World Electric Vehicle Journal

MDPI

Article

Electric Vehicle Range Estimation Using Regression Techniques

Moin Ahmed [1], Zhiyu Mao [2,*], Yun Zheng [1], Tao Chen [3] and Zhongwei Chen [1,*]

[1] Department of Chemical Engineering, University of Waterloo, 200 University Avenue West, Waterloo, ON N2L 3G1, Canada; m63ahmed@uwaterloo.ca (M.A.); yun.zheng@uwaterloo.ca (Y.Z.)
[2] College of Chemistry and Life Sciences, Zhejiang Normal University, Jinhua 321004, China
[3] Department of Economics, University of Waterloo, 200 University Avenue West, Waterloo, ON N2L 3G1, Canada; t66chen@uwaterloo.ca
* Correspondence: zhymao@zjnu.edu.cn (Z.M.); zhwchen@uwaterloo.ca (Z.C.)

Abstract: Electric vehicles (EVs) are an attractive alternative to conventional vehicles powered by internal combustion engines due to their low carbon footprint, low running cost, and higher energy efficiency. However, currently, they suffer from a lower range than conventional vehicles, which induces range anxiety for consumers. This work explores the EV parameters that strongly impact range using data-driven techniques. A detailed dataset of the technical specifications of commercial EV models manufactured from 2008 to 2021 was collected through web mining. Strong correlations were observed between range and battery capacity, top speed, curb weight, and acceleration (with Pearson coefficients of 0.90, 0.79, 0.70, and −0.84, respectively). Furthermore, regression algorithms were trained and tested on this dataset, with the lowest root-mean-squared error (*RMSE*) of 31.4 km obtained from support vector machine regression. With a mean EV range in the test set of 364.5 km, an *RMSE* of 31.4 km equates to around 8.6% accuracy. Additionally, simple linear relationships between EV range and EV model, battery, and performance parameters were determined that may be useful to EV consumers in calculating range.

Keywords: electric vehicles; linear regression; lithium-ion batteries

check for updates

Citation: Ahmed, M.; Mao, Z.; Zheng, Y.; Chen, T.; Chen, Z. Electric Vehicle Range Estimation Using Regression Techniques. *World Electr. Veh. J.* **2022**, *13*, 105. https://doi.org/10.3390/wevj13060105

Academic Editor: Vladimir Katic

Received: 14 March 2022
Accepted: 6 June 2022
Published: 14 June 2022

1. Introduction

Recent years have seen a governmental and industrial push toward the mass adoption of electric vehicles (EVs) due to their various benefits, including lower reliance on unsustainable and non-renewable fossil fuels [1–3]. Reduced global reliance on fossil fuels for transportation could mitigate some of their disastrous effects, including greenhouse gas emissions (GHG) and geopolitical tensions [4]. However, it is important to note that GHG reduction from EV mass adoption is only possible when these electrified means of transport are incorporated into a renewable and sustainable energy grid [2,3,5].

However, despite their numerous benefits and the improvements in lithium-ion battery (LIB) technology, EVs still generally suffer from a lower range than conventional internal combustion engines, long charging times, and limited fast-charging infrastructure in some geographical areas [6]. In combination, these limitations can induce "range anxiety" for EV users; hence, EV range estimations are an ongoing academic and industrial research area [7]. EV range is described as the distance an EV can travel before recharging or battery swapping [8]. The manufacturer-reported vehicle range is measured by subjecting the vehicle to a controlled drive cycle. Commonly used drive cycles for EV range estimation include multiple cycles of the urban dynamometer driving schedule (UDDS), the highway fuel economy test driving schedule (HWFET), the new European drive cycle (NEDC), and the worldwide harmonized light vehicle test procedure (WLTP) [9,10]. Since drive cycle protocols for multiple cycles of UDDS and HWFET are recommended by the U.S. Environmental Protection Agency (EPA), this drive cycle is referred to as EPA range throughout this work. EV design and performance specifications, external climate and road conditions, and consumer driving habits are critical parameters affecting EV range [11,12]. EV design and performance specifications can include battery specifications, motor efficiency, curb weight, and vehicular body aerodynamics.

Many relevant works on EV range estimation focus on determining the remaining EV range during a driving trip using physical or empirical models [13–18]. These models calculate the energy consumption of an EV based on the external conditions (e.g., road elevation and wind velocity) and EV parameters (e.g., weight, motor, braking efficiencies, etc.) and then determine the remaining EV range [13,19]. Moreover, for more accurate EV range estimations, if the velocity–time series of an EV is known (based on the drive cycle), the required EV energy is then obtained for each time point during the EV trip. In 2011, Hayes et al. used the simplified drive train model for the range estimation of a Tesla roadster and a Nissan Leaf subjected to the EPA-based drive cycle. Their model calculates fuel economy to estimate the EV range for a given drive cycle. The authors used the published vehicle parameters for unknown parameters, while assumptions and estimations were made for the rest of the model parameters [13]. Shibata et al. developed an EV power consumption model that included the energy consumption of HVAC and other auxiliary systems during an EV trip. Furthermore, coulomb counting with the Kalman filter was used to estimate the remaining battery capacity at any point in the drive cycle. The authors of this work explored the effects of different drive cycles on EV power consumption [17]. Abousleiman et al. used a similar model to compare model predictions with the energy consumption of a 2013 Fiat 500c driven on different routes. Depending on the route, the authors reported model discrepancies in the range of 0.66–2.9% [18]. While previously reported works assumed constant braking power regeneration, Fiori et al. improved on the power consumption model by incorporating the effects of EV deceleration on the braking power regeneration. In all the above works, the velocity–time characteristics and extensive vehicle parameters were required, and several model assumptions, such as constant external wind and road-related resistances, were made [16].

These EV energy consumption models can be improved by simulating EV battery pack voltage and capacity characteristics using various battery models, such as equivalent circuit models (ECM) [19,20]. These battery models require their own model parameters. For instance, ECMs, which simulate the voltage and SOC characteristics of the battery from the drive current using an electric circuit, require parameters pertaining to the values of the circuit components [20]. For instance, Barcellona et al. used an ECM to simulate the remaining battery capacity and battery pack voltage during an EV trip while using the EV power consumption model.

While these models can be implemented for real-time EV range estimation during a trip, these models are not easy to implement for consumers and manufacturers. EV driving range is an important decision factor for consumers when purchasing an EV [21]. This work attempts to find the relationships between the most relevant EV design parameters and marketed vehicle range using simple regression models [21].

This paper focuses on data-driven regression modeling of the impacts of EV characteristics, battery, and performance specifications on vehicle range. The data used for regression model training and testing included the specifications of commercial EVs; these data were collected using web-scraping and text-mining techniques. The first objective of this study was to ascertain significant parameters that affect EV range. To this end, correlations between range and other parameters were used. The second objective was to train the regression model to predict EV range from significant parameters. The correlations and the simple equation obtained from the regression-trained model can be used by EV consumers, designers, and manufacturers to ascertain how different options available (such as an EV's top speed, body style, etc.) can impact its overall range.

2. Methodology

2.1. Data Collection and Preprocessing

Figure 1 below outlines the flowchart of the range estimation workflow used in this work.

