Predicting Electric Vehicle Adoption in the EU: Analyzing Classification Performance and Influencing Attributes Across Countries, Gender, and Education Level

Mert Kumbasar
Computer Engineering Department
Çankaya University
Ankara, Turkey
c2111009@student.cankaya.edu.tr

Gül Tokdemir
Computer Engineering Department
Çankaya University
Ankara, Turkey
gtokdemir@cankaya.edu.tr

Gurdal Ertek
College of Business and Economics
United Arab Emirates University
Al Ain, UAE
gurdal@uaeu.ac.ae

Thouraya Gherissi Labben
College of Business and Economics
United Arab Emirates University
Al Ain, UAE
thouraya.labben@uaeu.ac.ae

***Abstract*—Electric vehicles (EVs) have been one of the trending technologies in recent decades, as they are expected to transform the current automotive technology and transportation systems. To this end, the scope of this study is analyzing survey data on European consumers' EV purchase decisions. The objective is comparing the predictive quality of various classification algorithms in predicting EV adoption, across country, gender and education level of the participants, as well as the analysis of the influencing attributes. Initially, the data is filtered for each value of the chosen categorical attribute (country, gender or education level) with the missing values being imputed. Then, several classification algorithms in the Python sklearn package are applied through 5-fold-cross validation and the performance of the algorithms are compared based on standard classification metrics. There are notable variations in classification performance and influencing attributes depending on the values of the selected categorical attributes.**

Keywords—Electric Vehicles (EVs), Market Adoption, Sustainable Development Goals (SDG), Machine Learning, Feature Ranking, Classification Algorithms

I. INTRODUCTION

Electric vehicles (EV) are considered as more environmentally friendly compared to traditional fuel-engine vehicles, as they significantly reduce the release of greenhouse gases like carbon dioxide and harmful gases [1]. However, the claim regarding eco-friendliness of EVs has also been questioned, resulting in contradicting evidence under certain conditions [2]. Challenges such as high cost, lack of necessary infrastructures, long waiting lists, short range [3] remain as critical obstacles for EV adoption. Consumers' emotions and sentiments play a crucial role in the adoption of EV technology [4]. Within the same research stream, a survey was conducted with 6108 respondents in four EU countries (Denmark, Germany, Hungary, Norway) to understand the relevance and significance of the attributes affecting European consumers' EV purchase decisions [5].

The adoption of electric vehicle (EV) technology has gained the interest of many research scholars who have analyzed the challenges of its adoption from various dimensions such as economics [6], infrastructure [7], consumers [8], and regulations [9].

Numerous attributes are mentioned that lead to the adoption of EVs among such as cost [10], performance [11], charging infrastructure [12]. The adoption of electric vehicles (EVs) is also hindered by barriers or challenges such as vehicles design and attributes [13], psychological characteristics [8], perceived risks [7] and environmental impact [14].

[15] presented a review of the literature on mathematical modeling of EV adoption. A number of studies have applied machine learning techniques to predict EV adoption behaviour [16][17][18] and but they did not focus the country and gender differences. The primary objective of this study is to compare the prediction of EV adoption in various EU countries and in different genders applying machine learning techniques on survey data shared by [5].

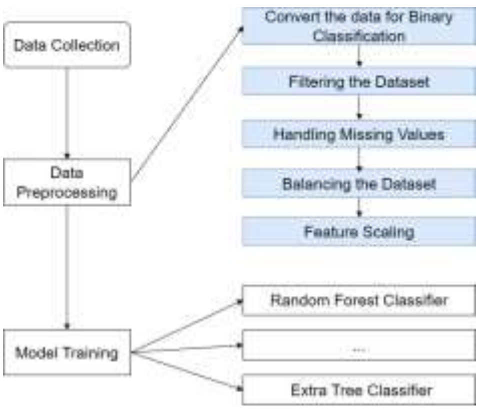

Fig. 1. The data analytics flow chart

The study is an extension of a recent earlier study which analyzed the same dataset [19] in the research study [18]. While [18] analyzed the complete data [19] derived from the data of [5], in this study, subsets of the same data [19] are analyzed separately, and the results are compared.

