EV Battery Degradation: A Data Mining Approach

Rui Rodrigues[1], Vitória Albuquerque[2], João Carlos Ferreira[1] and Miguel Sales Dias[1]

[1] ISCTE - University Institute of Lisbon, Lisbon 1649-026, Portugal
[2] NOVA IMS – Information Management School, Universidade Nova de Lisboa, 1070-312 Lisbon, Portugal

This is the Author Peer Reviewed version of the following chapter/ conference contribution published by Springer:

Rodrigues, R., Albuquerque, V., Ferreira, J. C., & Dias, M. S. (2022). EV Battery Degradation: A Data Mining Approach. In A. L. Martins, J. C. Ferreira, & A. Kocian (Eds.), Intelligent Transport Systems: 5th EAI International Conference, INTSYS 2021, Proceedings (pp. 177-191). (Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, LNICST; Vol. 426 LNICST). Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1007/978-3-030-97603-3_13

EV Battery Degradation: A Data Mining Approach

Rui Rodrigues[1[0000-1111-2222-3333]] Vitória Albuquerque[2[0000-0001-9684-968X]], João Carlos Ferreira[1[0000-0002-6662-0806]] and Miguel Sales Dias1[0000-0002-1445-2695][1]

[1] ISCTE - University Institute of Lisbon, Lisbon 1649-026, Portugal

2 NOVA IMS – Information Management School, Universidade Nova de Lisboa, 1070-312 Lisbon, Portugal

Abstract. The increase in greenhouse gas emissions into the atmosphere, and their adverse effects on the environment, has prompted the search for alternative energy sources to fossil fuels. One of the solutions gaining ground is the electrification of various human activities, such as the transport sector. This trend has fueled a growing need for electrical energy storage in lithium-ion batteries. Precisely knowing the degree of degradation that this type of battery accumulates over its useful life is necessary to bring economic benefits, both for companies and citizens. This paper aims to answer the current need by proposing a research question about electric motor vehicles. It focuses on habits EV owners practice, which could harm the battery life. This paper seeks to answer this question using a data science methodology. The results allowed us to conclude that all other factors had a marginal effect on the vehicles' autonomy decrease except for the car year. The biggest obstacle encountered in adopting electric vehicles was the insufficient coverage of the charging stations network.

Keywords: electric vehicles, charging process, behavior, data mining, machine learning

1 Introduction

The electrification of most human activities is nowadays a necessity. It is crucial to reduce greenhouse gas emissions – targeting the larger goal of decarbonizing human society. The application of energy storage technology in the transportation sector, mainly adopted in electric passenger vehicles, is a strategic step towards the widespread adoption of this type of mobile technology towards the mentioned decarbonization of society. This research on lithium-ion (Li-ion) batteries aims to know more about a subject still requiring broader understanding. Also, it aims to collect and obtain insights into the state-of-the-art on the topic and extract information from data relinquished by EV owners. This information is expected to discuss the degree of satisfaction EV users have with the current solutions available in Portugal, potentially repel consumers from purchasing this type of vehicle.

In recent decades, the increased occurrence of manifestations of intense and erratic climatic change has made it crucial to find alternative forms of energy consumption and conservation to conventional methods such as fossil fuel that significantly contribute to greenhouse gas emissions. Therefore, it is of particular importance to adopt these alternatives with the utmost celerity.

Battery aging is currently a problem that cuts across all sectors of activity that depend on them now or may depend on it soon, such as the transport sector in general and specifically in private vehicles. For example, electric mobility is an emerging, ever-growing mode of transport that causes an increased demand for Li-ion batteries in vehicles. However, these batteries have a limited useful life and are usually grouped in packs that make them difficult to replace. In addition, the recycling of batteries' toxic components has proven to be a hazard to the environment. Thankfully, there is a growing need to find methods that can extend the life of these battery packs to reduce their environmental footprint [1] and instead find non-toxic elements to their manufacturing process.

In the automotive industry, this premature aging of batteries is adverse in two ways: firstly, it limits the autonomy range of the private vehicle, and it also affects its general acceptance and adoption by the public. Therefore, the need to know the exact pace of battery degradation often motivates information campaigns for technology adoption, academic research, and industrial research and development to improve its performance and longevity [1].

Future potential owners of vehicles powered by Li-ion batteries are starting to require accurate information on how long their vehicle batteries will last [1]. Hence, consumers are interested in determining whether it is advantageous to invest in this new technology and pay extra fees for its early adoption.

