INSTITUTO SUPERIOR DE ENGENHARIA DE LISBOA

Departamento de Engenharia Eletrónica e de Telecomunicações e Computadores

Electric Vehicle X Driving Range Prediction – EV X DRP

David Alexandre Sousa Gomes Albuquerque

Licenciado

Projecto Final para obtenção do Grau de Mestre
em Engenharia Informática e de Computadores

Orientadores : Doutor David Pereira Coutinho
Doutor Artur Jorge Ferreira

Júri:

Presidente: Doutor Nuno Datia

Vogais: Doutor Gonçalo Duarte
Doutor David Pereira Coutinho
Doutor Artur Jorge Ferreira

September, 2022

Abstract

The electric vehicle use as a reliable and eco-friendly means of transportation has increased rapidly over the past few years. When choosing an electric vehicle, its driving range capacity is a decisive factor to be taken into account as it minimizes drive's anxiety while driving. An electric vehicle driving range depends on multiple factors that must be taken into account when attempting its prediction. Machine learning has become a widely used approach for highly complex problems, in which eRange prediction, being one of them, providing better accuracy enough solutions for the problem. This thesis compares, with standard metrics, implementations of machine learning based regression models when training with publicly available datasets.

Keywords: electric vehicle; range prediction; energy consumption; machine learning

Contents

List of Figures

Glossary

Average Energy Consumption The average consumption of energy of an electric vehicle by distance .

Big data Big data is a field that handles with large datasets that are too big and complex for traditional data processing .

Bushy tree A decision tree that has high variance results from high attribute splitting on with many values .

Dataset A structure containing data for a model.

Decision tree A supervised machine learning method for regression and classification, splitting the data into multiple branches .

Electric range Electric range, the maximum driving range of an electric vehicle using only power from its on-board battery pack to traverse a given driving cycle .

Electric vehicle A vehicle that depends on electric motors for its movement .

Ensemble .

Ensemble stacked generalization .

Extreme gradient boosting Also known as XGBoost is an open-source library that provides an efficient and effective implementation of the gradient boosting algorithm .

Full Battery Distance

.

Full Battery Energy

.

Gradient boosted regression tree The use of multiple regression tree models in which, each model predicts the error from the previous model .

K-means clustering

.

K-nearest neighbor

.

Light gradient boosted machine

.

Linear regression

.

Machine learning A branch of AI focused on learning from data.

Meaningful knowledge extraction

.

Multiple linear regression A statistical technique that is used to predict the outcome of a variable based on the value of more than one variables .

Neural network A collection of connected synapse nodes, simulating a biological brain .

Python A high level programming language.

Random forest .

Regression tree Decision tree that is used for the task of regression which can be used to predict continuous valued outputs instead of discrete outputs. .

Reinforcement learning An area of machine learning focused on managing intelligent agents that take actions with the intent of maximizing a reward .

Self-organizing map .

State of charge .

Time series A series of data points indexed by time.

Acronyms

AEC	Average Energy Consumption.
DT	Decision tree.
ERange	Electric range.
EV	Electric vehicle.
EVA	Electric Vehicles in Action.
FBD	Full Battery Distance.
FBE	Full Battery Energy.
GBRT	Gradient boosted regression tree.
JARI	Japan Automobile Research Institute.
LightGBM	Light gradient boosted machine.
MAE	Mean Absolute Error.
MAPE	Mean Absolute Percentage Error.
MLR	Multiple linear regression.
MSE	Mean Squared Error.
NDANEV	National Big Data Alliance of New Energy Vehicles.
R2	Coefficient of determination.

RF	Random forest.
RMSE	Root Mean Squared Error.
RT	Regression tree.
SOC	State of charge.
SOM	Self-organizing map.
XGBoost	Extreme gradient boosting.

1

Introduction

On today's day and age, the global concern on climate change has been a major focus on recent international agreements, such as the Paris Agreement (*Paris Agreement* 2015), incentivizing many car manufacturers to introduce electric vehicles (EVs) as the eco-friendly solution for sustainable transport for the future.

EVs have grown popularity in recent years and as a result, car manufacturers have increased competitiveness on vehicle's performance (Figenbaum *et al.*, 2015), namely the driving range capacity, as it is a decisive factor for consumers (Egbue and Long, 2012).

The EV's autonomy, designated here as electric range (eRange), allows consumers to know an estimate of the remaining driving distance for the existing EV battery power, easing driver's anxiety for the duration of a trip to a charging station (Smuts *et al.*, 2017; Song and Hu, 2021).

The eRange can be estimated through many driving data parameters, such as vehicle design, driver's behavior, weather, road inclination and state of charge (SOC) estimation and its accuracy allows consumers to rely on its vehicle for longer travel time and efficient charging plans. However, eRange estimation is a complex problem with multiple influencing factors (Varga *et al.*, 2019), fueling previous studies in the past to provide a solution for this challenge.

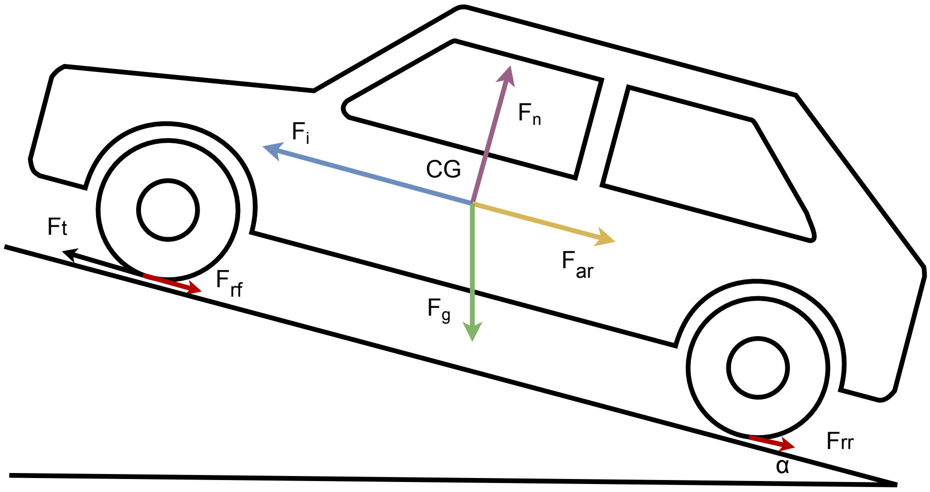

Figure 1.1: Main influencing forces on a moving vehicle. (F_i , inertial force; F_t , tractive force; F_g , gravitational force; F_{rr} , rear rolling resistance force; F_{fr} , front rolling resistance force; F_{ar} , aerodynamic (air) drag; F_n, normal force; CG, center Figure; α, the road slope)

1.1 Machine learning

The rise in popularity of machine learning (Amershi *et al.*, 2019) has demonstrated its effectiveness in the past with a variety of fields such as big data (Condie *et al.*, 2013; Zhou *et al.*, 2017), pattern recognition analysis and data mining (Bose and Mahapatra, 2001). This is due to its nature of learning from previous data to gradually achieve better results making it a widely recognized tool for complex problems (Mitchell, 2006).