II. METHODOLOGY

The methodology applied in the study is provided in Fig. 1. Upon collection of data, preprocessing was applied, starting with converting the target attribute to become binary. For each subset, filtering was needed, while imputing to handle missing values. In order to analyze the effect of balanced distribution SMOTE (Synthetic Minority Over-sampling Technique) method was applied [20]. For the numeric attributes standard scaling was applied to ensure optimal performance for all models.

Python sklearn [21] package was used to apply 25 different classification algorithms to explore country- gender- and education-level-based differences in the adoption of EVs. These classification algorithms are listed in the Supplement Document [22]. The sklearn package contains efficient and effective implementations of many of the most used methods, and thus was selected. To avoid performance issues related to packages themselves, rather than the algorithms, only the algorithms in sklearn were used. The code in Python 3.12 is implemented in Jupyter notebook platform. The developed and applied Python code that can be found as open source under GitHub [23].

~~We analyzed t~~The performance of the models was analyzed by 5-fold cross-validation and analyzed the results with respect to several classification performance metrics. While different number of folds can be chosen and tested, the choice of five folds has been encountered in many studies in the academic literature, supporting the choice. Furthermore, for each data subset analyzed, ~~we identified~~ the top ranking attributes were identified based on the Gini coefficient metric [24].

III. ~~PREPARE YOUR PAPER BEFORE STYLING~~RESULT~~D~~S OF THE EXPERIMENTS

A. Data

The data [19] consists of responses to a transnational survey amongst European consumers, regarding the attributes influencing consumer decisions regarding electric vehicle (EV) adoption [5][18]. The survey was conducted within the scope of the EU-funded proEME project (pro-eme.eu), which aimed at understanding "public knowledge and perception of EVs, and to identify the misconceived topics about EVs". The key areas covered in the survey included driving habits, environmental concerns, financial considerations and perceptions of EV technology.

The subset of the original survey results consists of 69 binary attributes, showing the Yes/No responses to the survey question choices, as well as one class attribute (Q16), which asks to choose the statement that suits most on EV choice. The responses have four alternatives which were "(A) If I had an EV, it would be my only car" (2883 samples), "(B) If I had an EV, it would be a supplement to a petrol or diesel car" (1526 samples), "(C) I would never buy an EV" (892 samples) and "(D) I do not know" (807 samples).

B. Data Preparation

When the original values of the target attribute (Q16) were analyzed, the unbalanced distribution of the different class labels was recognized. This was handled by converting the problem into a binary classification problem by combining responses A and B into a single class label, AB, representing the purchase intention and adoption which was encoded as 1. Similarly, responses C and D were merged under the class label CD, indicating a lack of purchase intention (No Adoption), which was ~~labeled~~labelled as 0 for classification. The original and transformed distribution among class labels are provided in Fig. 2. A major motivation for merging 4 classes to 2 was to increase the predictive performance and make the models practically usable. Yet, an even more important reason was to reduce complexity and vagueness that was introduced by the survey design, and to instead simplify the adoption decision by the consumers to a yes or no.

Fig 2. Original and transformed distribution of the class labels.

TABLE I. ~~TABLE 1.~~ ATTRIBUTES THAT PERFORMED IN TOP FIVE ACROSS THE DATA SUBSETS AND THE TOTAL COUNT OF ENTERING TOP FIVE~~.~~

Questions	Content	Count
Q12_3	Range	1
Q12_4	Purchase price	2
Q13_2	Positive effect on global climate	3
Q13_3	Less noise	1
Q14	Society must reward electric cars	8
Q15	Tried to drive	5
Q17	If had only an electric car, how good or bad it would suit	7
Q18_1	Electric cars are boring	3
Q18_3	Greater risk of bursting into flames	3
Q18_10	Can not drive fast enough on the highway	2
Q18_14	Only suitable as second car	2
Q18_19	More expensive in daily operation	1
Q21	How opinion changed during the past year	7

