
VEHICAL FUTURE PRICE PREDECTION USING MACHINE LEARNING

Prof. Anand Gyanbote¹, Iqbal Inamdar², Vikas Singh³

^{1,2,3}Brahmdevdada Mane Institute of Technology, Belati, Solapur.

DOI: <https://www.doi.org/10.58257/IJPREMS33541>

ABSTRACT

The price of a new car in the industry is fixed by the manufacturer with some additional costs incurred by the new Government in the form of taxes. So, customers buying a new car can be assured of the money they invest to be worthy. But, due to the increased prices of new cars and the financial capability of the customers to buy them, Used Car sales are on a global increase. Therefore, there is an urgent need for a Used Car Price Prediction system which effectively determines the worthiness of the car using a variety of features. Existing System includes a process where a seller decides a price randomly and buyer has no idea about the car and its value in the present day scenario. In fact, seller also has no idea about the car's existing value or the price he should be selling the car at. To overcome this problem we have developed a model which will be highly effective. Regression Algorithms are used because they provide us with continuous value as an output and not a categorized value. Because of which it will be possible to predict the actual price a car rather than the price range of a car. User Interface has also been developed which acquires input from any user and displays the Price of a car according to user's inputs.

Keywords -GPS - Global Positioning System GSM - Global System for Mobile

1. INTRODUCTION

In this project we are creating web application using python flask in machine learning and linear regression algorithm and random forest algorithm for accurate prediction for the vehicle price ML algorithms fabricate a model in light of test information, known as preparing information, to make forecasts without being expressly modified to do as such. AI contingent upon the idea of the sign or input accessible to the learning framework. Administered learning, solo learning, and Reinforcement learning are the three classifications of machine learning. ML is a part of artificial intelligence (AI) and software engineering which centers around the utilization of information and Algorithm to mimic the way that human learn, bit by bit working on its precision. ML algorithms fabricate a model in light of test information, known as preparing information, calculations are utilized in a wide assortment of utilizations, like medication, discourse acknowledgment and PC vision.

ML algorithms are customarily isolated into three general classifications, contingent upon the idea in this project we are **creating** application which predict the price 78% accuracy based on given input like vehicle name, purchase year, kilo meter driven, company, model, and is predict the accurate price. Regression, Ridge Regression, Lasso Regression, and Random Forest Regression ML algorithms in predicting the price of used cars. An important qualification of a price prediction tool is that depreciation can be represented to better utilize past data for current price prediction. The study has been conducted with a large public dataset of used cars. The results show that Random Forest Regression demonstrates the highest price prediction performance across all metrics used. It was also able to represent average depreciation much more closely than the other algorithms, at 13.7% predicted annual geometric depreciation for the dataset independent of vehicle age.

2. LITERATURE REVIEW

2.1) Design and Implementation of vehicle price prediction

This work will use a quantitative method to achieve the scientific goals. The evaluation of models will be done by collecting and comparing various performance metrics for each of the machine learning algorithms to be tested in the work. To ensure the highest possible accuracy for the various models, a result driven iterative process including data cleaning, model training, and model testing will be used to refine the models.

This work will be evaluated by how well the results derived from the method description are able to produce satisfactory answers to the research questions. The method should be able to produce conclusive answers to the first two research questions. It will be possible to train and test the Machine Learning models chosen, so long as they are viable for regression analysis. Through the creation of dummy variables, the categorical features in the dataset can be converted to continuous variables to be used as inputs in the regression models. This can however lead to a loss of information and reduced performance of the various ML models to different degrees.

2.2) Prediction The Price using linear regression algorithm and random forest algorithm.

The data, after being cleaned and normalized, is split into training and test data using a randomized 80-20 split. This is to ensure that the data used for testing does not contain any of the data used for training. Thus 20% of the data is reserved for testing purposes (see 4.4 Inference).

The training dataset was used to train the four price prediction ML models chosen: Multiple Linear Regression, Lasso Regression, Ridge Regression, and Random Forest Regression. All machine learning algorithms used in this report were imported from the sklearn library. Some models were provided input parameters to implement.

The motivations for the choice of input parameters are explained in this section for the models that require them. Random Forest was implemented with default parameters and `random_state=0`. The random state is necessary because it is a stochastic process that takes a seed value to begin. When testing different values, there was no noticeable performance increase. The random state throughout training was therefore set consistently to 0, minimizing stochastic behavior resulting from varying the random state.

2.3) Data Cleaning and Normalization.

A limitation on any ML algorithm's performance is the dataset. If the dataset does not include features that are strongly correlated to the price, the ML algorithm might not have access to enough information to accurately infer the price. Some strongly correlated features can be rendered redundant if another feature is included in the dataset and is strongly correlated to the redundant feature. It can only be opened when an authorized user is present, since there are no keys or combinations to be copied or stolen, or locks that can be picked.

The fingerprint-based lock therefore provides a wonderful solution to conventionally encountered inconveniences. This report focuses on the use of fingerprints. These were chosen somewhat arbitrarily but with the purpose of removing an appropriate percentage of uncommonly occurring extreme values in the dataset. This increases the performance of the models.

