International Journal of Advanced Technology in Engineering and Science
Vol. No. 11, Issue No. 04, April 2023
www.ijates.com
ijates
ISSN 2348 - 7550

Price Prediction of Used Vehicles UsingMachine Learning

Thanuja, Hemasai,Ch. Anuhya, Venkata Reddy

(Under the guidance of Mr. M. Chennakesavarao M.E, (Ph. D), Associate ***professor,***

Department of Computer Science and Engineering, Tirumala Engineering College)

Abstract

The production of vehicles has been consistently expanding in the previous decade, with more than 70 million traveler's vehicles being delivered in the year 2016. This has brought about the trade-in vehicle market, which all alone has become a roaring industry. The new approach of online gateways has worked with the requirement for both the client and the merchant to be better educated about the patterns and examples that decide the worth of a pre-owned vehicle on the lookout. Utilizing Machine Learning Algorithms like Linear Regression, Multiple Regression. we will attempt to foster a factual model which will actually want to anticipate the cost of a pre-owned vehicle, in light of past shopper information and a given arrangement of highlights. We will likewise be contrasting the forecast precision of these models to decide the ideal one.

Keywords:-Car Resale; data mining; Car Resale analysis; classification; prediction; system, linear Regression; Support Vector Machine.

I INTRODUCTION

In this project, we mainly focus on the analysis of the Vehicle Resale Predict and then predict the results through them using training data. The trade- in vehicle market is an always rising industry, which has nearly multiplied its fairly estimated worth over the most recent couple of years. Clients , merchants be better educated about the patterns and examples that decide the worth of the pre-owned vehicle on the lookout. AI calculations can be utilized to anticipate the retail worth of a vehicle, in light of a specific arrangement of highlights. Various sites have various calculations to create the retail cost of the trade-in vehicles, and subsequently there is certainly not a

brought together calculation for deciding the cost. Via preparing measurable models at foreseeing the costs, one can undoubtedly get a good guess of the cost without really entering the subtleties into the ideal site. The fundamental target of this paper is to utilize three distinct expectation models to anticipate the retail cost of a utilized vehicle and think about their degrees of precision. The informational collection utilized for the forecast models was made by Shonda Kuiper[1]. The information was gathered from the 2005 Focal Edition of the Kelly Blue Book and has 804 records of 2005 GM vehicles, whose retail costs have been determined. The

informational index fundamentally contains unmitigated qualities alongside two quantitative characteristics and then test data of academics not only external exams, but also the overall academic performance of each and every student. In a significant number of theuniversities, when we see the scholastic execution examination is done, however there is no framework that predicts the understudy's exhibition ahead of time. Of which if understudy fizzles in an Exam.

II PREVIOUS WORK:

As indicated by author Sameer Chand, they have done the forecasts of vehicle cost from the chronicled information that has been gathered from every day papers. They have utilized the administered AI strategies for foreseeing the cost of vehicles. Numerous different calculations like various straight relapse, k-closest neighbor calculations, gullible based, and some choice treecalculations additionally been utilized. Every one of the four calculations are looked at and tracked down the best calculation for forecast. They have confronted a few challenges in looking at the calculations, by one way or another they have overseen. As indicated by creators Pattabiraman, thispaper is more focused on the connection among vender and purchaser. To foresee the cost of four wheelers, more highlights are required like previously given value, mileage, make, model, trim, type, chamber, liter, entryways, voyage, sound, cowhide. Utilizing these highlights the cost of vehicle has been anticipated with the assistance of factual investigation framework for exploratory information examination. As per creators EnisGegic et al, in this paper the chiefly focus on gathering different information from web entryway by utilizing web scrap methods. Furthermore, those have been contrasted and the assistance of various AI calculations to foresee the vehicle cost in simple way. They arranged the value as per various scopes of value that is as of now given. Fake neural organization, support vector machine, arbitrary timberland calculations were utilized on various datasets to construct classifiers model. Another methodology was given by Richardson in his postulation work. In his hypothesis it states more strong vehicles will be delivered by vehicle maker. He looked at the crossover vehicles and conventional vehicles in scraper it really holds their incentive for longer time utilizing numerous relapse procedures. This works on the natural conditions, and furthermore it assists with giving colossal effectiveness of utilizing energizes. Wu et al, in this paper they have utilized neuro fluffy information- based framework to exhibit vehicle value forecast. By considering the accompanying ascribes like brand, year of creation and sort of motor they anticipated a model which has comparative outcomes as the basic relapse model. Additionally, they made a specialist framework named ODAV (Optimal Distribution of Auction Vehicles) as there is a popularity for selling the by vehicles toward the finish of the renting year by vehicle vendors. This framework gives experiences into the best costs for vehicles, just as the area where all that cost can be acquired. To anticipate a cost of vehicles, the K – closest neighbor AI calculation has been utilized which depends on relapse models. More number of vehicles has been traded through this framework so this specific framework is all the more effectively oversaw.