TABLE II. ~~TABLE 2.~~ ATTRIBUTE (FEATURE) RANKINGS FOR EACH COUNTRY~~.~~

Country	Top Five Attribute
Germany	Q21, Q14, Q18_1, Q12_3, Q15
Hungary	Q21, Q14, Q18_10, Q13_3, Q13_2
Norway	Q14, Q15, Q18_14, Q17, Q12_4
Denmark	Q18_1, Q17, Q21, Q15, Q18_3

TABLE III. ~~TABLE 3.~~ ATTRIBUTE (FEATURE) RANKINGS FOR GENDER~~.~~

Gender	Top Five Attributes
Male	Q21, Q17, Q13_2, Q15, Q14
Female	Q17, Q14, Q21, Q13_2, Q12_4

Formatted: table head

Formatted: table head

TABLE IV. ~~Table 4.~~ ATTRIBUTE (FEATURE) RANKINGS FOR EDUCATION LEVEL~~.~~

Education	Top Five Attributes
Low	Q14, Q18_1, Q21, Q17, Q18_14
Medium	Q21, Q17, Q14, Q18_3, Q18_19
High	Q17, Q15, Q18_3, Q14, Q18_10

TABLE V.
~~Table 5.~~ ALGORITHMS THAT PERFORMED IN TOP FIVE ACROSS ALL DATA SUBSETS AND THE TOTAL COUNT OF ENTERING TOP FIVE.

Algorithm	Code	Count
GradientBoostingClassifier	GBC	9
HistGradientBoostingClassifier	HGBC	9
RandomForestClassifier	RFC	9
ExtraTreesClassifier	ETC	9
SVC	SVC	3
MLPClassifier	MLPC	3
AdaBoostClassifier	ABC	2
BaggingClassifier	BC	1

TABLE VI. ~~Table 6.~~ TOP FIVE CLASSIFICATION ALGORITHMS AND THEIR CV ACCURACY FOR EACH COUNTRY~~.~~

Country	Top Five Classification Algorithms
Germany	RFC (0.887), HGBC (0.885), GBC (0.883), ETC (0.883), BC (0.863)~~, ETC~~
Hungary	ETC (0.932), HGBC (0.930), RFC (0.929), ~~MLPC,~~ GBC (0.919), MLPC (0.908)
Norway	RFC~~C,~~ (-0.877) ~~HGBC,~~ HGBC (0.875), GBC (0.871)~~, SVC~~, ETC (0.870), SVC (0.849)
Denmark	ETC (0.885), RFC (0.850), GBC (0.845)~~, SVC~~, HGBC (0.84)5, SVC (0.833)

TABLE VII. ~~Table 7.~~ TOP FIVE CLASSIFICATION ALGORITHMS AND THEIR CV ACCURACY FOR EACH GENDER~~.~~

Gender	Top Five Classification Algorithms
Male	HGBC (0.914), RFC (0. 910), ETC (0.908)~~,SVC,~~ GBC (0.903), SVC (0.891)~~ABC~~
Female	ETC (0.866), HGBC (0.862), RFC (0.861), ~~ABC,~~ GBC (0.858), ABC (0.845)

TABLE VIII. ~~Table 8.~~ TOP FIVE CLASSIFICATION ALGORITHMS AND THEIR CV ACCURACY FOR EACH EDUCATION LEVEL~~.~~

Education	Top Five Classification Algorithms
Low	ETC (0.866), RFC (0.863), GBC (0.863)~~, ABC~~, HGBC (0.858), ABC (0.853)
Middle	ETC (0.893), RFC (0.889), HGBC (0.887), ~~MLPC,~~ GBC (0.881), MLPC (0.865)
High	ETC (0.909), HGBC~~,~~ (0.903) RFC (0.901), GBC (0.897), MLPC (0.890)

In order to handle detected bias in the distribution of the class labels data balancing method namely SMOTE was applied that yielded equal number of class labels (i.e. class label values of 1 and 0) [20]. The distribution of values for Q16 in the original, transformed, and transformed and balanced datasets are presented in the Supplement Document [22].

