2021 5th International Conference on Electrical Engineering and Information & Communication Technology (ICEEICT)
Military Institute of Science and Technology (MIST), Dhaka-1216, Bangladesh

2021 5th International Conference on Electrical Engineering and Information & Communication Technology (ICEEICT) | 978-1-6654-9522-6/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICEEICT53905.2021.9667927

Driving Range Prediction of Electric Vehicles: A Machine Learning Approach

Shahid A. Hasib[1], Dip K. Saha[1], S. Islam[1], Mahib Tanvir[1], and Md. Shahinur Alam[1]
[1]Department of Mechatronics Engineering,
Rajshahi University of Engineering & Technology, Rajshahi-6204, Bangladesh
*shorifulislamtamal@gmail.com

***Abstract*—Due to the immense progress of green energy technology, the popularity of electric vehicle (EV) is increasing day by day. The rapid transition from internal combustion engine-based vehicle to battery-driven vehicle creates another issue that is limited storage capacity of batteries. Researchers are working hard to improve the storage capacity of battery through use of advanced materials. Meanwhile, the accurate prediction of driving range of EV has become a topic of interest for the researchers. In this paper, multiple regression machine learning algorithms are used to predict the electric vehicle range. Among the models, Multiple Linear Regression (MLR) gives the best R squared value of 0.973 and the lowest RMSE value of 39.67 in predicting the EV range. The result is compared with other machine learning models.**

***Index Terms*—Electric vehicle, range prediction, machine learning, multiple linear regression**

I. Introduction

Electric vehicles (EV) are the vehicles that run totally on electric motors powered by battery. EVs are getting more importance in the vehicle framework as a potentially effective medium for suitable transportation [1]. The necessity of more reliable, efficient and renewable medium of transport makes the development of EVs an important topic for both vehicle manufacturers and academic researchers [2]. Like other vehicles EVs do have a limited range and also takes longer time to get recharged. EVs have a range issue compared to the conventional internal combustion engines. Beside this the charging stations are not available like fuel pumps and charging station takes long time to recharge the battery [3]. That's why it is necessary to measure the range of the vehicle to maintain a proper battery management system.

Vehicle range is one of the main parameters for the EVs specially when a customer goes to buy a car. EVs are supposed to be a solution to reduce the greenhouse gas emissions and air pollution . However, electric vehicle range has been considered as a major problem in electric vehicles acceptance [4]. EV range is limited by the available energy storage and affected by various external factors such as temperature, vehicle conditions and other systems present in the vehicle. Moreover, average vehicle speed of driving cycle has impact on energy consumption. EV range prediction is very much necessary in terms of determining when a vehicle is getting maximum range and how it can be improved. Also this prediction result can be helpful to the customers to decide which car to buy as the prediction result provides a comparison between pricing and range of electric vehicles from different manufacturers. Although many reasons affect a person while buying a car and pricing is an important factor for a person to choose one vehicle over another. Total lifetime is one of the top factors affecting purchase decisions for many users [5]. So it is necessary to come to a optimal point where a vehicle can provide great range with better lifetime and affordable pricing.

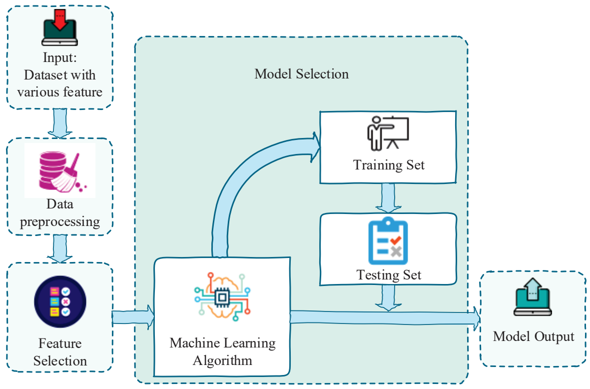

Fig. 1: Illustration of electric vehicle range prediction process using Machine Learning.