As a result, machine learning has been chosen in the past as one of the adopted solutions for the eRange estimation problem, as it will be introduced in section 2.3, making it a more accurate solution.

refence state of art eRange + Machine learning approaches

1.2 The approach

This project addresses the eRange estimation problem through the use of a machine learning based model, being comprised by three distinct phases: the Dataset generation phase, the Learning phase and the Estimation phase.

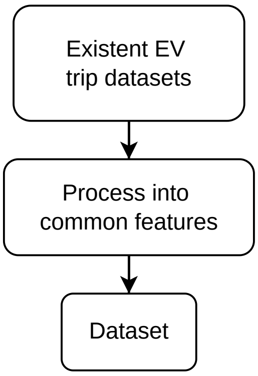

Figure 1.2: System overview - Dataset generation.

On the Dataset generation phase (Figure 1.2) a dataset will be created from historical traffic data from personally recorded vehicle trips, as well as external existing datasets such as *VED* (Oh *et al.*, 2019) or *Emobpy*, an external library that provides the generation of EV trip and consumption time series (Gaete-Morales *et al.*, 2021). The resulting dataset contains multiple trips with their respective vehicle power consumption [kW] and the vehicle speed [km/h] in a time series format.

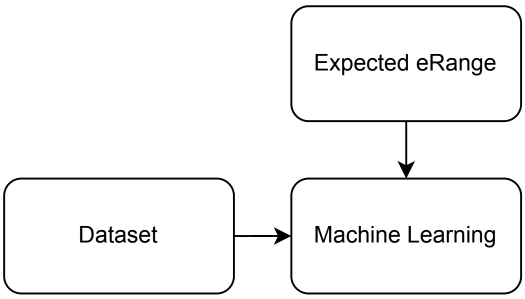

Figure 1.3: System overview - Learning phase.

The generated dataset will then be used to train the model on the Learning phase through machine learning, allowing it to learn the estimation of the eRange for each trip on the dataset as depicted in Figure 1.3.

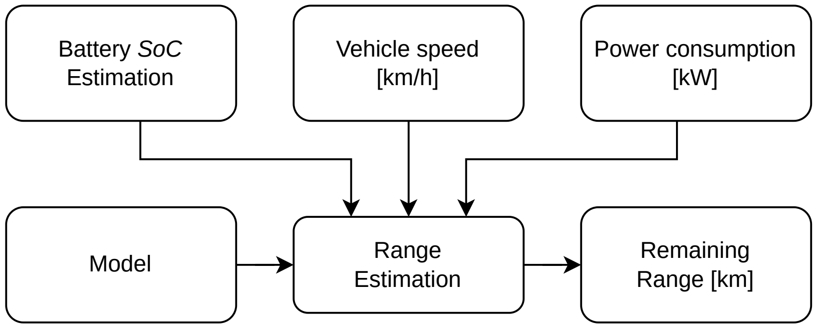

Figure 1.4: System overview - Estimation phase.

After training and learning the model with the dataset, the Estimation phase performs the eRange prediction on live SOC monitoring of a driving EV (Figure 1.4).

However, as the comparison between existing models is an unavoidable necessity to ascertain the model effectiveness, a Python application was also developed with the goal to demonstrate the different results between the implemented model and the different existing solutions provided by other authors, giving the user training options for the different selected algorithms.

The remainder of this document is structured as follows: Chapter 2 refers the state of the art on existing eRange estimation solutions and their reliance on available datasets while mentioning machine learning and its usage on existing eRange estimation solutions. Chapter 3 will focus on detailing the resulting application and model implementation on this project's solution. On Chapter 4 it is presented the project's main implemented features and observed results. The last chapter 5 will end the document with a retrospective on the project development and its accomplishments regarding its main goals.

2

State of the Art

Nowadays, EVs have motivated multiple studies concerning related problems in this field, such as statistical measurement of charging (Brighente *et al.*, 2021), regenerative braking (Yoong *et al.*, 2010), charging topologies (Yilmaz and Krein, 2013) and eRange prediction (Varga *et al.*, 2019).

The eRange prediction is an important EV feature to present to consumers as it reduces driver's anxiety while driving. This has been previously studied before, prompting multiple ways to tackle the problem.

When proving a solution to the eRange prediction problem, valid EV driving data in the form of a dataset is required to learn and evaluate the proposed model, and compare it to existing alternatives, making it indispensable in determining the effectiveness of the chosen solution.

This chapter will detail the literature and resources for the eRange prediction problem, the availability of public datasets as model training data (Section 2.1), eRange estimation solutions without machine learning (Section 2.2) and finally machine learning and its applicability in eRange prediction solutions (Section 2.3).

2.1 Datasets

To this end, existing datasets are available for use in supplying the needed training features for an eRange prediction model, some of them being instant such: as SOC; speed; acceleration; battery energy consumption and road inclination (or elevation). While others being constant for the duration of the trip: vehicle information as battery capacity, AEC (Average Energy Consumption), max eRange, weight; trip information such as commute type city or highway; total energy consumption and total distance.

Datasets like *VED* dataset (Oh *et al.*, 2019) that although providing 54 different EV driving trip data for estimation, lack trip and vehicle information as well as sufficient EV model variety, with only three distinct EVs present on its dataset, all from the same model 2013 *Nissan leaf*.

