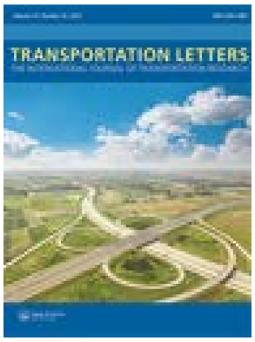

Transportation Letters

The International Journal of Transportation Research

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ytrl20

An interpretable machine learning approach to understanding the impacts of attitudinal and ridesourcing factors on electric vehicle adoption

Javier Bas, Zhenpeng Zou & Cinzia Cirillo

TRANSPORTATION LETTERS
https://doi.org/10.1080/19427867.2021.2009098

An interpretable machine learning approach to understanding the impacts of attitudinal and ridesourcing factors on electric vehicle adoption

Javier Bas [a], Zhenpeng Zou [b] and Cinzia Cirillo [c]

[a]Department of Economics, Universidad de Alcalá, 0.7 Facultad de Ciencias Económicas, Madrid, Spain; [b]The National Center for Smart Growth, University of Maryland, College Park, MD, USA; [c]Department of Civil and Environmental Engineering, University of Maryland, College Park, MD, USA

ABSTRACT

The global electric vehicle (EV) market has been experiencing an impressive growth in recent times. Understanding consumer preferences on this cleaner, more eco-friendly mobility option could help guide public policy toward accelerating EV adoption and sustainable transportation systems. Previous studies suggest the strong influence of individual and external factors on EV adoption decisions. In this study, we apply machine learning techniques on EV stated preference survey data to predict EV adoption using attitudinal factors, ridesourcing factors (e.g., frequency of Uber/Lyft rides), as well as underlying socio-demographic and vehicle factors. To overcome machine learning models' low interpretability, we adopt the innovative Local Interpretable Model-Agnostic Explanations (LIME) method to elaborate each factor's contribution to the predicting outcomes. Besides what was found in previous EV preference literature, we find that the frequent usage of ridesourcing, knowledge about EVs, and awareness of environmental protection are important factors in explaining high willingness of adopting EVs.

KEYWORDS

Electric vehicles; attitudes; ridesourcing; machine learning; local interpretable model-agnostic explanations (lime)

Introduction

The widespread adoption of electric vehicles (EVs) has been seen as an effective tool to address the global challenge of climate change, provided that electricity is harnessed from a clean energy source, such as water, wind, and solar (MacInnis and Krosnick 2020). EV adoption has also become a viable pathway toward sustainability for a major carbon emitter like the United States, where 92% of its CO_2 emission comes from fossil fuel combustion that is largely related to the transportation sector (The United States Environmental Protection Agency (EPA) 2021). In European countries, EV adoption is considered a national strategy for combating climate change as a substantial reduction in carbon emission could be achieved via EV's mass uptake (Canals Casals et al. 2016). With such visions, the global electric vehicle market experienced a spectacular rise in sales and popularity during the past decade: In 2019, EVs accounted for 2.6% of global car sales and 1% of global car stock – an impressive 40% year-on-year increase (The International Energy Agency (IEA) 2020). We can expect EVs to play a significant role in helping many countries meet their global net zero emission goals by the mid-21st century.

The accelerating global EV adoption is attributed to multiple factors. From the supply side, the cost of EV production steadily declines as battery cost has decreased more than 85% since 2010 (The International Energy Agency (IEA) 2020). This helps lower EV prices in the competition against traditional gasoline cars. At the same time, EV manufacturers have improved battery capacity to accommodate long-range driving and developed the battery charging technology that significantly shortens the recharge time (The International Energy Agency (IEA) 2020). From the policy side, governments provide financial incentives, such as tax rebates to EV manufacturers and consumers, and build the essential infrastructure, such as public charging stations, to boost EV adoption (Liao, Molin, and van Wee 2017). From an international perspective, many countries are embracing EV technologies as a part of the global initiative of reducing energy consumption and greenhouse gas emissions, combating climate change, and achieving a more sustainable and eco-friendly growth (UNEP, n.d.). Last but not least, consumers start to see EV as a robust alternative to gasoline cars now that the production technologies have matured. Consumer willingness to purchase an EV can be influenced by their familiarity with the technology (Jensen, Cherchi, and Mabit 2013), their attitudes toward its multitudes of environmental and social benefits (Axsen, Bailey, and Castro 2015), their social network's influence (Kim, Rasouli, and Timmermans 2014), their sociodemographic characteristics, and their job/home locations (Sovacool et al. 2019).

Despite the auto industry's enthusiasm and consumer's optimism toward EV adoption, we should not neglect certain barriers to EV adoption that are yet to be cleared, such as retail relationships of the EV market through dismissive auto dealership (Zarazua de Rubens, Noel, and Sovacool 2018), concerns about the technology and infrastructure readiness (Vassileva and Campillo 2017), as well as the mobility justice discourse regarding who should be held accountable for the emissions from EV production and disposal processes – usually the wealthy nations who lead in both energy consumption and EV adoption (Henderson 2020).

In this study, we also examine an emerging factor that may influence the EV adoption besides the aforementioned ones: the usage of ridesourcing services (Uber, Lyft, and their equivalents). Both shared mobility and electric vehicles are umbrellaed under the global concept of 'new mobility', updating the transportation system with emerging technologies and business models. Ridesourcing services, supported by transportation network companies (TNCs) like Uber and Lyft, are well covered in urban areas. More than one-third of the U.S. adults said to have used ridesourcing services, according to a Pew Research Center survey conducted in fall 2018

CONTACT Javier Bas Vicente javier.bas@uah.es Department of Economics, Universidad de Alcalá 0.7 Facultad de Ciencias Económicas, Empresariales y Turismo, Alcalá de Henares, Madrid 28802, Spain

(Jiang 2019). It is safe to claim that while not many people have experienced EVs, many have tried ridesourcing services in recent times. We are interested in empirically understanding whether a consumer's experience with ridesourcing could influence her decision to purchase an EV. The affinities for shared mobility and EVs are closely related to one another (Burghard and Elisabeth 2019). In particular, both shared mobility and EV mobility visions address the sustainability challenges of today's transportation system. In addition to ridesourcing factors, we examine a list of attitudinal factors that could explain one's willingness to adopt the EV technology.

We adopt innovative machine learning (ML) approaches in this study. Thanks to their capacity to process big data, ML techniques have been widely applied to the information and technological fields, such as fraud detection, robotics, spam filtering, translation services, preventive health care, and object detection. In recent years, ML has also been applied in transportation research on topics like congestion reduction, safety improvement, environmental impact studies, and energy consumption optimization. One of the distinct characteristics of ML is the non-parametric modeling structure, meaning that the predictor does not take a predetermined form (such as linear or quadratic form). The data-driven nature relaxes a traditional parametric model's assumptions on its modeling structure and relationships between features. However, there are also concerns over such non-parametric approach, including the requirement of large-size data to train ML models, the issue of overfitting (where the model fits too well for the training data and loses predictability on new data), and low interpretability of the results. Regarding interpretability, the majority of ML algorithms used today are black boxes. The final throughputs – the performance metrics of an ML algorithm – focus on predictability. Although predictability prevails over interpretability in many ML applications, many researchers find it meaningful to elaborate the underlying relations between ML predictors and prediction. With recent attempts to develop interpretable ML models, researchers have made significant progress toward building ML models with both predictability and interpretability, such as (Zhao et al. n.d.; Lakkaraju et al. 2017).

To address the limited interpretability of ML models, we adopt a novel ML technique called Local Interpretable Model-Agnostic Explanations (LIME). Specifically, we hope to understand each predictor's contribution to the prediction outcomes. The predictors of interest include the usage of ridesourcing services and an individual's attitudes toward promoting the EV technology and protecting the environment. We believe that providing interpretation to these factors by applying this novel technique is a qualitative leap. For the first time, it is possible to offer valuable insights to researchers who can benefit from the high predictability of non-parametric ML models when solving similar empirical problems, while being able to interpret the modeling results.

The rest of the paper is organized as follows: Section 2 reviews literature on EV adoption and ML methods; Section 3 provides an overview of the technical aspects of supervised ML and the LIME method. Section 4 summarizes the data used in this study; Section 5 exhibits the results. Finally, Section 6 provides concluding remarks.