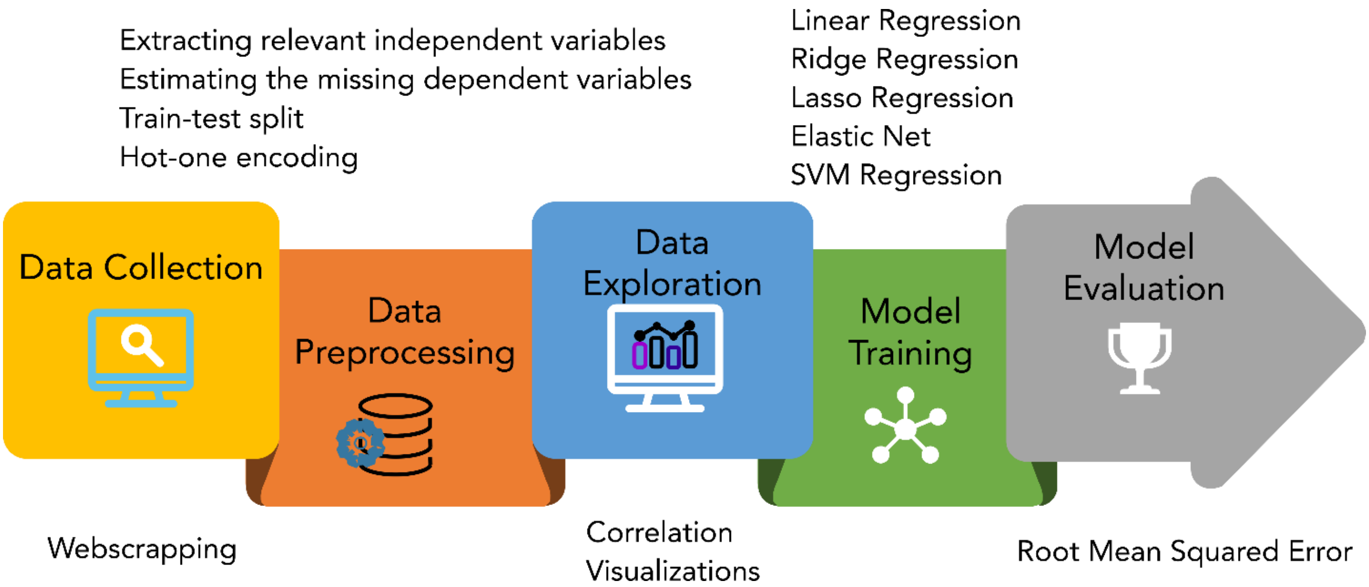

Figure 1. Flowchart of regression process.

The electric vehicle (EV) specification data were collected from a relevant website (EV Specifications) [22] through web mining (using custom web-mining code in Python). The model year of the EVs ranged from 2008 to 2021. Data collected included the EV model, battery, dimensions, and performance specifications. At the time of analysis, the website (mentioned above) was missing information on Volkswagen's e-Gold 2018 model and, hence, the information for this EV was extracted from another website (Edmunds) [23]. At that point, there were 317 EVs in the extracted dataset. Some of the information contained in the dataset (column names in the dataset are referred to as the features in this work) is outlined in Table S1. For simplicity in further analysis, only the features in the dataset that have a maximal effect on EV range and that were not sparse were considered. These features pertained to EV model information, battery characteristics, and performance metrics (Figure 2). The range estimated by multiple UDDS and HWFET range drive cycles is referred to as the EPA range in this article.

For many vehicles, EV range was expressed with reference to EPA, NEDC, or WLTP standards, with several vehicles missing EPA range values. Therefore, the EPA ranges of these vehicles were estimated using the WLTP and NEDC ranges by means of normal linear regression along with bootstrapping (Section S2). Bootstrapping is a resampling technique; within the linear regression application, it can be used for more realistic linear regression parameter estimation in cases of missing data.

Once the missing EPA ranges were estimated, the correlations between the features and the EPA range were determined by the Pearson correlation coefficient using the following equation [24]:

$$r_{xy} = \frac{\sum (x_i - \overline{x}) \sum (y_i - \overline{y})}{\sqrt{\sum (x_i - \overline{x})^2} \sqrt{\sum (y_i - \overline{y})^2}} \tag{1}$$

In our case, x and y refer to the relevant independent variable and EV range, respectively. x_i and y_i denote the individual data points, while $\overline{x}$ and $\overline{y}$ denote their mean values.

Next, a randomized train-test split of 80–20% was performed on the dataset. After further data analysis using visualizations, the independent variables that did not show much effect on EV range were discarded. Furthermore, features containing qualitative information (type of battery, EV body type, and battery cooling system) were plotted for distribution, and categories within a column were merged for simplicity. In addition, hot-code encoding was also performed on these qualitative features.

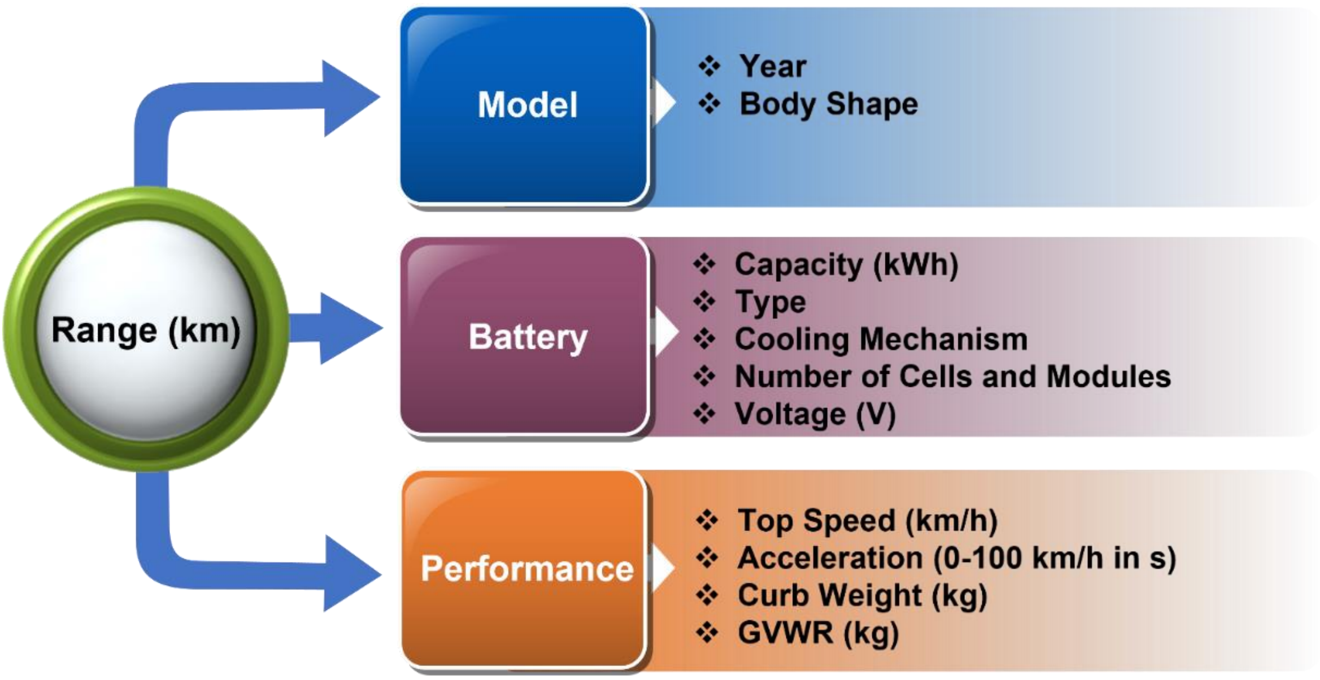

Figure 2. EV model, battery, and performance features considered from the dataset.