Our study focuses on EV owners driving habits. Data for this study were retrieved from a public inquiry to Tesla vehicles owners. The Tesla dataset was published [2], an electrical news website, on Apr. 14, 2018. It depicted a downward trend curve for vehicle degradation that stabilized at a deficit of ten percent of battery total capacity after one hundred and sixty thousand miles, which was promising news. Furthermore, confirming the Tesla findings, in early 2020, Tesla Inc. published a report stating their batteries would retain 90 percent of their original capacity after 200,000 miles of usage [3].

We applied the Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology to the Tesla dataset by creating a classification model based on the variables available in the raw file obtained by the Tesla article [2]. The end goal was to answer the following research question "Which factors have the most impact on the battery degradation?"

2 Literature Review

Actions[4]–[7]to reduce greenhouse gases (GHG) emissions have been implemented, even before the covid-19 pandemic, resulting in CO2 emissions decreasing by 1.8 percent from 2018 to 2019. The USA Environmental Protection Agency (EPA) [5] as well as the European Green Deal [6] explain that this was primarily due to a drop in total energy used and improved energy efficiency in 2019 compared to 2018, brought down by fossil fuel emissions reduction [5]. In tandem with CO2 emissions reduction, an ongoing shift from coal to natural gas (the least harmful greenhouse gas-emitting fossil fuel) occurs in the energy sector.

One alternative solution to the burning of fossil fuels in the transportation sector is adopting Li-ion batteries. This type of battery is currently empowering EVs. Technological improvements have been implemented in the last ten years to increase these batteries' energy capacity and efficiency [8]. However, because its capacity is finite, any factor that decreases its energy retention ability is crucial. The degradation of the energy capacity of this type of battery, which is observed, for instance, in cell phones, is one of the main problems faced by energy experts.

Battery early aging often depends on the Li-ion battery materials' chemical composition, namely its anode, cathode, and electrolyte. In addition, external factors, such as voltage, discharge intensity, temperature, and the number of charging-discharging cycles performed, are also considered important factors. However, the reference literature does not quantify how relevant these factors are to the overall battery longevity. For instance, the Tesla manufacturer applies solutions to mitigate premature battery aging; all its vehicles have a management system whose primary function is to control the battery's temperature to remain below 55 degrees Celsius [1].

However, behavioral factors associated with the operation and charging of electric cars and their storage are significantly considered to impact the degradation of batteries [1]. As already observed with mobile phones, car batteries are subjected to premature aging if left unused. This concern regards that both cases use the same technology and materials. On the other hand, their continued use also leads to progressively shorter service life. May [9] suggest that both technologies' similarities would not stop at that point, and the EV would be as prevalent as the mobile phone. May also envisions that one day, everyone would be able to have one.

The evaluation of Li-ion batteries' performance is still an ongoing process. This technology continues to be studied and matured iteratively by the scientific community that seeks different methods to measure its capacity, internal resistance, and voltages and its influence in charge and discharge cycles [10]–[12].

According to Yun, [11] the high complexity of practical solutions brings difficulty in measuring the variables mentioned above, especially in controlling the internal variables related to the consistency of the manufacturing quality of the various components of the batteries. Thus, it becomes necessary to assess batteries' health status or State of Health (SoH).

The SoH of the battery, expressed as a percentage, represents its current capacity in Watts, concerning its original capacity. This value weighs various parameters of Li-ion batteries, such as their voltage, current, and capacity. Currently, few articles [13]–[18] can accurately predict the actual value of SoH.

There are two types of battery capacity forecasting methods to determine the SoH: model-based methods and data-based methods. Model-based methods were always related to the chemical composition of batteries, and there is plenty of reference literature available on this subject. However, most authors did not focus on this area of scientific knowledge.

There are two types of battery capacity forecasting methods to determine the SoH: model-based methods and data-based methods. Model-based methods were always

related to the chemical composition of batteries, and there is plenty of reference literature available on this subject.

Regarding data-based prediction models, these sometimes use the parameters referred to earlier [18]–[23] to monitor the SoH and forecast the state of the RUL [19], [20]. Compared to prediction methods based on chemical models, these data-based methods [21] are faster, more convenient, and less complex [22]. Moreover, Machine Learning (ML) methods can be used, resulting in the accuracy improvement of these models. These prediction methods have raised a growing interest in verifying the SoH of batteries [23].

3 Data Analytics

CRISP-DM [24] is a methodology widely used by data science specialists to develop solutions for business problems based on data [24]. CRISP-DM can be understood as a cross-industry standard process for data mining and envisages transforming the company's data into knowledge and helpful information for management and decision-making.

Data Mining is part of Data Science, which uses statistics, mathematics, and ML approaches as a basis for crossing data, using induction techniques to propose hypotheses and solve business issues.

