academy website - "https://www.sas.com/en_in/insights/analytics/machine-learning.html"
[7] Dr. M. J. Garbade – "Clearing the Confusion:AI vs Machine Learnings Deep LearningDifferences" Available: https://towardsdatascience.com/clearing-the-confusion-ai-vs-machine- learning-vs deep-learning differences-fce69b21d5eb
[8] A. Wilson – "A Brief Introduction to Supervised Learning" Available: https://towardsdatascience.com/a-brief-introduction-to-supervised-learning54a3e3932590
[9] J. Rocca – "Ensemble methods: bagging, boosting and stacking" Available: https://towardsdatascience.com/ensemble-methods-bagging-boosting- and stacking-c9214a10a20
[10] J. Brownlee – "Bagging and Random ForestEnsemble Algorithms for machine Learning" Available: https://machinelearningmastery.com/bagging-and-random-forest- ensemble- algorithms for-machine learning/
[11] T. Yiu –"Understanding Random Forest Available: https://towardsdatascience.com/understanding-random-forest- 58381e0602d2
[12] N. Donges –" A COMPLETE GUIDE TO THE FOREST
[13] Kaggle - Used Cars Details | Kaggle [14] Google Colab –
[14] MyCars123.ipynb - Colaboratory (google.com)
[15] Dataset, https://www.kaggle.com/valentynsichkar/traffi c- signs-classification-withcnn
[16] Causes of road accidents, https://www.bajajfinservmarkets.in/insu rancemotor- insurance/traffic-rules-signs and-violations/common-causes- of-roadaccidents-in-india.html



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)