The *Charge Car* project of the CREATE Lab at Carnegie Mellon University (*ChargeCar Database* n.d.) publicly supplies crowd-sourced data that has served previous eRange prediction models before (Zheng *et al.*, 2016). This dataset has a high vehicle diversity due to the open nature of the platform, allowing any user being able to upload combustion engine based vehicle information as well the location data, speed, weather as well other parameters, battery information could be supplied through CREATE RAV4-recorder box (*ChargeCar's CREATE RAV4-recorder box* n.d.). The location, trip and vehicle information are then used to determine the simulated EV consumption for each trip. As of the time of writing, a total of 373 unique trips are openly available as a dataset option.

Another dataset solution is the *Emobpy* Python tool (Gaete-Morales *et al.*, 2021) that focuses on EV trip and charge data generation through empirical mobility statistics and customizable assumptions. Although this approach has by definition an infinite supply of EV trips was well as proper vehicle information, this dataset is lacking in some trip parameters such as speed, elevation, trip and commute type.

A dataset collected through probe data from nearly 500 BEVs by the Japan Automobile Research Institute (JARI) between February 2011 to January 2013 by the JARI (*Project Consigning Technology Development for Rational Use of Energy (Innovative Manufacturing Process Technology Development)* n.d.) allegedly consists of the following features: time; location; vehicle state (driving, normal charging or fast charging); speed; air-conditioner and heater state; SOC. Although useful and featured in some papers (Sun *et al.*, 2015; Sun *et al.*, 2016; Liu *et al.*, 2017; Liu *et al.*, 2018), this study was unable to acquire this dataset from (*JARI research database* n.d.) perhaps due to the language barrier.

Previous studies in of eRange prediction (De Cauwer *et al.*, 2017) have depended on *EVteclab*'s Electric Vehicles in Action (EVA) platform, a Flemish Living Labs project (*EVteclab's eva platform* n.d.). The platform supplies a dataset containing monitoring data of 30 different model *Ford Connect* EVs for the duration of a year. This dataset although, supplying a few useful parameters like timestamp, latitude, longitude and vehicle speed, was inaccessible at the time this of writing.

The cloud based EV dataset supplied by the National Big Data Alliance of New Energy Vehicles (NDANEV) (*National Big Data Alliance of New Energy Vehicles* n.d.) has been used in a similar eRange prediction approaches (Zhao *et al.*, 2020). The data was collected from Controller Area Network (CAN) of five different EVs of an undisclosed model through with T-BOX, later uploading it to NDANEV. The dataset distinguishes from the others by including battery cell temperature information, a vital feature for includes many valuable parameters, including cell temperature information being one of them which has been previously used in the measure of battery cell inconsistency.

As some datasets are do not directly supply vehicle information, the ev-database (*Electric Vehicle Database* n.d.) website supplies an easily accessible database for existing EVs, displaying Average Energy Consumption (AEC), total eRange and usable battery energy. Through this information, some otherwise information lacking datasets can be used in eRange prediction models.

Table 2.1 describes most features present in currently accessed public datasets, the * character indicates the feature is not always present.

Table 2.1: Dataset features.

	VED dataset	Emobpy	Classic EV X Project	ChargeCar	NDANEV
Trips	507	Unlimited	1	373	2372
EV Models	1	102	1	?	1
EVs	3	N/A	1	?	5
Features	timestamp, speed, location, battery SOC, battery voltage, battery current, AC power, heater power, OAT	timestamp, distance, IEC, consumption, average power, state	timestamp, IEC, RBE, speed	timestamp, elevation, planar distance, adjusted distance, speed, acceleration, model power, actual power*, current*, voltage*	timestamp, speed, total voltage, total current, battery SOC, temp. range, motor voltage, motor current, mileage

2.2 EV Range prediction

Existing work has demonstrated the use of eRange estimation on EVs, showing the need for different types of accuracy on eRange estimation depending on the SOC state, the proposed approach by (Zhang *et al.*, 2012) minimizes the performance impact of minimum cost route searching from high accuracy eRange prediction.

Other studies have focused on delivering higher eRange estimation accuracy, making use of more complex models. The use of an adaptive history based model approach was proposed by (Coutinho, 2021), which relies on past information about vehicle's instant consumption energy, to determine an adaptive average energy consumption.

2.3 The use of machine learning

The use of machine learning for a multitude of cases (Amershi *et al.*, 2019) in fields such as big data (Condie *et al.*, 2013; Zhou *et al.*, 2017) and data mining (Bose and Mahapatra, 2001) has proven its robustness on solving complex problems.

As a result, some approaches for the eRange problem have already applied machine learning, most notably supervised learning to achieve the estimation. The use of decision trees (DTs), random forest (RF), and K-nearest neighbor in ensemble stacked generalization approach, through the *JARI* dataset (in 2.1) (Ullah *et al.*, 2021) has proved its effectiveness in yielding more acceptable values in overfitting cases for proposed evaluation metrics.

Recent models using gradient boosted regression trees (GBRTs) have combined extreme gradient boosting (XGBoost) and light gradient boosted machine (LightGBM) to provide better predictive performance from these ensemble methods (Zhao *et al.*, 2020) with the NDANEV (in 2.1) dataset, the approach classified four different the driving patterns from three different parameters: speed; motor current; change rate of motor current, through K-means clustering algorithm and thus influencing result eRange due to their different energy consumption rates, (source code in *Machine Learning-Based Method for Remaining Range Prediction of Electric Vehicles - source* n.d.).

Approaches using unsupervised clustering of self-organizing maps (SOMs) have been used for clustering big data into driving patterns, prior to range estimation (Lee and Wu, 2015). The hybrid version of SOM with regression trees (RTs) has taken advantage of SOM neurons storage feature of nearing related neighbor information being kept closely together, performing RTs locally, avoiding bushy trees and improving upon previous solutions by keeping meaningful knowledge extraction on bushy trees (Zheng *et al.*, 2016) both approaches used different datasets from undisclosed monitored data sources.

Reinforcement learning in the form of neural networks has also been used for external energies disturbances on the speed profile of a driving profiles so that it could then be combined with multiple linear regression (MLR) for the estimation (De Cauwer *et al.*, 2017), using *EVteclab*'s dataset (in section 2.1).

