IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 7.105 ⁜ Vol. 9, Issue 4, April 2022

DOI: 10.17148/IARJSET.2022.9457

Car Price Prediction using Machine Learning

Aditya Nikhade[1], Rohan Borde[2]

Dr. D Y Patil School of Engineering Academy Ambi, Pune, Maharashtra, India

Abstract: The main goal of this project is to predict used car prices, compare prices, and estimate the lifespan of a certain car based on a variety of facts about that vehicle. A new car is reported to lose 10% of its value the moment it is driven out of the dealership. In this circumstance, the number of kilometers the car has been driven is the most important factor in determining its price. As a second consideration, it's important to remember that different car manufacturers price their vehicles differently, which results in price discrepancies between models. In other words, the primary goal of this project is to ensure that the money spent on the car is a good investment for the company itself. We employed supervised machine learning techniques to make predictions about used automobile prices. Kaggle's website is the primary source of data for the predictions. A variety of methods have been employed to predict the outcomes, including multiple linear regression, decision trees, and k-nearest neighbors. They then rank and compare each other's forecasts. Data Which we've gathered in order to determine the greatest performers among them. From this, it's clear that although though this seems like a simple problem, it turned out to be really challenging to solve accurately. All four of these techniques were effective and comparable. We plan to apply more complex algorithms in the foreseeable future to improve the accuracy of our forecasts.

INTRODUCTION :

Predicting the price of a used car was a difficult undertaking due to the numerous elements that influence the price of a used vehicle on the market. In the future, this research will develop machine learning models that can almost forecast a car's pricing based on its features and overall performance, allowing consumers to make informed decisions. We are utilizing and evaluating several learning algorithms on a dataset that contains the sale prices of various makes and models in this research. This project will compare all of this data to all regression techniques as well as the performance of several machine learning algorithms such as Linear Regression, Ridge Regression, and Decision Tree Regressor, in order to determine which one is the best. The project will assess the price of a car based on several parameters and compare the prices of old and new cars. This study will also determine how long an automobile will last, taking into account government restrictions as well as company claims. Regression Algorithms produce output with a continuous value rather than a category value, making it more predictable to obtain the real price of a car rather than a price range, which is why they are utilized. According to user input, a user interface has been created that accepts input from any user and displays the car's pricing. Predicting a car's resale value is not as straightforward as it appears. The value of a used car is determined by a number of things. The car's age, make (and model), origin (the manufacturer's original country), mileage (the number of kilometers it has travelled), and horsepower are the most crucial factors to consider. The cost of gasoline is steadily rising. As a result, fuel economy is the most important factor to consider. In practice, practically everyone has no idea how much fuel their automobile uses each kilometer travelled. Other factors include the type of fuel it uses, the interior style, the braking system, acceleration, the volume of its cylinders (measured in cc), safety index, the car's size, number of doors, paint color, weight, consumer reviews, prestigious awards won by the car manufacturer, whether it is a sports car, whether it has cruise control, its physical condition, whether it belonged to an individual or a company, whether it is automatic or manual transmission, and whether it is automatic or manual transmission. The car's appearance and feel are also major factors in its price. As we can see, the price is influenced by a variety of factors. Unfortunately, information regarding these criteria is not readily available, and the buyer must make a decision to buy a car at a specific price based on only a few factors.

METHODOLGY :

In the system, there are two basic phases: 1. Training phase: The system is trained by using the data in the data set and fitting a model (line/curve) according to the algorithm chosen. 2. Testing phase: the system is given inputs and its functionality is tested. The accuracy has been verified. As a result, the data used to train or test the model must be appropriate. The purpose of the technology is to detect and forecast the price of a used car. To do this, an appropriate algorithm must be employed to complete the two tasks. The accuracy of various algorithms was compared. Prior to the algorithms being chosen for further use. The best candidate for the job was chosen.

OBJECTIVES :

I. To create an efficient and effective model that forecasts the price of a used car based on the inputs of the user II. To reach a high level of precision III. Create a user-friendly User Interface (UI) that takes the user's input and forecasts the price.