Gradient boosting decision tree (GBDT) algorithm was introduced in [6], which requires many feature variables where traditional regression can work with less. However, GBDT is novel and provides better accuracy along with reliability when predicting battery EV driving range. As stated by the authors. GBDT model can only provide importance distribution of the feature variables, not specifying the interconnection and interaction between the feature variables. A blended machine learning model was proposed in [7] to calculate the driving range of EVs considering real world historical driving data. The proposed model integrates two advanced machine learning algorithms which are Extreme Gradient Boosting Regression Tree (XGBoost) and Light Gradient Boosting Regression Tree (LightGBM). The model was prepared and trained to create a co-relation between the feature variables. Moreover, the authors stated to have the unbalanced distribution of data set eliminated by introducing an anchor based strategy.

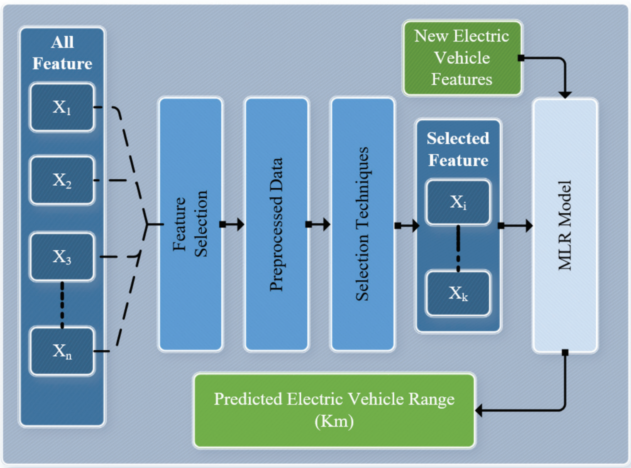

Fig. 2: Electric vehicle range prediction process flow diagram for multiple linear regression model.

In this paper, the EV range is predicted with machine learning models. Figure 1 gives the illustration of electric vehicle range prediction process with machine learning techniques. Initially the data is processed and important features are selected. When the data is ready, it was split into training and testing set. Afterwards the data is fed into the machine learning algorithm. The analysis of result is shown in graphically and also the result has been compared with other machine learning models. The variables with highest impact on predicting the EV range are also analysed. In section II the electric vehicle model and range is discussed. In the section III the proposed approach for range prediction is given. The methodology and model results are analysed in the later sections.

II. Electric Vehicle Range Prediction & Dataset Description

A. EV Range Prediction

Three main types of EVs are available which are battery electric vehicles (BEV), plug-in hybrid electric vehicles (PHEV), and hybrid electric vehicles (HEV). However, the main focus is on the BEVs and BEVs are powered only by batteries. BEVs use electric motor to run the wheels and zero emissions are recorded. The prediction approaches of EV range can be categorized into two: simulation based range prediction and data-driven range prediction method. Simulation based methods may provide higher accuracy but also needs greater model fidelity [8]. On the other hand, data-driven models are not dependent on real world models but require enough feature extraction and data preprocessing to filter out irrelevant data from the dataset [9]. However, many factors can affect the EV range such as the driverâs driving skills and habit, the vehicleâs performance and other parameters, road conditions and vehicle maintenance etc. Among all these factors very few can be accounted for. As a result, there are less amount of data processed to consider as prediction parameters.

TABLE I: Electric Vehicle Dataset Attributes

No.	Attribute	Description
1	Brand	Brand name of the EV manufacturer.
2	Model	Model name provided by the EV manufacturer.
3	Rapid Charge	Rapid charging is the fastest way to charge an EV which will charge the vehicle to 80% rapidly.
4	Power Train	Driving power train i.e. front, rear, or all wheel drive.
5	Body Style	Basic body size or style i.e. SUV, Sedan, Hatchback, Pickup.
6	Seats	Number of seats in the vehicle
7	Price (Euro)	Price in Germany in Euro
8	Acceleration (Sec)	Acceleration as 0-100 km/h
9	Top Speed (Km/h)	The top speed in km/h
10	Efficiency (Wh/Km)	Efficiency of the vehicle
11	Fast Charge (Km/h)	Fast charging capacity of the vehicle
12	Battery Pack (KWh)	battery pack of the vehicle
12	Range (Km)	The electric range provided by the vehicle

B. Dataset Description

Data do play an important role in prognostic modeling. In the data acquisition process, pocket-full of monitoring data are analyzed and collected from many sensors. Now a days, the availability of data is very much increased which can be used for software simulation [10]. Although there is rapid growth in the EV market, a few datasets are publicly available on EVs. The dataset used in this work was collected from EV Database [11]. The EV database aims to collect real world data, also claims the dataset is completely independent (not subsidized in any way), and this data was processed by [12]. The dataset is very simple and provides better support while analyzing and running visualizations. The dataset contains various attributes of Electric Vehicle manufactured by different brand. The table 1 describes the data attributes:

