Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 5, May : 2023

PRE-OWNED CAR PRICE PREDICTION

[1] Dr. Ahemad Sajjad Khan [2] Saiyyad Shafaq [3] Shreya Bamane [4] Pooja Mahant [5] Trupti Sahare

[1 2 3 4 5] Department of Electronics and Telecommunication

Anjuman college of Engineering and Technology, Sadar , Nagpur

Abstract

The goal of this project is to create software that can reliably forecast a used car's price based on its qualities so that consumers may make educated decisions. On a dataset made up of the sale prices of various brands and models, we put various learning techniques into practise and evaluate their effectiveness. To determine which machine learning algorithm performs the best overall, we will compare its performance to other machine learning algorithms including Linear Regression and Random forest. The cost of the car will be determined based on a number of factors. Regression algorithms are employed because they provide us a continuous value as an output rather than a categorised value, making it easier to anticipate the actual cost of an automobile rather than the estimated cost.

Introduction

Numerous Indian consumers often sell their used cars to new owners, sometimes known as second, third, and so forth. These purchasers can sell their used cars on a number of websites, including cars24.com, cardekho.com, and OLX.com, but what should the automobile be worth? Algorithms for machine learning could be able to solve this issue. I predicted the purchase price of the used car using machine learning methods like Random Forest and Extra Tree Regression and the potent Python package Scikit- Learn using a history of previous used car sale data and machine learning techniques like Supervised Learning. No matter how big or small the dataset is, the results demonstrate that both approaches have excellent prediction accuracy.. The current system involves a procedure where a vendor chooses a price arbitrarily and the buyer is unaware of the car and its current market value. In actuality, neither the seller nor the price at which he ought to sell the car have any notion of the current value of the vehicle. We have created a model that will be quite effective in resolving this issue. Regression algorithms are employed because, as opposed to categorical values, their output is a continuous value. As a result, it will be feasible to forecast the exact cost of an automobile rather than its price range. The price of an automobile according to input from any user is displayed on a user interface that has also been designed.Identifying whether a used car's quoted price is reasonable Determining a used car's market value is a difficult undertaking because of the myriad factors that influence it. This project's primary goal is in order to make wise purchases, machine learning algorithms are being developed that can precisely estimate a used car's price based on its qualities. On a dataset made up of the sale prices of various brands and models, we put various learning techniques into practise and evaluated their effectiveness.

Problem statement

For automakers, the used automobile industry is sizable and crucial. It's also extremely likely that new automobile sales are correlated with the used car market. Car manufacturers must participate in the used car market in order to handle lease returns from car rental businesses, sell used cars at new car retail, and handle fleet returns.The used market presents a number of challenges for automakers. There are several variables that make it challenging to sell used automobiles on the used car market, which lowers sales margins, including the dire state of the world, the general issue of more people, more rivalry from other manufacturers, and the trend towards electronic cars. Because of this, automakers need effective decision support systems to preserve used car profit.

Literature Review

The recent introduction of online portals has allowed buyers and The factors that determine the market price of a used car can be properly assessed by sellers. Some examples of machine learning algorithms include regression trees, multiple regression, and lasso regression. We'll aim to create a statistical model that predicts the value of a used car using information from previous clients and other vehicle characteristics.[1] This study compares the accuracy of various models' predictions in order to determine the most suitable one.Several earlier research on the topic of used car price prediction have been carried out.Pudaruth used naive Bayes, k-nearest neighbours, multiple linear regression, and decision trees to predict the value of used cars in Mauritius. But because there weren't as many cars[2]. In order to produce more objective findings, we employ Price Based on Supervised Machine Learning algorithms to predict the price of used cars with less human participation. The procedure involves preprocessing the dataset using Python's Pycaret package and comparing the effectiveness of each algorithm using the function for algorithm comparison.[3]. This study focuses on creating an effective machine model using mathematics that might estimate the cost of a used car based on its present qualities. From the viewpoint of someone who is selling, it is difficult to anticipate the price of a used car with any degree of accuracy.[4].

Methodology

The process of building a model involves two stages:

Training: The dataset is used to train the model, which then fits a model using the model algorithm of choice.

Testing: The model's correctness is evaluated when the inputs are submitted. The data that is utilised to train or test the model must then be appropriate. Good models must be chosen because the model is designed to identify and estimate the cost of a used car.

