CAR PRICE PREDICTION SYSTEM USING MACHINE LEARNING

Ahamad Khan[1], Sasmita Mishra[2], Pradeep Kumar Mishra[3], Shiwangi[4], Sudhanshu[5], Gouri Shankar Mishra[6] Tarun Maini[7]

[1,3,4,5] Department of Computer Science & Application, Sharda School of Engineering & Technology,
Sharda University, Greater Noida, UP, India

[2] Dept. Of Business Management, C. V. Raman Global University, Bhubaneswar, Odisha, India

[6,7] Department of Computer Science & Engineering, Sharda School of Engineering & Technology,
Sharda University, Greater Noida, UP, India

[1] 2018012274.ahamad@ug.sharda.ac.in , [2] sasmitamishra@cgu-odisha.ac.in , [3] pradeepkumar.mishra@sharda.ac.in

[4] 2018008503.shiwangi@ug.sharda.ac.in , [5] 2018013169.sudhanshu@ug.sharda.ac.in ,

[6] gourisankar.mishra@sharda.ac.in , [7] tarun.maini@sharda.ac.in

Abstract:
Determining whether or not the indexed price of a vehicle has increased or decreased is a difficult undertaking because of the numerous factors that influence a used car's market value. This project intends to develop machine learning models that can exactly anticipate a used car's price depending on its features, helping customers to make intelligent conclusions. The supervised system learning approach is used in this research to build a system for predicting car prices. The investigations employ a few linear regressions as part of the system learning and prediction approach, which provided 98% prediction accuracy. Multiple unbiased variables are employed in multiple linear regressions, the true and predicted values of only one dependent variable, yet, are compared in order to assess the accuracy of the effects. The knowledge came from classified ads for automobiles used. Regression approaches of all kinds were employed by the researchers, the researchers used a range of regression methods, such as desire tree regression, linear regression, manual vector regression, and random regression, woodland regression, to get the highest level of accuracy. This research paper's objective is to examine whether auto charges correspond to the information provided (previous customer information like engine capability, distance travelled of manufacture, and so forth). After that, it was decided which method best suited the available fact set by comparing the relative performances of the several algorithms. Right now, a Java utility makes use of the earlier prediction model. Additionally, test data were used to assess the model's accuracy performance, and a result of 87.38% was obtained.

Keywords: multiple Vehicle rate, regression version, linear regression machine learning, Correlation

1. Introduction

Market prices for second hand cars have nearly doubled over the past few years, and the industry is expanding. The proliferation of online resources and other tools has made it easier for buyers and dealers to gain a better understanding of the factors that affect a used car's market value. The rate of any vehicle can be predicted using device learning algorithms based on a specified set of variables. A critical and significant project is car charge forecast, particularly when the vehicle is utilized and no longer fresh from the factory. More and more car buyers are looking into options of buying new vehicles outright due to surge in demand for used cars and up to 8% drop in demand for new cars in 2013. People choose to rent cars instead of purchasing them, which is a violation of the law between the parties involved, the direct provider, 0.33 birthday party, business entity, or coverage enterprise categories of vendors. The art of car price prediction is both thrilling and notoriously difficult. 921.456 cars were registered in 2014, with 84% of those being used for personal transportation, according to information from the BiH agency [1].

Future predictions predict that this tendency will continue and that there will be a rise in the number of cars. The number is raised by 2.7% in year 2013 and predicted that it will increased. Navigation may even effect the cost of the vehicle. In this study, we used a number of approaches and strategies to improve the precision of the used automobile price prediction.

In this study, we expanded a price model for used cars in a comparison study using lasso regression and linear regression. Information acquired via Kaggle for each set of rules. Finding the best prediction model for estimating the cost of a used car is the primary objective of this study.

2. Literature review

Pricing forecast for second hand cars the first paper discusses the application of system studying techniques. They examine the potential applications of supervise system learning techniques for estimating the cost of used cars in Mauritius. The forecasts are based on more extensive data collected from regular courses.

Multiple methodologies, including multiple linear regression evaluations, have been used to make the predictions. According to author [3] car rate estimations for historical events came from daily newspapers. They employed supervised device studying methods to estimate the cost of automobiles.

Numerous decision tree algorithms, okay-nearest neighbor algorithms, multiple linear regression, and nave based are some of the other techniques that have been used. [4] This essay seeks to identify the best model prediction by assessing the effectiveness of the various ones.

Peerun et al. ran a study to assess how well the neural community as a whole predicted used car prices. However, the anticipated value isn't very accurate, especially for more expensive vehicles. They discovered that assist vector device regression performed somewhat better than neural networks and linear regression in predicting the cost of a secondhand car [4].