The table represents some information on the EV models. For a specific brand all their models don't support rapid charging. An EV can be charged in three charging methods: slow, fast and rapid. Currently, rapid charging is the fastest way to charge an electric vehicle. However, itâll only charge the vehicle to 80% rapidly, in order to preserve vehicleâs battery life, before dropping to a slower rate of charging. The aim is to select the most effective attributes among all the attributes of dataset that are described previously to establish a MLR model that allows to predict Electric Vehicle Range.

III. Methodology

A. Problem description

The task of prediction is done by forecasting the unknown output depending on the input variables values [13]. In electric vehicle the output that referred to Electric Vehicle Range is in form of kilometre (Km), numerical values (regression problem).

The multiple linear regression (MLR) method helps to determine the relationship between a dependent variable and multiple independent variables. The other regression process such as simple linear regression, support vector regression, random forest regression and others help to relate one dependent variable with one independent variable. The MLR is focused to model the corelations between one dependent variables and multiple independent variables by fitting a liner equation into the observed data. The MLR model theoretically assumes that the unit change in independent variables results in a uniform change in the output or dependent variable. The MLR model can be expressed as [14]:

$$Z = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3 + ... + \alpha_k X_k + \mu \quad (1)$$

Here, Z is the dependent variable, also known as response variable or output variable. X_n for n = 1, 2, 3âŠ...k, are the regressors or dependent variables, also known as predictors. α_0 is an intercept also known as Zâs value when every X_n equals to 0. α_n for n = 1, 2, 3âŠ...k, are the regression coefficients, α_n is the change in Z depending on unit change in X_k. μ symbolizes the change in the linear model and also known as a random error term. This is a particular observed value for Y. The dataset includes 15 features and 102 records. Figure 2 describes the EV range prediction process for MLR model. The features which were having probability score (p-value) less than 0.05 were selected. As an insignificant p value recommends that the predictor are not linked with the changes in the response.

B. Primary Analysis

The box plot was analysed to see the relations of data features in prediction of EV range. Figure 3 shows the box plot of EV Range which EV range for every encoded electric vehicle brand. The brand categorical variable was replaced with dummy variable to analyse the EV range.

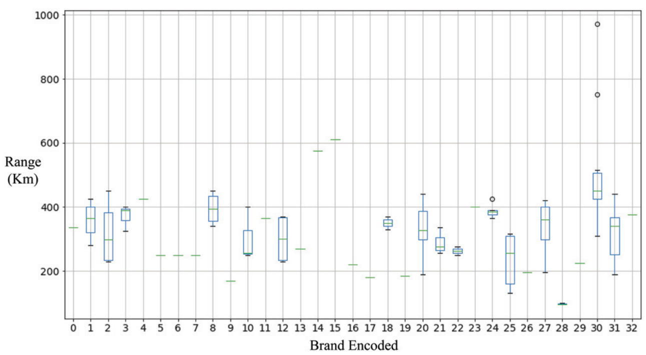

Fig. 3: Box plot of EV Range.

After configuring the outlier observations, the null and alternate hypotheses were applied. The variables were tested for the following:

Null Hypothesis H0: Population correlation coefficient can be considered as zero (no significant difference). Also no significant relation in the process between the dependent and independent variables were found.

Alternative Hypothesis Ha: Population correlation coefficient can not be considered as zero (significant difference found). Significant relation were found between the dependent and independent variables.

The hypotheses stated above can be verified using the p-value. The verification can be tested when the p-value gets less than a previously determined significance level, the null hypothesis will be rejected.

C. Principal Component Analysis

Principal component analysis (PCA) is a dimension reduction tool is used to reduce numerous variables set to a small number of set but containing all the information from the large set. It is a mathematical procedure that can transform correlated variables into uncorrelated variables called principal component. PCA is a useful mathematical technique that has can be applied in the sectors such as facial recognition and image compression. It can be used to find patterns high dimensional data. However, PCA is responsive to scaling the feature components. The first step of this work was to implement principal component analysis (PCA).