Literature review

In this section, we focus on the existing literature on EV adoption and the ML applications in new mobility research. EV research has grown substantially in volume as EV's popularity has risen. We will summarize key findings from both quantitative and qualitative research on the influential factors for EV adoption. Furthermore, we will discuss literature's findings on the comparative advantage of ML approaches over traditional statistical models in new mobility studies. Finally, we will briefly cover the limited research evidence linking EV adoption to shared mobility.

Factors influencing EV adoption

The consumer preferences for EV are influenced by both individual factors and external factors (Liao, Molin, and van Wee 2017). On individual factors, we learn from literature that EV ownership is significantly positively associated with income level (Sovacool et al. 2019), familiarity with EV technology (Jensen, Cherchi, and Mabit 2013), the pro-environment attitude (Axsen, Bailey, and Castro 2015), Hands-on EV experience (Rezvani, Jansson, and Bodin 2015), and peer influence/social norms (Carley, Siddiki, and Nicholson-Crotty 2019). Other factors are either insignificant or bring mixed results toward EV adoption, such as education level (Sierzchula et al. 2014), the number of owned vehicles (Javid and Neja, 2017), and population density (Sierzchula et al. 2014). Due to the novelty of EV and the differences in EV market, the individual heterogeneities in EV preferences are inevitable.

Consumers are also responsive to external factors on EV adoption. Some of them are related to the vehicle itself, including price, battery range, and charging time barriers (Carley, Siddiki, and Nicholson-Crotty 2019). Some of the external factors are related to financial and non-financial incentives offered by government on EV adoption. Financial incentives, including cash refund or tax rebate on EV purchases, have a positive impact on EV adoption (Rezvani, Jansson, and Bodin 2015). Similarly, non-financial incentives, such as state funded EV infrastructure, discounted parking cost, toll fee/ road charge waiver, and licensing incentives, all encourage EV adoption to various degrees (Hardman 2019; He et al. 2021). Without diving deep into the wealth of literature on various factors that influence EV adoption, we aim to address a unique angle on the relationship between usage (and awareness) of shared mobility on the potential EV purchases in this study.

We find few empirical studies that attempt to build the connection between EV adoption and shared mobility. (Jenn, Laberteaux, and Clewlow 2018) quantitatively examine the factors that influence shared mobility (car-sharing and ridesourcing) usage. They identify a positive correlation between the frequency of ridesourcing rides and their willingness to buy an EV. Our study focuses on potential EV consumers using EV survey data, whereas theirs focuses on car-sharing and ridesourcing user characteristics and preferences. Therefore, we can describe more nuanced information on EV preference attributes. Internationally, (Burghard and Elisabeth 2019) find evidence that car-sharing users are likely to purchase EV in a survey study about electric car-sharing in Germany. The other connection between shared mobility and EV lies in the possible mobility vision of using shared electric vehicles (SEVs) as a smart mobility solution for congestion and emission reduction (Taiebat and Xu 2019). In summary, our study is unique in that it directly addresses the connection between ridesourcing usage and EV adoption.

Machine learning research on new mobility

ML techniques have been applied to research on either new mobility options or alternative energy vehicles for ML's high predictability. Researchers use ML models to predict battery life, energy consumption, and mileage range for alternative fuel vehicles (Zahid et al. 2018; Fukushima et al. 2018). Furthermore, ML techniques are used in research on the intelligent transportation system (ITS), which often involves big data analytics. Such research topics include stop delivery times prediction (Hughes et al. 2019), traffic flow and mobility profiles estimation (Liu et al. 2019; Parsa et al. 2020; Sun, Leurent, and Xie 2021), driving behavior recognition (Yi et al. 2019), and parking occupancy prediction (Yang et al. 2019). However, ML approaches are oftentimes adopted in academic research solely for the purpose of comparing their predictability against traditional statistical models. For instance, (Martín-Baos, García-Ródenas, and Rodriguez-Benitez 2021) compare predictability between a random utility model that uses the kernel logistic regression (KLR, an ML approach) for utility specification and other models like the multinomial logit (MNL) method, support vector machines (SVMs), and random forest (RF). Many studies compare prediction accuracy among multiple ML algorithms, such as using cellphone accelerometer and gyroscope data to predict transportation mode choice (Jahangiri and Rakha 2015), using speed and acceleration data to predict driving conditions (Huang, Tan, and He 2011), using EV data to predict EV distance range (Sun et al. 2019), and using survey data to classify potential EV buyers (Bas, Cirillo, and Cherchi 2021a). In many cases, the more structurally complexed algorithms would outperform the more conventional ones. Nevertheless, researchers point out that predictability significantly depends on data quality and processing. It is difficult to claim that certain ML algorithm is a clear winner for all types of prediction problems. In general, ML shows promising outcomes in predictability as compared to traditional statistical models.

ML techniques have also been adopted in new mobility research using survey data. (Lee et al. 2019) apply the gradient boosting machine (GBM) method to understand the user preferences related to autonomous vehicles (AVs). They include a number of attitudinal factors in the survey study, such as pro-AV sentiments, environmental concerns, interest in AV technology, and attitudes toward public transit. (Zarazua de Rubens 2019) applies a K-means clustering algorithm in a market segmentation analysis on potential EV adopters using EV survey data. Based on socio-economic characteristics, vehicle preference, and stated interest in EV, the author derives six market segments. Finally, ML methods have also been adopted to understand ridesourcing behavior. For instance, (Chen, Zahiri, and Zhang 2017) use an ensemble tree learning method in a combined classification problem to determine a ridesourcing trip's service type.

Apart from identifying the ML model that best predicts EV adoption using attitudinal and ridesourcing factors, another contribution of this paper to current literature is adopting the novel LIME method to interpret highly abstract ML results. To the best of our knowledge, LIME has not been as widely adopted in transportation studies as in other disciplines, such as medical informatics (Pan et al. 2019) and neuroscience (Wang et al. 2019). We believe that using LIME to explain the predictions will help users understand the underlying ML models being estimated in this study.

In summary, there still lacks a well-recognized ML approach that makes a reliable prediction on EV adoption and is able to explain the relative importance of the influencing factors. This paper aims to combine EV preference survey data, a unique data-driven ML modeling framework, and the application of LIME to interpret prediction outcomes to shed light on the role of individual ridesourcing usage and attitudes in EV adoption decisions.

Methodological framework

As Figure 1 illustrates, the methodological framework comprises two steps. In the first step, we conduct a feature engineering process that generates the final dataset and run a number of ML models that predict EV adoption. The leading model in terms of predictability is selected. The second step is to make use of the LIME method on the leading model to elaborate on the contribution of each feature to the prediction. This step helps us identify the most important features on a case-by-case basis. We further describe the methodological framework in the following subsections.

An overview of supervised learning

In this study, one of our objectives is to train a machine learning model that can accurately classify EV adopters and non-adopters based on attitudinal factors, ridesourcing usage factors, vehicle-related factors, and sociodemographic factors. This is a *supervised learning* process, where an ML algorithm uses a set of predictors to classify an observed response (called '*label*'). A subset of the entire dataset (called '*training dataset*') serves the purpose to train a ML model and evaluate its performance such that the predicted results resemble labels as closely as possible. A second subset (called '*testing dataset*') is then utilized to test how well the model performs when classifying observations outside of the training process. Since

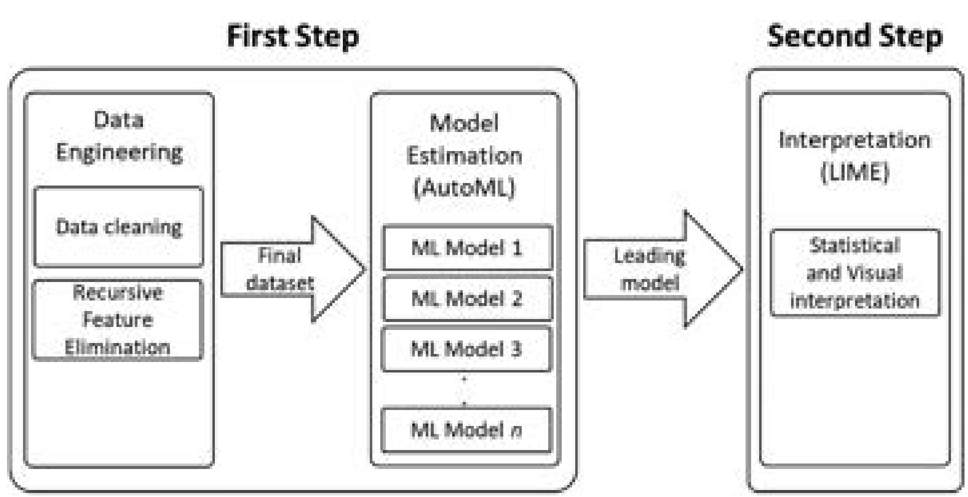

Figure 1. Methodological framework of the study.

both subsets are drawn from the same original data, we apply the *K-Fold Cross Validation* method to avoid any unintentional selection bias when splitting the full dataset.