Table1: Summary of literature review

Year of paper Publication	Title of Paper	Findings
2019	Car price prediction system using machine learning techniques	In 2014, 921.456 cars were registered in BiH, with 84% of those being used for personal transportation, according to statistics agency data.
2021	Car price prediction using machine learning	His guiding philosophy was that automakers should build cars that are more durable
2021	Prediction used car price	The value of a secondhand car can be predicted statistically using only data about the previous owner and characteristics about the vehicle.
2022	Machine learning for used car price prediction	To estimate the cost of four-wheelers, other features are required, such as the price, mileage, specification, model, fuel type , door, tyre , etc.
2021	Machine learning for used car price prediction	A wider variety of this technology, which was used to transport cars, is much better managed

Richardson's paintings for his thesis [3]. His guiding philosophy was that automakers should build cars that are more durable. Richardson conducted a few regression analyses and found that hybrid motors maintain their value for a longer .period of time conventional motors This is related to climate change and environmental challenges, and it improves fuel performance. Using a system based on neuro-fuzzy knowledge, Wu et al. [4] tested the ability to anticipate car prices. They paid attention to the following characteristics: the logo, the year of production, and the type of engine.

Since the straightforward regression version gave similar results, their prediction model did too. Due to the high demand for selling the automobiles through car dealers at the end of the lease year, they also created a specialized mechanism known as ODAV (premiere Distribution of Auction Vehicles). This technique offers facts in a convenient location at the finest automotive rates.. The price of vehicles was estimated using the okay- nearest neighbour system learning method, which is based entirely on regression models. Because a wider variety of this technology, which was used to transport cars, is much better managed.

3. Proposed Methodology

With some additional costs incurred by the government in the form of taxes, the price of new motors in the market is set by the manufacturer. Customers may be confident that their investment will be profitable when they buy a new car. However, used car sales are increasing globally as a result of the greater availability of new cars and consumers' reluctance to buy new vehicles due to a lack of affordable options.

To determine a used car's worthiness using a variety of capabilities, a used car price prediction machine is required. There are websites that provide this service, but they might not use the most accurate prediction algorithm. But there may also be a difference between evaluating the power of the real market price of a used car and using specific models and structures. Knowing their true market value is essential when both buying and selling.

Any company can easily charge their new cars based solely on the production and advertising costs they include. However, it can be challenging to define a price for a used car because it depends on numerous factors, including the car's brand, the year it was made, and more. Our objective is to forecast the premium price for a used car in the India. To compete with its foreign rivals, the company opened a production plant there and began manufacturing vehicles there. Chinese automaker Geely Automobile wants to enter the US market and compete with its EU and US counterparts. In order to better understand the elements that influence vehicle pricing, they have scaled back the size of a vehicle consultant company. They are especially curious to learn about the variables affecting car costs in the American market because those could be very different from those in the Chinese market. The objective of the business is: "Which factors can accurately forecast a vehicle's price?"

The consulting firm has built a significant dataset of different automobile models sold in the American market based on numerous market surveys, revealing how well those traits accurately define a car's pricing.

4. Experimentation and result

This work seeks to enhance a superb regression version in order to provide an accurate prediction of car pricing. To do this, we need some historical information on previously owned vehicles, for which we use pricing and a few other popular features. In addition to other features being considered as unbiased variables, the vehicle's price is taken into account as the established variable [1].
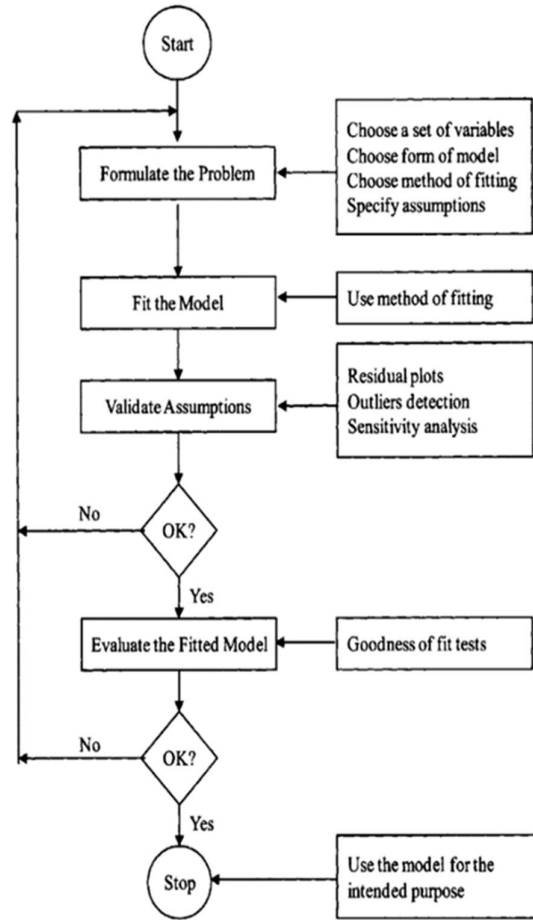
Fig.1: Flowchart of Regression Working.