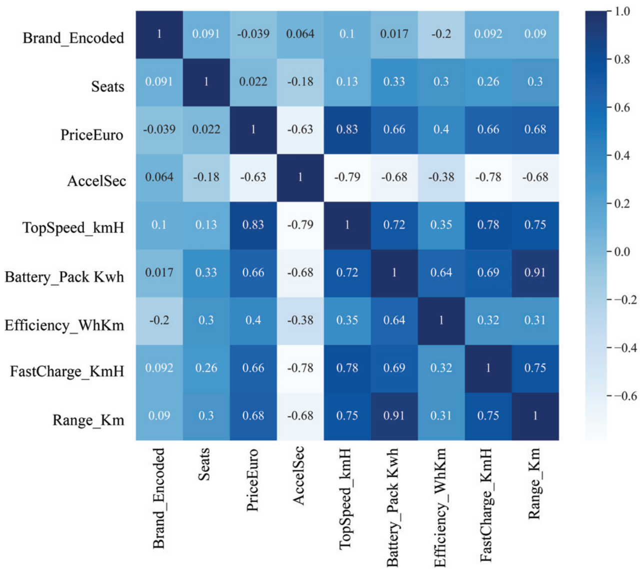

Fig. 4: EV dataset feature co-relation.

From figure 4 it can be seen that there are some good amounts of correlation between data features. From figure 5 we can see that the first principal component describes around 65% of the total variance in the data and the second component describes further 14%. Considering this, the first two components together describe about 79% of the total variance. However, in this work various machine learning (ML) models are applied and selected the most important variables. The model with the highest performance will be selected as the best model.

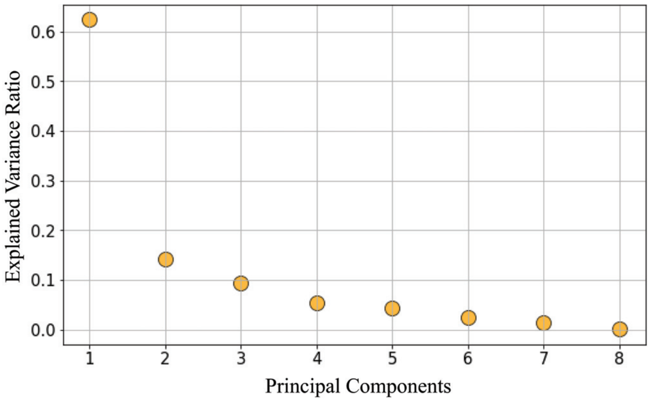

Fig. 5: Explained variance ratio of the fitted principal component vector.

TABLE II: OLS Regression Results

Dependent Variable	y
Model	OLS
No. of Observations	102
Df Residuals	93
Covariance Type	Non robust
R-squared	0.973
Adjusted R-squared	0.971
F-statistic	424.9
Log-likelihood	-453.18
AIC	924.4
BIC	948.0
Omnibus	4.382
Prob (Omnibus)	0.112
Skew	-0.464
Kurtosis	3.210
Durbin-Watson	2.018
Jarque-Bera (JB)	3.849
Prob (JB)	0.146
Condition No.	5.02

IV. Result and Discussion

The MLR model was applied and OLS regression result was analysed. If the value of p was less than 5%, the variable was kept. After running the python code specific to this step, the result was obtained and the results are discussed later in this section.

A. Cross Validation Result

From MLR model OLS regression result we obtained the R-squared value of 97.3% and adjusted R-squared value of 97.1%. After obtaining the result we applied k-fold cross validation to our dataset, the cross-validation generator value was set to 10. The cross-validation result shows the accuracy around 93% and RMSE of 39.67.

TABLE III: MLR Model Results

	Coefficient	**Standard Error**	**t**	**P>|t|**	**[0.025**	**0.975]**
Constant	338.6275	2.133	158.725	0.000	334.391	342.864
x1	-4.7656	2.187	-2.179	0.032	-9.108	-0.423
x2	-13.5350	2.934	-4.613	0.000	-19.361	-7.709
x3	-6.9614	2.333	-2.984	0.004	-11.594	-2.329
x4	5.1840	2.502	2.072	0.041	0.215	10.153
x5	23.9997	3.922	6.120	0.000	16.212	31.788
x6	119.7420	4.678	25.595	0.000	110.452	129.032
x7	-50.2471	2.972	-16.908	0.000	-56.148	-44.346
x8	14.9095	3.621	4.118	0.000	7.719	22.099

