WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Price Prediction of Electric Vehicles

Prof. Sejal Thakkar[1], Mital Donga[2], Johny Panjwari[3], Dharmil Kotak[4], Kresha Savalia[5]

[1] Assistant Professor. CE Department, Indus University, Ahmedabad, India.
[2,3,4,5] Student, B.Tech CE, Indus University, Ahmedabad, India.

ABSTRACT –

Demand for electric vehicles has increased significantly over the past decade, and is projected to increase substantially with the onset of Covid-19. Therefore, to increase reliability in the expansion of the electric car market, a model is developed that can forecast the current market price of an electric car based on various criteria. This analysis can be used to study the trends in the industry, provide better insight into the market, and assist the community in its smooth operation. The objective of this research is to predict the price of electric vehicles based on a dataset.

***Key Words*: Price Prediction, Linear Regression, sklearn, manufacture, range.**

1. INTRODUCTION

The electric vehicle market is a significant part of the automotive industry and plays an important role in the economy. Predicting the price of an electric vehicle is a difficult task as it depends on various factors such as make, model, manufacture, range_km and price of the electric vehicle. Predicting the price of an electric vehicle is crucial for both buyers and sellers as it helps them make informed decisions.

A popular technique for predicting electric vehicle prices is linear regression. This technique involves establishing a linear relationship between the target variable (i.e., the price of the vehicles) and one or more independent variables (i.e., the factors that influence the price). This review paper provides an overview of electric vehicle price prediction using linear regression, including an overview of existing research and an assessment of the strengths and limitations of the technique. It also identifies avenues for further research in this area.

In this project, we will use the Kaggle dataset, which consists of a total of 5500 samples. Sklearn will be used for model building and all code will be implemented in Google Colab.

Linear regression

Linear regression is a statistical technique commonly used to model the relationship between two variables. In the context of predicting the price of electric vehicles, linear regression can be used to predict the price of an electric vehicle based on its characteristics, such as manufacture, range_km, and condition.

In linear regression, a straight line is fitted to a series of data points representing the relationship between the independent variable (e.g., manufacture or range_km) and the dependent variable (e.g., price). The line is then used to make predictions about the value of the dependent variable for new values of the independent variable.

There are two types of linear regression, namely:

1) Simple linear regression

2) Multiple linear regression.

Simple linear regression predicts a single dependent variable based on a single independent variable, while multiple linear regression predicts a single dependent variable based on multiple independent variables.

Linear regression is a popular technique for electric vehicle price prediction because it is easy to implement and can provide accurate predictions with the right characteristics. However, it is important to carefully select the features used in the model and consider other factors such as market trends and demand in the predictions.

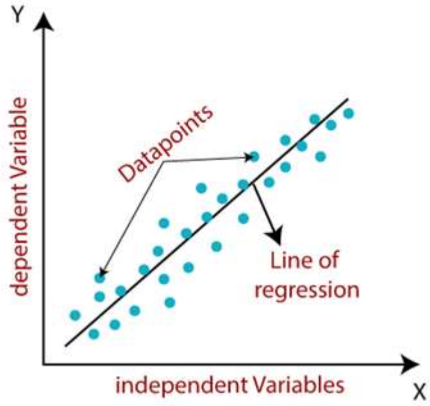

Fig 1. Basic Linear Regression graph

2. Dataset and Data Augmentation

The dataset used is from Kaggle and consists of 5500 rows. The dataset is divided into a training and a test set. The training dataset makes up about 80% and the remaining 20% is the test dataset. There are several factors, such as car_name, battery_packKwh, range_km, fastCharging_Kmh, price, plugtype, engine, seats etc.

	Brand	Model	AccelSec	TopSpeed_KmH	Range_Km	Battery_Pack Kwh	Efficiency_WhKm	FastCharge_KmH	RapidCharge	PowerTrain	PlugType	BodyStyle	Segment	Seats	PriceEuro
0	Tesla	Model 3 Long Range Dual Motor	4.6	233	460	70.0	161	940	Yes	AWD	Type 2 CCS	Sedan	D	5	55480
1	Volkswagen	ID.3 Pure	10.0	160	270	45.0	167	250	Yes	RWD	Type 2 CCS	Hatchback	C	5	30000
2	Polestar	2	4.7	210	400	75.0	181	620	Yes	AWD	Type 2 CCS	Liftback	D	5	56440
3	BMW	iX3	6.8	180	360	74.0	206	560	Yes	RWD	Type 2 CCS	SUV	D	5	68040
4	Honda	e	9.5	145	170	28.5	168	190	Yes	RWD	Type 2 CCS	Hatchback	B	4	32997
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
97	Nissan	Ariya 63kWh	7.5	160	330	63.0	191	440	Yes	FWD	Type 2 CCS	Hatchback	C	5	45000
98	Audi	e-tron S Sportback 55 quattro	4.5	210	335	86.5	258	540	Yes	AWD	Type 2 CCS	SUV	E	5	96050
99	Nissan	Ariya e-4ORCE 63kWh	5.9	200	325	63.0	194	440	Yes	AWD	Type 2 CCS	Hatchback	C	5	50000
100	Nissan	Ariya e-4ORCE 87kWh Performance	5.1	200	375	87.0	232	450	Yes	AWD	Type 2 CCS	Hatchback	C	5	65000
101	Byton	M-Byte 95 kWh 2WD	7.5	190	400	95.0	238	480	Yes	AWD	Type 2 CCS	SUV	E	5	62000

Fig 2: Dataset Structure

Data augmentation techniques are used to increase the size of the dataset by generating new samples from existing ones. Data augmentation is particularly useful when the size of the dataset is limited. Some common data augmentation techniques include rotation, translation, scaling, and flipping of manufactures.