2.2. Regression Techniques

Regression analysis is a simple yet versatile tool for exploring and mapping relationships between variables. In this work, a multiple linear regression model was used to model the EPA range from multiple factors that affect it. Furthermore, regularized linear regression models were utilized and explored to prevent data over-fitting, namely ridge regression, lasso regression, and elastic net [25].

Support vector machines (SVMs), another popular machine learning (ML) technique, can perform classification, regression, and outlier detection. Moreover, they are applicable to small and medium datasets. When used for classification problems, the objective of an SVM is to separate different classes with the highest possible difference between them. However, when SVMs are used for regression problems, the objective changes to achieving the lowest possible difference between different classes [26]. In this work, linear regression and SVMs were explored mainly due to their suitability for small datasets. Below, a mathematical overview of linear regression and SVMs is presented.

2.2.1. Linear Regression

A linear regression model was trained to predict the EV EPA range from independent variables and can be concisely represented as [27]:

$$\hat{y} = \boldsymbol{\theta} \cdot \boldsymbol{x} \tag{2}$$

where $\hat{y}$ denotes the model predicted value, while $\boldsymbol{\theta}$ and $\boldsymbol{x}$ denote the model's parameter vector and instance's feature vector, respectively. The normal equation used to solve for $\boldsymbol{\theta}$ is given by [27]:

$$\hat{\theta} = \left(x^T x\right) x^T y \tag{3}$$

The mean squared error was used as the cost function for the linear regression and is expressed as [27]:

$$MSE = \frac{1}{m}\sum_{i=1}^{m}\left(\boldsymbol{\theta}^T \boldsymbol{x}^i - y^i\right)^2 \tag{4}$$

where m is the total number of data points. In order to prevent the possibility of over-fitting the model, regularized forms of linear regression were also trained. These forms included ridge, lasso, and elastic regressions. The cost functions for ridge, lasso, and elastic regressions are presented below, in the order listed [25]:

$$J(\boldsymbol{\theta}) = MSE(\boldsymbol{\theta}) + \alpha\left(\frac{1}{2}\right)\sum\nolimits_{i=1}^{n}\theta_i^2 \tag{5}$$

$$J(\boldsymbol{\theta}) = MSE(\boldsymbol{\theta}) + \alpha\sum\nolimits_{i=1}^{n}|\theta_i| \tag{6}$$

$$J(\boldsymbol{\theta}) = MSE(\boldsymbol{\theta}) + r\,\alpha\sum\nolimits_{i=1}^{n}|\theta_i| + \alpha\left(\frac{1-r}{2}\right)\sum\nolimits_{i=1}^{n}\theta_i^2 \tag{7}$$

where α is the regularization parameter used for ridge, lasso, and elastic net regressions, and r is the additional mix-ratio parameter used in elastic net regression.

2.2.2. Support Vector Machine

Support vector machines (SVMs) can be used for regression problems and are suitable for small sample sizes [26]. The mapping function of a dataset $\{(x_1,y_1), (x_2,y_2), \ldots, (x_n,y_n)\}$ is expressed as [28,29]:

$$f(x) = \sum_{i=1}^{n} w_i \boldsymbol{\phi}(x_i) + b \tag{8}$$

where w_i and b are the weight vector and a constant, respectively, and are found by first defining the following optimization problem. Also, the function $\boldsymbol{\phi}(x_i)$ denotes the mapping (can be non-linear for generality) in the feature space. This optimization problem ensures that the function above, *f(x)*, is as flat as possible [28,30].

$$Min\frac{1}{2}||w||^2, \tag{9}$$

$$s.t.\begin{cases} y_i - \phi(w, x_i) - b \leq \varepsilon \\ \phi(w, x_i) + b_i - y \leq \varepsilon \end{cases} \tag{10}$$

By adding slack variables (ξ_i and ξ_i^*), the optimization problem becomes [28,31]:

$$Min\frac{1}{2}||w||^2 + \sum_{i=1}^{n} C_i(\xi_i + \xi_i^*) \tag{11}$$

$$s.t.\begin{cases} y_i - \phi(w, x_i) - b \leq \varepsilon + \xi_i \\ \phi(w, x_i) + b_i - y \leq \varepsilon + \xi_i^* \\ \xi_i \xi_i^* \geq 0 \end{cases} \tag{12}$$

The solution to this problem is expressed as [28,31]:

$$f(x) = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)K(x_i, x) + b \tag{13}$$

where the variables α_i and α_i^* are the Lagrange multipliers, and the function K is the so-called kernel function. The radial basis function is used for the kernel function in this work.

2.2.3. Model Training

The models were trained using the LinearRegression and SVC classes in Python's SciKit Learn package [30]. A randomized train-test split of 20–80% was performed on the dataset, and hot-code encoding was also performed on qualitative features (battery cooling mechanisms, type of lithium-ion battery, and body style). Due to the small size of the dataset, the hyperparameters of the regression learning techniques were kept constant; hence, the dataset was not split for validation.

2.2.4. Model Evaluation

The root-mean-squared error (*RMSE*) was used to evaluate the model and is expressed by [32]:

$$RMSE = \frac{\sqrt{\sum_{i=1}^{m} \hat{y}_i - y_i}}{m} \tag{14}$$

where $\hat{y}_i$ and y_i denote the model predicted and actual values, respectively, and m is the total number of observations/data points.

3. Results and Discussion

The missing values for EPA range (EPA range refers to the range estimated from multiple cycles of UDDS and HWFET drive cycles) were estimated using the WLTP and NEDC range values. Figure 3 shows EPA range values plotted against WLTP and NEDC range values. The Pearson correlation coefficient between EPA and WLTP ranges is 0.96, while it is 0.98 between EPA and NEDC ranges. In the dataset, out of 317 entries, only 176 entries had an EPA range. In order to obtain estimates for the missing EPA range values, the bootstrapped linear regression parameters (Figure 3A,B) between EPA and WLTP ranges and between EPA and NEDC ranges were used (see Supporting Information Section S2). The bootstrap method was performed for 2000 iterations until the bootstrapped linear regression parameters (coefficient and the y-intercept) stabilized (Figure 3C,D). The relationships between the ranges are outlined in Section S4.

Next, the independent variables that were considered to have the highest impact on EPA range were retained (see the Methodology section for the specific variables considered), while other variables were removed for further consideration.