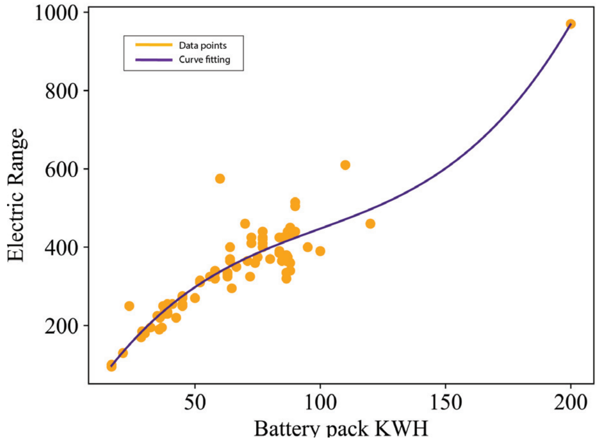

Fig. 6: Battery pack vs Electric Range (Training set) for Polynomial Regression.

B. Other Machine Learning Models

These are some other machine learning models applied to estimate the EV range.

1) Polynomial regression: The regression model was applied for the dataset and the figure 6 and figure 7 represents the result. Though the curve fits quite nicely to the data points but still many data points are far from the curve both in training and testing set. This resulted in R-squared value of 85% and RMSE value of 63.13.
2) Random forest regression: The model was applied and tested. The figure 8 and figure 9 represents the result. As it can be seen from the figure that the model fits good in the data points, resulting the second best model after MLR. Battery pack gives R squared value of 89% and RMSE of 44.35.
3) Simple linear regression: After training and testing the model, battery pack gives R squared value 82.5% and RMSE of 57.98. The figure 10 and figure 11 represents the training and testing set results respectively. The model gives a liner prediction line across the data points and as a result, it fails to cover many of the data points. However, due to linear relationship of data, the model is able to give a good result.
4) Support vector regression: Support vector machine (SVM) is a supervised learning method. Generally, this method is used for various classification and regression

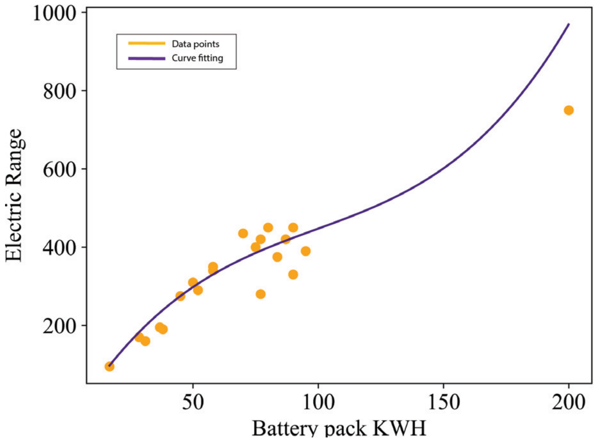

Fig. 7: Battery pack vs Electric Range (Testing set) for Polynomial Regression.

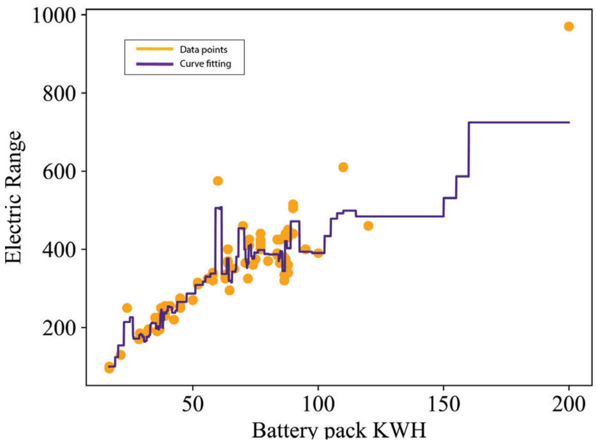

Fig. 8: Battery pack vs Electric Range (Training set) for Random Forest Regression.