The Information: This project's dataset was downloaded from Kaggle Data Cleaning: Cleaning data with a data cleaning library like NumPy, pandas for Datasets, and NumPy for undesired data.

Pre-Processing Data: Encoder for Labels: There are 12 categorical variables and four numerical variables in our dataset (the price column is not included). In order to use the ML models, we must convert these categorical variables to numerical variables. The sklearn module Label Encoder was used to tackle this problem. Data Normalization: The dataset does not have a normal distribution. Each feature has its own set of parameters. Because their impact will be so minimal compared to the high value, the ML model will try to ignore coefficients of features with low values without normalization. Data for Training and Testing: In this process, 10% of the data was split for testing purposes and 90% of the data was used for training.

Architecture

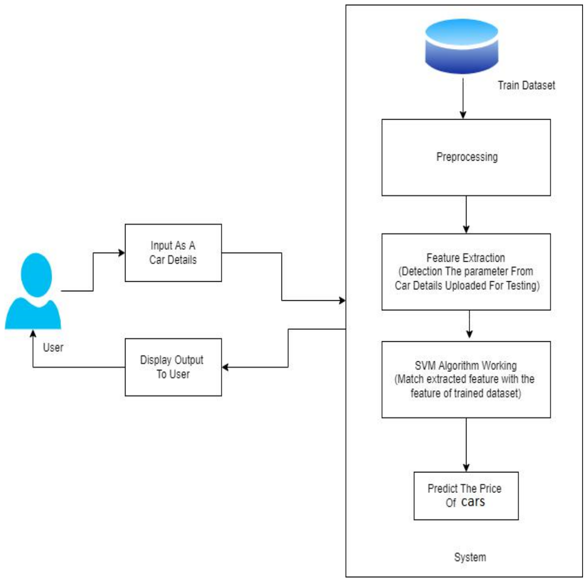


Figure 1

As shown in the above figure, the process starts by collecting the dataset. The next step after this is to do Data Preprocessing which includes Data cleaning, Data reduction, Data Transformation. Then, we will predict the price using various machine learning algorithms. The algorithms involve Linear Regression, Ridge Regression and Lasso Regression. The best model is selected which predicts the most accurate price. After selecting the best model, the predicted price will be display to the user according to user's inputs. User can give input through website to for used car price prediction to machine learning model

Linear Regression By fitting a linear equation to observable data, linear regression attempts to model the

relationship between two variables. The other variable is referred to as a dependent variable. For example, a

modeler might want to use a linear regression model to match people's weights to their heights. In order to determine the association between many continuous variables, linear regression is used. There are several independent variables as well as a single independent variable. y = m1X1+m2X2+......+b m1, m2, m3.... y intercept X1,

X2, X3...... independent variables y dependent variables.

Regression of the Ridge

A regularized form of LinearRegressor is a Ridge regressor. The regularized term with the parameter 'alpha' is used to adjust the regularization of the model, which helps to reduce the variance of the estimations. Step 5: Fitting the SVM classifier model.

LITERATURE SURVEY :

Prediction of Prices for Used Car by Using Regression Models[1] Abstract —For this work, we compared the performance of regression models based on supervised machine learning models. Each model is trained using data from a German e-commerce website on the used automobile market. As a consequence, the best performance comes from gradient boosted regression trees, with a mean absolute error (MSE) of 0.28.
Random forest regression with MSE = 0.35 and multiple linear regression with MSE = 0.55 were followed by random forest regression and multiple linear regression, respectively.

Keywords—comparative study, multiple linear regression, random forest, gradient boosting, supervised learning
Prediction car prices using quantify qualitative data and knowledge-based system [2] Abstract —The process of knowledge acquisition for expert systems has a strong association with car pricing utilizing machine learning. The time-consuming practice of recommendation, posting for automobile purchasing or selling on internet market websites has recently become the key method for knowledge acquisition. We can divide the data into two sorts after discovering it: structured and unstructured, both of which require knowledge-based analysis. The approaches for extracting meaning, data inference, and rules for qualitative data will be covered in this paper. The current study's main goal is to investigate various types of automotive data in order to develop an automated technique for predicting car pricing.
Index Terms—Prediction; Car Pricing; Entity Embedding; Quantify Qualitative Data; Knowledge-based System
The third paper uses BP neural networks to develop a price evaluation model for a second-hand automotive system. [3] Abstract --The price evaluation model based on big data analysis is proposed in this research utilizing the optimized BP neural network method, which takes advantage of widely circulated vehicle data and a large number of vehicle transaction data to analyze the pricing data for each vehicle type. Its goal is to create a model for determining the price of a used car that best reflects the car's condition.