Although more complex than previous solutions, the use of machine learning for the eRange estimation problem had reduced the prediction error, and thus further justifying its usage in this project for the eRange prediction problem.

tabela

3

Methodology

This work aims to provide an eRange prediction machine learning model based on the best performing sub-model when training from publicly available datasets. To this end, a dataset was constructed on the preprocessing phase of this project to be used for training by two machine learning algorithms: Ensemble stacked generalization (Ullah *et al.*, 2021) and Linear regression. Each of the sub-models follow the standard supervised machine learning two stage process of first learning the dataset features with the expected eRange result (fitting), and then the prediction of eRange is obtained from real-time parameters. One of the predictors is then selected based on its performance, through different performance metrics (3.2) and used for future execution in a real-time trip as detailed in Figure 3.1.

The project features a Python application capable of configuring the EV X DRP dataset for training the machine learning algorithms. *training datasets*; *Minimum time-series time-step* controlling the minimum time interval for when datasets don't have it fixed; and *Minium trip execution time* to help reduce training bias to trips with lower execution times.

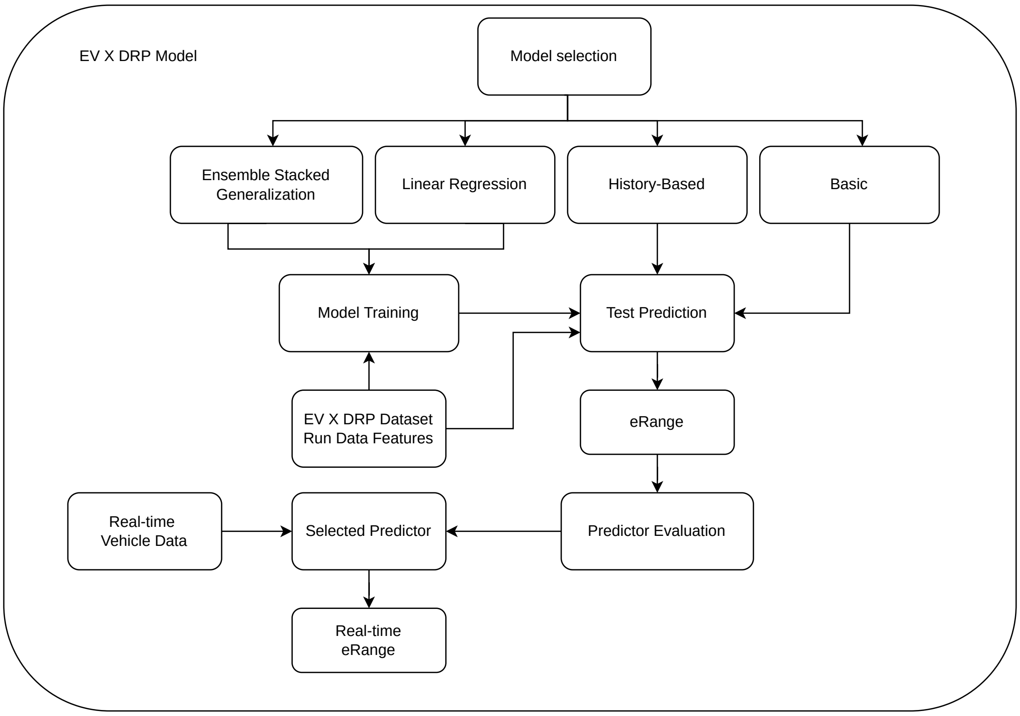

Figure 3.1: EV X Driving Range project.

The ensemble stacked generalization model follows the (Ullah *et al.*, 2021) implementation however, some differences in this model such as the model's original application being the EVs energy consumption prediction and not eRange, as well as the lack of availability of its *JARI* dataset (in 2.1) made its implementation more challenging integration with other datasets.

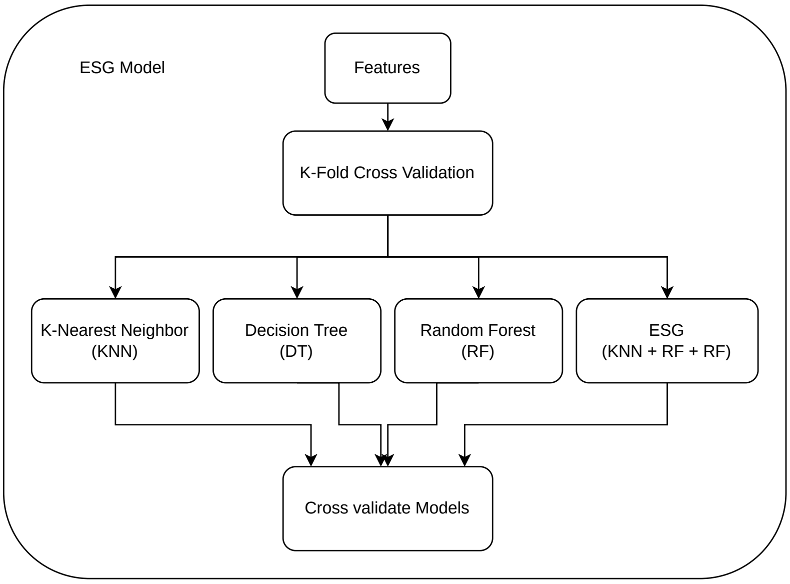

Figure 3.2: ESG model.

As the eRange prediction problem of determining the vehicle's maximum distance is approached in a supervised learning implementation, the emphHistory-based adaptive algorithm in Coutinho, 2021 was chosen for the training phase's target for the estimators, as it did not use machine learning for its deemed optimistic results and less frequent updates. This approach estimates real-time AEC estimation, that relies on vehicle's past $N = 10$ minute window of the trip's energy consumption history as well as real-time SOC value. This algorithm was designed as better alternative to its *Basic algorithm* in which the estimation uses the vehicle's (constants) AEC and FBD provided by the manufacturer and a real-time SOC value, and thus, was also included as comparable algorithm.

3.1 Data collection

When training machine learning model for regression problems, the accuracy of the results on test data will depend on the diversity of the data. To ensure model effectiveness on different vehicles, we opted for a diverse EV model dataset built from existing available datasets *VED* (Oh *et al.*, 2019); *ChargeCar* (*ChargeCar Database* n.d.) *Classic EV X Project DRP* (Coutinho, 2021), thus avoiding overfitting of few EV models.