We further trim the dataset through feature engineering. As the size and dimensionality of the dataset grow, the training process may be computationally intensive and may incur data sparsity that could cause a decline in predictability. We first eliminate six highly collinear features out of 42 features. We then opted for a *recursive feature elimination* (RFE) process to identify the most irrelevant features to the prediction. RFE follows a recursive procedure where it starts with fitting a given ML algorithm using all features in the training dataset (a subset of the entire dataset), followed by ranking features by importance, removing the least important features, and re-fitting the model using a subset of the features. The process is repeated until the optimal list of predictors is reached. In our case, 32 features remain after RFE.

Another difficulty in building a proper ML application is to find the best algorithm from a large number of candidate algorithms. Although researchers may know, based on their experience or domain knowledge, the ML algorithms that would work best, their performances vary significantly by case. Moreover, each algorithm has its own hyperparameters – a characteristic external to the model itself whose value cannot be estimated from data (e.g., k in k-Nearest Neighbors) – that are pre-set before the learning process. Until recently, the complexity of hyperparameters implies that researchers needed to fine-tune ML models through an iterative process that took a different combination of hyperparameters in each iteration (i.e., *grid search*). Given the large number of possible combinations, this process may not be easily done thoroughly and is subject to human errors. To automatize this process, we utilize an open-source artificial intelligence platform, *H2O*, in this study (H2O.ai 2020a). H2O allows users to build a large number of ML models and evaluate their performances through an application called *AutoML*[1] (H2O.ai 2020b). AutoML automatically runs through a set of ML algorithms with hyperparameters and produce a leaderboard of models based on predictability. The current version of AutoML trains and validates a set of pre-defined algorithms, as well as performs grid searches.[2]

Local Interpretable Model-Agnostic Explanations (LIME)

After training ML models and evaluating their performance, we try to understand how each predictor contributes to the prediction for the ML model with the best performance ('*the leading model*'). In particular, we hope to gain insights on the relationship between attitudinal/ridesourcing factors and EV adoption. We apply the LIME method for this task. (Ribeiro, Singh, and Guestrin 2016) describe the LIME technique that, '(*it*) *presents textual or visual artifacts that provide qualitative understanding of the relationship between the instance's components [. . .] and the model's prediction*.' LIME has three distinct characteristics: (1) *interpretable*, meaning that it provides qualitative measurement between the input feature and the response; (2) *local fidelity*, meaning that it must correspond to how the model behaves in the vicinity of the observation being predicted; and (3) *model-agnostic*, meaning that it should be able to explain any model. In short, the LIME method samples observations from the dataset locally, makes predictions based on an underlying ML model, and evaluates a prediction by its proximity to the observation being explained. The word *proximity*, in this context, is equivalent to *similarity*: the extent to which a case being predicted matches its original representation in the feature space. Therefore, LIME can explain the predictions of any classifier in a faithful way, by approximating a highly abstract model locally with an interpretable model instead. In practice, the output of LIME is a visual representation that shows the weight of each feature supporting or contradicting the prediction on a label. For us researchers, we can visually observe the importance of predictors and make inferences about their role in determining the prediction – the EV adoption decision in our case.

Data

The data used in this paper are a part of survey data collected to study EV preferences and the role of an individual's social network structure played in such preferences (Bas et al., 2021b). The survey includes a stated choice experiment (SCE), in which respondents are asked to evaluate the importance of different electric and gasoline vehicle attributes – such as price, range, and charging times – in their preferences toward one or the other. In addition to the SCE, we also ask respondents to provide information on their social network, car ownership, socioeconomic statuses, usage of ridesourcing apps, and their attitudes toward environmental protection, transportation technology, and EV itself. Table 1 describes the list of 32 features used to build the ML models. The descriptive statistics are shown in Table A1 in the Appendix.

We surveyed current residents of the State of Maryland, U.S., who are 18-year-old or older with a valid driver's license. The survey was conducted online in 2019 and was taken by 380 respondents. Six responses were removed due to inconsistency in their answers. Each respondent faced six or nine choice tasks, providing us a total of 3,183 pseudo-observations regarding EV adoption preferences. It is worth mentioning that although this sample size may seem small for a standard ML application, the problem at hand is not structurally complicated that would require to train big data. Nor are the data imbalanced toward a majority class (i.e., particularly large proportion of adoption of EV or no adoption) that would require a large dataset with enough observations of both classes for training and validation. In addition, we use ensemble methods to improve the training process with a relatively small dataset. The ensemble methods combine several basic ML models to produce one optimal predictive model. These methods have been proven to work well with small datasets in previous ML applications (e.g., Greene and Cunningham 2006).

Results

In this section, we first present ML models' performance in a scoreboard with a focus on the top-performance model. We then compare the leading model with a benchmark model – the binomial family of generalized linear model (GLM). We choose GLM as a benchmark since the logistic regression model is a traditional parametric model of reference that estimates the choices made by an individual among a set of alternatives. Finally, we visualize the cases of LIME outcomes, which demonstrate LIME's capabilities to provide interpretability to ML learning models whose outputs lack it. ML models and showcase the influence of attitudinal and ridesourcing usage factors on the decision of EV adoption.

Model performance

The first part of our results is a performance comparison among different ML models. ML algorithms predict the probability of an observation pertaining to each class and assign the observation to one of classes based on a probability threshold. The classification can be wrong, producing false positives and false negatives, which leads to two important metrics: *Sensitivity* and *Specificity*.

Table 1. Features resulting from the RFE process.

Feature	Description	Value
CHOICE	Choice made in each scenario	Binary feature, 1 means 'Adoption'
Ridesourcing		
APP_USEFREQ	Frequency of rides	1 to 5 scale, 1 means higher frequency
APP_STRFREQ*	Frequency of shared rides with strangers	
APP_PURP_WRK	Main usage of the app is for *Work*	Binary feature, 1 means 'Yes'
APP_PURP_LEI	Main usage of the app is for *Leisure*	
APP_PURP_SOC	Main usage of the app is for *Social*	
APP_PURP_AIRP	Main usage of the app is for *Airport*	
APP_PURP_NEVER*	I never used a ridesourcing app	
Attitudes		
ATT_EC1	*I do what I can to contribute to reduce global climate changes, even if it costs more and takes time.*	0 to 4 Likert scale, 0 means strong disagreement and 4 means strong agreement
ATT_EC2†	*The authorities should NOT introduce legislation that forces citizens and companies to protect the environment.*	
ATT_TI1†	*It is NOT important for me to follow technological development.*	
ATT_TI2	*I often purchase new technology products, even though they are expensive.*	
ATT_TI3	*I am optimistic about the future of shared mobility (such as carshare and rideshare).*	
ATT_TI4†	*New technologies create more problems than they solve.*	
ATT_PROEV1	*Electric vehicles should play an important role in our mobility systems.*	
ATT_PROEV2†	*If I use an electric vehicle instead of a conventional vehicle, I will have to cancel some activities.*	
ATT_PROEV3	*Electric vehicles are more reliable than conventional vehicles.*	
ATT_PROEV4†	*I am concerned that EVs are not powerful enough to make a safe takeover.*	
ATT_PROEV5	*When forced to change daily activity arrangement, I don't feel anxious.*	
Sociodemographic		
AGE*	*Age of the individual*	Numeric
MALE	*Individual is male*	Binary feature, 1 being 'Yes'
EGFT†	*Individual works for government full time*	
EGPT	*Individual works for government part time*	
EPCFT	*Individual works for private company full time*	
EPCPT	*Individual works for private company part time*	
ERET	*Individual is retired*	
ESELF	*Individual is self employed*	
STU†	*Individual is a student*	
UEMPL	*Individual is unemployed*	
EOTHER	*Individual is in another employment category*	
EDUDGR	*Individual´s educational degree*	1 to 5 scale, the higher, the more educated

(*Continued*)

Table 1. (Continued).