C. Results of Ranking

Table 1 presents the count of how many times each attribute appeared as one of the top five ranking attributes according to Gini coefficient. Tables 2-4 show, in ranked order of importance, the top five ranking attributes based on the Gini coefficient.

D. Results of Classification

Table 5 presents the count of how many times each classification algorithm appeared as one of the top five ranking algorithms. The algorithms that perform among the top five most for the different subsets are GradientBoostingClassifier, HistGradientBoostingClassifier, and RandomForestClassifier. Tables 6-8 show, in ranked order of performance with respect to Classification Accuracy (CA), the top five ranking attributes based on the Gini coefficient. Fig. 3. shows the ROC (Receiving Operating Characteristic) curves for each data subset, showing a high performance overall across all subsets.

The detailed classification results are provided in the Supplement Document [22].

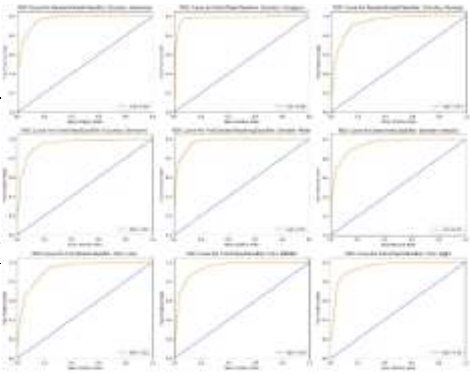

Fig 3. ROC (Receiving Operating Characteristic) curves for each data subset.

IV. DISCUSSION

The differences in the top ranking attributes and performance of different classification algorithms imply that such a segmented analysis is needed in understanding attributes behind EV adoption. The observations also call for customized strategies to address concerns in each market segment to be able to increase adoption. For different market segments, marketing strategies can be customized to address the specific concerns of that market segment.

If two-dimensional market segmentation is carried out, such as both country and gender levels, either that two-dimensional sub-segment can be custom analyzed by filtering the observations, or attributes common to the two components of the sub-segment can be chosen. As an example to the latter strategy, let us consider a marketing strategy to be developed for Male customers in Germany. In this case, the following approach can be followed: Germany has Q21, Q14, Q18_1, Q12_3, Q15 as the top-ranking attributes. Male customers, on the other hand, have Q21, Q17, Q13_2, Q15, Q14 as the top-ranking attributes across Europe. The common attributes for these two segments are Q21 and Q14. Therefore, a marketing strategy targeted for Male customers in Germany would focus on Q21 and Q14, which are how opinion changed during past year (Q21) and the idea that “the society must reward electric cars instead of petrol and diesel cars” (Q14).

Formatted: table head

Formatted: table head

Formatted: table head

Formatted: table head, Left

Regarding the algorithms, an implication of the presented research is that ensemble methods and top-performing algorithms identified in this study should be included as potential best performers and at the least, benchmarks, in predictive analytics studies on EV adoption.

V. CONCLUSIONS

The approach of identifying top-ranking attributes and top-performing classification algorithms is not new and can be observed in many studies. However, the current study, to the best knowledge of authors, is the first study where the approach is applied to predict EV adoption, identify the top-ranking attributes, and compare top-performing algorithms. The full data used [18], full analysis results [22], and the source code [23] for the study are available online.

Future research on the topic can focus on two-dimensional or even three-dimensional segmentation. Such an approach would yield results specific to selected segment combinations (ex: Germany and Male), while may suffer from accuracy losses due to reduced sub-sample size.

Another future area of research could be running automated machine learning (AutoML) [25] on the current dataset and sub-samples based on the segments. AutoML can be run through packages/libraries for programming languages, such as AutoGluon-Tabular [26] and Auto-Sklearn for the Python language [27], using custom AutoML desktop tools, such as Auto Model by Altair AI Studio [28], and/or through online services, such as AutoTrain of HuggingFace [29].