The CRISP-DM approach gathers the best practices so that the DM is as productive and efficient as possible, analyzing financial data, human resources, production, customer habits, and other data sources to propose data-based models for improvement or problem-solving. It defines a Data Mining project's life cycle, dividing it into the six phases, shown below in Fig. 1, and following a linear progression.

It is essential to emphasize the theoretical character of this research. The application of the CRISP-DM phases to our case is therefore limited. For this reason, there are phases of this methodology that are less explored than others.

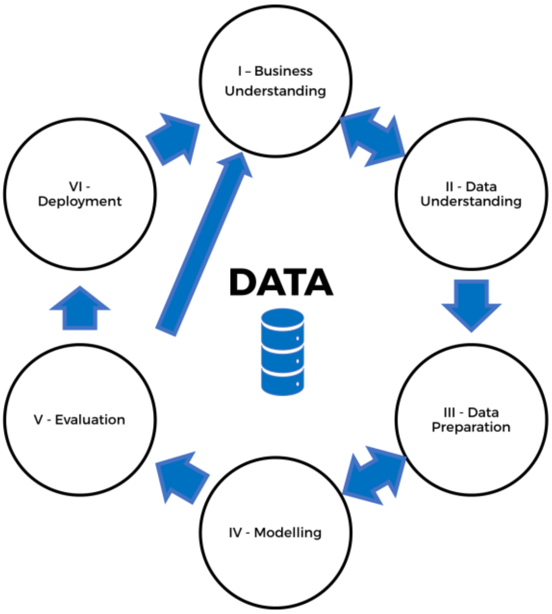


Fig. 1. – CRISP-DM methodology flowchart.

3.1 Business Understanding

The first phase of CRISP-DM aims to understand the entity's primary needs and business requirements where the methodology will be implemented. It looks for all the details about its internal organization, terminology, marketing strategies, target audience, and available products. The conclusion of this step defines guidelines for the steps that follow, such as the selection, cleaning, and interpretation of the information retrieved for the implemented data mining project. In the case of this study, the final client is an individual electric car driver. For example, a potential EV car buyer needs estimates of what an EV car's range would be. The EV car buyer would be interested in predicting the range of a car on a full charge based on its attributes. More precisely, the EV car buyer would need to answer the business question related to our research question: "Which factors have the most impact on battery degradation?".

3.2 Data Understanding

The second step of CRISP-DM consists of organizing and documenting all the available data sources relevant to the institution or client. This documentation implies the identification of a target audience and the selection of sources of data. This stage is an iterative process that includes searching for data sources and data essential for its selection. It is expected to obtain an extensive dataset with the potential of obtaining meaningful information about EV users and their vehicles. Ideally, it would reach a diversity of responses high enough to ensure a comprehensive analysis of the problem in focus.

However, in this paper, the original dataset was obtained through a single source. It came from an international news blog called Elektrek [2]. This blog shares a dataset compiling answers to a survey from a forum of Tesla users, who registered their range entries and other data in an excel spreadsheet, collecting a total of 1425 observations, structured in 43 variables.

3.3 Data Preparation

The data preparation phase aims to transform the information collected into clean, structured, and integrated data. It comprised procedures performed with the Python programming language [25], using the Jupyter Notebook tool [26]. Our data preparation included foursteps: first, removal of out-of-scope variables, followed by the elimination of blank records, elimination of outliers and finally the matching of variable formats. These four steps are described as follows:

Removal of out-of-scope variables: the original dataset (with 43 variables) had out-of-scope variables for this study. Table 2 below lists all the variables present in the Tesla dataset, as collected, before any data cleaning operations, their data type, exclusion status from the study, and the reason behind the exclusion

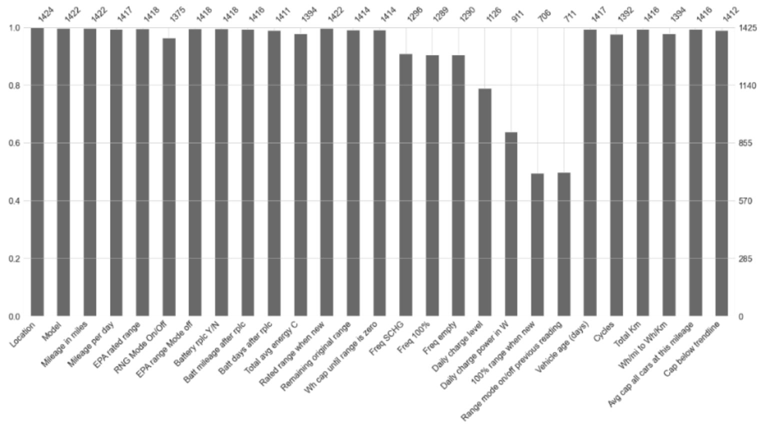
Table 1. Tesla dataset variable list by type.