If X and Y are the input and output, respectively, then the linear regression correlation can be written as:

$$Y = \beta_1 X + \beta_0 \qquad (1)$$

In the above equation:

β_0 and β_1 are the coefficient of regression. X is independent variable or input variable.
Y is dependent variable (depend on X) or outputvariable, whose value depend on X.
The equation 1 represents the relation in caseof single input. For multiple input:

$$Y = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots\ldots + \beta_{n-1} X_{n-1} + \beta_n X_n \quad (2)$$ In the above equation 2:
$X_1, X_2, X_3, X_4, \ldots\ldots\ldots, X_n$ are the inputs.

The regression plot reveals whether or not the residual values are normally distributed. After applying a regression model to the data, the plots shown in parents 1 and 2 were obtained. The daily probability plot of residuals vs % is depicted inside the result as upward sloping line (Fig. 1).

An ensemble learning-based regressionmodel called random wooded area. As the call indicates, it uses a version known as the selectiontree to create the ensemble version, which when combined with other decision trees yields a forecast. This variant has the

advantage that the wood is made in, because they operate in parallel and are extremely uncorrelated, each tree is not affected by the specific errors of other bushes, leading to exact results. Bootstrap Aggregation orbagging, which supply the randomness necessary to provide sturdy and uncorrelated bushes, help topartially ensure this uncorrelated behavior due tothe large number of features in the dataset, this version was chosen to evaluate a bagging strategyusing the following gradient boosting methods.
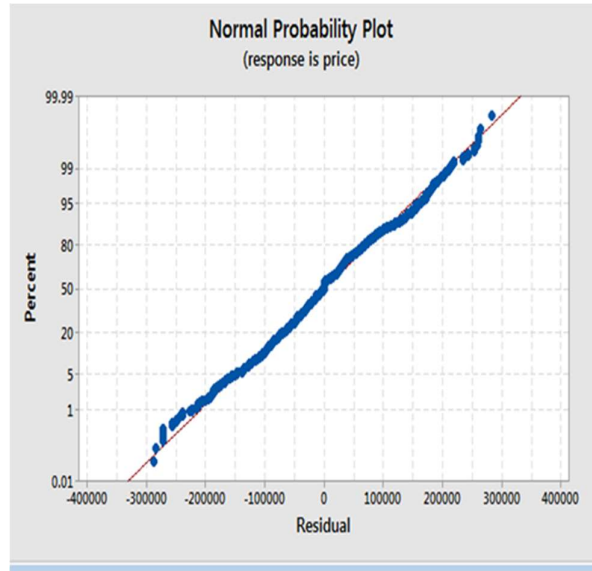
Fig.2: Result obtained by using linear regression.

The MLP regression [2], which uses a deep neural network perceptron regression version, was utilised to add mode complexity to the model. This model uses stochastic gradient descent or LBFGS to reduce squared losses. Two concealed layers with a width of 200 have been utilized, acquiring knowledge charge was set to 0.001 and the batch size was 200.relu is now used as the feature for activation.
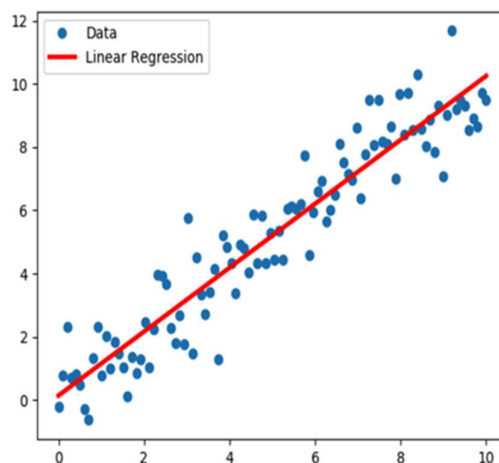
Fig.3 : Linear regression relation between independent and dependent variable

The coefficients of multiple regression models can be estimated using ridge regression when linearly independent variables have a high degree of correlation. Numerous academic fields

have used it, including as engineering, chemistry, and econometrics. A little amount of bias is added to linear regression using the bias regression method of version regularization, also known as ridge regression, in order to enhance long-term predictions. A request for L2 regularization has also been made.

In this method, the charge feature is altered by adding a penalty window. The amount of bias incorporated into the model is the ridge regression penalty. By multiplying the squared weight of each man and woman of each man or woman label by lambda we may determine it.

5. Conclusion.

The data set used in this study could be extremely valuable for carrying out similar investigations using various prediction techniques. By using this data set with the same or alternative prediction tools, it is possible to predict car prices. For training and function-checking purposes, the SVM classifier divided the data into parts (assist vector device). In particular, 75% of the data were used for system learning exams and 25% for device learning courses. The three network learning models' accuracy was evaluated and compared to one another. This evaluation is essential. firms that provide learning algorithms for both single and numerous devices. So, this model will help in predicting the car's charge.