Feature	Description	Value
HHMEM*	*Number of household members*	Numeric
HHMEM_EMP*	*Number of household members employed*	Numeric
HHINC_BRA	*Household income bracket*	1 to 7 scale, the higher, the wealthier
INIDINC_BRA*	*Individual´s income bracket*	1 to 7 scale, the higher, the wealthier
Vehicle-related		
EV_PRICE	*Electric vehicle price shown in SCE*	Numeric
EV_PROPCOST†	*Electric vehicle propulsion cost shown in SCE*	
EV_FASTCHARGE	*Electric vehicle fast charge time shown in SCE*	
EV_TAXDEDAM	*Electric vehicle purchase tax deduction shown in SCE*	
GAS_PRICE	*Gasoline vehicle price shown in SCE*	
GAS_PROPCOST†	*Gasoline vehicle propulsion cost shown in SCE*	
GAS_RANGE	*Gas vehicle range shown in SCE*	

*Columns removed to reduce pair-wise linear correlation (correlation threshold 0.5).
†Features expressed in negative terms.

Sensitivity is the proportion of actual positives identified as such. Specificity is the proportion of actual negatives correctly identified as such. Varying the probability threshold used for the classification will produce different values of sensitivity and specificity. The different values of these two metrics for each threshold can be plotted in a chart, called *Receiver Operator Characteristic* (ROC). Finally, the value of the area under the ROC curve (known as AUC) provides a measure of an ML algorithm's predictability: If the ROC curve reaches the top-left corner (i.e., varying the probability threshold would not affect all classifications), then all positive and negative classes would be classified correctly. AUC equals to 1 in such case.

As aforementioned, we train more than 60 ML models using different algorithms in H2O AutoML and rank them according to their AUC value. It is worth mentioning that, in addition to each individual model, AutoML also estimates *stacks* of models. A stack is a combination of ML models that have almost no interpretability, but usually provide more accurate prediction (for instance, the combination of all Gradient Boost Machine models). Since our goal is to achieve both predictability and interpretability, we will not show the performance metric of stacks of ML models, which is only marginally better than that of our leading model. In Table 2, only the top – and bottom-performance models are listed.

Table 2. Model performance scoreboard using the cross-validation dataset.

Top-performance models	AUC
Gradient Boost Machine (GBM), grid 1, model #19	0.9285
Gradient Boost Machine (GBM) model #3	0.9277
Gradient Boost Machine (GBM) model #4	0.9270
Gradient Boos Machine (GBM), grid 1, model #5	0.9269
Gradient Boos Machine (GBM), grid1, model #28	0.9266
. . ..	
Bottom-performance models	AUC
Deep Learning (DNN), grid 2, model #4	0.8301
Generalized Linear Model (GLM) model #1 (*benchmark*)	0.8214
Deep Learning (DNN), grid 2, model #5	0.8029
Deep Learning (DNN), grid 3, model #4	0.7340
Gradient Boost Machine (GBM), grid 1, model_#30	0.7088

Clearly, the GBM models have the best overall performance. GBM is an ensemble-type algorithm that combines several decision tree algorithms into a hybrid algorithm to achieve high prediction accuracy. In fact, the Top 13 models are GBM with different hyperparameters, followed by a distributed random forest (DRF) model (which also uses an ensembled algorithm). At the bottom of the list, the deep neural network (DNN) family do not perform as well as the GBM family for this dataset. DNN algorithms are more suitable for highly complex, non-linear classification problems, such as image recognition. Additionally, the benchmark model (binomial GLM) ranks fourth from the bottom. In the next subsection, we will compare the leading model and the benchmark model in details.

The leading model

In this subsection, we first summarize the global feature importance on the prediction. For the leading ML model, the feature importance is measured by relative importance. The average impact of each feature across all decision trees within GBM on the mean squared error (MSE) loss function is calculated. The feature of the largest impact is of the most importance, while the impact of all other features is provided relative to the most important feature. Hence, in Figure 2(a), we observe one feature (*ATT_PROEV1*) with an importance scale of '1' and other features with a fraction of the importance scale. For the benchmark model, the coefficient values are equivalent to relative importance.[3] The results are displayed in Figure 2(b). The significantly positive coefficients are in blue color and the significantly negative ones are in orange color.

When taking a closer look at feature importance, we find that the most important feature in the leading model is an attitudinal factor (*ATT_PROEV1*), which reveals a favorable view on EV in the mobility system ('*electric vehicles should play an important role in our mobility systems*'). The second most important factor is household income level (*HHINC_BRA*), suggesting that income plays a critical role in EV adoption decision. In addition, we find other highly considered attitudinal factors, such as the Pro-EV factors (*ATT_PROEV2, ATT_PROEV3, ATT_PROEV5*), environmental concerns (*ATT*_EC1, *ATT_EC2*), and an indicator of EV technology inclinations (*ATT_TI2*).

In addition, we find evidence that ridesourcing usage (*APP_USEFREQ*) and an optimistic attitude toward the shared mobility (*ATT_TI3*) are relatively important in explaining EV adoption decisions. On the other hand, ridesourcing trip purposes are not crucial to EV preferences. These results are consistent with our hypothesis that familiarity and enthusiasm toward the shared mobility technology is likely to positively influence the individual willingness to adopt EV.

It is noteworthy that the policy factor of tax deduction on EV purchase (*EV_TAXDEDAM*), EV price and battery range (*EV_PRICE, EV_RANGE*), EV fast-charging time (*EV_FASTCHART*), the comparable gasoline vehicle's price and mileage (*GAS_PRICE, GAS_RANGE*), as well as an individual's educational level (*EDUDGR*) are relevant in determining the EV adoption decision, which is consistent with what we find in previous literature.

If we compare these findings with those from the benchmark model, we can come up with somewhat similar results. The GLM model suggests that the most important features in explaining EV adoption are vehicle price (EV and the comparable gasoline vehicle) and the pro-EV, pro-environment, pro-technology attitudinal factors. High education level and private-sector employment status are the significant socioeconomic factors in supporting EV adoption decision. Again, the frequent uses of ridesourcing apps also help explain an inclination in EV adoption.

The main differences between the leading model and the benchmark model are their modeling performances. The AUC of the leading model (0.9285) is significantly higher than that of the benchmark model (0.8214), suggesting better predictability of the GBM algorithm. We provide in Table 3 and Table 4 the detailed classification performance of the leading and benchmark model, respectively, using a confusion matrix, a convenient way of displaying both the correct predictions and the errors. In this case, the leading model misclassifies 37 individuals as EV *adopters* when they actually are not in the dataset (Type I errors), and 128 as EV non-*adopters* when they are (Type II errors). These figures lead to sensitivity and specificity values of 0.73 and 0.9, respectively. On the other hand, the benchmark model yields the same sensitivity and a lower specificity value. It means that although the GLM misclassifies a similar number of false-negative cases as the GBM (127 vs. 128), the GLM misclassifies a significantly higher number of false-positive cases than the GBM (64 vs. 37). That is, the issue of overestimation of EV adopters is more significant in the benchmark model than the leading model.

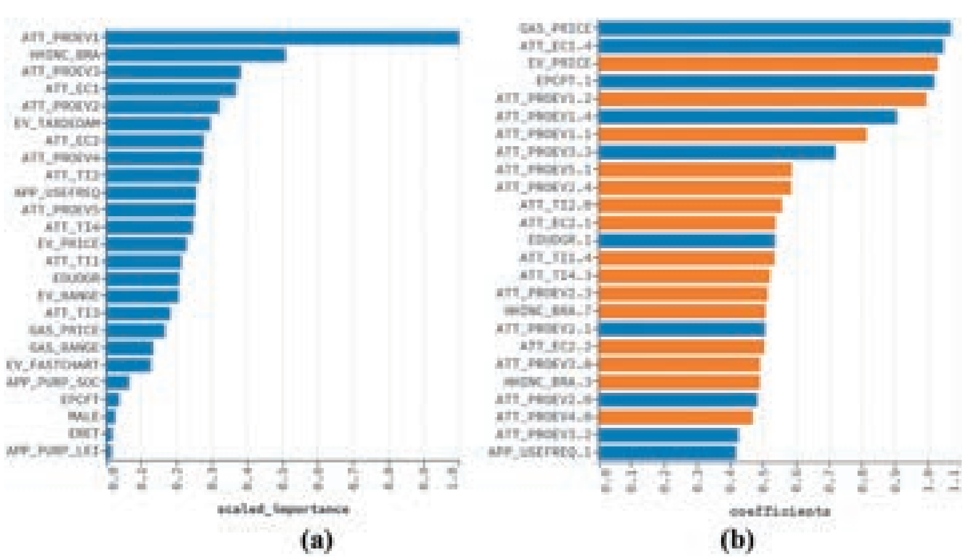

Figure 2. Feature importance for the leading model (a) and the benchmark model (b).