Variable Name	Data type
Location	String
Model	String
Mileage in Miles	Float
Mileage per day	Float
EPA rated range	Float
RNG Mode On/Off	Float
EPA range Mode off	Float
Battery Replacement Y/N	Boolean
Battery mileage after replacement	Float
Battery days after replacement	Float
Total Average Energy Consumption	Float
Rated range when new	Float
Remaining original range	Float
Wh capacity until range is zero	Float
Freq. SCHG	String
Freq. 100%	String
Freq. empty	String
Daily charge level	Float
Daily charge power in Watts	Float
100% range when new	Float
Range mode on/off previous reading	Float
Vehicle age (days)	Float
Cycles	Float
Total Km	Float
Wh/mi to Wh/km	Float
Avg. Cap. All cars at this mileage	Float
Cap below trendline	Float

Elimination of blank records: some of the observations from the dataset had missing fields. A visual representation of each variable's number of null values was created (see Fig. 1). The higher the bars, the more complete the variable was. Given that these missing values can lead to bias, leading to wrong conclusions, all missing responses had to be excluded from the dataset for further analysis.

Elimination of outliers: when existing variable values were too far from the remaining observations, they were considered outliers and needed to be removed. The method used to detect outliers was based on percentiles. With the percentile's method, all data variables outside an interval formed by the 5th and 95th percentiles were considered potential outliers and removed (see Fig. 3).

Matching of variable formats: the normalization of the dataset variables with formats (e.g., dates, distance units in the Imperial system). Some of the variables were in object format (mostly text ones), which had to be transformed from their original format to a numerical form. Variable encoding was performed as follows: for the location variable, integers were attributed to the name of the countries; for the model variable, each car model was given a distinct integer; for the Freq. SCHG, Freq. 100%, and Freq. empty, their values were encoded by employing a Likert scale, whose values ranged from 1 to 8. This step processed the variables across all observations (see Table 1).



- **Fig. 1.** – Visual representation of null values by each variable.

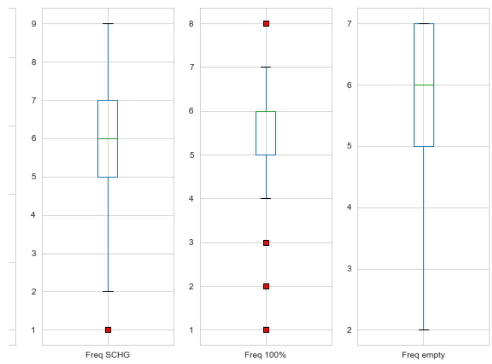


Fig. 2. – Example of outliers' identification of the three Frequency variables.

Table 2. Exclusion of the Tesla dataset variables.

Variable	Type	Excluded	Reason
Username	String	Yes	Irrelevant
Location	String		
Vehicle manufacture date	Date		
Date of range reading	Date	Yes	No insights were found.
Model	String		
Mileage in miles	Int		
Mileage per day	Float	Yes	Duplicate variable.
EPA rated range at 100% charge in miles	Int	Yes	No correlations were found with other variables.

Range mode on/off at the time of reading?	Bool-ean	Yes	No correlations were found with other varia-bles.
EPA range after correction if range mode was off	Float	Yes	No correlations were found with other varia-bles.
Did you have a battery replacement?	Bool-ean	Yes	Very few vehicles had their battery replaced.
What happened to the EPA range after re-placement?	String	Yes	Excessive number of null values.
At what miles did you replace the battery?	Int	Yes	Excessive number of null values.
Mileage in mi after correction if the bat-tery was replaced	Int	Yes	Dependent on battery re-placement.
Battery age (days) after correction if the battery was replaced	Int	Yes	Dependent on battery re-placement.
Lifetime average energy consumption at the time of reading Wh/mi	Int	Yes	Irrelevant for this study.
Rated range of this model when new	Int	Yes	Like variable mileage in miles.
Remaining original range	Float	Yes	Variable replaced by Av-erage Capacity
Remaining usable Wh capacity until typi-cal range shows zero		Yes	No correlations were found with other varia-bles.
Unanswered questions	Int	Yes	Majority of no answers.
Frequency of supercharging	String		
Frequency of 100% charge	String		
Frequency of almost empty (5mi or less)	String		
Daily charge level	Float	Yes	Excessive number of null values.
Daily charge power in watts	Float	Yes	Excessive number of null values.
What was the 100% rated range when the car was new?	Int	Yes	Excessive number of null values.
Range mode on/off at the time of reading the previous column?	Bool-ean	Yes	Excessive number of null values.
Rated range at the beginning of the trip	Int	Yes	Excessive number of null values.
Rated range at the end of the trip	Int	Yes	Excessive number of null values.
Consumption for this trip	Float	Yes	Excessive number of null values.
Range mode on/off when reading these trip numbers?	Bool-ean	Yes	Excessive number of null values.
Typical range consumption for the trip	Float	Yes	Excessive number of null values.
Typical range after correction if range mode was off	Int	Yes	Excessive number of null values.
Remaining usable capacity until typical range shows zero according to trip data	Float	Yes	Excessive number of null values.

