IFERP

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 9, Issue 11, November 2022

Electric Car Price Prediction using Machine Learning Techniques

[1] Vishal Khandare, [2] Manish Pandey

[1][2] Maulana Azad National Institute of Technology, Bhopal, M.P., India
Email: [2] vishal.khandare272@gmail.com, [3] contactmanishpandey@yahoo.co.in

Abstract*— Electric cars are becoming increasingly popular and are now the talk of the town. Accurate price forecasts are important because electric cars will be more economical to produce than conventional petroleum vehicles by the end of the decade. Rising costs for crude oil, its extraction and use are a burden on the environment. Electric vehicles, which are less expensive than conventional petroleum vehicles, are considered a great success in the world's transition to burning petroleum. In this research paper, we will use machine learning models. First, we use a linear regression model that achieves accuracy between 67% and 70%. Second, we will use an XGBoost model that achieves an accuracy of 90%. Finally, we will use the random forest model, which achieves 91% to 95% accuracy. In general, the results suggest that the price difference for larger cars will narrow by the middle of the decade, with further technological progress needed to make small electric vehicles cheaper compared to gasoline vehicles by the end of the decade.

***Keywords*— Electric cars, Linear regression, Random Forest, XGBoost**

I. INTRODUCTION

An electric vehicle (EV)[9] consists of a battery pack, which is the main power source that stores the electrical energy in the form of direct current. The inverter it contains converts direct current into alternating current and vice versa for further use. Another important component is the controller which regulates the amount of electrical energy supplied to the motor by the battery. The motor which will convert electrical energy to mechanical energy. Finally, DC-DC converter to convert battery voltage to desired level to drive auxiliary including power brake, steering and air conditioning and other (including all electronics, communication and sensors). Electric vehicle is a possible substitution for traditional crude oil vehicles to address the problem of increasing pollution, global warming, depletion of natural resources, etc.

Electric vehicle price prediction demands high knowledge and expertise as it depends on many distinctive attributes like motors, batteries, aerodynamics, weight, and driving range. The most expensive part in an electric vehicle is lithium-ion battery when compared to their other conventional counterparts. Various approaches are applied in this research paper to achieve high precision in electric car price prediction.

Not all of the electric vehicles we see on the road are powered entirely by electricity. Many of them are hybrid or plug-in hybrid vehicles. Battery electric vehicles (BEVs) are powered entirely by electricity from a lithium-ion battery. This type of vehicle does not use the traditional internal combustion engine and therefore emits no pollutants into the atmosphere. Hybrid electric vehicles have both an gasoline engine and a electrically operated engine, but the cell cannot be recharged by inserting a plug into an electric charging station. Unlike hybrid vehicles plug-in hybrid vehicles can be charged from an electrical outlet as well as powered by braking, where the vehicle is powered by both a gasoline engine and an electric motor.

This paper is organized as follows: Section II consists of related work in the field of used car price prediction. Section III explains the materials and methods used in the study of our work. Section IV presents different types of machine learning algorithms and examines their respective performance in predicting used car prices. Finally, Section V concludes our work and presents a plan for future work.

II. RELATED WORK

In his paper, Chuyang Jin [1] presents various regression methods including linear regression, polynomial regression, support vector regression, decision tree regression and random forest regression to achieve the highest precision. Assumed R-squared for five models in the paper. For all five models in the study, the author noted that support vector regression requires additional work. The process of support vector regression is to build a level that is furthest from the data points. The Random Forest regression has a maximum R-squared greater than 0.9, and is the only regression that reaches 0.9. The lowest of the five models is linear regression with an R-squared of 0.7234. This makes sense because linear regression works best with simple linear data. Support vector regression achieves an R-squared of 0.83545, much higher than linear regression and a little higher than polynomial regression. The decision tree regression is similar. With a minimal margin, but higher than the support vector regression, it reaches a value of 0.85

In his work, Shengquiang [2] proposed five regression models that are formed and evaluated using the data information acquired by feature engineering for training. To obtain a new regression model, Random Forest and XGBoost are mixed and the performance of the model is finer than the

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 9, Issue 11, November 2022

five regression models. WMR is used for the training datasets. It is better than Random Forest, XGboost, multiple linear regressions, GBDT, LightGBM and regression models in predicting used car prices.