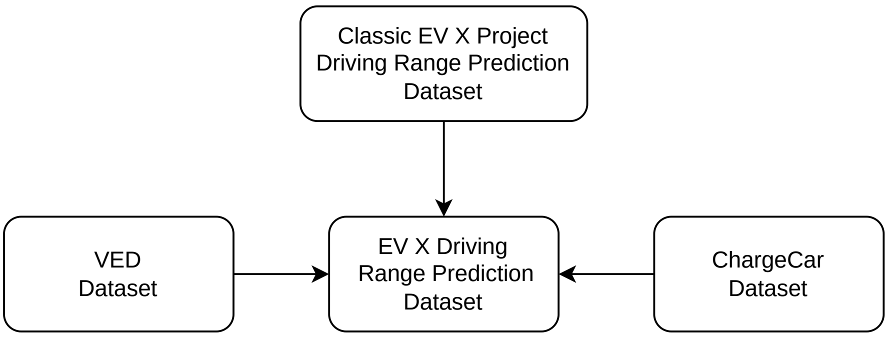

Figure 3.3: EV X Driving Range Prediction dataset sources.

The algorithm integrates EV trip datasets for training, thus requiring EV trips time-series with the following features: SOC (percentage), power consumption (kW/h), distance (km) and speed (km/h), as well as vehicle information: AEC (KWh), Full Battery Energy (FBE) (KWh), and Full Battery Distance (FBD) (km). For this reason, both *VED* and *ChargeCar* datasets *ChargeCar Database* n.d. were chosen. When configuring the algorithm's training, different datasets can be selected, as well as a minimum trip type and minimum driving time, as these variables have been tested and found to highly influence ML methods performance. During the preprocessing phase, some features such FBE, FBD and AEC are sometimes not available on certain datasets. These however can be obtained from existing static EV information datasets as *Electric Vehicle Database* n.d. Other features such as power variation and distance, are trip dependent and must therefore be calculated.

The power consumption delta ΔP measures the power consumption variation between the previous i trip instant and the next f, using the DC power formula of current A times voltage V.

$$1)\ \ \Delta P = P_f - P_i = V_f * A_f - V_i * A_i$$

The acceleration is needed to calculate the distance feature of the dataset and for its calculation, the difference of the two trip instance velocity values is divided by the time variation.

$$2)\ \ a = \frac{v_f - v_i}{\Delta t}$$

For the distance, we take into account the previously calculated acceleration a between trip instants and apply it to the initial velocity for the time variation.

$$3)\ \ D = v_i * \Delta t + \tfrac{1}{2} * a * \Delta t^2$$

3.2 Evaluation metrics

As prediction accuracy must be measured for each eRange prediction algorithm, five evaluation metrics were chosen for this task: *Mean Absolute Error (MAE) Mean Squared Error (MSE)*, *Mean Absolute Percentage Error (MAPE)*, *Root Mean Squared Error (RMSE)* and *Coefficient of determination (R2)*, as demonstrated in formulas 4 to 7 ($\mathbf{y_i}$ represents observed value, $\mathbf{\hat{y}_i}$ the predicted value and $\mathbf{\bar{y}_i}$ the average observed value) (Chicco *et al.*, 2021).

MAE represents the average of the absolute difference between the actual and predicted values, increasing weights for values closer to the average, mitigating outliers effects on this metric.

$$4)\ \ MAE = \tfrac{1}{n} \sum_{i=1}^{n} \left| y_i - \hat{y}_i \right|$$

MAPE indicates the same information as MAE, however, makes it more clear to compare between models due to its value is represented in a percentage. Unfortunately, this formula has complications for when the observed value is close to zero, due to the division by error.

$$5)\ \ MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right| * 100\%$$

MSE measures both bias and variance of the residuals, averaging the squared difference between the original and predicted values.

$$6)\ \ MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

RMSE measures the standard deviation of residuals with the square root of MSE.

$$7)\ \ RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}$$

R2 represents the proportion of the variance in the dependent variable, and contrary to other the other evaluation metrics, where lower values are perceived as better, higher values for R2 are desirable, as lower values indicate redundant or inconsequential variables.

$$8)\ \ R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y}_i)^2}$$

In this work, the prediction model is compariatatbly considered more accurate when it has lower MAE, MAPE, MSE, RMSE values and a higher R2 value.

3.3 Developed application

The focus of the application is to allow the comparison of existing eRange prediction models with the project's implemented model. For this reason, two additional prediction models (besides "basic" and "history based" from Coutinho, 2021) were integrated into the application: ensemble stacked generalization approach of Ullah *et al.*, 2021; combined XGBoost and LightGBM of Zhao *et al.*, 2020, as well as individual linear and tree based machine learning algorithms.

The developed application features algorithm training and trip execution customization settings. The selection of enabled eRange prediction algorithms for training allows multi algorithm comparison for the same test trip.

4

Results

The application displays different eRange prediction results for the selected trip and prediction algorithms, allowing for an easy overview of the different dataset parameters, allowing initial input dataset configuration to depend on multiple datasets.

On a conventional 47 minutes trip from the VED dataset (Oh *et al.*, 2019) for a Nissan Leaf model (2013) was run, with training data from the same dataset in Figure 4.1, with a minimum trip time of 9 minutes and a timestamp of 5 seconds.