Remaining original capacity	Float	Yes	Excessive number of null values.
Trip based battery capacity calculation explained	Float	Yes	Excessive number of null values.
100% range when the car was new after range mode adjustment	Int	Yes	Dependent on range mode.
Vehicle age (days)	Int		
Cycles	Int		
Mileage in miles	Int		
Wh/mi to wh/km	Float	Yes	Unit conversions made in Python and SPSS.
Average capacity of all cars at this mileage according to chart trendline	Float	Yes	Irrelevant for this study.
Your capacity minus chart trendline at this mileage	Float	Yes	Irrelevant for this study.

Most vehicles from the sample traveled a few kilometers because the vast majority have a range below 100,000 kilometers. This occurrence aligns with the fact that most vehicles in the sample are less than ten years old.

3.4 Modeling

In this phase, ML techniques were applied to the Tesla dataset to understand how several factors affected the range of the EVs present in the dataset, as stated in our research question. Therefore, it was necessary to take the following steps to know which variables affect a car maximum range with a full charge:

Classification analysis: Several ML supervised classification algorithms from the Scykit-learn package [27] were exploited to classify which variables influenced the cars' maximum range on a full charge, addressing the research question. The objective of these classifiers was to label the current range that each car had in comparison to the original range value announced by their respective manufacturers. We created a new discrete variable called "Degradation," containing two labels: "Normal" and "Abnormal." These two labels represented the batteries' degree of energy capacity loss in a percentage of the original maximum range. Degradation levels lower than ten percent were labeled "Normal" and higher than ten percent as "Abnormal." The ten percent threshold was used based on the Elektrek article claiming that ten percent was the average capacity degradation of EVs after 160,000 miles [2]. The following ML classification models were chosen and assessed to perform the labeling task:

- K-Nearest Neighbours (KNN) [28]
- Logistic Regression [29]
- Naïve-Bayes [30]
- Support Vector Machine – Linear (L-SVM) [31]
- Support Vector Machine – Radial (R-SVM) [32]

These models were selected because they are of low complexity, and as such, they should generalize better when dealing with small datasets. Since the Tesla dataset is a small dataset with just 1,425 observations, the decision boundary of complex models such as Decision Trees or the Random Forest would change wildly and, therefore, be inappropriate to tackle our problem. As a result, the results from those more complex models would have high degrees of variance. Simpler models such as those chosen above were believed to perform better as they have more minor degrees of freedom.

Following best practices from the literature, each classification model was preceded by a split of the original dataset into a training set and a test set, the split chosen was 80 percent - 20 percent, respectively, [33][34], so that fitted models would be evaluated regarding their performance and compared using a confusion matrix.

Cross-Validation: We employed a standard iterative cross-validation technique [35] to compare model performance, obtain the best model, and avoid overfitting [36]. Overfitting happens when a model obtains near-perfect scores after being trained with known training and testing sets but cannot make accurate predictions when using new data. In our case, the training set was split into ten smaller equal sets. A model was trained using nine sets as training data and judged its results against the tenth set. Then, a loop was created to switch the testing set between all ten sets. The average of the values computed in the loop reports the global performance of the model.

3.5 Evaluation

The evaluation phase aims to assess the validity of the results of our automatic labeling process. First, the cars' current range was checked to see if the vehicles retained more or less than ten percent of the initial total battery capacity. Subsequently, the classification models and the cross-validation technique labeled the vehicles with an abnormal or regular decline of energy capacity. Finally, the factors that had the most negative impact on the batteries were pinpointed, answering our research question. Generally, the models' performance presented accuracy results ranging between 57.18 percent and 62.84 percent, as seen in Table 3.

3.6 Modeling Results

Table 3 shows test accuracy models by median and standard deviation. The median accuracy was calculated by using the accuracy formula shown in Equation 1.

Table 3. Test Accuracy results from the Cross-validation.