In this project, data augmentation can be used to create new samples by modifying existing ones. For example, new samples can be generated by varying the range_km or condition of the electric vehicles or by combining features from different samples.

Data augmentation can improve the accuracy and robustness of the model by increasing the diversity and quantity of the data. However, care should be taken to ensure that the augmented data remains representative of the original dataset and does not introduce bias or noise into the model.

In summary, the choice of dataset and effective data augmentation techniques are critical for building an accurate and robust model. The dataset should contain diverse and representative samples, and data augmentation should be used to increase the quantity and diversity of the data.

```
Data columns (total 12 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   Name            180 non-null    object
 1   KWH             180 non-null    float64
 2   Acceleration    180 non-null    float64
 3   TopSpeed        180 non-null    int64
 4   Range           180 non-null    int64
 5   Efficiency      180 non-null    int64
 6   FastChargeSpeed 180 non-null    int64
 7   Drive           180 non-null    object
 8   NumberofSeats   180 non-null    int64
 9   PriceinGermany  180 non-null    int64
 10  PriceinUK       180 non-null    int64
 11  Manufacturer    180 non-null    object
dtypes: float64(2), int64(7), object(3)
memory usage: 17.0+ KB
```

Table -1: Rows description in dataset

3. Model Architecture

The primary model architecture used for electric vehicle price prediction is linear regression. Linear regression is a statistical model that predicts a numerical output variable based on one or more input variables. In the context of electric vehicle price prediction, the input variables are typically characteristics such as electric vehicle manufacture, range_km, and condition, as well as make and model.

The general formula for linear regression is:

y = b0 + b1x1 + b2x2 + ... + bnxn, where y is the predicted output variable, b0 is the intercept, b1-bn are the coefficients for each input variable (x1-xn), and xn are the input variables.

The model architecture for predicting the price of electric vehicles using linear regression typically involves training the model on a dataset of historical electric vehicle prices and characteristics such as manufacture, range_km, and condition. The model is then used to predict the price of a new or old vehicle based on its features.

Feature engineering has also been used in some studies to improve the accuracy of the models. Feature engineering involves selecting or transforming input variables to create new features that have better predictive power for the output variable. For example, one could create a feature that combines electric vehicle manufacture and range_km to better capture the overall wear and tear of the vehicle.

Overall, the model architecture for electric vehicle price prediction using linear regression is relatively simple, focusing on selecting appropriate input variables and optimizing the coefficients during the training process.

4. Model compilation and training

The process of model compilation involves the selection of appropriate features to be used in the linear regression model. Common features include electric vehicle manufacture, range_km, and condition, as well as make and model. Feature engineering techniques can also be used to create additional features that can improve the accuracy of the model. Once the features are selected, they are used to create the regression equation that is used to predict the price of an electric vehicle.

The training process involves using historical data to train the linear regression model. The data usually contains information about the characteristics used in the model as well as the actual prices of the electric vehicles. The data is divided into two groups: Training groups and Test groups. The training set is used to train the model, and the test set is used to measure the accuracy of the model.

During the training process, the model adjusts its parameters to minimize the error between predicted prices and actual prices. This is done using Python libraries such as Pandas and Sklearn. The process is repeated until the model obtains the minimum error. After the model is trained, it can be used to make predictions based on new data. The accuracy of the model can be evaluated using metrics such as mean absolute error, root mean square error, or R-squared.Overall, the model compilation and training process for electric vehicles price prediction using linear regression involves selecting appropriate features and using historical data to train the model. The process can be iterative and may involve feature engineering and optimization techniques to improve the accuracy of the model.

```
MSE:  107902152.72013463
```

```
RMSE:  10387.596099200942
```

```
R2:  0.815987092758551
```

```
MAE:  7155.050407804233
```

Table -2: Minimal error of model

5. Model Evaluation

Model evaluation is an essential step in the process of developing a predictive model. Evaluation involves assessing the performance of the model against a set of known data points or a holdout data set. The evaluation process aims to determine the accuracy, precision, recall, and other performance metrics of the model.

A common approach to model evaluation is to use a metric called the coefficient of determination or R-squared (R^2). The R^2 metric measures the proportion of variation in the dependent variable (EV price) that is explained by the independent variables (top speed, range_km, plug type, engine, and model). A high R^2 value indicates that the model fits the data well, while a low R^2 value indicates that the model does not fit well.