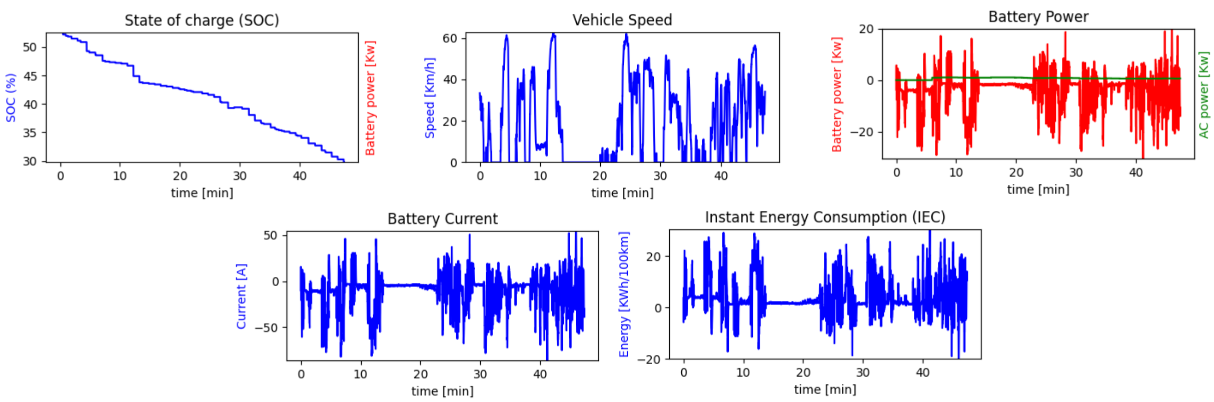

Figure 4.1: Execution trip parameters

The comparison of the different selected eRange prediction models can be seen on the following graph with predictions all implemented algorithms: "basic" and "history based" non machine learning approaches from Coutinho, 2021 (in blue and red, respectively); decision tree based approaches like the ensemble stacked generalization of Ullah *et al.*, 2021, (in green); and the linear regression approach (in purple).

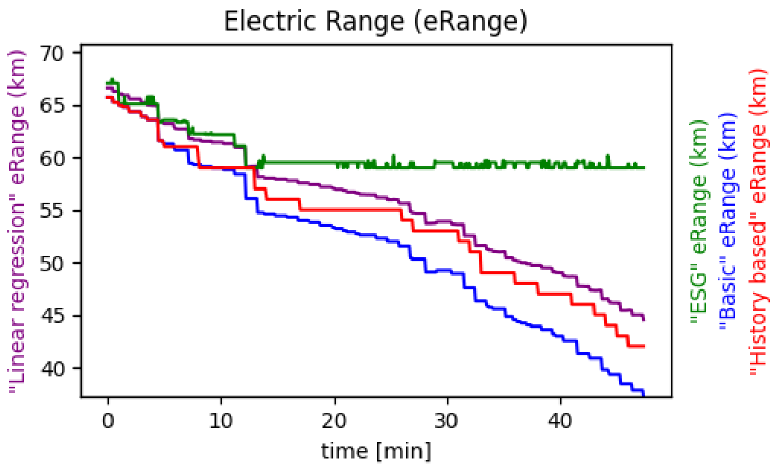

Figure 4.2: eRange: All machine learning algorithms

The history-based approach shows an increase in eRange when regenerative breaking is charging the battery and "plateau" sections when the minimum instance energy is not enough to trigger a recalculation for the eRange. These sections have been smoothed by the LR algorithm, which shows a smooth evolution on the prediction values. The LR algorithm improves on the History-based approach that exhibits a “stair-case effect", which may cause anxiety on the driver, each time the estimated value drops in a step. Thus, LR seems to be the best ML approach addressed so far. On the other hand, the ESG algorithm provides more optimistic estimate, yielding larger eRange predicted values. One possible cause for this performance may be the missing original dataset training features such as elevation, however the negative value of R2 also indicates poor fitting between the selected dataset features and the prediction value. It is worth noting the resemblance with the "basic" approach from the same paper in which is history-based approach is based off. The evaluation metrics are reported in Table 4.1.

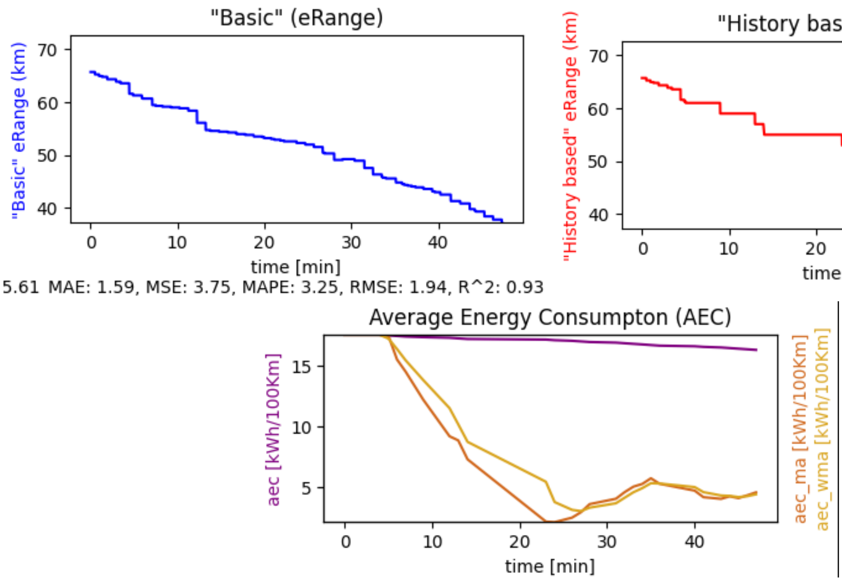

Figure 4.3: Basic and History-Based approaches with AEC values.

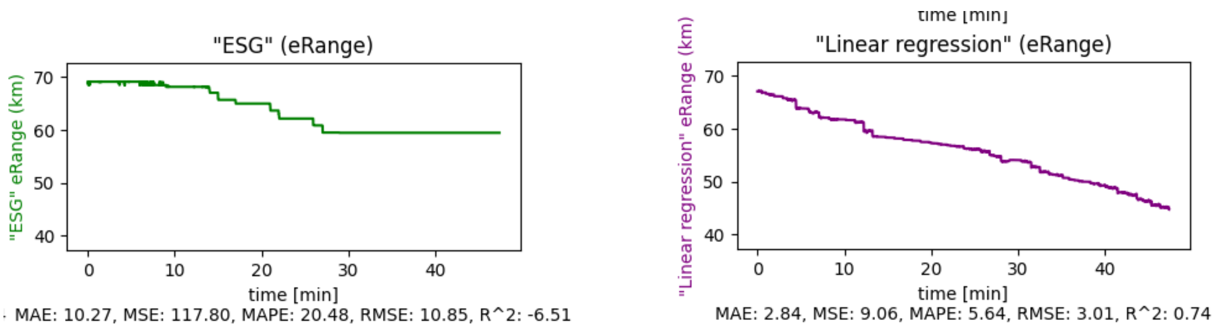

Figure 4.4: ESG and Decision Tree.