III PROPOSED SYSTEM:

In view of the differing highlights and factors, and furthermore with the assistance of master information the

vehicle value forecast has been done precisely. The most essential elements for forecast are brand and model, period use of vehicle, mileage of vehicle, gear type and fuel type utilized in the vehicle just as fuel utilization per mile profoundly influence cost of a vehicle because of continuous changes in the cost of a fuel. Various highlights like (discretionary) outside shading, entryway number, sort of transmission, measurements, security, cool, inside, if it has route will likewise impact the vehicle cost. In this, we applied distinctive methods(like relapse, grouping, bunching and so forth) and techniques(like regulated, solo, semi managed) to accomplish higher accuracy of the pre-owned carvalue expectation.

They are three modules present in our project there are

1. Registration Module
2. Login Module
3. Resale value Prediction Module

Registration Module:

In this module web are giving some enlistment fields like name, email id, secret key, affirm secret word and address. In the wake of filling every one of the fields and snap on register button by client his/her enlistment was finished. at whatever point enlistment was finished, their subtleties will be put away in data set. Presently client having login qualifications.

Login Module:

In this module,after enrollment client will get login qualifications, by utilizing that login certifications client will login, if the login was effective, he will beexplored to another page or, more than likely he will get mistake message like username/secret key isn't right.Prior to understanding what direct relapse is, let us get ourselves acclimated with relapse. Relapse is a strategy for demonstrating an objective worth dependent on free indicators. This strategy is generally utilized for spreading and discovering circumstances and logical results connection between factors. Relapse methods generally vary dependent on the quantity of autonomous factors.

In figure1 it is very clear how the linear regression algorithm will work.

Cost Function:

The cost function helps us to figure out the bestpossible values for a_0 and a_1 which would provide the best fit line for the data points. Since we want the best values for a_0 and a_1, we convert thissearch problem into a minimization problem where we would like to minimize the error between the predicted value and the actual value[5].

Resale Value Prediction Module: Minimize $\frac{1}{n}\sum_{i=1}^{n}(pred_i - y_i)^2$

Here the student can only view his marks with his respective username and password and then they get the grade to their marks[4], then have better knowledge on how to improve in subjects.

IV Algorithms:

We used to algorithms in this paper

1. Linear Regression
2. Random Forest Regression
3. Decision Tree Regression

Linear Regression:

It is an AI calculation dependent on administered learning. It plays out a relapse task. It is utilized to assess genuine qualities (cost of houses, number of calls, absolute deals and so forth) in view of nonstop variable(s). Here, we set up connection among free and ward factors by fitting a best line. This best fit line is known as relapse line and spoke to by a straight condition Y= a *X + b.**J**= $\frac{1}{n}\sum_{i=1}^{n}(pred_i - y_i)^2$

Cost function(J) of Linear Regression is the Root Mean Squared Error (RMSE) between predicted y value (prediction) and true y value (y).

Random Forestalgorithm:

Random Forest is an adaptable, simple to utilize AI calculation that produces, even without hyper- boundary tuning, an incredible outcome more often than not. It is likewise perhaps the most utilizedcalculations, in view of its effortlessness and variety (it very well may be utilized for both order and relapse errands). In this post we'll figure out how the arbitrary woodland calculation functions, how it contrasts from different calculations and how to utilize it and is diagrammatically represented in Fig 1.

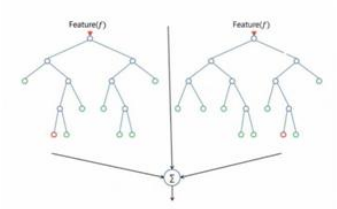


Fig 1. Random Forest

The Following Code will give a brief idea how we can use the algorithm is python.

```
import numpy as nm

import matplotlib.pyplot as mtpimport pandas as  pd #importing datasets

data_set= pd.read_csv('user_data.csv')

#Extracting Independent and dependent Variablex= data_set.iloc[:, [2,3]].values

y= data_set.iloc[:, 4].values

# Splitting the dataset into training and test set.

From sklearn.model_selection import train_test_split

x_train, x_test, y_train, y_test= train_test_split(x, y, test_size= 0.25, random_state=0)
```

Python provides the predefined library from SKLEARNS we can import all the algorithm and train them.