Nitis Monburinon [3] led a relative study of the execution of regression models based on supervised machine learning. Every model was trained with used car market data obtained from a e-business website. With a mean absolute error of 0.30, gradient boosted regression trees provide the best result. Similarly, the mean absolute error of random forest regression is 0.36 and the mean absolute error of multiple linear regressions is 0.56.

III. MATERIALS AND METHODS

Electric car price prediction[8] consists of various steps. Firstly, data [5] is collected from a local webpage. The data consists of the following features: brand, model, acceleration in seconds, top speed in kilometer per hour, range in kilometer, efficiency in Watt hour per kilometer, fast charging speed in kilometer per hour, number of seats, price in Euros. The range is one of the most important criteria when buying an electric vehicle. Compared to the combustion engine, the range is shorter. Do not depend on driving range provided by the manufacturer, but look nearby and ask friends or family for reviews to get an idea about the driving range of the EV in real-world conditions. As already mentioned, the battery is the most expensive component in electric vehicles, the efficiency of which is increased by reducing its size. Therefore, an electric vehicle with a smaller battery size is more efficient. The price of lithium-ion batteries will fall in the coming years as battery manufacturing will take place outside of China and cheaper nickel is used instead of cobalt

Once the data is collected it needs some preprocessing and cleaning. We will remove symbols and units so that they get easily trained by our model as it is easy for models to read numeric values. Since many values are missing from the data, they are initially filled with zeros.

We will process the data. With the data preparation complete, we will focus on the data and data analysis. We find the correlation between the data using the heat map.

Processed data is as follows-:

	Brand	Model	AccelSec	TopSpeed_KmH	Range_Km	Efficiency_WhKm	FastCharge
0	Tesla	Model 3 Long Range Dual Motor	4.6	233	450	161	940
1	Volkswagen	ID.3 Pure	10.0	160	270	167	250
2	Polestar	2	4.7	210	400	181	620
3	BMW	iX3	6.8	180	360	206	560
4	Honda	e	9.5	145	170	168	190

Figure 1:Processed dataset of different cars

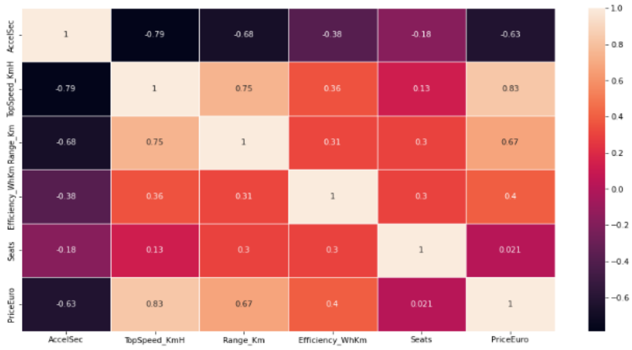

Figure 2: Correlation shown using heat map

The darker the color, the lesser the similarity between the two data points. For example, we can see that top speed and acceleration have a similarity of "-0.79",which is negative and not a fine score. Also the similarity between top speed and range is ".75" which is a fine score. So, the similarity in data has many real-life similarities.

IV. MODEL IMPLEMENTATION

The prices of the vehicles are predicted, which is derived from all parameters and data. Many data points are missing from the dataset.The dataset consists of prices in Euros. Test set at 1/3rd of the data and 2/3rd of the data is train set.

Linear regression[10] technique is applied to our model. It is a statistical way of measuring the relationship between variables. The variables are categorized into two parts dependent variable and non dependent variable. It is a supervised machine learning algorithm that gives continuous output. After applying R2 squared accuracy metrics we are getting accuracy of 67% to 70% after.