Table 3. Confusion matrix for the leading model, testing data.

		Predicted class		
		Adoption	No Adoption	Specificity
Actual class	Adoption	280	128	0.9
	No Adoption	37	352	
	Total	389	408	
	Sensitivity	0.73		

Table 4. Confusion matrix for the benchmark model, testing data.

		Predicted class		
		Adoption	No adoption	Specificity
Actual class	Adoption	250	127	0.8
	No adoption	64	349	
	Total	413	377	
	Sensitivity	0.73		

Finally, the ROC curves (the blue curve in both plots) for both models are presented in Figure 3, with AUC of 0.9285 and 0.8214, suggesting that our leading model outperforms the benchmark model in the prediction of EV adoption.

Interpreting ML results using LIME

As aforementioned, the modeling complexity of some ML models is both the reason for their high predictability and their low interpretability. To overcome the issue of abstract ML model results, we apply the Local Interpretable Model-Agnostic Explanation (LIME) method as described in the methodology section. LIME produces visual representations in which one can observe a feature's influence, either supporting or contradicting, on the prediction.

Figure 4 shows two observations that have been classified by our leading model as adopter (left), and non-adopter (right). For the adopter, we can observe that the pro-environmental attitudes (*ATT_EC1* = 4, ATT_EC2 = 3) support the EV adoption classification. Similarly, the pro-EV attitude (*ATT_PROEV3* = 3) and the frequent usage of ride-sourcing apps (*APP_USEFREQ* = 3) also support the adoption classification. As expected, a high EV driving range and high household income level (*EV_RANGE, HHINC_BRA* = 5) would also contribute to EV adoption decisions. On the flip side, a low pro-EV sentiment (*ATT_PROEV1* = 2, '*Electric vehicles should play an important role in our mobility systems*') and high EV prices (*EV_PRICE*) could contradict the adoption classification.

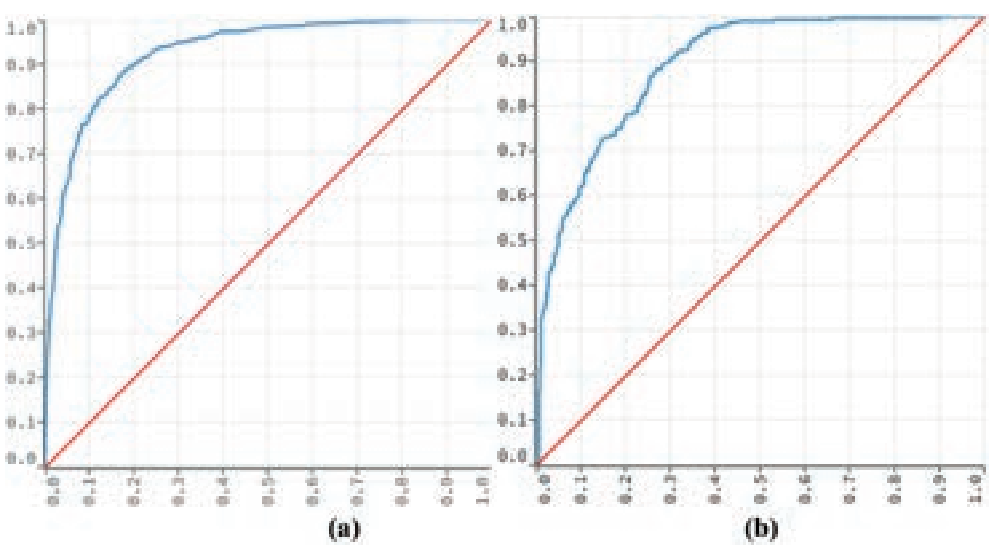

Figure 3. The ROC curve and the AUC value for the leading model (a) and the benchmark model (b), testing data.

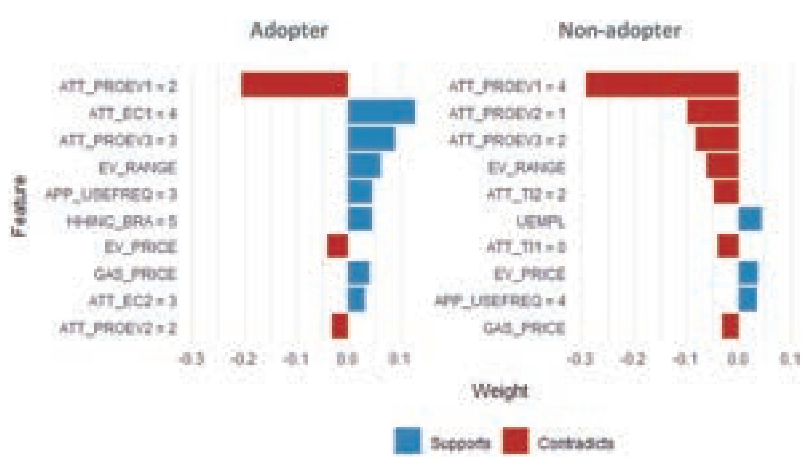

Figure 4. Contribution of top ten predictors to the prediction for two cases using LIME.

We can make similar inferences by interpreting feature importance for the case of a non-adopter. A strongly positive view on EV's contribution to the transportations system (*ATT_PROEV1* = 4) would contradict the non-adopter classification. The analogous interpretation applies to the other attitudinal features as well as vehicle features (range and price). On the other hand, being unemployed (*UEMPL*), a low frequency of ridesourcing usage (*APP_USEFREQ* = 4) and EV prices all support the no-adoption classification for this case.

LIME allows us to understand the predictor–prediction relationship on a case-by-case basis. Through the interpretation, we can make inferences that the favorable attitudes toward EV, environmental protection, and transportation technology, as well as the frequent usage of ridesourcing apps may lead to a higher chance of EV adoption or a lower chance of non-adoption (and the vice-versa). It is possible that the inferences made from these two cases are subjective to a case selection bias, therefore we further include the visual representation of another six randomly selected observations (three adopters and three non-adopters) in Figure A1 in the appendix. Two takeaways can be made: (1) the important features that can explain an individual's adoption/no adoption decision vary by case; and (2) despite individual heterogeneity, we can identify the common impact of a feature on the adopter/non-adopter classification across multiple cases. Together, these two takeaways demonstrate the unique advantage of LIME in interpreting ML modeling results.

Figure 5 provides another type of visual representation of the LIME method – a heatmap that depicts the contribution of the attitudinal and ridesourcing features of interest for four cases that are classified as EV adoption. Each column represents one case and each row represents one feature. The darker blue color indicates a stronger support to such classification, whereas the darker red color indicates a stronger contradiction against such classification. Lighter/white colors indicate a feature's lower level of contribution.

Noticeably, pro-EV attitudes, pro-environment attitudes, and frequent uses of ridesourcing apps greatly contribute to the adoption classification. In addition, we find that if a ridesourcing user takes a trip unrelated to work (*APP_PURP_WRK* = 0) or social occasions (*APP_PURP_SOC* = 0), then her ridesourcing trip purpose would contradict EV adoption. On the other hand, if a ridesourcing user takes a trip unrelated to the airport commutes (*APP_PURP_AIRP* = 0) or leisure (*APP_PURP_LEI* = 0), then she is more likely to be classified as an adopter. To put it more concisely, frequent ridesourcing users for work and social trips could become EV adopters in the future. This is a reasonable inference as frequent ridesourcing users tend to be high-income yet car-less (Zou and Cirillo 2021) who could potentially later become EV owners. Consistent with our previous findings, infrequent uses of ridesourcing apps and indifferent/negative attitudes toward EV and environmental concerns significantly contradict to the 'adoption' classification.