Decision Tree algorithm:

Decision Tree calculation has a place with the group of directed learning calculations. In contrast to other managed learning calculations, the choice tree calculation can be utilized for taking care of relapse and order issues as well.The objective of utilizing a Decision Tree is to make a preparation model that can use to foresee the class or worth of the objective variable by taking in straightforward choice standards induced from earlier data(training information).

In Decision Trees, for anticipating a class mark for a record we start from the foundation of the tree. We analyze the upsides of the root characteristic with the record's quality. Based on correlation, we follow the branch relating to that worth and leap to the following hub.

Accuracy (e.g. classification accuracy) is a measure for classification, not regression so we can't calculate accuracy for a regression model.

For regression, one of the matrices we've to get the score (ambiguously termed as accuracy) is R-squared (R2).

You can get the R2 score (i.e accuracy) of yourprediction using [12] the score(X, y, sample_weight

= None) function R2 score(accuracy) of our project is 76.65341530515258 %

Regression model accuracy calculated in following ways

1.R-Squared

2.Mean Absolute Error3.Mean Squared Error

1.R-Squared of our project is 76.65341530515258 %

2. mean absolute error of our project is3412.8022022817163

3. mean square error of our project is47719623.3816024

REGRESSION	R2SCORE	RMSE	MAE
Linear	0.7665341530515258	6907.939155898986	3412.8022022817163
Decision	0.6693132403821778	8221.389576449885	4398.592078169706
Random	0.8614440569888814	5321.6879374429045	1725.187967350853

Table 1. Accuracy results of linear, decision andrandom forest algorithms

The above Table 1 contains the results of all three Regression algorithms Linear, Decision and Randomfor all the above algorithms we are calculating the R2SCORE, RMES and MAE errors.

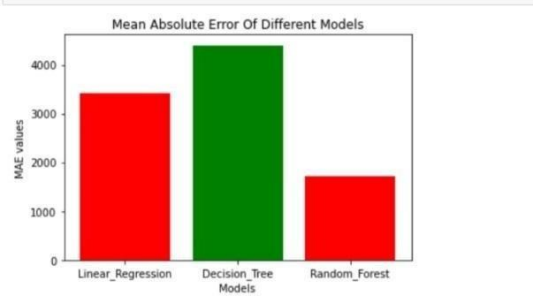


Fig 2. Mean Absolute Error of three models

The above Fig 2 shows the results of the mean absolute error of three models that is linear regression, decision tree and Random Forest models.

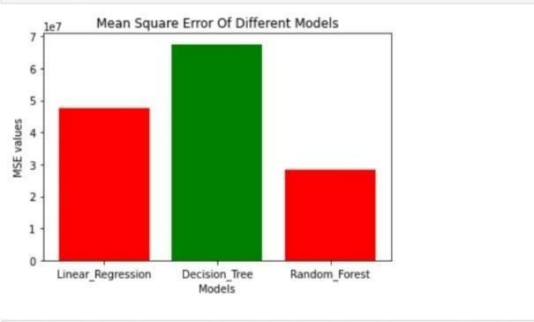


Fig 3. Mean Square Error of three models

The above Fig 3 shows the results of the mean square error of different models that is linear regression, decision tree and random forest algorithm.

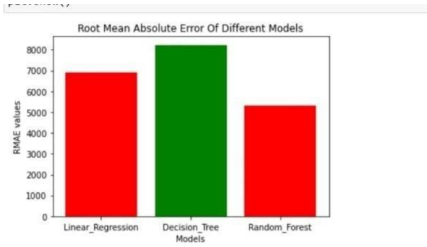
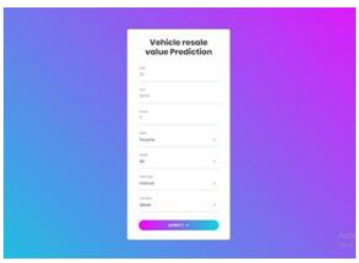


Fig 4. Root Mean Absolute error of three models

The above Fig 4 shows the results Root Mean absolute Error of three results Linear Regression,Decision Tree and Random Forest Algorithm.