This reduction was made to simplify the analysis. Table 1 summarizes the correlations between the independent variables and the dependent variable (EPA range). Battery capacity, top speed, and curb weight seem to show strong positive correlations to EPA range, while acceleration shows a strong negative correlation. This negative correlation can be explained by the physics-based energy consumption models of an EV. When all operating conditions and vehicle parameters are held constant, higher acceleration requires higher power output from the battery pack; this drains the battery state-of-charge more quickly [16,33].

Figure 4 shows the plots for all the independent variables against EPA range. While model year seems to have a low impact on EPA range according to the correlation (Table 1) and plot (Figure 4A), it was kept for further analysis because it is well known that the range of EVs has been steadily increasing over the years [34]. This low correlation and unclear linear trend may arise from differences in EPA range between different EV body shapes (Figure 5A). Since batteries provide the energy source for the EV propulsion system and other auxiliary systems, the positive linear relationship between battery capacity and range is expected and demonstrated by the high correlation in Table 1 and Figure 4B.

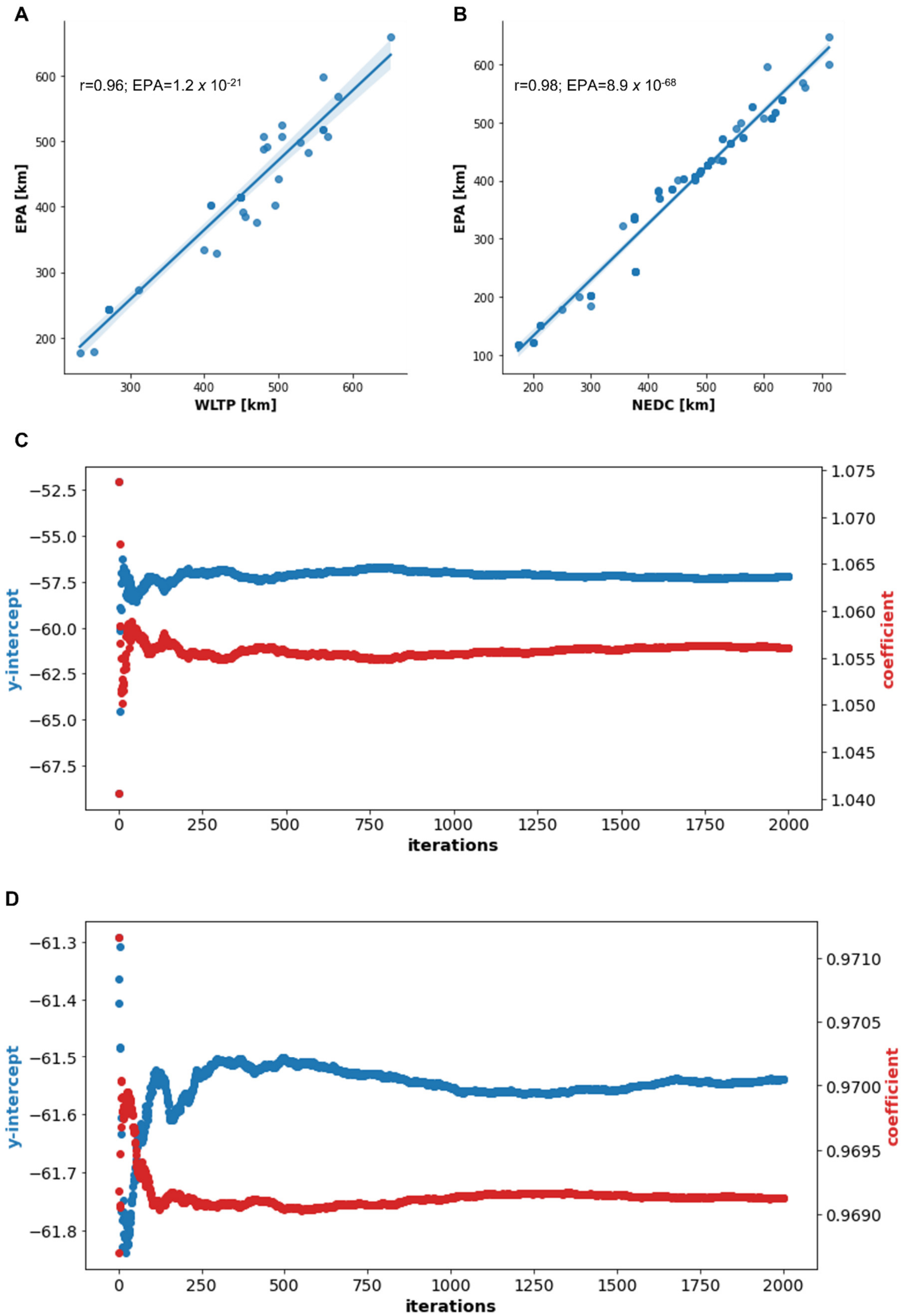

Figure 3. A plot, along with the lines of best fit and Pearson correlation coefficients between (**A**) EPA and WLTP ranges and (**B**) EPA and NEDC ranges. All units are in km. Furthermore, the bootstrapped linear regression parameters for each iteration are presented for (**C**) EPA and WLTP ranges and (**D**) EPA and NEDC ranges.