Lasso regression[7] is applied after linear regression. It uses regularization to improve the prediction accuracy and explainability of the resulting statistical model. It's just an addition of linear regression that adds a regular penalty to the loss function during training. After applying R2 squared accuracy metrics we are getting accuracy of 70% to 72%

Table 1: Comparison between models

Regression Model	R2 squared	
1	Linear Regression	.67
2	Lasso Regression	.70
3	Random Forest	.90
4	XGBoost	.89

For improving accuracy in our model, we applied the random forest[4] technique. It is a collection of many random decision trees and is much less sensitive to the training data. It is called random because we use two random processes bootstrapping and random feature selection

Classification and regression problems can also be solved using this technique. The concept of ensemble learning means to combine multiple learning algorithms for same task to

IFERP connecting engineers... developing research

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 9, Issue 11, November 2022

make better predictions than individual learning model. After applying R2 squared accuracy metrics we are getting accuracy of 90% to 92% after.

For checking the further increment in accuracy, we applied the XGBoost[6] technique. XGBoost is the most powerful machine learning algorithm out there today. XGBoost stands for extreme gradient boost trees. It is an ensemble method in which each tree boosts attributes that led to misclassification of previous tree. It can prevent overfitting and can handle missing values automatically. It enables early stopping, finding optimal number of iterations

V. CONCLUSION

Electric car price depends on various factors which play a crucial role in predicting the price. Data preparation and cleaning are important steps before training our model. Applying linear regression on the data set accuracy was 67%. Therefore, the random forest machine learning algorithm has been proposed and this method gains an accuracy of 90.38%.It gives far better results than linear regression. The disadvantage of using Random Forest is that it consumes a lot of computational power since a combination of models is used instead of a single model. So, we applied the XGBoost technique to the model which gave an accuracy of 89%.

Electric cars play a central role in the future of mobility. Electric vehicles do not cause pollution and are environmentally friendly. In the present situation purchasing an electric c ar is not a good option, where the charging infrastructure is still in its infancy. The driving range is very less as compared to traditional crude oil engine. The conventional car takes much less time to fill up its petrol or diesel tank while charging electric vehicle takes lot of time. However, if you are familiar with buying an electric car and consider all the factors, you should go for it. In the coming years, people will use more and more electric cars, then we will train our model with different datasets to check the performance.

REFERENCES

[1] Chuyang Jin “Price prediction of used car using machine learning”, 2021 IEEE International Conference on EmergencyScience and Information Technology (ICESIT), 978-1-6654-3531-4/21/$31.00 ©2021 IEEE, **DOI:**10.1109/ICESIT53460.2021.9696839

[2] Shengqiang “Second-hand Car Price Prediction Based on a Mixed-Weighted Regression Model”,2022 7th International Conference on Big Data Analytics (ICBDA), **DOI:** 10.1109/ICBDA55095.2022.9760371

[3] Nitis Monburinon “Prediction of prices for used cars by using regression models”,2018 5th International Conference on Business and Industrial Research (ICBIR), **DOI:** 10.1109/ICBIR.2018.8391177

[4] sklearn.ensemble.RandomForestClassifier — scikit-learn 0.19.2 documentation. (n.d.). Retrieved from: http://scikitlearn.org/stable/modules/generated/sklearn.ensemble .RandomForestClassifier.html [accessed: June 30, 2022].

[5] Used cars database. (n.d.) Retrieved from: https://www.kaggle.com/datasets/kkhandekar/cheapest-electric-cars [accessed: June 04, 2022].

[6] Dinglong Yu,” House Price prediction based on a machine learning model”. 2021 2nd International Seminar on Artificial Intelligence, Networking and Information Technology (AINIT)

[7] Pattabiraman Venkatasubbu. “Used Cars Price Prediction using Supervised Learning Techniques”. International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958

[8] Prateek Majumder.” Data Analysis and Price Prediction of Electric Vehicles”. Retrieved from: https://www.analyticsvidhya.com

[9] Szilard Enyedi.” Electric Cars challenges and trends”. 2018 IEEE International Conference on Automation, Quality and Testing, Robotics (AQTR)

[10] S. Peerun, N. H. Chummun, and S. Pudaruth, “Predicting the Price of Second-hand Cars using Artificial Neural Networks”