After calculating all the above results we understand that the Random Forest algorithm is more suitable for us to calculate the prediction of vehicles.

V Project Screens:



Our project is generating the following results.

Fig 5. Home page

The screen1 represents the main page. It containsthe Prediction of cars.

```
RFR = RandomForestRegressor()
# Number of trees in random forest
n_estimators = [int(x) for x in np.linspace(start = 200, stop = 2000, num = 10)]
# Number of features to consider at every split
max_features = ['auto', 'sqrt']
# Maximum number of levels in tree
max_depth = [int(x) for x in np.linspace(10, 110, num = 11)]
max_depth.append(None)
# Minimum number of samples required to split a node
min_samples_split = [2, 5, 10]
# Minimum number of samples required at each leaf node
min_samples_leaf = [1, 2, 4]
# Method of selecting samples for training each tree
bootstrap = [True, False]
# Create the random grid
random_grid = {'n_estimators': n_estimators,
               'max_features': max_features,
               'max_depth': max_depth,
               'min_samples_split': min_samples_split,
               'min_samples_leaf': min_samples_leaf,
               'bootstrap': bootstrap}
rf_random = RandomizedSearchCV(estimator = RFR, param_distributions = random_grid, n_iter = 100, cv = 3, verbose=2, random_state
# Fit the random search model
rf_random.fit(X_train, y_train)
```

Fig 6. Fitting the model

In Fig 6, we represented the fitting the DatabaseModel.

The resale value of a vehicle is: [205944.]

Fig 7. Results of linear regression model

In Fig 7 the Linear Regression algorithm will predictthe total cost of the car.

VI CONCLUSION:

In this paper, four distinctive AI procedures have been utilized to figure the cost of pre-owned vehicles in Mauritius. The mean blunder with direct relapse was about Rs 51,000 while for kNN it was about Rs 27,000 for Nissan vehicles and about Rs 45,000 for Toyota vehicles. J48 and NaiveBayes exactness hung between 60-70% for various blends of boundaries. The primary shortcoming of choice trees and credulous bayes is their powerlessness to deal with yield classes with numeric qualities. Consequently, the value quality must be ordered into classes which contained a scope of costs yet this clearly presented further justification for errors. The primary limit of this examination is the low number of records that have been utilized. As future work, we plan to gather more information and to utilizes further developed methods like counterfeit neural organizations, fluffy logic and hereditary calculations to foresee vehicle costs.

VII Future Enhancement:

Hence, this investigation utilized various models to foresee utilized vehicle costs. Nonetheless, there was a generally little dataset for making a solid inductionas a result of the quantity of perceptions. Assembling more

information can yield more powerful expectations. Furthermore, there could be more highlights that can be acceptable indicators. For instance, here are a few factors that may work on the model: number of entryways, gas/mile (per gallon), shading, mechanical and restorative reconditioning time, used-to-new proportion, examination to-exchange proportion. Another point that has space to improve is that the information cleaning cycle should be possible all the more enthusiastically with the assistance of morespecialized data. For instance, rather than utilizing the 'fill' technique, there may be pointers that assistance to fill missing qualities all the more genuinely.

References

[1] Kanwal Noor, 2017, Vehicle Price Prediction System using Machine Learning Techniques International Journal of Computer Applications. Volume 167 - Number 9

[2] Mariana Lusitania et al, (2009). Support vectorregression analysis for price prediction in a vehicle leasing application

[3] Richardson, M. S. (2009). Determinants of usedvehicle resale value.

[4] Listiani, M. (2009). Support vector regression analysis for price prediction in a car leasing application (Doctoral dissertation, Master thesis, TU Hamburg-Harburg).

[5] Richardson, M. S. (2009). Determinants of used car resale value. Retrieved from: https://digitalcc.coloradocollege.edu/islandora/object

[6] /[6] Wu, J. D., Hsu, C. C., & Chen, H. C. (2009). An expert system of price forecasting for used cars using adaptive neuro-fuzzy inference. Expert Systems with Applications, 36(4), 7809-7817

[7] Gongqi, S., Yansong, W., &Qiang, Z. (2011,January). New Model for Residual Value Prediction of the Used Car Based on BP Neural Network an Nonlinear Curve Fit. In Measuring Technology and Mechatronics Automation (ICMTMA), 2011 Third International Conference on (Vol. 2, pp. 682.

[8]. Pudaruth, S. (2014). Predicting the price of used cars using machine learning techniques. Int. J.Inf.Comput. Technol, 4(7), 753-764.