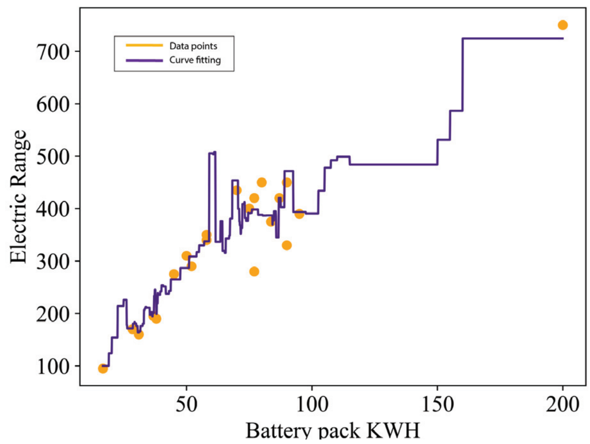

Fig. 9: Battery pack vs Electric Range (Testing set) for Random Forest Regression.

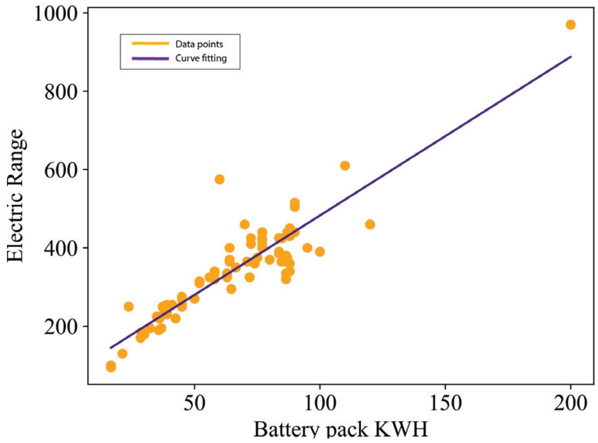

Fig. 10: Battery pack vs Electric Range (Training set) for Simple Linear Regression.

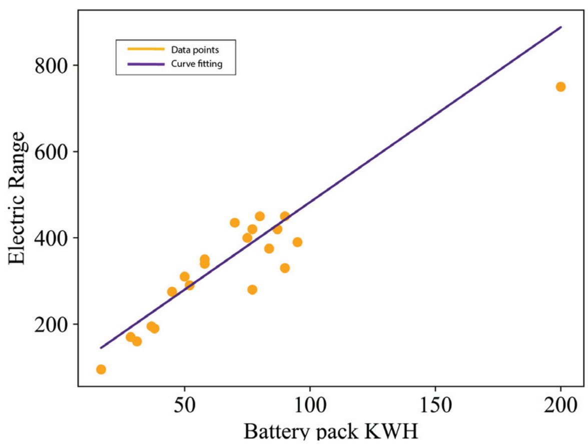

Fig. 11: Battery pack vs Electric Range (Testing set) for Simple Linear Regression.

problems. However, the benefit of SVM is its effectiveness in large dimensional spaces. It's effectiveness also can be seen in large dimensions with lower number of samples. Therefore, applying support vector regression model in the chosen dataset, battery pack gives R squared value 79.6% and RMSE of 363.4. The figure 12 shows the result for support vector regression. The model misses many points to fit the data points under the curve. This model gives a huge RMSE error for that and many data points are far from the prediction line.

The results are summarized in the table 4. Analyzing the result it can be said that this procedure can be a proper alternative to other complex methods.

TABLE IV: Comparison of Result between different ML models

ML Model Name	R-squared value	RMSE
Multiple Linear Regression	97.3%	39.67
Support Vector Regression	79.6%	363.46
Polynomial Regression	85%	63.13
Random forest Regression	89%	44.35
Simple Linear Regression	82.5%	57.98

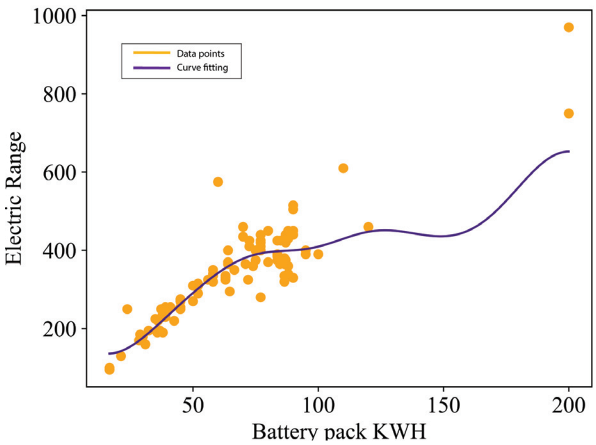

Fig. 12: Battery pack vs Electric Range for Support Vector Regression.