Overall, model evaluation is an essential step in the development of a predictive model. The evaluation process involves assessing the performance of the model against a set of known data points or a holdout data set using various performance metrics such as R^2, MSE, MAE, and RMSE. The selection of the appropriate evaluation metric depends on the specific objectives of the model and the available data.

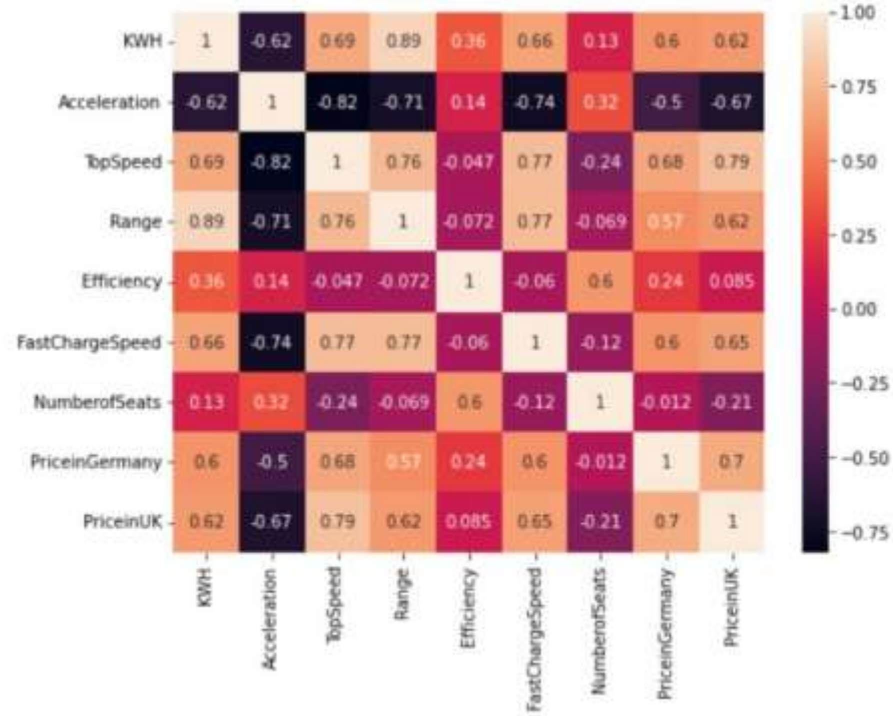

Fig 3: Correlation Data Points

The lighter the colour, the stronger the correlation between the two data points. For example, we can see that the price in Germany and the price in the UK have a correlation of "0.7", which is a good value. So the correlation data has many similarities with real life.

After rigorous training and testing of the model with the aim of minimising the possible errors, we finally reached the maximum accuracy of our model.

Omnibus:	2512.925	Durbin-Watson:	2.005
Prob(Omnibus):	0.000	Jarque-Bera (JB):	20749.749
Skew:	2.838	Prob(JB):	0.00
Kurtosis:	12.293	Cond. No.	75.6

Table -4: Model accuracy

Final accuracy of 75.6% was obtained on the testing data.

6. CONCLUSION

In conclusion, the review highlights the importance of predicting electric vehicle prices and the effectiveness of linear regression in achieving this goal. The studies reviewed in this paper show that linear regression is a reliable and accurate technique for predicting electric vehicle prices, with accuracy ranging from 78% to 86.5%.

The characteristics used in the studies, such as range, efficiency, and condition of the electric vehicles, as well as make and model, are commonly known and have been shown to be effective. However, the report also suggests that further research can explore the use of additional features and advanced techniques, such as Deep Learning, to improve the accuracy of electric vehicle price prediction.

Overall, this review provides valuable insight into the topic of electric vehicle price prediction using linear regression and highlights the potential for further research in this area.

7. REFERENCES

[1] Sameerchand Pudaruth, “Predicting the Price of Used Cars using Machine Learning Techniques,” International Journal of Information & Computation Technology, vol. 4, no. 7, pp. 753–764, 2014.

[2] Zamar Khan “Used Car Price Evaluation using three Different Variants of Linear Regression” International Journal of Critical Infrastructures, vol.1, issue 1 (Jan-March 2022).

[3] Nitis Monburinon, Prajak Chertchom, Thongchai Kaewkiriya, Suwat Rungphenung, Sabir Buya and Pitchayakit Boonpou “Prediction of Prices for Used car using Regression Models” 2018 5th International Conference on Business and Industrial Research (ICBIR), Bangkok, Thailand

[4] Muhammad Asghar, Khalid Mehmood, Samina Yasin and Zimal Mehboob Khan “Used cars price prediction using machine learning with optimal features” Pakistan Journal of Engineering and Technology, PakJET, Vol.4, No. 2, 2021

[5] A’aeshah Alhakamy, Areej Alhowaity, Anwar Abdullah Alatawi and Hadeel Alsaadi “Are Used Cars More Sustainable? Price Prediction Based on Linear Regression” MDPI Sustainability 2023, 15, 911

[6] Rodrigo S. Pereira “Simple EDA and Linear Regression” 2023