The second option is available online, where a specific platform enables the user to determine the price he might receive if he decides to sell.

Kilometers travelled: We are aware that a vehicle's mileage plays a significant part when it is being placed up for sale. The age of the car increases with mileage.

Fiscal power: It is the vehicle's output of power. A vehicle has a higher value when it produces more.

Year of registration: The year the vehicle was registered with the Road Transport Authority is known as the "year of registration." The automobile will have a higher value the newer it is. Every year that goes by, the value will decrease.

Fuel Type: There were two types of fuel types present in the dataset that we had. Petrol and Diesel. It was relatively less dominant.

Random Forest Algorithm

Popular machine learning algorithm andom Forest pertains to the supervised learning methodology. It can be applied to ML issues involving both classification and regression. It is built on the idea of ensemble learning, which is a method of integrating various classifiers to address difficult issues and enhance model performance.According to what its name implies, "Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset." Instead than depending on a single decision tree, the random forest uses forecasts from each tree and predicts the result based on the votes of the majority of predictions. The value of each element on the hypothesis can also be determined using the Random Forest Algorithm. The irregular forests strategy is very simple to comprehend. Each tree in Random Forest is arbitrarily picked from a subset of highlights. Because of the significant amount of change, there is less link between trees and more variability.

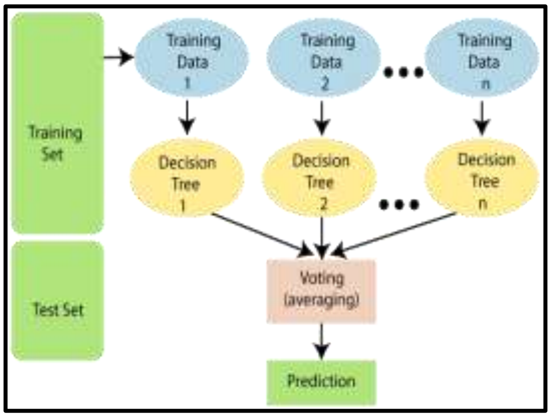

Fig 1: Random forest algorithm

Random Forest Hyper Parameters

The accuracy and speed of the expectation model are improved by the hyperparameters. The Sk Learn Library has the following restrictions. To choose the number of trees the computation builds before determining the most extreme democratic or expectations midpoints, use the number N assessors. Higher tree counts generally increase speed and stabilise conjectures, but they also come with some serious drawbacks.

Linear regression

One of the simplest and most widely used Machine Learning techniques is linear regression. It is a statistical technique for performing predictive analysis. For continuous/real/numeric variables like sales, salary, age, and product price, among others, linear regression makes predictions.The linear regression algorithm, often known as linear regression, demonstrates a linear relationship between a dependent (y) and one or more independent (y) variables. Given that linear regression demonstrates a linear relationship, it may be used to determine how the dependent variable's value changes as a function of the independent variable's value.The link between the variables is represented by a sloping straight line in the linear regression model. The Goodness of fit determines how the line of regression fits the set of observations. The process of finding the best model out of various models is called optimization.

$y= a_0+a_1x+ \varepsilon$

Here, Y= Dependent Variable (Target Variable), X= Independent Variable (predictor Variable), a0= intercept of the line (Gives an additional degree of freedom), a1 = Linear regression coefficient (scale factor to each input value)., ε = random error

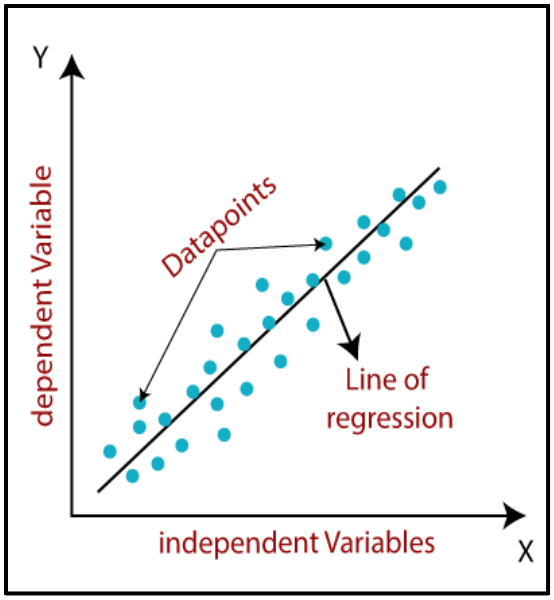

Fig 2: Linear regression

Implementation

Technologies used

Since Python has a large number of built-in methods in the form of bundled libraries, it was the primary technology utilised for the implementation of machine learning concepts. The well-known libraries and tools we used for this project are listed below.