Our objective for the upcoming studies is to test this device's ability to work well with different statistics sets, notwithstanding the device's spectacular success in the automobile rate prediction challenge. Using the aid of the OLX and eBay used automobile information modules. We can expand the size of our test data and verify the advised course of action.

References

[1] B. I. D. K. Z. M. J. K. Enis Gegic, Car Price Prediction using Machine Learning, TEM Journal., 2019.

[2] S. J. Kanwal noor, Vehicle price prediction system using machine learning techniques, International Journal of Computer Applications, 2017.

[3] A. C. M. K. M. S. Ketan Agrahari, Car Price Prediction Using Machine Learning, IJIRT, 2021.

[4] P. G. a. V. N. Kshitij Kumbar, Predicting Used Car Prices, CS 229 Project report.

[5] N. C. N. ,. A. K. J. Mr. Ram Prashath R, Price Prediction of Used Cars Using Machine learning, International Journal for Research in Applied Science & Engineering Technology (IJRASET), 2022.

[6] N. K. N. C. S. K. Madhuvanthi, Car Sales Prediction Using Machine Learning Algorithmns, International Journal of Innovative Technology and Exploring Engineering (IJITEE), 2019.

[7] S. R. Mohammed Munawwar Rangila, Car Price Prediction using Linear Regression, International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET), 2022.

[8] C. Jin, Price Prediction of Used Cars Using Machine Learning, 2021 IEEE International Conference on Emergency Science and Information Technology (ICESIT), 2021.

[9] A. Pandey, Car's Selling Price Prediction using Random Forest Machine Learning Algorithm, 5th International Conference on Next Generation Computing Technologies (NGCT-2019), 2020.

[10] Y. P. Y. S. K. G. S. P. Mrs. T Veda Reddy, Car Price Prediction using Machine Learning, Journal of Emerging Technologies and Innovative Research (JETIR), 2022

[20] Gouri Sankar Mishra, Pradeep Kumar Mishra, Parma Nand, Rani Astya, and Amrita, "User Authentication: A Three Level Password Authentication Mechanism", in Journal of Physics: Conference Series 1712, 2020, 012005 doi:10.1088/1742-6596/1712/1/012005

[21] Rachna Jain, Abhishek Sharma, Gouri Sankar Mishra, Parma Nand, and Sudeshna Chakraborty, "Named Entity Recognition in English Text", in Journal of Physics: Conference Series 1712, 2020, 012013, doi:10.1088/1742-6596/1712/1/012013

[22] G. S. Mishra, P. Nand and Pooja, 'English text to Indian Sign Language Machine Translation: A Rule Based Approach', International Journal of Innovative Technology and Exploring Engineering (IJITEE), 2019, 8(10)

[23] Gouri Sankar Mishra, K. K. Ravulakollu and A. K. Sahoo, "Word based statistical machine translation from English text to Indian Sign Language", in Journal of Engineering and Applied Sciences,12(2), 2017, p.481-488.

[24] S. Tyagi, and Gouri Sankar Mishra, "POS Tagging Using Support Vector Machines and Neural Classifier", in International Journal of Computer Science And Technology, Vol. 7, Issue 2, 2016, ISSN : 2229-4333

[25] Pradeep Kumar Mishra, Ali Imam Abidi and Gouri Sankar Mishra, "Improved Methodology For Personality Assessment Using Handwritten Documents", in Journal of Positive School Psychology, Pages 3263-3273,pub 2022/6

[26] Pradeep Kumar Mishra, Ali Imam Abidi and Gouri Sankar Mishra, "Baseline and its Slant Based Personality Assessment from Handwritten Documents", in International Journal of Mechanical Engineering PP 2947-2953, pub 2022/5

[27] Samiksha Kumari, Karan Kumar Singh, Gouri Sankar Mishra and Parma Nand, "A Comparative Study Of Security Issues And Attacks On Underwater Sensor Network", Accepted, to be published in Lecture Notes in Networks and Systems, 2021, LNEE: https://www.springer.com/series/15179

[28] S. Parhar, A. Roy, K. Kumar, A. Kumar and G. S. Mishra, "Lung Field Segmentation of X-ray Images by Normalized Gradient Gaussian Filter and Snake Segmentation," 2021 2nd International Conference for Emerging Technology (INCET), Belagavi, India, 2021, pp. 1-5, doi: 10.1109/INCET51464.2021.9456146.

[29] Syed Faraz Ali, Gouri Sankar Mishra and Ashok Kumar Sahoo, "Domain bounded English to Indian sign language translation model", In International Journal of Computer Science and Informatics, 2013, p. 41-45