Classifiers	Median	Standard Deviation
K-Nearest Neighbours	0.5718	0.0682
Logistic Regression	**0.6284**	**0.1068**
Naïve-Bayes	0.5872	0.0847
L-SVM	0.5921	0.1109
R-SVM	**0.6175**	**0.1058**

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

Eq.1. – Accuracy formula used for determining the classifiers' performance. TP-True Positive, TN-True Negative, FP -False positives, FN-False Negatives.

The two best-fitted models identified using cross-validation are highlighted in bold. The key takeaway from the cross-validation evaluation results was the Logistic

Regression and the Radial Support Vector Machine (Radial SVM). These were the two best performing models from the group, reaching 62.84% and 61.75% of median accuracy, respectively, for the labeling task. Moreover, it shows that from the 500 observations (vehicles), the best classification model, the Logistic Regression, correctly guessed 62.84% of the degradation labels.

At this point in the analysis, each independent variable's weight over the *Degradation* dependent variable remains unknown to answer the research question. However, t below depicts just that, showing both the weight and significance of each variable.

Table 4. Degradation - Coefficients of the independent variables

Degradation	Regression Coefficient	Chi-Squared	p-Value
Country	0.384	4.52	0.052
Year	2.191	12.62	0.037
Maker	0.137	0.03	0.184
Model	0.012	0.01	0.982
Freq. Fast	-0.02	0.90	0.814
Freq. Full	0.09	0.55	0.555
Freq. Empty	0.027	0.28	0.370
Mileage	0.059	0.10	0.627
Max Range	0.061	0.88	0.837

From Table 4, it is possible to conclude that there were two predominant independent variables: the *Year* and *Country*. Furthermore, these two predictors seem to have positive and strong correlations with the *Degradation* dependent variable. Lower values for the *Year* variable correspond to older vehicles, and an "Abnormal" value of the *Degradation* variable means a higher than usual degrading of the battery.

There is a correlation regarding the variable Country, possibly due to a large volume of vehicles from Asia and the Pacific regions, causing a dataset imbalance. However, it would be necessary to have more data to draw more elaborated conclusions to explain this phenomenon. As for the remaining variables listed, there seems to be no relationship between them and *Degradation*.

In conclusion, the variable *Year* was the most impactful factor on the rate of degradation. However, none of the behavioral factors were significant, which would aid in answering the research question *Which factors have the most impact on battery degradation?*

3.7 Deployment

In the deployment phase, knowledge extracted from the data is delivered and applied. From this moment, the processes within the organization might be changed or new products created. Our Logistic Regression ML classification algorithm is, in essence, our final prototype. It aims to determine which behavioral habits from the EV drivers negatively impact the Li-ion battery capacity of the cars, answering the research question.

The chosen model allowed us to gather and label the degree of battery degradation in two categories and get a general sense of the degradation trend of all vehicles. For the personas to whom this study is aimed, i.e., the owners of EVs, this information can be crucial. It sheds light on the current rate of degradation of the car batteries and anticipates the need for future maintenance events, such as a complete battery replacement.

4 Conclusions

This paper aimed to investigate the effect of capacity degradation on electric vehicles' batteries by following a data science and analytics approach. The objective was attained by answering the research question "*Which factors have the most impact on the battery degradation"?*

As shown in Table 4, charging and parking habits were negligible at best and almost irrelevant to the decay of the cars' Li-ion batteries. EVs regularly charged at fast-charging stations did not display significantly lower values from the *Max Range* variable than those that avoided that practice. Additionally, the results have revealed that the cars' model year (expressed by the *Year* variable) was the only variable that significantly impacted the batteries' capacity. In short, the CRISP-DM methodology answered the research question by identifying the vehicle's year of release as the determining factor for battery degradation, without any of the identified behavioral factors having a meaningful part in the decay effect.

Our work relied on a dataset made public by a forum of EV users, limited to its participants. This approach can be improved by reaching more platforms of EV users, obtaining more responses, and a larger dataset. This could improve the model's accuracy score. In addition, future work could include other factors that affect battery performance, such as battery replacements and data extracted from the vehicles' Battery Management Systems, such as voltage, temperature, and current going in and out of the batteries.

Acknowledgments

This study was performed in the scope of ISCTE collaboration with Santos e Vale, who financed the research.

References

[1] G. Saldaña, J. I. S. Martín, I. Zamora, F. J. Asensio, O. Oñederra, and M. González, "Empirical Electrical and Degradation Model for Electric Vehicle Batteries," *IEEE Access*, vol. 8, pp. 155576–155589, 2020, doi: 10.1109/ACCESS.2020.3019477.

[2] F. Lambert, "Tesla battery degradation at less than 10% after over 160,000 miles, according to latest data.," *Elektrek*, 2018. https://electrek.co/2018/04/14/tesla-battery-degradation-data/ (accessed May 04, 2021).