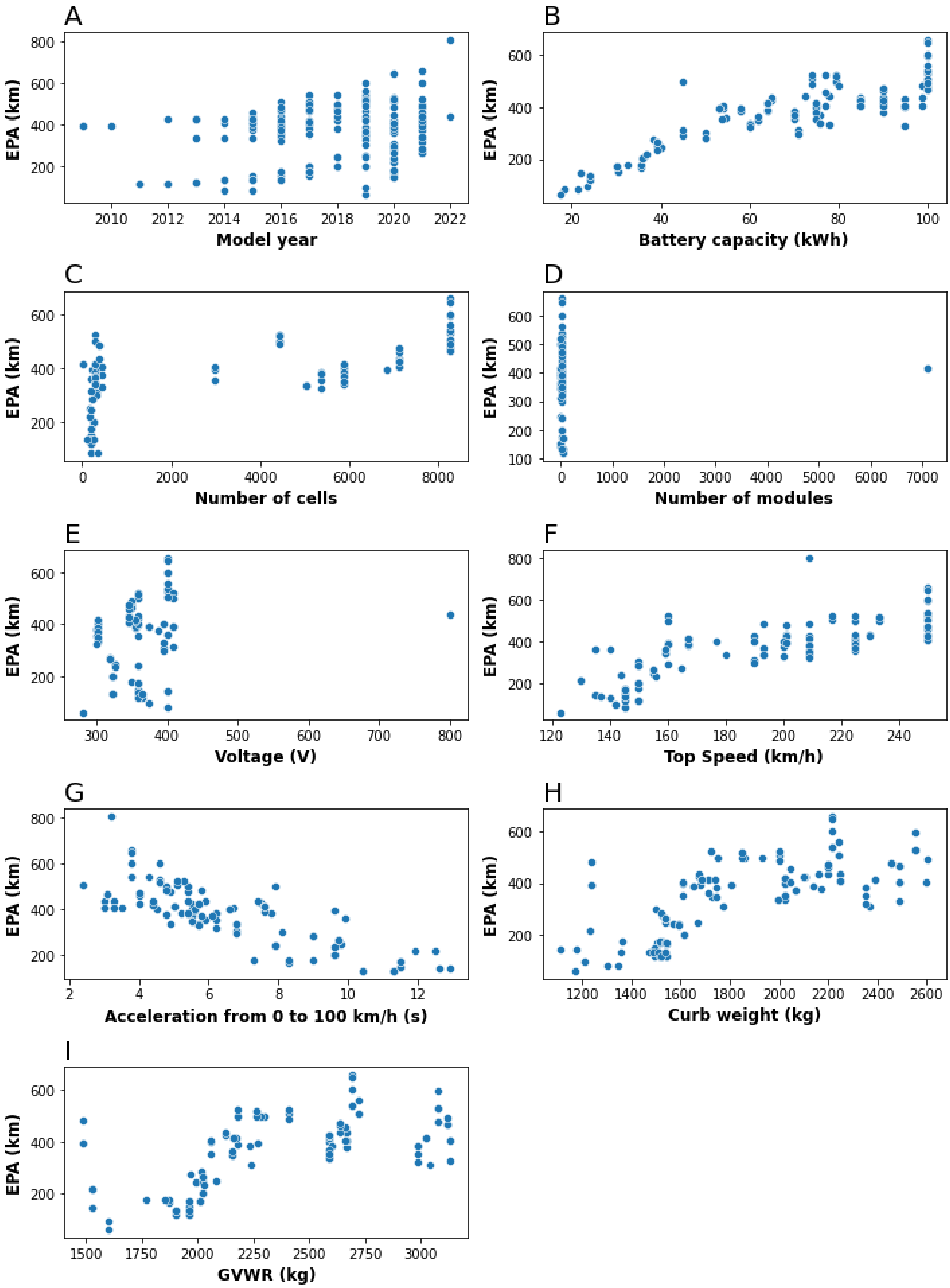

Figure 4. Plots of independent variables against EPA range, where the independent variables are (**A**) model year, (**B**) battery capacity (in KWh), (**C**) number of battery cells in an EV, (**D**) number of battery modules, (**E**) EV battery pack voltage, (**F**) EV top speed (in km/h), (**G**) EV acceleration from 0 to 100 km/h (in s), (**H**) EV curb weight (in kg), and (**I**) gross vehicle weight rating (in kg).

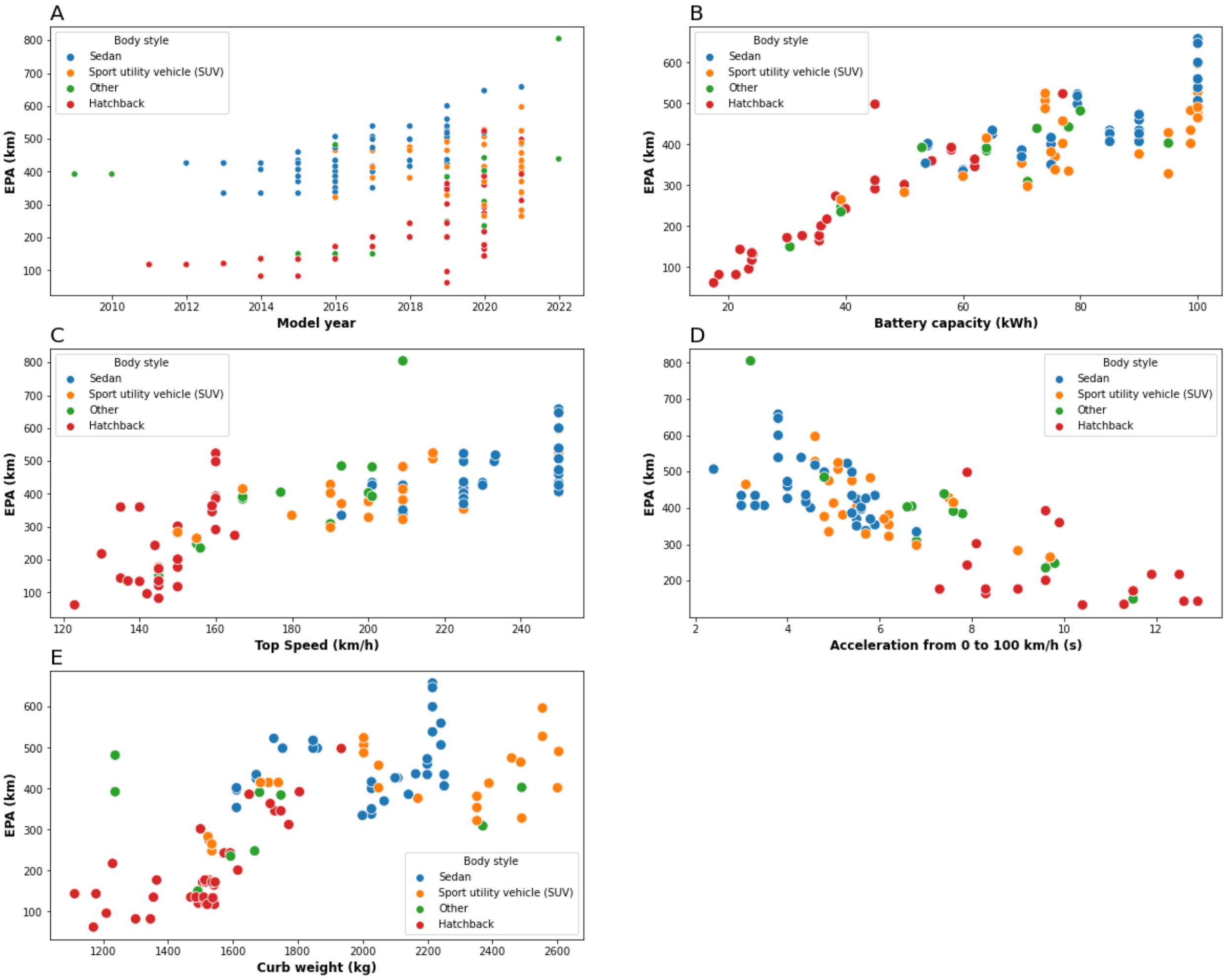

Figure 5. Scatter plot of independent variables, (**A**) EV model year, (**B**) battery capacity (in KWh), (**C**) top speed, (**D**) acceleration from 0 to 100 km/h (in s), and (**E**) EV curb weight (in kg), with EPA range, with scatter points colored differently based on EV body style.

Table 1. Correlations between independent variables and EPA.

Variables	Correlations
Battery Capacity (kWh)	0.903860
Top Speed (mph)	0.794916
Curb Weight (lb)	0.700247
Number of Battery Cells	0.689241
GVWR (lb)	0.660293
Model Year	0.262226
Voltage (V)	0.187742
Number of Battery Modules	0.041696
Acceleration from 0 to 100 km/h (s)	−0.839825

Usually, parameters for battery design, e.g., the number of battery cells and modules, correlate with single-cell capacity and voltage to obtain battery pack capacity (kilowatt-hours) for EV powertrain requirements. In this work, instead of single-cell specifications for each EV model, battery capacity is provided as an entire battery pack kilowatt-hour in EV level. Therefore, the number of cells, number of modules, and voltage show an uncertain correlation to battery capacity and have a low correlation to EPA range and an uneven distribution, as Table 1 and Figure 4C–E show separately. These parameters were not considered in further calculations or analyses.