Contribution of Real-Time Pricing to Impacts of Electric Cars on Distribution Network: [4]Abstract—The problem scenario of a huge number of electric cars is depicted in this study. The distribution network is put under a lot of strain by concurrent charging. Distribution transformers' thermal limitations and distribution

feeders' voltage limits have been exceeded. Rather than building distribution infrastructure, demand side management aids in the alleviation of difficulties. Customers modify their electricity consumption in response to pricing signals that change over time. The role of real-time pricing in the consequences of charging electric cars is investigated in this study. The simulations are based on real-world data from Thailand, including load demand and electricity costs for residential consumers. Customers who respond to real-time pricing signals spend less money on power bills than those who choose flat-rate or TOU tariffs, according to the findings. Furthermore, real-time pricing lessens the distribution network's load.

Index Terms— Electric car, Flat rate, Price-based demand response (DR), Real-time pricing (RTP), Time of use (TOU)

A Comprehensive Study of Machine Learning algorithms for Predicting Car Purchase Based on Customers' Demands: [5] Abstract--The automobile sector is a significant contributor to the national economy. Cars are becoming increasingly popular as a mode of private transportation. When a buyer wants to acquire the correct vehicle, especially a car, he needs to do some research. Because it is an extremely expensive automobile. Before purchasing a new car, several criteria and aspects must be considered, such as spare parts, cylinder volume, headlamp, and, most importantly, pricing. So, before making any decisions, it is critical for the buyer to make the best purchase possible that meets all of the criteria. Our purpose is to assist the consumer in making an informed decision about whether or not to purchase a vehicle. As a result, we intended to develop a technique for making decisions in a car purchase system. That is why, in this research, we suggest some well-known methods for improving car buying accuracy. We used those methods on 50 data points in our dataset. The Support Vector Machine (SVM) produces the greatest results, with a prediction accuracy of 86.7 percent. In this study, we also present comparison findings for all data samples using various methods for precision, recall, and F1 score.

Index Terms—Supervised Machine Learning, Naive Bayes, Random Forest tree, Support Vector Machine, KNN, Accuracy, Cosine Similarity

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 7.105 ⁜ Vol. 9, Issue 4, April 2022

DOI: 10.17148/IARJSET.2022.9457

FUTURE SCOPE:

This Initiative Machine learning models will be linked to a variety of datasets and websites that can supply real-time data

for price prediction. Will be kept on their website or on GitHub. We may also upload a large amount of car price data to help improve the machine learning model's accuracy. We're also working on an Android app as a user interface for connecting with and interacting with users. We also intend to employ a neural network to improve the model's performance.

CONCLUSION:

With the rising prices of new automobiles on the market, there is a need for used car sales at every Taluka level for those who cannot afford to acquire high-priced new cars.

As a result, an automobile Price Prediction system is required, which will predict the car's value based on a range of factors. The application of this model approach will aid in determining an accurate used automobile price estimate. We develop a model using the linear regression algorithm with the help of most survey papers, and we may create a UI application for it.