REFERENCES

[1] N.C. Onat, G.M. Abdella, M. Kucukvar, A.A. Kutty, M. AlNuaimi, G. Kumbaroğlu, M. Bulu, "How eco-efficient are electric vehicles across Europe? A regionalized life cycle assessment-based eco-efficiency analysis," Sustainable Development, vol. 29, no. 5, pp. 941-956, 2021. https://doi.org/10.1002/sd.2186

[2] P. Lu, S. Hamori, L. Sun, S. Tian, "Does the Electric Vehicle industry help achieve sustainable development goals? -- Evidence from China," Frontiers in Environmental Science vol. 11, 1276382, 2023. https://www.frontiersin.org/articles/10.3389/fenvs.2023.1276382/

[3] F. Alanazi, "Electric Vehicles: Benefits, Challenges, and Potential Solutions for Widespread Adaptation," Applied Sciences, vol. 13, no. 10, 6016, 2023. https://doi.org/10.3390/app13106016

[4] G. Schuitema, J. Anable, S. Skippon, N. Kinnear, "The role of instrumental, hedonic and symbolic attributes in the intention to adopt electric vehicles," Transportation Research A: Policy Pract., vol. 48, pp. 39-49, 2013. https://doi.org/10.1016/j.tra.2012.10.004

[5] M. Westerhof, J.R. Reyes García, S. Haveman, G.M. Bonnema, "Transnational survey data on European consumers' attitude and perceived knowledge about electric vehicles," Data in Brief, vol. 49:109378. 2023. https://doi.org/10.1016/j.dib.2023.109378

[6] H. Quak, N. Nesterova, T. van Rooijen, "Possibilities and barriers for using electric-powered vehicles in city logistics practice". Trans. Res. Procedia, vol. 12, pp. 157-169, 2016. https://doi.org/10.1016/j.trpro.2016.02.055

[7] N. Berkeley, D. Jarvis, A. Jones, "Analysing the take up of battery electric vehicles: an investigation of barriers amongst drivers in the UK," Transp. Res. D Transp. Environ, vol. 63, pp. 466-481, 2018. https://doi.org/10.1016/j.trd.2018.06.016

[8] L. He, M. Wang, W. Chen, G. Conzelmann, "Incorporating social impact on new product adoption in choice modeling: a case study in green vehicles," Transp. Res. Part D, vol. 32, pp. 421-434, 2014. https://doi.org/10.1115/detc2012-71123

[9] K.Y. Bjerkan, T.E. Nørbech, M.E. Nordtømme, "Incentives for promoting battery electric vehicle (BEV) adoption in Norway," Transp. Res. D Transp. Environ, vol. 43, pp. 169-180, 2016. https://doi.org/10.1016/j.trd.2015.12.002

[10] K. Palmer, J.E. Tate, Z. Wadud, J. Nellthorp, "Total cost of ownership and market share for hybrid and electric vehicles in the UK, US and Japan," Appl. Energy, vol. 209, pp. 108-119, 2018. https://doi.org/10.1016/j.apenergy.2017.10.089

[11] B.K. Sovacool, W. Abrahamse, L. Zhang, J. Ren, "Pleasure or profit? Surveying the purchasing intentions of potential electric vehicle adopters in China," Transp. Res. A Policy Pract., vol. 124, pp. 69-81, 2019. https://doi.org/10.1016/j.tra.2019.03.002

[12] S.M. Skippon, N. Kinnear, L. Lloyd, J. Stannard, "How experience of use influences mass-market drivers' willingness to consider a battery electric vehicle: a randomised controlled trial," Transp. Res. A Policy Pract., vol. 92, pp. 26-42, 2016. https://doi.org/10.1016/j.tra.2016.06.034

[13] M. Mohamed, C.D. Higgins, M. Ferguson, W.J. Requia, "The influence of vehicle body type in shaping behavioural intention to acquire electric vehicles: a multi-group structural equation approach," Transp. Res. A Policy Pract., vol. 116, pp. 54-72, 2018. https://doi.org/10.1016/j.tra.2018.05.011