The machine learning approaches trained with trips where the vehicle drove only few minutes or stayed most of the time stopped have been noted to caused high deviation rates for all algorithms, presumably due to the low sample size of these occurrences, thus for the first example above, a minimum driving time of 10 minutes was assured in all training trips. This however has proven difficult for these algorithms to predict accurately for these kinds of trips.

Table 4.1: ESG and Linear Regression metrics.

ML approach	**MAE**	**MSE**	**MAPE**	**RMSE**	**r^2**
ESG	6.51	64.40	0.13	8.03	-0.73
LR	**2.38**	**8.23**	**0.05**	**2.87**	**0.78**

These experimental results show that eRange prediction can be achieved with ML techniques, overcoming the existing basic and History-based approaches. The LR approach has fast training and achieves adequate results with a smooth varying curve on the prediction values. The ESG approach needs to be further explored and its parameters need to be fine tuned.

5

Conclusion

This thesis proposes a machine learning model approach for the eRange prediction problem, using data from publicly available datasets with more than 9 different EVs in more than 3252 trips and while implementing different eRange prediction approaches for comparison. This study has also demonstrated the impact of different datasets on existing machine learning models due to different datasets and training configurations. For further comparison options, additional machine learning approaches could be added, as well as more data from previously unaccessible datasets from section 2.1 into the implemented Python application.

As future work, we plan to perform the integration of this work with the real-time information of the EV (as depicted in Figure 3.1) to continuously provide eRange estimations on real-time. We also plan to include more features, such as driving patterns and road elevation.

References

[1] Murat Yilmaz and Philip T. Krein, "Review of battery charger topologies, charging power levels, and infrastructure for plug-in electric and hybrid vehicles", *IEEE Transactions on Power Electronics*, vol. 28, no. 5, pages 2151–2169, 2013. DOI: 10.1109/TPEL.2012.2212917.

[2] Indranil Bose and Radha K. Mahapatra, "Business data mining - a machine learning perspective", *Information & Management*, vol. 39, no. 3, pages 211–225, 2001, ISSN: 0378-7206. DOI: https://doi.org/10.1016/S0378-7206(01)00091-X. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S037872060100091X.

[3] *Chargecar database*, http://www.chargecar.org/data, Accessed: 2022-07-03.

[4] *Chargecar's create rav4-recorder box*, http://chargecar.org/participate/lending, Accessed: 2022-07-03.

[5] David Coutinho, "Classic ev x project driving range prediction, technical report", Draft version, Jul. 2021. [Online]. Available: https://www.researchgate.net/publication/353210805_Classic_EV_X_Project_Driving_Range_Prediction_TECHNICAL_REPORT_draft_version.

[6] Yang Song and Xianbiao Hu, "Learning electric vehicle driver range anxiety with an initial state of charge-oriented gradient boosting approach", *Journal of Intelligent Transportation Systems*, vol. 0, no. 0, pages 1–19, 2021. DOI: 10.1080/15472450.2021.2010053. eprint: https://doi.org/10.1080/15472450.2021.2010053. [Online]. Available: https://doi.org/10.1080/15472450.2021.2010053.

[7] Ona Egbue and Suzanna Long, "Barriers to widespread adoption of electric vehicles: An analysis of consumer attitudes and perceptions", *Energy Policy*, vol. 48, pages 717–729, 2012, Special Section: Frontiers of Sustainability, ISSN: 0301-4215. DOI: https://doi.org/10.1016/j.enpol.2012.06.009. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0301421512005162.

[8] Carlos Gaete-Morales, Hendrik Kramer, Wolf-Peter Schill, and Alexander Zerrahn, "An open tool for creating battery-electric vehicle time series from empirical data, emobpy", *Scientific Data*, vol. 8, no. 1, page 152, 2021, ISSN: 2052-4463. DOI: 10.1038/s41597-021-00932-9. [Online]. Available: https://doi.org/10.1038/s41597-021-00932-9.

[9] Y. Zhang, W. Wang, Y. Kobayashi, and K. Shirai, "Remaining driving range estimation of electric vehicle", in *2012 IEEE International Electric Vehicle Conference*, ser. 2012 IEEE International Electric Vehicle Conference, 2012, pages 1–7. DOI: 10.1109/IEVC.2012.6183172. [Online]. Available: https://doi.org/10.1109/IEVC.2012.6183172.

[10] Martin Smuts, Brenda Scholtz, and Janet Wesson, "A critical review of factors influencing the remaining driving range of electric vehicles", in *2017 1st International Conference on Next Generation Computing Applications (NextComp)*, 2017, pages 196–201. DOI: 10.1109/NEXTCOMP.2017.8016198.

[11] Irfan Ullah, Kai Liu, Toshiyuki Yamamoto, Muhammad Zahid, and Arshad Jamal, "Electric vehicle energy consumption prediction using stacked generalization: An ensemble learning approach", *International Journal of Green Energy*, vol. 18, no. 9, pages 896–909, 2021. DOI: 10.1080/15435075.2021.1881902. eprint: https://doi.org/10.1080/15435075.2021.1881902. [Online]. Available: https://doi.org/10.1080/15435075.2021.1881902.

[12] Chung-Hong Lee and Chih-Hung Wu, "A novel big data modeling method for improving driving range estimation of evs", *IEEE Access*, vol. 3, pages 1980–1993, 2015. DOI: 10.1109/ACCESS.2015.2492923.

[13] Cedric De Cauwer, Wouter Verbeke, Thierry Coosemans, Saphir Faid, and Joeri Van Mierlo, "A data-driven method for energy consumption prediction and energy-efficient routing of electric vehicles in real-world conditions", *Energies*, vol. 10, no. 5, 2017, ISSN: 1996-1073. DOI: 10.3390/en10050608. [Online]. Available: https://www.mdpi.com/1996-1073/10/5/608.

[14] Davide Chicco, Matthijs J. Warrens, and Giuseppe Jurman, "The coefficient of determination r-squared is more informative than smape, mae, mape, mse and rmse in regression analysis evaluation", *PeerJ Computer Science*, vol. 7, e623, Jul. 2021, ISSN: 2376-5992. DOI: 10.7717/peerj-cs.623. [Online]. Available: https://doi.org/10.7717/peerj-cs.623.