2.4) Machine Learning Dataset Limitations.

A limitation on any ML algorithm's performance is the dataset. If the dataset does not include features that are strongly correlated to the price, the ML algorithm might not have access to enough information to accurately infer the price. Some strongly correlated features can be rendered redundant if another feature is included in the dataset and is strongly correlated to the redundant feature. and mobile say 10 meters.

3. WORK CARRIED OUT

Main purpose of this project to help dealers To save the time and money of people whose know about the how dealership handle and buy and purchases the vehicle its also help to sellers also how don't now about real price of ours vehicle its predict the price given user input like company, model, year of purchases, kilometre driven, and give the accurate output.

4. APPLICATION DEVELOPMENT

The implementation aims to user easily use the our web application and its friendly to use its very simple ask for input and give the out put basis of input and in this we python flask which is python framework which proived real time Ui friendly application and user easye to use and solve their problem user also give new vehicle price price pridction also and its really works.predict price of used car and hence appropriate algorithms must be used to do the two different tasks. Before the algorithms are selected for further use, different algorithms were compared for its accuracy. The well-suited one for the task was chosen.

4.1) Training and Testing Accuracy Comparison

The three performance metrics that were taken for machine learning algorithm are R-squared, RMSE, and MAPE. These measurements were taken for both the training and testing inferences and averaged across all five iterations of inferences taken to produce a table of metrics.

Therefore, the dataset was split into four price categories and MAPE measurements taken for the inferences each algorithm. The four price categories were each approximately 25% of the dataset each, with the first one being the lowest priced cars and the last being the highest priced cars. In other words, the MAPE was taken for cars belonging to the 0th-25th price percentile, 25th-50th price percentile, 50th-75th price percentile, and 75th-100th price percentile. This serves to demonstrate the performance of each algorithm across different price categories of cars.

4.2) IMPLEMENTATION DETAILS

1. Data Sets.

Kaggle Dataset: The dataset was sourced from Kaggle and includes 122,144 car listings from the years 2018, 2019, and 2020 from all areas in the United States. It is available publicly. It includes all types of road-going consumer and also used, we also collect data from car dekho.com and local dealers list also

Dataset features:

	name	company	year	Price	kms_driven	fuel_type
0	Hyundai Santro Xing	Hyundai	2007	80000	45000	Petrol
1	Mahindra Jeep CL550	Mahindra	2006	425000	40	Diesel
2	Hyundai Grand i10	Hyundai	2014	325000	28000	Petrol
3	Ford EcoSport Titanium	Ford	2014	575000	36000	Diesel
4	Ford Figo	Ford	2012	175000	41000	Diesel
...
810	Maruti Suzuki Ritz	Maruti	2011	270000	50000	Petrol
811	Tata Indica V2	Tata	2009	110000	30000	Diesel
812	Toyota Corolla Altis	Toyota	2009	300000	132000	Petrol
813	Tata Zest XM	Tata	2018	260000	27000	Diesel
814	Mahindra Quanto C8	Mahindra	2013	390000	40000	Diesel

Fig 1: Dataset

Linear Regression

Linear Regression is a technique to estimate the linear relationship between each of a number of independent variables and a dependent variable. Linear Regression fits a linear model with coefficients $w = (w_1, \dots, w_p)$ to minimize the residual sum of squares between the observed targets in the dataset, and the targets predicted by the linear approximation. [5]

Random Forest Algorithm

Random Forest is an ensemble learning technique for classification and regression tasks. The algorithm makes use of Decision Trees. They consist of a set of independent binary trees, each stochastically trained on random subsets of data. Although these trees individually may be over trained, the randomness in the process of training results in the trees producing independent estimates, which are then combined to produce a result

5. RESULTS

This app predicts the price of a car you want to sell. Try filling the details below:

Select the company:
Hyundai

Select the model:
Hyundai Grand i10

Select Year of Purchase:
2016

Select the Fuel Type:
Petrol

Enter the Number of Kilometres that the car has travelled:
30000

Predict Price

Prediction: ₹388028.64

6. CONCLUSIONS

In this paper we discuss the challenges faced by both buyers and sellers to address this issue we are creating this web application, to solve this problem its application aimed to demonstrate skills machine learning and web development this approach involves problem solving of and predict the price of vehicle and they get better results .

7. REFERENCES

- [1] Karl Storchmann, "On the depreciation of automobiles: An international comparison". Department of Economics, Yale University, 2004. pp. 372-373.
- [2] Sri Totakura, Harika Kosuru, "Comparison of Supervised Learning Models for predicting prices of Used Cars". C.S. Bachelors Thesis, Blekinge Institute of Technology, Karlskrona, 2021.
- [3] Enis Gegic, Becir Isakovic, Dino Keco, Zerina Masetic, Jasmin Kevric, "Car Price Prediction Using Machine Learning"; (TEM Journal 2019).
- [4] Jaideep A Muley, "Prediction of Used Cars' Prices by Using SAS EM", Oklahoma State University.
- [5] Enis gegic, Becir Isakovic, Dino Keco, Zerina Masetic, Jasmin Kevric, "Car Price Prediction Using Machine Learning"; (TEM Journal 2019)