Discussion and conclusion

In this paper, we uncovered a novel perspective on EV adoption decisions that shows how people's attitudes toward the environment and technology advancement, as well as their usage of the currently available new mobility, ridesourcing, could impact their preference to adopt the EV technology. Instead of relying on the conventional approach, such as a discrete choice analysis, we try out a number of machine learning techniques to achieve the best prediction outcomes based on the EV survey data. As our results reveal, machine learning models are able to produce highly accurate predictions on EV adoption/no adoption decisions. In addition to providing the global importance of the features in prediction, we adopt the LIME method to explain the contribution of each feature to the prediction outcome on a case-by-case basis. The major advantage of LIME is its graphical interpretation of the factors that drive the prediction outcome: In a binary prediction – 'whether or not a person is willing to adopt EV', the visual representations of LIME results show the extent to which a factor contribute to the 'adoption' and 'no adoption' classification in a case-by-case manner. It is an elegant way to unfold the mysterious behind-the-scenes prediction process of the highly complex ML models. In this case, the pro-EV, environmentally conscious attitudes, and frequent uses of ridesourcing apps are largely associated with a high likelihood of willingness to adopt an EV. Our finding is consistent with what (Jenn, Laberteaux, and Clewlow 2018) find in their research that the current use of shared mobility positively impacts the possibility to own an electric vehicle. We also find the results on attitudinal factors consistent with previous literature (e.g., Axsen, Bailey, and Castro 2015).

In terms of what the results indicate to advocates of EVs and policymakers who see EVs as a useful tool to build green transportation, the biggest takeaway from our study is that the more people are in favor of EV technology, are aware of environmental protection, and utilize new mobility, the more likely they are willing to adopt electric vehicles

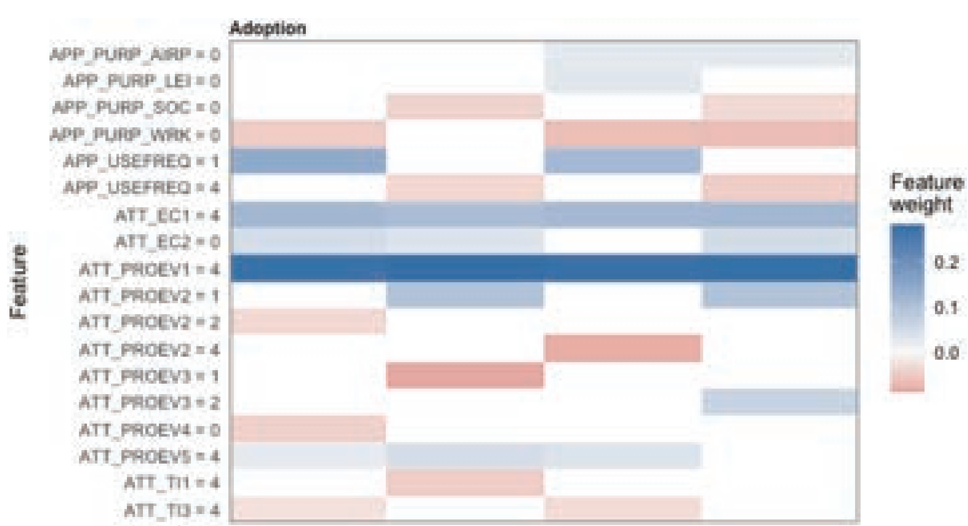

Figure 5. A heatmap showing features' contribution to the adoption classification.

in the future. In addition, the new mobility options, including shared mobility, EVs, autonomous vehicles, are oftentimes branded as 'sustainable' solutions to reduce carbon emissions. Although it is debatable how credible such claim is empirically, our findings show the potential for consumers to try these mobility options as a result of the sustainability argument. Thus, in the global effort of combating climate change and achieving sustainable transportation, policymakers can come up with strategies to entice the large base of ridesourcing users to gradually adopt the EV technology, possibly through a partnership with TNCs.

We also acknowledge the limitations of this study. Firstly, we could not exhaust all the factors that may influence EV adoption. Richer information on the ridesourcing factors (beyond frequency and trip purposes) could help in achieving higher levels of predictability. Secondly, the small sample size means that our results are subject to idiosyncratic errors, which may not be easily identified. In addition, our sample is based on residents in the State of Maryland, so the predictors may not work universally for a more global sample, say, the entire United States, or for a sample in another U.S. state/ metropolitan area or some other country. Nonetheless, we argue that the methodological framework is transferrable to another context. In particular, we hope the use of open-source ML platforms and tools, such as H2O, and LIME could inspire research on other sustainable transportation technologies and mobility options.

Notes

1. We run AutoML in the R environment using the application programming interface (API) developed by H2O. We run the ML models in an Amazon Web Service instance with 16 vCPU and 64 GiB of memory.
2. Specifically, the following algorithms are included in AutoML: Five pre-specified Gradient Boosting Machine (GBM), three pre-specified Extreme Gradient Boosting Machine (XGBoost GBM), a default Random Forest (DRF), a near-default Deep Neural Network (DNN), an Extremely Randomized Forest (XRT), a fixed grid of Generalized Linear Model (GLM), a random grid of XGBoost GBMs, a random grid of GBMs, and a random grid of DNNs.
3. The value of the coefficients is equivalent to their relative importance since the coefficient with the highest value can be interpreted as the most important (and therefore being normalized to 1), and then the rest can be scaled accordingly (which would not alter the shape of the graph shown in Figure 2 (b).). In this case, for the sake of a more traditional interpretation of the GLM model, we keep the value and sign (blue/orange color) of the coefficients without transforming them into their importance counterpart.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Javier Bas http://orcid.org/0000-0003-2699-9273
Zhenpeng Zou http://orcid.org/0000-0003-1789-7638
Cinzia Cirillo http://orcid.org/0000-0002-5167-0413

Author contributions

The authors confirm contribution to the paper as follows: study conception and design: Javier Bas; data collection: Javier Bas, Cinzia Cirillo; analysis and interpretation of results: Javier Bas, Zhenpeng Zou, Cinzia Cirillo; draft manuscript preparation: Javier Bas, Zhenpeng Zou. All authors reviewed the results and approved the final version of the manuscript.

References

Axsen, J., J. Bailey, and M. A. Castro. 2015. "Preference and Lifestyle Heterogeneity among Potential Plug-in Electric Vehicle Buyers." *Energy Economics* 50: 190–201. doi:10.1016/j.eneco.2015.05.003.

Bas, J., C. Cirillo, and E. Cherchi. 2021a. "Classification of Potential Electric Vehicle Purchasers: A Machine Learning Approach." *Technological Forecasting and Social Change* 168. doi:10.1016/j.techfore.2021.120759.

Bas, J., C. Cirillo, and E. Cherchi (2021b). A stated choice experiment for considering social conformity in the adoption of electric vehicles (Working Paper).

Burghard, U., and D. Elisabeth. 2019. "Who Wants Shared Mobility? Lessons from Early Adopters and Mainstream Drivers on Electric Carsharing in Germany." *Transportation Research Part D* 71: 96–109. doi:10.1016/j.trd.2018.11.011.

Canals Casals, L., E. Martinez-Laserna, B. Amante García, and N. Nieto. 2016. "Sustainability Analysis of the Electric Vehicle Use in Europe for Co2 Emissions Reduction." *Journal of Cleaner Production* 127: 425–437. doi:10.1016/j.jclepro.2016.03.120.

Carley, S., S. Siddiki, and S. Nicholson-Crotty. 2019. "Evolution of Plug-in Electric Vehicle Demand: Assessing Consumer Perceptions and Intent to Purchase over Time." *Transportation Research Part D* 70: 94–111. doi:10.1016/j.trd.2019.04.002.

Chen, X., M. Zahiri, and S. Zhang. 2017. "Understanding Ridesplitting Behavior of On-demand Ride Services: An Ensemble Learning Approach." *Transportation Research Part C: Emerging Technologies* 76: 51–70. doi:10.1016/j.trc.2016.12.018.

Fukushima, A., T. Yano, S. Imahara, H. Aisu, Y. Shimokawa, and Y. Shibata. 2018. "Prediction of Energy Consumption for New Electric Vehicle Models by Machine Learning." *IET Intelligent Transport Systems* 12 (9): 1174–1180. doi:10.1049/iet-its.2018.5169.