[15] Erik Figenbaum, Nils Fearnley, Paul Pfaffenbichler, Randi Hjorthol, Marika Kolbenstvedt, Reinhard Jellinek, Bettina Emmerling, G. Maarten Bonnema, Farideh Ramjerdi, Liva Vågane, and Lykke Møller Iversen, "Increasing the competitiveness of e-vehicles in europe", *European Transport Research Review*, vol. 7, no. 3, page 28, 2015, ISSN: 1866-8887. DOI: 10.1007/s12544-015-0177-1. [Online]. Available: https://doi.org/10.1007/s12544-015-0177-1.

[16] Alessandro Brighente, Mauro Conti, Denis Donadel, and Federico Turrin, "Evscout2.0: Electric vehicle profiling through charging profile", *CoRR*, vol. abs/2106.16016, 2021. arXiv: 2106.16016. [Online]. Available: https://arxiv.org/abs/2106.16016.

[17] Xiao-Hui Sun, Toshiyuki Yamamoto, and Takayuki Morikawa, "Stochastic frontier analysis of excess access to mid-trip battery electric vehicle fast charging", *Transportation Research Part D: Transport and Environment*, vol. 34, pages 83–94, 2015, ISSN: 1361-9209. DOI: https://doi.org/10.1016/j.trd.2014.10.006. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1361920914001473.

[18] Xiao-Hui Sun, Toshiyuki Yamamoto, and Takayuki Morikawa, "Fast-charging station choice behavior among battery electric vehicle users", *Transportation Research Part D: Transport and Environment*, vol. 46, pages 26–39, 2016, ISSN: 1361-9209. DOI: https://doi.org/10.1016/j.trd.2016.03.008. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1361920916000377.

[19] Kai Liu, Toshiyuki Yamamoto, and Takayuki Morikawa, "Impact of road gradient on energy consumption of electric vehicles", *Transportation Research Part D: Transport and Environment*, vol. 54, pages 74–81, 2017, ISSN: 1361-9209. DOI: https://doi.org/10.1016/j.trd.2017.05.005. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1361920917303887.

[20] Kai Liu, Jiangbo Wang, Toshiyuki Yamamoto, and Takayuki Morikawa, "Exploring the interactive effects of ambient temperature and vehicle auxiliary loads on electric vehicle energy consumption", *Applied Energy*, vol. 227, pages 324–331,

2018, Transformative Innovations for a Sustainable Future Part III, ISSN: 0306-2619. DOI: https://doi.org/10.1016/j.apenergy.2017.08.074. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0306261917310929.

[21] Tyson Condie, Paul Mineiro, Neoklis Polyzotis, and Markus Weimer, "Machine learning on big data", in *2013 IEEE 29th International Conference on Data Engineering (ICDE)*, 2013, pages 1242–1244. DOI: 10.1109/ICDE.2013.6544913.

[22] Lina Zhou, Shimei Pan, Jianwu Wang, and Athanasios V. Vasilakos, "Machine learning on big data: Opportunities and challenges", *Neurocomputing*, vol. 237, pages 350–361, 2017, ISSN: 0925-2312. DOI: https://doi.org/10.1016/j.neucom.2017.01.026. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925231217300577.

[23] Saleema Amershi, Andrew Begel, Christian Bird, Robert DeLine, Harald Gall, Ece Kamar, Nachiappan Nagappan, Besmira Nushi, and Thomas Zimmermann, "Software engineering for machine learning: A case study", in *2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP)*, 2019, pages 291–300. DOI: 10.1109/ICSE-SEIP.2019.00042.

[24] Liang Zhao, Wei Yao, Yu Wang, and Jie Hu, "Machine learning-based method for remaining range prediction of electric vehicles", *IEEE Access*, vol. 8, pages 212 423–212 441, 2020. DOI: 10.1109/ACCESS.2020.3039815.

[25] *Machine learning-based method for remaining range prediction of electric vehicles - source*, https://github.com/liangzhao123/range_prediction, Accessed: 2022-07-18.

[26] B. Zheng, P. He, L. Zhao, and H. Li, "A hybrid machine learning model for range estimation of electric vehicles", in *2016 IEEE Global Communications Conference (GLOBECOM)*, ser. 2016 IEEE Global Communications Conference (GLOBECOM), 2016, pages 1–6. DOI: 10.1109/GLOCOM.2016.7841506. [Online]. Available: https://doi.org/10.1109/GLOCOM.2016.7841506.

[27] Tom Mitchell, "The discipline of machine learning", Tech. Rep. CMU ML-06 108, 2006.

[28] *National big data alliance of new energy vehicles*, http://www.ndanev.com, Accessed: 2022-07-18.

[29] *Electric vehicle database*, https://ev-database.org/car/1011/Nissan-Leaf, Accessed: 2022-07-03.

[30] *Evteclab's eva platform*, http://proeftuin-ev.be/content/eva-platform, Accessed: 2022-07-19.

[31] *Jari research database*, https://www.jari.or.jp/research-database/, Accessed: 2022-07-03.

[32] *Project consigning technology development for rational use of energy (innovative manufacturing process technology development)*, https://web.archive.org/web/20130402003134/http://www.meti.go.jp/english/press/2012/0502_01.html, Accessed: 2022-07-17.

[33] *Paris agreement*, UN Treaty, United Nations, Dec. 2015. [Online]. Available: https://treaties.un.org/pages/ViewDetails.aspx?src=TREATY&mtdsg_no=XXVII-7-d&chapter=27&clang=_en.

[34] Bogdan Ovidiu Varga, Arsen Sagoian, and Florin Mariasiu, "Prediction of electric vehicle range: A comprehensive review of current issues and challenges", *Energies*, vol. 12, no. 5, 2019, ISSN: 1996-1073. DOI: 10.3390/en12050946. [Online]. Available: https://www.mdpi.com/1996-1073/12/5/946.

[35] M.K Yoong, Y.H Gan, G.D Gan, C.K Leong, Z.Y Phuan, B.K Cheah, and K.W Chew, "Studies of regenerative braking in electric vehicle", in *2010 IEEE Conference on Sustainable Utilization and Development in Engineering and Technology*, 2010, pages 40–45. DOI: 10.1109/STUDENT.2010.5686984.

[36] G. S. Oh, David J. Leblanc, and Huei Peng, *Vehicle energy dataset (ved), a large-scale dataset for vehicle energy consumption research*, 2019. arXiv: 1905.02081 [physics.soc-ph].