NUMPY A general-purpose library for handling arrays is called NumPy[1]. It offers a multidimensional array object with outstanding speed as well as capabilities for interacting with these arrays. It is the cornerstone Python module for scientific computing. Beyond its apparent applications in science, NumPy also functions well as a multi-dimensional container of general data. Numpy's ability to declare any data-types makes it possible for

NumPy to quickly and easily interact with a wide range of databases.

SCIPY SciPy is a Python library that is available for free and open source and is used for technical and scientific computing. SciPy includes modules for a variety of common tasks in science and engineering, including optimisation, linear algebra, integration, interpolation, which special functions, FFT, signal and image processing, and ODE solvers. SciPy is a component of the NumPy stack, which also includes programmes like Matplotlib, pandas.

SCIKIT-LEARN Through a standardised Python interface, Scikit-learn offers a variety of supervised and unsupervised learning techniques. It is distributed under several Linux distributions and is available under a liberal simplified BSD licence, which promotes both academic and commercial use.

An open-source web tool called the Jupyter Notebook enables you to create and share documents with real-time code, equations, visuals, and text. It comprises data translation and cleaning, mathematical simulation, statistical modelling, data visualisation, and much more.

Pre - Owned Car Price Prediction

Fig 3: Implemented Project model

Linear regression

Using the value of another characteristic, LR is used to forecast the value of a variable. The dependent variable is the aspect that you are attempting to forecast. The independent variable is the label that is used to forecast the value of another feature. A = m + nB, where B is the independent variable, A is the dependent variable, and is the intercept y, and n is the slope of the line, makes up the LR equation. Accuracy: 85%

Random Forest

A Supervised Machine Learning technique called Random Forest Regression uses an ensemble model for regression. An ensemble learning technique combines the results of various machine learning models to produce more accurate predictions than a single model. Accuracy: 90%

Conclusion

A used car price prediction framework that can accurately assess a vehicle's worth given a range of characteristics is thus urgently needed. The suggested method will assist in determining an exact cost expectation. We didn't need to compute training accuracy because we utilised a Supervised machine learning model, which has high and complete training accuracy. However, the testing accuracy we received was 90 percent. The instances from the projected test set are also seen above. For the output of the desirable car pricing, we have provided the attribute inputs. This ML model may eventually be linked to a number of websites that provide ongoing data for cost forecasting. To increase the AI model's precision, we may also include a tonne of historical data on car costs. We are able to create an Android application as a UI. To improve execution, we suggest using moveable learning rates and training on subsets of data rather than the entire dataset. Due to escalating new vehicle prices and consumers' inability to afford them, used vehicle sales are growing internationally.

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, Thongchai Kaewkiriya, Suwat Rungpheung, Sabir Buya, Pitchayakit Boonpou, “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] M. Antonakakis, T. April, M. Bailey, M. Bernhard, E. Bursztein, J. Cochran, Z. Durumeric, J. A. Halderman, L. Invernizzi, M. Kallitsis, D. Kumar, C. Lever, Z. Ma, J. Mason, D. Menscher, C. Seaman, N. Sullivan, K. Thomas, and Y. Zhou, "Understanding the mirai botnet," in Proc. of USENIX Security Symposium, 2017

[7] Iman Sharafaldin, Arash Habibi Lashkari, and Ali A. Ghorbani, ―Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization‖, 4th International Conference on Information Systems Security and Privacy (ICISSP), Portugal, January 2018

[8] Hossein Hadian Jazi, Hugo Gonzalez, Natalia Stakhanova, and Ali A. Ghorbani. "Detecting HTTP-based Application Layer DoS attacks on Web Servers in the presence of sampling." Computer Networks, 2017

[9] A. Shiravi, H. Shiravi, M. Tavallaee, A.A. Ghorbani, Toward developing a systematic approach to generate benchmark datasets for intrusion detection, Comput. Security 31 (3) (2012) 357–374.

[10] Abhishek pandey1, Vanshika Rastogi2, Shanika Singh “ Car selling price prediction using random forest Machine learning Algorithm, MAY 2021.