RESULTS AND DISCUSSION:

After the successful training of the model then we can apply the trained model for prediction of car prices from manual input of various data parameters of the car such as the Car name, selling price, owner type, fuel type, etc. we get the output as follows:- This figure 2 below is the interface representation of the application for predicting the prices:

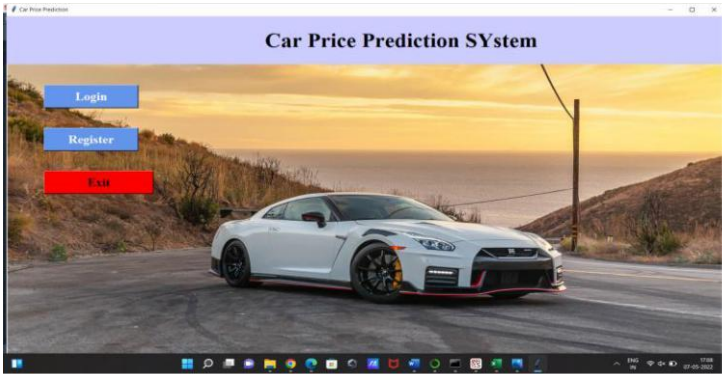
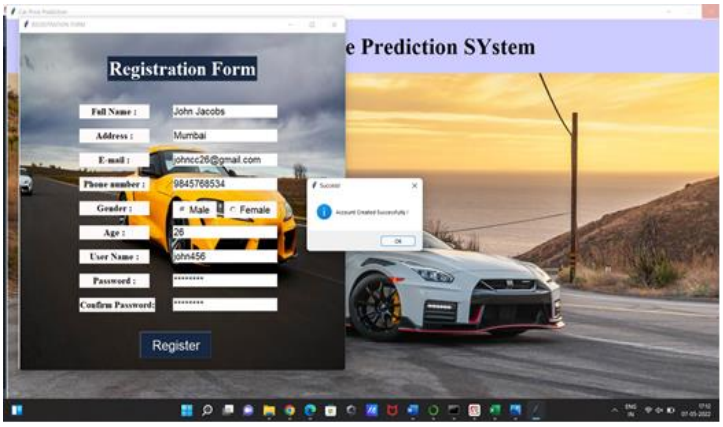


Figure 2



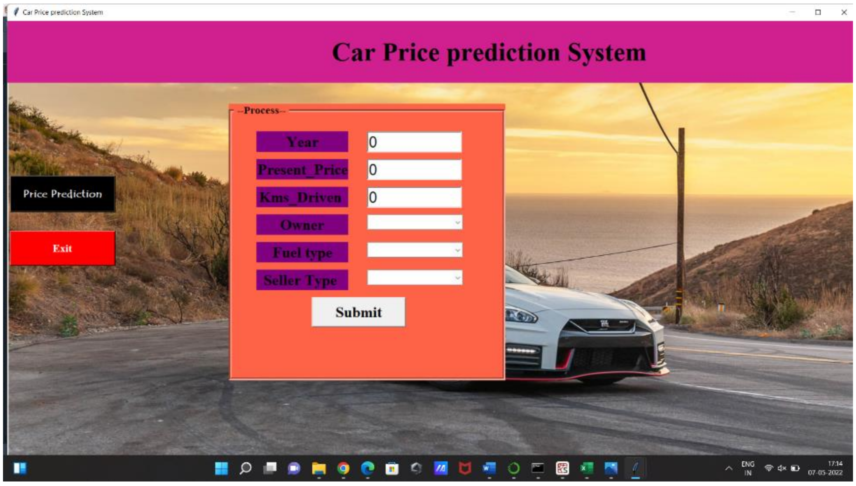
Full Name : John Jacobs
Address : Mumbai
E-mail : johncc26@gmail.com
Phone number : 9845768534
Gender : ● Male ○ Female
Age : 26
User Name : john456
Password : ********
Confirm Password: ********

Account Created Successfully !

Figure 3

DOI: 10.17148/IARJSET.2022.9457

Figure 3 represents the Registration Form where the user is asked to fill in the required information to register to the application.