Generally, higher battery capacity delivers higher energy. It is reasonable to expect that EVs with an increased battery capacity will have a higher top speed and acceleration, as Figure S1A,B shows; this is further supported by a higher EPA range, as indicated in Table 1 and Figure 4F,G. Moreover, EVs with higher battery capacity also have higher curb weights and GVWRs, as Figure S1C,D shows. This explains the positive correlations between the EPA range and both curb weight and GVWR (Table 1 and Figure 4H,I).

In order to understand the impact of EV body shape on EPA range, plots of various independent variables against EPA range with data points colored differently based on body shape are provided (Figure 5). A general trend, where sedans have the highest and hatchbacks have the lowest EPA range, is observed. Sedans generally have higher battery capacities than hatchbacks (Figure S1B–D and Figure 6B), which might explain the EPA range difference. Furthermore, hatchbacks experience more drag forces from their increased weight compared to sedans (the weight difference between hatchbacks and sedans is shown in Figure 6E) when external conditions are held constant [35]; thus, additional battery capacity is utilized to overcome this drag.

Figure 6 presents the distributions (histogram) and the impact on EPA range (box plot) for qualitative information. It should be noted that some categories within these features were merged to even out the distributions. For example, there were nine categories of battery cooling mechanisms, eight of which were merged into a category 'Other'. Similarly, eight categories of EV body types were condensed into four categories. From the box plot, it can be seen that each category has a noticeable difference in mean and range of EPA range values. Consistent with previous observations in Figure 5 above, the median and interquartile range (IQC) for EV range with respect to body style follows the following order: IQC(Sedan) > IQC(SUV) > IQC(Other) > IQC(Hatchback), as observed in Figure 6A. EVs with water-based cooling mechanisms have higher medians and IQCs than other circulation mechanisms, which consist mainly of air-based cooling systems. This thermal observation is consistent with the literature [36]. Furthermore, given that the IQC for EVs using lithium-ion and lithium-polymer batteries is roughly the same, and the data are unbalanced (i.e., there are many more data points for lithium-ion batteries than for lithium-polymer), it is difficult to ascertain the effect of different lithium-ion battery types on the EV range from Figure 6C alone. However, based on the linear regression equation (Equation (S6)), using lithium-ion batteries has a detrimental effect on EV range compared to lithium-polymer batteries.

Based on the findings from data exploration, the independent variables, such as battery capacity, model year, top speed, acceleration, curb weight, GVWR, body style, battery cooling mechanism, and type of lithium battery, were used for model training and evaluation. The models were trained on the training set, and the features were preprocessed as outlined in the methods section. Table 2 shows the model parameters used in this work.

Table 3 shows the *RMSE* values obtained by model evaluation on the test set. It can be seen that SVM regression demonstrates the lowest *RMSE* of 31.428 km. The mean EPA range on the test set was around 364.5 km, and an error of 31.428 km equates to around 8.6% accuracy. Interestingly, the *RMSE* increases to 32.13 km when regularization in the SVM model increases (C = 100). While *RMSE* was considered as the evaluation metric in this work, other metrics (i.e., mean absolute error (MAE), mean absolute percent error (MAPE), and R-squared (R^2) error) were also calculated and are presented in Section S7. Similar to the trend observed in the *RMSE* model evaluations, SVM regression outperforms other regression methods in MAE and MAPE evaluations (Table S2). The R^2 error is a measure of the portion of variability in the dependent variable predicted from the independent variable, and when its value is close to 1, it indicates a good fit between the model and the data [37,38]. The R^2 value was almost 0.9 (Table S2), suggesting that the models can predict the variability of EV range from the independent variables.

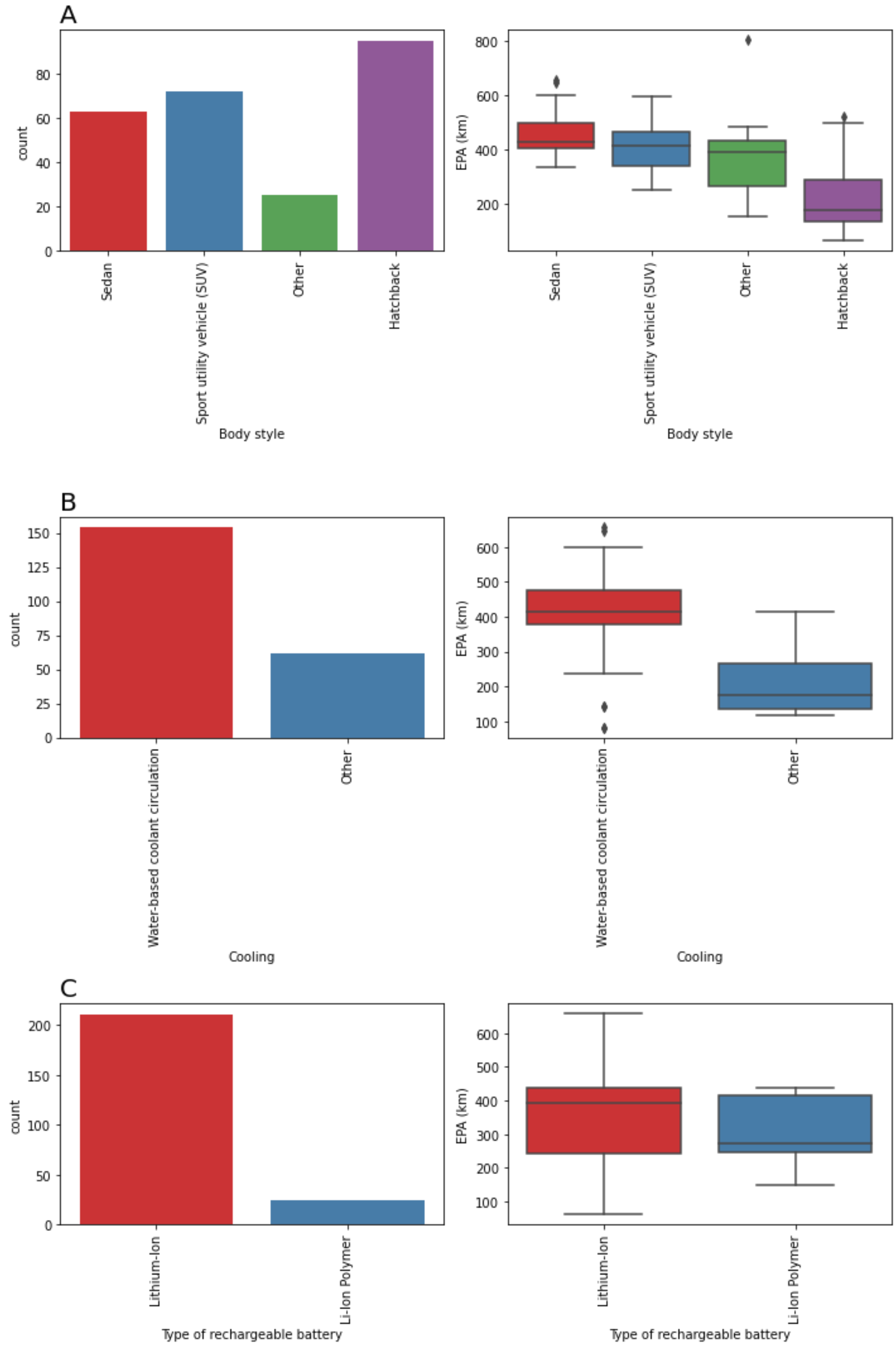

Figure 6. Histograms and boxplots of (**A**) EV body style, (**B**) battery cooling mechanism, and (**C**) type of lithium-ion battery used. The histograms plot the frequency (referred to as the count) of data points against the qualitative independent variables. Boxplots display the average and quartiles of the EPA range against the qualitative independent variables.