[3] I. Tesla, “2020 Tesla Impact Report,” 2020. https://www.tesla.com/ns_videos/2020-tesla-impact-report.pdf (accessed Aug. 16, 2021).

[4] R. Y. and B. Z. (eds. . Masson-Delmotte, V., P. Zhai, A. Pirani, S. L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M. I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J. B. R. Matthews, T. K. Maycock, T. Waterfield, O. Yelekçi, “Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change,” Cambridge, 2021. [Online]. Available: https://www.ipcc.ch/report/ar6/wg1/downloads/report/IPCC_AR6_WGI_Full_Report.pdf.

[5] EPA, “United States Environmental Protection Agency. Inventory of US Greenhouse Gas Emissions and Sinks : 1990–2019,” *Report No. EPA 430-R-15-004*, 2019. https://www.epa.gov/ghgemissions/inventory-us-greenhouse-gas-emissions-and-sinks (accessed Feb. 19, 2021).

[6] European Comission, “European Green Deal,” 2021. https://ec.europa.eu/info/strategy/priorities-2019-2024/european-green-deal_en (accessed Aug. 18, 2021).

[7] “Plano Nacional de Energia e Clima 2021-2030,” 2019. Accessed: Aug. 16, 2021. [Online]. Available: https://www.portugalenergia.pt/setor-energetico/bloco-3/],.

[8] C. C. Chan, “The state of the art of electric, hybrid, and fuel cell vehicles,” *Proc. IEEE*, vol. 95, no. 4, pp. 704–718, 2007, doi: 10.1109/JPROC.2007.892489.

[9] G. May and A. El-Shahat, “Battery-degradation model based on the ANN regression function for EV applications,” in *GHTC 2017 - IEEE Global Humanitarian Technology Conference, Proceedings*, 2017, vol. 2017-Janua, pp. 1–3, doi: 10.1109/GHTC.2017.8239301.

[10] M. Singh Ceng and K. Janardhan Reddy, “Predicting Life-Cycle Estimation of Electric Vehicle Battery Pack through Degradation by Self Discharge and Fast Charging,” *SAE Tech. Pap.*, no. 2020, 2020, doi: 10.4271/2020-28-0435.

[11] Z. Yun and W. Qin, “Remaining Useful Life Estimation of Lithium-Ion Batteries Based on Optimal Time Series Health Indicator,” *IEEE Access*, vol. 8, pp. 55447–55461, 2020, doi: 10.1109/ACCESS.2020.2981947.

[12] X. Tan *et al.*, “Real-Time State-of-Health Estimation of Lithium-Ion Batteries Based on the Equivalent Internal Resistance,” *IEEE Access*, vol. 8, pp. 56811–56822, 2020, doi: 10.1109/ACCESS.2020.2979570.

[13] X. S. Hu, H. Yuan, C. F. Zou, Z. Li, and L. Zhang, “Co-Estimation of State of Charge and State of Health for Lithium-Ion Batteries Based on Fractional-Order Calculus,” *Ieee Trans. Veh. Technol.*, vol. 67, no. 11, pp. 10319–10329,

2018, doi: 10.1109/tvt.2018.2865664.

[14] Y. Jiang, J. Zhang, L. Xia, and Y. Liu, "State of Health Estimation for Lithium-Ion Battery Using Empirical Degradation and Error Compensation Models," *IEEE Access*, vol. 8, pp. 123858–123868, 2020, doi: 10.1109/ACCESS.2020.3005229.

[15] M. Aaboud *et al.*, "Search for long-lived, massive particles in events with displaced vertices and missing transverse momentum in root S=13 TeV pp collisions with the ATLAS detector," *Phys. Rev. D*, vol. 97, no. 5, 2018, doi: 10.1103/PhysRevD.97.052012.

[16] J. Kim *et al.*, "Data-Driven State of Health Estimation of Li-Ion Batteries With RPT-Reduced Experimental Data," *IEEE Access*, vol. 7, pp. 106987–106997, 2019, doi: 10.1109/ACCESS.2019.2932719.

[17] K. Li, F. Wei, K. J. Tseng, and B.-H. Soong, "A Practical Lithium-Ion Battery Model for State of Energy and Voltage Responses Prediction Incorporating Temperature and Ageing Effects," *IEEE Trans. Ind. Electron.*, vol. 65, no. 8, pp. 6696–6708, 2018, doi: 10.1109/TIE.2017.2779411.

[18] D. Liu, X. Yin, Y. Song, W. Liu, and Y. Peng, "An On-Line State of Health Estimation of Lithium-Ion Battery Using Unscented Particle Filter," *IEEE Access*, vol. 6, pp. 40990–41001, 2018, doi: 10.1109/ACCESS.2018.2854224.