Car Price prediction System

--Process--

Year: 0

Present_Price: 0

Kms_Driven: 0

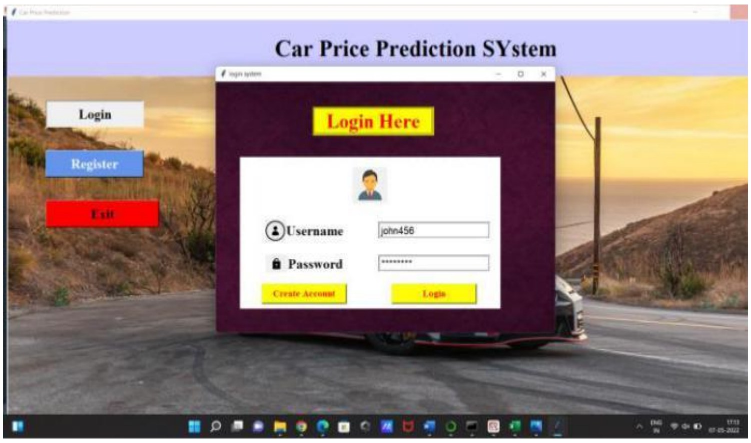
Owner:

Fuel type:

Seller Type:

Figure 4

Figure 4 represents the log in page where the user is asked to fill in the username and password respectively.



Car Price Prediction SYstem

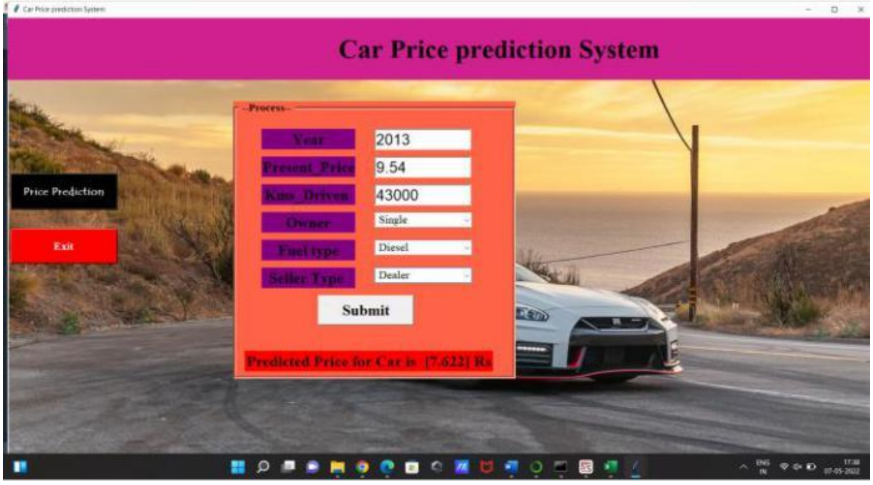
Login Here

Username: john456

Password: ********

Figure 5

Figure 5. represents the price prediction system in which the user is asked to fill in the required information of car such as year Present Price ,Kilometers Driven,Owner Type ,Fuel Type,And Seller Type.



Car Price prediction System

--Process--

Year: 2013

Present_Price: 9.54

Kms_Driven: 43000

Owner: Single

Fuel type: Diesel

Seller Type: Dealer

Predicted Price for Car is [7.622] Rs

Figure 6

DOI: 10.17148/IARJSET.2022.9457

Figure 6 shows OUTPUT of the predicted price of car based on the Data that the User provides to the prediction Model.

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 Machine Learning"; (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, ThongchaiKaewkiriya, 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, NguyenNhat Anh, Nguyen Thi Ngoc Anh, "Prediction carpricesusing qualify qualitative data and knowledge-based system" (Hanoi National University

[6] A. K. Elmagarmid, P. G. Ipeirotis, and V. S. Verykios, "Duplicate Record Detection: A Survey," IEEE Transactions on Knowledge and Data Engineering, vol. 19, no. 1, pp. 1–16, jan 2007.

[7] M.C.Newman,"Regression analysis of log-transformed data: Statistical bias and its correction," Environmental Toxicology and Chemistry, vol. 12, no. 6, pp. 1129-1133, 1993. [Online]. Available: http://dx.doi.org/10.1002/etc.5620120618

[8] F. Pedregosa, G. Varoquaux, A Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vander- (5)plas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duches-nay, "Scikit-learn: Machine Learning in fPgython," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.

[9] J. Morgan, "Classification and Regression Tree Analy-sis," Bu.Edu, no. 1, p. 16, 2014. [Online]. Available: http://www.bu.edu/sph/files/2014/05/MorganCART.pdf