The equation for estimating the EPA range of an EV using linear regression (without standardization with an *RMSE* of 33.20 km) is given in Section S6. It should be noted that while the proposed linear regression equation yields a higher error rate than estimating the EV range using a power consumption model in the literature (8.6% vs. approximately 1%), it requires significantly fewer parameters. This regression equation requires eight

EV parameters that can be easily found in the literature or EV manuals, while the energy consumption model requires more than 20 EV model-specific parameters [16–18].

Table 2. The ML model parameters.

Model	Parameters
Normal Linear Regression	-
Ridge Regression	$\alpha = 1$
Lasso Regression	$\alpha = 0.1$
Elastic Net	$\alpha = 0.1$ $r = 0.5$
SVM Regression	$\varepsilon = 0.1$ $C = 1$

Table 3. *RMSE*s from model evaluations.

Model	*RMSE*
Normal Linear Regression	33.200
Ridge Regression	33.540
Lasso Regression	33.084
Elastic Net	37.423
SVM Regression	31.428

4. Conclusions

EVs have several potential merits, such as low- environmental impact and low operating cost, but they suffer from a lower range than conventional vehicles. This work attempted to explore the EV design specifications that influence range using data-driven techniques. A detailed dataset for commercial EV models manufactured from 2008 to 2021 was generated by extracting information from the internet using web-scraping techniques. This dataset contained information about EV model, battery, dimensions, performance specifications, and range. EPA range was considered in this study; in cases with missing EPA ranges, a line of best fit was utilized to estimate it from NEDC and WLTP ranges.

Parameters (such as battery capacity, battery voltage, top speed, and curb weight) that seem to have the most significant effect on range were retained (refer to Table 2). Battery capacity, top speed, and curb weight showed strong positive correlations with range, while acceleration showed a strong negative correlation. Furthermore, other parameters were dropped from further analysis and calculations based on low correlations, outliers, and poor apparent relationship with range. Moreover, categories within parameters (body style, EV battery cooling mechanism, and LIB type) were merged to distribute the categories and prevent over-fitting on one category evenly. Next, linear regression and SVR machine learning techniques were employed for model training and testing. Normal linear regression models seem to have the lowest *RMSE* of 31.428 km.

Furthermore, the multi-linear regression model equation, without a standardization step, can estimate an EV's EPA range from its battery, design, and performance specifications. Newer EV model year and higher battery capacity seem to impact EPA range positively. From this equation, EVs with water-based cooling mechanisms have higher EPA ranges than those with other cooling mechanisms. Furthermore, EVs with sedan and SUV body styles have higher EPA ranges when all other design, battery, and performance specifications are kept constant.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/wevj13060105/s1, Table S1: Information about the dataset used in this work; Figure S1: Scatter plots exploring different relations in the dataset. (A) top speed against its EPA range with different battery capacities, (B) acceleration against battery capacity with different body styles, (C) curb weight against battery capacity with different body styles, and (D) GVWR against battery capacity with different body styles; Table S2: Model evaluation from MAE, MAPE, and R^2.

Author Contributions: Conceptualization, M.A.; Methodology, M.A. and T.C.; Software, M.A.; writing—original draft preparation, M.A.; writing—review and editing, Z.M. and Y.Z.; supervision, Z.C. and T.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Dataset used in this work is available at https://github.com/m0in92/EV_dataset (accessed on 14 March 2022).

Conflicts of Interest: The authors declare no competing interest.

References

1. Rietmann, N.; Hügler, B.; Lieven, T. Forecasting the Trajectory of Electric Vehicle Sales and the Consequences for Worldwide CO_2 Emissions. *J. Clean. Prod.* **2020**, *261*, 121038. [CrossRef]
2. Ellingsen, L.A.W.; Singh, B.; Strømman, A.H. The Size and Range Effect: Lifecycle Greenhouse Gas Emissions of Electric Vehicles. *Environ. Res. Lett.* **2016**, *11*, 054010. [CrossRef]
3. Ma, H.; Balthasar, F.; Tait, N.; Riera-Palou, X.; Harrison, A. A New Comparison between the Life Cycle Greenhouse Gas Emissions of Battery Electric Vehicles and Internal Combustion Vehicles. *Energy Policy* **2012**, *44*, 160–173. [CrossRef]
4. Perkins, J.H. *Changing Energy: The Transition to a Sustainable Future*; University of California Press: Oakland, CA, USA, 2017; ISBN 9780520287792.
5. Wu, Z.; Wang, M.; Zheng, J.; Sun, X.; Zhao, M.; Wang, X. Life Cycle Greenhouse Gas Emission Reduction Potential of Battery Electric Vehicle. *J. Clean. Prod.* **2018**, *190*, 462–470. [CrossRef]
6. Adnan, N.; Nordin, S.M.; Rahman, I.; Vasant, P.M.; Noor, A. A Comprehensive Review on Theoretical Framework-Based Electric Vehicle Consumer Adoption Research. *Int. J. Energy Res.* **2017**, *41*, 317–335. [CrossRef]
7. Ahmed, M.; Zheng, Y.; Amine, A.; Fathiannasab, H.; Chen, Z. The Role of Artificial Intelligence in the Mass Adoption of Electric Vehicles. *Joule* **2021**, *5*, 2296–2322. [CrossRef]
8. Good, D. EPA Test Procedures for Electric Vehicles and Plug-in Hybrids DRAFT Summary-Regulations Take Precedence. 2017. Available online: www.fueleconomy.gov (accessed on 14 March 2022).
9. Mock, P.; Kühlwein, J.; Tietge, U.; Franco, V.; Bandivadekar, A.; German, J. The WLTP: How a New Test Procedure for Cars Will Affect Fuel Consumption Values in the EU. *Int. Counc. Clean Transp.* **2014**, *9*, 3547. [CrossRef]
10. How Does The EPA Calculate Electric Car Range? | CleanTechnica. Available online: https://cleantechnica.com/2020/08/18/how-does-epa-calculate-electric-car-range/ (accessed on 11 September 2021).
11. Mruzek, M.; Gajdáč, I.; Kučera, Ľ.; Barta, D. Analysis of Parameters Influencing Electric Vehicle Range. In *Procedia Engineering*; Elsevier Ltd.: Amsterdam, The Netherlands, 1 January 2016; Volume 134, pp. 165–174.
12. Rastani, S.; Yüksel, T.; Çatay, B. Effects of Ambient Temperature on the Route Planning of Electric Freight Vehicles. *Transp. Res. Part D Transp. Environ.* **2019**, *74*, 124–141. [CrossRef]
13. Hayes, J.G.; de Oliveira, R.P.R.; Vaughan, S.; Egan, M.G. Simplified Electric Vehicle Power Train Models and Range Estimation. In Proceedings of the 2011 IEEE Vehicle Power and Propulsion Conference, VPPC 2011, Chicago, IL, USA, 6–9 September 2011; IEEE: Piscataway, NJ, USA, 2011. [CrossRef]
14. Bolovinou, A.; Bakas, I.; Amditis, A.; Mastrandrea, F.; Vinciotti, W. Online Prediction of an Electric Vehicle Remaining Range Based on Regression Analysis. In Proceedings of the 2014 IEEE International Electric Vehicle Conference, IEVC 2014, Florence, Italy, 17–19 December 2014; IEEE: Piscataway, NJ, USA, 2014. [CrossRef]
15. Minett, C.F.; Salomons, A.M.; Daamen, W.; Van Arem, B.; Kuijpers, S. Eco-Routing: Comparing the Fuel Consumption of Different Routes between an Origin and Destination Using Field Test Speed Profiles and Synthetic Speed Profiles. In Proceedings of the 2011 IEEE Forum on Integrated and Sustainable Transportation Systems, FISTS 2011, Vienna, Austria, 29 June–1 July 2011; IEEE: Piscataway, NJ, USA, 2011; pp. 32–39. [CrossRef]
16. Fiori, C.; Ahn, K.; Rakha, H.A. Power-Based Electric Vehicle Energy Consumption Model: Model Development and Validation. *Appl. Energy* **2016**, *168*, 257–268. [CrossRef]
17. Shibata, S.; Nakagawa, T. Mathematical Model of Electric Vehicle Power Consumption for Traveling and Air-Conditioning. *J. Energy Power Eng.* **2015**, *9*, 269–275. [CrossRef]
18. Abousleiman, R.; Rawashdeh, O. Energy Consumption Model of an Electric Vehicle. In Proceedings of the 2015 IEEE Transportation Electrification Conference and Expo, ITEC 2015, Dearborn, MI, USA, 14–17 June 2015; IEEE: Piscataway, NJ, USA, 2015. [CrossRef]