[19] A. El Mejdoubi, H. Chaoui, H. Gualous, P. Van Den Bossche, N. Omar, and J. Van Mierlo, "Lithium-ion batteries health prognosis considering aging conditions," *IEEE Trans. Power Electron.*, vol. 34, no. 7, pp. 6834–6844, 2019, doi: 10.1109/TPEL.2018.2873247.

[20] J. W. Wei, G. Z. Dong, and Z. H. Chen, "Remaining Useful Life Prediction and State of Health Diagnosis for Lithium-Ion Batteries Using Particle Filter and Support Vector Regression," *Ieee Trans. Ind. Electron.*, vol. 65, no. 7, pp. 5634–5643, 2018, doi: 10.1109/tie.2017.2782224.

[21] C. Vidal *et al.*, "Hybrid Energy Storage System State-Of-Charge Estimation Using Artificial Neural Network for Micro-Hybrid Applications," in *2018 IEEE Transportation and Electrification Conference and Expo, ITEC 2018*, 2018, pp. 868–873, doi: 10.1109/ITEC.2018.8450251.

[22] J. Qu, F. Liu, Y. Ma, and J. Fan, "A Neural-Network-Based Method for RUL Prediction and SoH Monitoring of Lithium-Ion Battery," *IEEE Access*, vol. 7, pp. 87178–87191, 2019, doi: 10.1109/ACCESS.2019.2925468.

[23] C. Vidal, P. Malysz, P. Kollmeyer, and A. Emadi, "Machine Learning Applied to Electrified Vehicle Battery State of Charge and State of Health Estimation: State-of-the-Art," *IEEE Access*, vol. 8, pp. 52796–52814, 2020, doi: 10.1109/ACCESS.2020.2980961.

[24] "Cross Industry Standard Process for Data Mining (CRISP-DM)," 1996. https://www.datascience-pm.com/crisp-dm-2/ (accessed Aug. 18, 2021).

[25] “Python.org.” https://www.python.org/ (accessed Aug. 24, 2021).

[26] “Project Jupyter.” https://jupyter.org/ (accessed Aug. 24, 2021).

[27] F. Pedregosa, G. Varoquaux, and A. Gramfort, “Scikit-learn: Machine Learning in Python,” *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.

[28] · N S Altman and N. S. Altman, “BU-1065MA An Introduction to Kernel and Nearest Neighbor Nonparametric Regression An Introduction to Kernel and Nearest Neighbor Nonparametric Regression,” 1991.

[29] S. H. Walker and D. B. Duncan, “Estimation of the Probability of an Event as a Function of Several Independent Variables,” *Biometrika*, vol. 54, no. 1/2, p. 167, Jun. 1967, doi: 10.2307/2333860.

[30] G. H. John and P. Langley, “Estimating Continuous Distributions in Bayesian Classifiers,” Accessed: Aug. 20, 2021. [Online]. Available: http://robotics.

[31] “sklearn.svm.LinearSVC — scikit-learn 0.24.2 documentation.” https://scikit-learn.org/stable/modules/generated/sklearn.svm.LinearSVC.html#sklearn.svm.LinearSVC (accessed Aug. 20, 2021).

[32] “RBF SVM parameters — scikit-learn 0.24.2 documentation.” https://scikit-learn.org/stable/auto_examples/svm/plot_rbf_parameters.html (accessed Aug. 20, 2021).

[33] P. Petersen, A. T. Thorgeirsson, S. Scheubner, S. Otten, F. Gauterin, and E. Sax, “Training and validation methodology for range estimation algorithms,” *VEHITS 2019 - Proc. 5th Int. Conf. Veh. Technol. Intell. Transp. Syst.*, pp. 434–443, 2019, doi: 10.5220/0007717004340443.

[34] S. Pokharel, P. Sah, and D. Ganta, “Improved Prediction of Total Energy Consumption and Feature Analysis in Electric Vehicles Using Machine Learning and Shapley Additive Explanations Method,” *World Electr. Veh. J. 2021, Vol. 12, Page 94*, vol. 12, no. 3, p. 94, Jun. 2021, doi: 10.3390/WEVJ12030094.

[35] “3.1. Cross-validation: evaluating estimator performance — scikit-learn 0.24.2 documentation.” https://scikit-learn.org/stable/modules/cross_validation.html (accessed Aug. 19, 2021).

[36] “The Danger of Overfitting Regression Models.” https://blog.minitab.com/en/adventures-in-statistics-2/the-danger-of-overfitting-regression-models (accessed Aug. 20, 2021).