[14] N.C. Onat, M. Kucukvar, O. Tatari, "Well-to-wheel water footprints of conventional versus electric vehicles in the United States: a state-based comparative analysis," J. Clean. Prod., vol. 204, pp. 788-802, 2018. https://doi.org/10.1016/j.jclepro.2018.09.010

[15] L. Maybury, P. Corcoran, L. Cipcigan, "Mathematical modelling of electric vehicle adoption: A systematic literature review," Transportation Research Part D: Transport and Environment, vol. 107, 103278, 2022. https://doi.org/10.1016/j.trd.2022.103278

[16] J. Bas, C. Cirillo, E. Cherchi, "Classification of potential electric vehicle purchasers: A machine learning approach," Technological Forecasting and Social Change, vol. 168, 120759, 2021. https://doi.org/10.1016/j.techfore.2021.120759

[17] J. Jia, B. Shi, F. Che, H. Zhang, "Predicting the regional adoption of Electric Vehicle (EV) with comprehensive models," IEEE Access, vol. 8, pp. 147275-147285, 2020, https://10.1109/ACCESS.2020.3014851

[18] G. Ertek, "Survey Data on European Consumers' Adoption Attitude Towards Electric Vehicles (EVs)," Mendeley Data, V1, 2024. https://doi.org/10.17632/rz8pctwrbv.1

[19] S. AlRashdi, A. AlHassani, F. Haile, R. AlNuaimi, T. Labben, G. Ertek, "Comparing the Performance of Classification Algorithms for Predicting Electric Vehicle Adoption," ICBAP 2024, Sharjah, UAE. Lecture Notes in Operations res., Ali Emrouznejad et al: Business Analytics and Decision Making In Practice, 978-3-031-61588-7. Springer. https://doi.org/10.1007/978-3-031-61589-4_17

[20] N.V. Chawla, K.W. Bowyer, L.O. Hall, W.P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," Journal of Artificial Intelligence Research, vol. 16, pp. 321-357, 2002. https://doi.org/10.1613/jair.953

[21] scikit-learn: Machine Learning in Python. https://scikit-learn.org

[22] M. Kumbasar, G. Tokdemir, G. Ertek, T.G. Labben. "Supplement to 'Predicting Electric Vehicle Adoption in the EU: Analyzing Classification Performance and Influencing Attributes Across Countries, Gender, and Education Level'". https://ertekprojects.com/ftp/supp/26.pdf (to be uploaded upon the acceptance of the paper)

[23] electric-vehicle-adoption, Python code to analyze electric-vehicle (EV) adoption (code developed by Mert Kumbasar). GitHub. https://github.com/gurdalertek/electric-vehicle-adoption/tree/main

[24] L. Ceriani, P. Verme, "The origins of the Gini index: extracts from Variabilità e Mutabilità (1912) by Corrado Gini," J Econ Inequal, vol. 10, pp. 421–443, 2012. https://doi.org/10.1007/s10888-011-9188-x

[25] P. Gijsbers, M.L. Bueno, S. Coors, E. LeDell, S. Poirier, J. Thomas, B. Bischl, J. Vanschoren, "AMLB: an AutoML benchmark," Journal of Machine Learning Research, vol. 25, no. 101, pp. 1-65, 2024. https://www.jmlr.org/papers/volume25/22-0493/22-0493.pdf

[26] N. Erickson, J. Mueller, A. Shirkov, H. Zhang, P. Larroy, M. Li, A. Smola, "AutoGluon-Tabular: Robust and accurate AutoML for structured data," arXiv preprint arXiv:2003.06505.

[27] Auto-Sklearn. GitHub. https://github.com/automl/auto-sklearn

[28] Auto Model, Altair AI Studio. https://docs.rapidminer.com/latest/studio/guided/auto-model/

[29] AutoTrain. Hugging Face. https://huggingface.co/autotrain/