19. Barcellona, S.; de Simone, D.; Grillo, S. Real-Time Electric Vehicle Range Estimation Based on a Lithium-Ion Battery Model. In Proceedings of the ICCEP 2019—7th International Conference on Clean Electrical Power: Renewable Energy Resources Impact, Otranto, Italy, 2–4 July 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 351–357. [CrossRef]
20. Miri, I.; Fotouhi, A.; Ewin, N. Electric Vehicle Energy Consumption Modelling and Estimation—A Case Study. *Int. J. Energy Res.* **2021**, *45*, 501–520. [CrossRef]
21. Kim, S.; Lee, J.; Lee, C. Does Driving Range of Electric Vehicles Influence Electric Vehicle Adoption? *Sustainability* **2017**, *9*, 1783. [CrossRef]
22. EVSpecifications-Electric Vehicle Specifications, Electric Car News, EV Comparisons. Available online: https://www.evspecifications.com/ (accessed on 11 September 2021).
23. Used 2018 Volkswagen E-Golf Specs & Features | Edmunds. Available online: https://www.edmunds.com/volkswagen/e-golf/2018/features-specs/ (accessed on 11 September 2021).
24. Rodgers, J.L.; Nicewander, W.A. Thirteen Ways to Look at the Correlation Coefficient. *Am. Stat.* **1988**, *42*, 59. [CrossRef]
25. Geron, A. Regularized Linear Models. In *Hands-On Machine Learning with Scikit-Learn and TensorFlow*; O Reilly: Springfield, MO, USA, 2019; pp. 134–141.
26. Geron, A. Support Vector Machines. In *Hands-On Machine Learning with Scikit-Learn and TensorFlow*; O Reilly: Springfield, MO, USA, 2019; pp. 153–174.
27. Geron, A. Linear Regression. In *Hands-On Machine Learning with Scikit-Learn and TensorFlow*; O Reilly: Springfield, MO, USA, 2019; pp. 112–117.
28. Gao, C.; Bompard, E.; Napoli, R.; Cheng, H. Price Forecast in the Competitive Electricity Market by Support Vector Machine. *Phys. A Stat. Mech. Its Appl.* **2007**, *382*, 98–113. [CrossRef]
29. Gu, J.; Zhu, M.; Jiang, L. Housing Price Forecasting Based on Genetic Algorithm and Support Vector Machine. *Expert Syst. Appl.* **2011**, *38*, 3383–3386. [CrossRef]
30. Scikit-Learn: Machine Learning in Python—Scikit-Learn 0.24.2 Documentation. Available online: https://scikit-learn.org/stable/ (accessed on 8 June 2021).
31. Vapnik, V. SVM for Estimating Regression Function. In *The Nature of Statistical Learning Theory*; Springer: Berlin, Germany, 1995; pp. 183–188.
32. Ma, Y.; Zhang, Z.; Ihler, A.; Pan, B. Estimating Warehouse Rental Price Using Machine Learning Techniques. *Int. J. Comput. Commun. Control* **2018**, *13*, 235–250. [CrossRef]
33. Plett, G. Simulating Battery Packs. In *Battery Management Systems, Volume 2*; Artech House: Norwood, MA, USA, 2015; pp. 31–67.
34. CER—Market Snapshot: Average Electric Vehicle Range Almost Doubled in the Last Six Years. Available online: https://www.cer-rec.gc.ca/en/data-analysis/energy-markets/market-snapshots/2019/market-snapshot-average-electric-vehicle-range-almost-doubled-in-last-six-years.html (accessed on 3 June 2021).
35. Yakkundi, V.; Mantha, S.; Sunnapwar, V. Hatchback versus Sedan—A Review of Drag Issues. *Int. J. Mech. Eng.* **2017**, *4*, 5–13. [CrossRef]
36. Lu, M.; Zhang, X.; Ji, J.; Xu, X.; Zhang, Y. Research Progress on Power Battery Cooling Technology for Electric Vehicles. *J. Energy Storage* **2020**, *27*, 101155. [CrossRef]
37. Chicco, D.; Warrens, M.J.; Jurman, G. The Coefficient of Determination R-Squared Is More Informative than SMAPE, MAE, MAPE, MSE and *RMSE* in Regression Analysis Evaluation. *PeerJ Comput. Sci.* **2021**, *7*, e623. [CrossRef] [PubMed]
38. Saunders, L.; Russell, R.; Crabb, D. The Coefficient of Determination: What Determines a Useful R2 Statistic? *Investig. Ophthalmol. Vis. Sci.* **2012**, *53*, 6830–6832. [CrossRef] [PubMed]