Greene, D., and P. Cunningham. 2006. "Practical Solutions to the Problem of Diagonal Dominance in Kernel Document Clustering." In *ICML '06: Proceedings of the 23rd International Conference on Machine Learning*, Vol. 2006, 373–384. Pittsburgh, PA. doi:10.1145/1143844.1143892.

H2O.ai. (2020a) H2o: R Interface for H2O. R Package Version 3.30.0.6. Available online: https://github.com/h2oai/h2o-3

H2O.ai (2020). AutoML: Automatic Machine Learning. Published Online. Available online: https://docs.h2o.ai/h2o/latest-stable/h2o-docs/automl.html

Hardman, S. 2019. "Understanding the Impact of Reoccurring and Non-financial Incentives on Plug-in Electric Vehicle Adoption - a Review." *Transportation Research Part A* 119: 1–14. doi:10.1016/j.tra.2018.11.002.

He, Z., Y. Zhou, X. Chen, J. Wang, W. Shen, M. Wang, and W. Li. 2021. "Examining the Spatial Mode in the Early Market for Electric Vehicles Adoption: Evidence from 41 Cities in China." *Transportation Letters* 1–11. published online:. doi:10.1080/19427867.2021.1917217.

Henderson, J. 2020. "Evs are Not the Answer: A Mobility Justice Critique of Electric Vehicle Transitions." *Annals of the American Association of Geographers* 110 (6): 1993–2010. doi:10.1080/24694452.2020.1744422.

Huang, X., Y. Tan, and X. He. 2011. "An Intelligent Multifeature Statistical Approach for the Discrimination of Driving Conditions of a Hybrid Electric Vehicle." *IEEE Transactions on Intelligent Transportation Systems* 12 (2): 453–465. doi:10.1109/TITS.2010.2093129.

Hughes, S., S. Moreno, W. F. Yushimito, and G. Huerta-Cánepa. 2019. "Evaluation of Machine Learning Methodologies to Predict Stop Delivery Times from GPS Data." *Transportation Research Part C: Emerging Technologies* 109: 289–304. doi:10.1016/j.trc.2019.10.018.

Jahangiri, A., and H. A. Rakha. 2015. "Applying Machine Learning Techniques to Transportation Mode Recognition Using Mobile Phone Sensor Data." *IEEE Transactions on Intelligent Transportation Systems* 16 (5): 2406–2417. doi:10.1109/TITS.2015.2405759.

Javid, R. J., and A. Nejat. "A Comprehensive Model of Regional Electric Vehicle Adoption and Penetration." *Transport Policy* 54 (2017): 30–42.

Jenn, A., K. Laberteaux, and R. Clewlow. 2018. "New Mobility Service Users' Perceptions on Electric Vehicle Adoption." *International Journal of Sustainable Transportation* 12 (7): 526–540. doi:10.1080/15568318.2017.1402973.

Jensen, A. F., E. Cherchi, and S. L. Mabit. 2013. "On the Stability of Preferences and Attitudes before and after Experiencing an Electric Vehicle." *Transportation Research Part D* 25: 24–32. doi:10.1016/j.trd.2013.07.006.

Jiang, J. (2019). More Americans are using ride-hailing apps. Published on 4 January 2019. Available at: https://www.pewresearch.org/fact-tank/2019/01/04/more-americans-are-using-ride-hailing-apps/

Kim, J., S. Rasouli, and H. Timmermans. 2014. "Expanding Scope of Hybrid Choice Models Allowing for Mixture of Social Influences and Latent Attitudes: Application to Intended Purchase of Electric Cars." *Transportation Research. Part A: Policy and Practice* 69: 71–85. doi:10.1016/j.tra.2014.08.016.

Lakkaraju, H., E. Kamar, R. Caruana, and J. Leskovec (2017). "Interpretable & Explorable Approximations of Black Box Models." *ArXiv*: 1707.01154 [Cs]. http://arxiv.org/abs/1707.01154

Lee, D., J. Mulrow, C. J. Haboucha, S. Derrible, and Y. Shiftan. 2019. "Attitudes on Autonomous Vehicle Adoption Using Interpretable Gradient Boosting Machine." *Transportation Research Record: Journal of the Transportation Research Board* 2673 (11): 865–878. doi:10.1177/0361198119857953.

Liao, F., E. Molin, and B. van Wee. 2017. "Consumer Preferences for Electric Vehicles: A Literature Review." *Transport Reviews* 37 (3): 252–275. doi:10.1080/01441647.2016.1230794.

Liu, Z., Y. Liu, Q. Meng, and Q. Cheng. 2019. "A Tailored Machine Learning Approach for Urban Transport Network Flow Estimation." *Transportation Research Part C: Emerging Technologies* 108: 130–150. doi:10.1016/j.trc.2019.09.006.

Maclnnis, B., and J. Krosnick (2020). Climate Insights 2020: Electric Vehicles. Resources for the Future Report. Retrieved online: https://media.rff.org/documents/Climate_Insights_2020_Electric_Vehicles.pdf

Martín-Baos, J. Á., R. García-Ródenas, and L. Rodriguez-Benitez. 2021. "Revisiting Kernel Logistic Regression under the Random Utility Models Perspective. An Interpretable Machine-learning Approach." *Transportation Letters* 13 (3): 151–162. doi:10.1080/19427867.2020.1861504.

Pan, L., G. Liu, X. Mao, H. Li, J. Zhang, H. Liang, and X. Li. 2019. "Development of Prediction Models Using Machine Learning Algorithms for Girls with Suspected Central Precocious Puberty: Retrospective Study." *Jmir Medical Informatics* 7 (1): e11728. doi:10.2196/11728.

Parsa, A. B., R. Shabanpour, A. (Kouros) Mohammadian, J. Auld, and T. Stephens. 2020. "A Data-driven Approach to Characterize the Impact of Connected and Autonomous Vehicles on Traffic Flow." *Transportation Letters*. 1–9. doi:10.1080/19427867.2020.1776956.

Rezvani, Z., J. Jansson, and J. Bodin. 2015. "Advances in Consumer Electric Vehicle Adoption Research: A Review and Research Agenda." *Transportation Research Part D* 34: 122–136. doi:10.1016/j.trd.2014.10.010.

Ribeiro, M. T., S. Singh, and C. Guestrin. 2016. "'Why Should I Trust You?': Explaining the Predictions of Any Classifier." In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135–1144. San Francisco, CA. 10.1145/2939672.2939778 .

Sierzchula, W., S. Bakker, K. Maat, and B. van Wee. 2014. "The Influence of Financial Incentives and Other Socio-economic Factors on Electric Vehicle Adoption." *Energy Policy* 68: 183–194. doi:10.1016/j.enpol.2014.01.043.

Sovacool, B. K., J. Kester, L. Noel, and G. Z. de Rubens. 2019. "Income, Political Affiliation, Urbanism and Geography in Stated Preferences for Electric Vehicles (Evs) and Vehicle-to-grid (V2g) Technologies in Northern Europe." *Journal of Transport Geography* 78: 214–229. doi:10.1016/j.jtrangeo.2019.06.006.

Sun, D., F. Leurent, and X. Xie. 2021. "Discovering Vehicle Usage Patterns on the Basis of Daily Mobility Profiles Derived from Floating Car Data." *Transportation Letters* 13 (3): 163–171. doi:10.1080/19427867.2020.1861505.

Sun, S., J. Zhang, J. Bi, Y. Wang, and M. H. Y. Moghaddam. 2019. "A Machine Learning Method for Predicting Driving Range of Battery Electric Vehicles." *Journal of Advanced Transportation* 2019. doi:10.1155/2019/4109148.

Taiebat, M., and M. Xu. 2019. "Synergies of Four Emerging Technologies for Accelerated Adoption of Electric Vehicles: Shared Mobility, Wireless Charging, Vehicle-to-grid, and Vehicle Automation." *Journal of Cleaner Production* 230: 794–797. doi:10.1016/j.jclepro.2019.05.142.

The International Energy Agency (IEA). (2020). Global EV Outlook 2020: Entering the Decade of Electric Drive? Published online in June 2020. Available at: https://www.iea.org/reports/global-ev-outlook-2020

The United States Environmental Protection Agency (EPA). (2021). Inventory of U.S. Greenhouse Gas Emissions and Sinks: 1990-2019. Retrieved online: https://www.epa.gov/sites/production/files/2021-04/documents/us-ghg-inventory-2021-main-text.pdf

Vassileva, I., and J. Campillo. 2017. "Adoption Barriers for Electric Vehicles: Experiences from Early Adopters in Sweden." *Energy* 120: 632–641. doi:10.1016/j.energy.2016.11.119.