V. Conclusion

The main goal this work is to prepare a model that will estimate the range of electric vehicles and to provide a comparison of results between some machine learning approaches. Initially, an overview of the electric vehicles and their range was provided and methodology of the range prediction was discussed later. The EV range is predicted using various input parameters. The results of EV range prediction is also compared with other machine learning approaches. The figure 1 and figure 2 will help understand the overall prediction process. A dataset was used to predict the battery range using different machine learning models and later the results were discussed and compared. Table 1 has the dataset attributes, table 2 and table 3 visualizes the OLS regression result of the MLR model and finally table 4 compares the result between various machine learning models. In the future, research can be done on the variables correlation within the model. The results obtained in this research is satisfactory but further optimization is required to better the prediction accuracy in the days to come.

References

[1] S. Barcellona, D. De Simone, and S. Grillo, "Real-time electric vehicle range estimation based on a lithium-ion battery model," in *2019 International Conference on Clean Electrical Power (ICCEP)*. IEEE, 2019, pp. 351–357.

[2] Y. Qin, X. Tang, T. Jia, Z. Duan, J. Zhang, Y. Li, and L. Zheng, "Noise and vibration suppression in hybrid electric vehicles: State of the art and challenges," *Renewable and Sustainable Energy Reviews*, vol. 124, p. 109782, 2020.

[3] M. Mruzek, I. Gajdáč, L. Kučera, and T. Gajdošík, "The possibilities of increasing the electric vehicle range," *Procedia engineering*, vol. 192, pp. 621–625, 2017.

[4] T. Franke, I. Neumann, F. Bühler, P. Cocron, and J. F. Krems, "Experiencing range in an electric vehicle: Understanding psychological barriers," *Applied Psychology*, vol. 61, no. 3, pp. 368–391, 2012.

[5] J. Neubauer, A. Brooker, and E. Wood, "Sensitivity of battery electric vehicle economics to drive patterns, vehicle range, and charge strategies," *Journal of Power Sources*, vol. 209, p. 269â277, 07 2012.

[6] S. Sun, J. Zhang, J. Bi, and Y. Wang, "A machine learning method for predicting driving range of battery electric vehicles," *Journal of Advanced Transportation*, vol. 2019, pp. 1–14, 01 2019.

[7] L. Zhao, W. Yao, Y. Wang, and J. Hu, "Machine learning-based method for remaining range prediction of electric vehicles," *IEEE Access*, vol. 8, pp. 212 423–212 441, 12 2020.

[8] K. N. Genikomsakis and G. Mitrentsis, "A computationally efficient simulation model for estimating energy consumption of electric vehicles in the context of route planning applications," *Transportation Research Part D: Transport and Environment*, vol. 50, pp. 98–118, 2017.

[9] J. G. Hayes, R. P. R. De Oliveira, S. Vaughan, and M. G. Egan, "Simplified electric vehicle power train models and range estimation," in *2011 IEEE vehicle power and propulsion conference*. IEEE, 2011, pp. 1–5.

[10] S. Ahamed Hasib, S. Islam, R. Chakrabortty, M. Ryan, D. Saha, H. Ahamed, S. Moyeen, S. Das, M. Ali, M. Robiul Islam, Z. Tasneem, and F. Badal, "A comprehensive review of available battery datasets, rul prediction approaches, and advanced battery management," *IEEE Access*, vol. PP, pp. 1–1, 06 2021.

[11] "Ev database," Aug 2020. [Online]. Available: https://ev-database.org/

[12] "Evs-one electric vehicle dataset-smaller," Aug 2020. [Online]. Available: https://www.kaggle.com/geoffnel/evs-one-electric-vehicle-dataset

[13] O. Aissaoui, Y. Madani, L. Oughdir, A. Dakkak, and Y. EL ALLIOUI, *A Multiple Linear Regression-Based Approach to Predict Student Performance*, 01 2020, pp. 9–23.

[14] E. E. Services, *Data science and big data analytics: discovering, analyzing, visualizing and presenting data*. Wiley, 2015.