Wang, X., D. Wang, Z. Yao, B. Xin, B. Wang, C. Lan, Y. Qin, S. Xu, D. He, and Y. Liu. 2019. "Machine Learning Models for Multiparametric Glioma Grading with Quantitative Result Interpretations." *Frontiers in Neuroscience* 12: 1046. doi:10.3389/fnins.2018.01046.

Yang, S., W. Ma, X. Pi, and S. Qian. 2019. "A Deep Learning Approach to Real-time Parking Occupancy Prediction in Transportation Networks Incorporating Multiple Spatio-temporal Data Sources." *Transportation Research Part C: Emerging Technologies* 107: 248–265. doi:10.1016/j.trc.2019.08.010.

Yi, D., J. Su, C. Liu, M. Quddus, and W.-H. Chen. 2019. "A Machine Learning Based Personalized System for Driving State Recognition." *Transportation Research Part C: Emerging Technologies* 105: 241–261. doi:10.1016/j.trc.2019.05.042.

Zahid, T., K. Xu, W. Li, C. Li, and H. Li. 2018. "State of Charge Estimation for Electric Vehicle Power Battery Using Advanced Machine Learning Algorithm under Diversified Drive Cycles." *Energy* 162: 871–882. doi:10.1016/j.energy.2018.08.071.

Zarazua de Rubens, G., L. Noel, and B. K. Sovacool. 2018. "Dismissive and Deceptive Car Dealerships Create Barriers to Electric Vehicle Adoption at the Point of Sale." *Nature Energy* 3 (6): 501–507. doi:10.1038/s41560-018-0152-x.

Zarazua de Rubens, G. 2019. "Who Will Buy Electric Vehicles after Early Adopters? Using Machine Learning to Identify the Electric Vehicle Mainstream Market." *Energy* 172: 243–254. doi:10.1016/j.energy.2019.01.114.

Zhao, J., J. Ye, M. Xu, and C. Xu. n d. "Practical Model with Strong Interpretability and Predictability: An Explanatory Model for Individuals' Destination Prediction considering Personal and Crowd Travel Behavior." *Concurrency and Computation-Practice & Experience* e6151. doi:10.1002/cpe.6151.

Zou, Z., and C. Cirillo. 2021. "Does Ridesourcing Impact Driving Decisions: A Survey Weighted Regression Analysis." *Transportation Research Part A* 146: 1–12. doi:10.1016/j.tra.2021.02.006.

APPENDIX

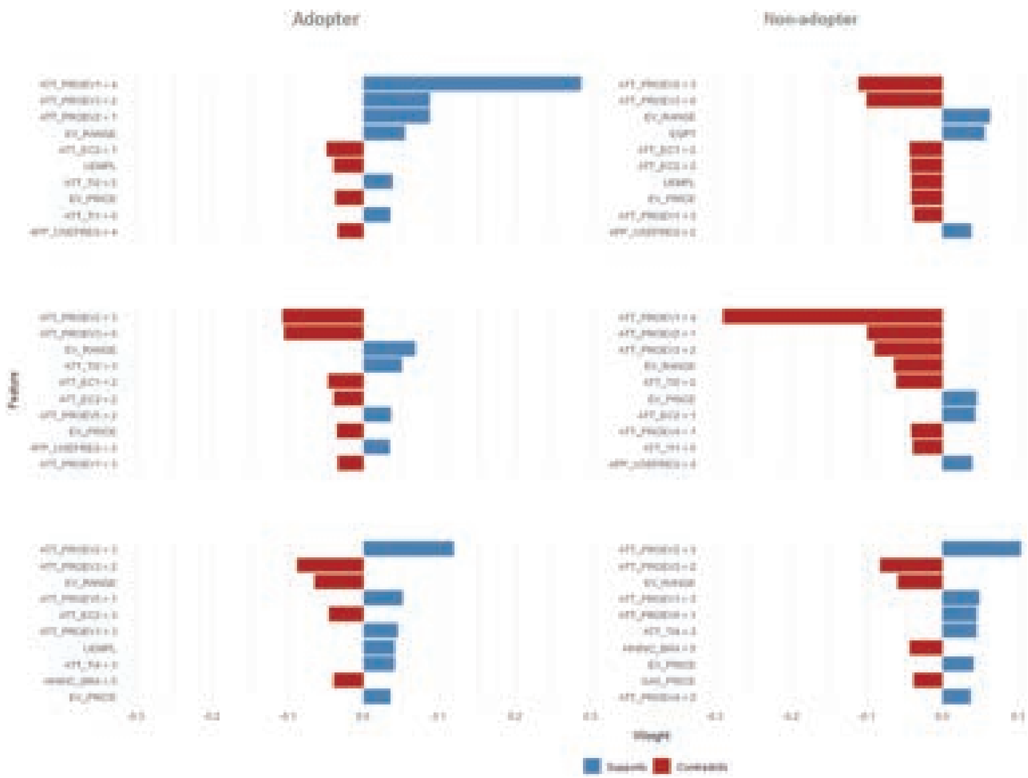

FIGURE A1. Contribution of top ten predictors to the prediction for six random cases using LIME.

TABLE A1. Descriptive statistics of the features.

Feature	Mean	Std. dev.	Min.	25%	Median	75%	Max.
APP_USEFREQ	3.68	1.10	1	3	4	5	5
APP_STRFREQ	3.68	1.15	1	3	4	5	5
APP_PURP_WRK	0.05	0.23	0	0	0	0	1
APP_PURP_LEI	0.11	0.32	0	0	0	0	1
APP_PURP_SOC	0.21	0.41	0	0	0	0	1
APP_PURP_AIRP	0.11	0.32	0	0	0	0	1
APP_PURP_NEVER	0.49	0.5	0	0	0	1	1
ATT_EC1	2.64	1.02	0	2	3	3	4
ATT_EC2	1.58	1.33	0	0	1	3	4
ATT_TI1	1.44	1.17	0	1	1	2	4
ATT_TI2	1.89	1.26	0	1	2	3	4
ATT_TI3	2.30	1.11	0	2	2	3	4
ATT_TI4	1.59	1.12	0	1	2	2	4
ATT_PROEV1	2.85	0.97	0	2	3	4	4
ATT_PROEV2	1.69	1.25	0	1	1	3	4
ATT_PROEV3	1.82	1.04	0	1	2	2	4
ATT_PROEV4	2.17	1.15	0	1	2	3	4
ATT_PROEV5	2.10	1.15	0	1	2	3	4
CHOICE	0.39	0.48	0	0	0	1	1
AGE	45.77	17.02	18	32	45	60	86
MALE	0.38	0.48	0	0	0	1	1
EGFT	0.07	0.25	0	0	0	0	1
EGPT	0.01	0.09	0	0	0	0	1
EPCFT	0.37	0.48	0	0	0	1	1
EPCPT	0.07	0.26	0	0	0	0	1
ERET	0.17	0.38	0	0	0	0	1
ESELF	0.06	0.24	0	0	0	0	1
STU	0.06	0.24	0	0	0	0	1
UEMPL	0.11	0.32	0	0	0	0	1
EOTHER	0.09	0.29	0	0	0	0	1
EDUDGR	3.52	1.02	1	3	4	4	5
HHMEM	2.79	1.86	0	2	2	4	25
HHMEM_EMP	1.46	1.12	0	1	1	2	8
HHINC_BRA	3.60	1.78	1	2	4	5	7
INDINC_BRA	2.50	1.64	0	1	2	3	7
EV_PRICE	3.75	1.08	2	3	3.3	5	5.5
EV_PROPCOST	0.04	0.01	0.015	0.03	0.04	0.05	0.09
EV_RANGE	2.59	0.71	1.6	1.76	2.5	3	3.95
EV_FASTCHART	31.34	10.82	15	25	30	35	60
EV_TAXDEDAM	3.20	2.48	0	1.5	2.5	5	7.5
GAS_PRICE	3.19	1.01	1.7	2.4	2.8	4.3	5
GAS_PROPCOST	0.08	0.03	0.05	0.055	0.06	0.1	0.16
GAS_RANGE	4.48	0.48	3.25	4.25	4.5	4.78	5.26