A Novel Two-level Causal Inference Framework for On-road Vehicle Quality Issues Diagnosis

Qian Wang
Ford Motor Company
Dearborn, MI48124
qwang435@gatech.edu

Huanyi Shui
Ford Motor Company
Dearborn, MI48124
hshui@ford.com

Thi Tu Trinh Tran
Ford Motor Company
Dearborn, MI48124
ttran98@ford.com

Milad Zafar nezhad
Ford Motor Company
Dearborn, MI48124
MZAFARNE@ford.com

Devesh Upadhyay
Ford Motor Company
Dearborn, MI48124
dupadhya@ford.com

Kamran Paynabar
Georgia Institute of Technology
Atlanta, GA30332
kamran.paynabar@isye.gatech.edu

Anqi He
Ford Motor Company
Dearborn, MI48124
AHE6@ford.com

Abstract

In the automotive industry, the full cycle of managing in-use vehicle quality issues can take weeks to investigate. The process involves isolating root causes, defining and implementing appropriate treatments, and refining treatments if needed. The main pain-point is the lack of a systematic method to identify causal relationships, evaluate treatment effectiveness, and direct the next actionable treatment if the current treatment was deemed ineffective. This paper will show how we leverage causal Machine Learning (ML) to speed up such processes. A real-word data set collected from on-road vehicles will be used to demonstrate the proposed framework. Open challenges for vehicle quality applications will also be discussed.

1 Introduction

On-road quality issues of automobile vehicles not only cause financial burdens for the Original Equipment Manufacturer (OEM) but also impose repair based costs and delays to the vehicle owner. If unchecked these issues can adversely impact customer loyalty. Efficient root cause identification and mitigation can help alleviate some of these impacts.

To the best of our knowledge, an iterative approach for managing on-road vehicle quality issues systematically is not found in literature. In an industrial setting, the process of identifying and eliminating root causes for a given concern can take weeks since domain experts typically investigate for root causes from a large signal corpus, and recommend a fix. If the deployed treatment is found to be ineffective, the cycle repeats until convergence to zero faults. There are mainly three characteristics in data that are bottlenecks for providing solutions for vehicle quality issues efficiently:

1) imbalanced data: as would be expected, the number of reported failure examples are relatively small, typically ranging from 1%-10% of the entire vehicle population.

2) high dimensional data: there are a multitude of factors that can directly or indirectly lead to failures, such as differences in vehicle attributes, vehicle usage, etc.

3) dynamic data set: vehicle usage is not static. If not fully covered in the original data, out-of-distribution changes in these factors can impact the treatment effectiveness analysis. Additionally, there is no guarantee that data before and after the treatment on all vehicles can be collected due to varying levels of user authorization for data connectivity. This can often make assessment difficult.

To overcome these challenges, this paper focuses on developing a causality based analysis framework to improve the efficiency of such find-fix-track process to address on-road vehicle quality issues. The framework mainly consists of two parts - root cause analysis and treatment effectiveness analysis. Traditional root cause analysis - filtering methods focus on correlations but fail to consider interactions between feature and the influence of confounders. Example algorithms include Information Gain [TJ06], Symmetrical Uncertainty [WFH11], Correlation-based Feature Selection [Hal99], Chi-squared statistic [UW95] and Fisher Score [GLH12]. Machine learning based root cause analysis - embedded methods investigate the interactions between features, but still fail to eliminate effects of confounders. Some example algorithms are Random Forest [Bre01], XGBoost [CG16], and Logistic Regression [HLS13]. In this paper, we propose to employ causal machine learning methods to not only investigate feature interactions, but also eliminate effects from confounders.

The analysis of treatment effectiveness can be conducted using A/B tests given a fully randomized experimental design. However, real-world vehicle data is fully observational, and outcomes can be influenced by many external factors. Therefore, causal inference methods need to be employed to estimate the efficacy of treatment under the influence of confounders. State-of-the-art methods in this area include Propensity Score Matching [RR83], Inverse Propensity Weighting [RRZ94], Doubly Robust Estimator [BR05], Causal Forest [WA18], and Meta-learners (for example X-learner [Kün+19]). However, these methods only focus on estimation rather than the statistical testing of the significance of effects. In addition, current methods do not provide additional information for improving treatments.

To overcome the above difficulties and fill the gaps between current literature and practical needs, this paper proposes a two-level causal inference framework. The first level identifies the potential causes of a quality issue of interest, calculates the causal effects of any prior treatments applied and builds statistical tests for the significance of treatment effect at both entire population and individual vehicle levels. The second level employs causal inference on top of the causal effect estimation from the first level to interpret the direction and significance of each feature's impact on the treatment. This information can be used to understand the following: given the current treatment, what is the next potential causes to improve the treatment.

The main contribution of this paper is that to our knowledge this may be the first paper in the literature that focuses on iteratively resolving on-road vehicle quality issues by using causal inference methods. A complete framework is proposed for handling the full mitigation cycle including the identification of root causes, the statistical tests on the treatment effectiveness based on fully observational data, and the causal analysis of the critical features for the further improvement of the treatment. Comparing to ad-hoc methods, the proposed framework overcomes the difficulties of considering the interaction between features and confounders using causal inference. Also, this may be the first attempt at implementing an iterative statistical testing procedure for sequential improvement of treatments from observations before and after treatments at both entire vehicle population level and individual vehicle level. Our approach continues to improve treatments until convergence to zero concerns. Finally, we also discuss results from a real-world use case. This use case data is collected from on-road vehicles, is high-dimensional, highly-imbalanced with binary vehicle attribute features, continuous covariates - vehicle usage features and binary outputs.

2 Methodology

Assume we have data of the form $\{(x^{(k)}, y^{(k)}, T^{(k)}, z^{(k)}), k = 1, 2, \cdots, N\}$ where $x = (x_1, x_2, x_i, ..., x_n)$ are binary features indicating whether a vehicle has the corresponding attribute x_i (can also be referred as vehicle packages or modules). One or more features can be a potential cause for some quality issue; $y = (y_1, y_2, ..., y_m)$ are continuous co-variate features representing vehicle usage (i.e. weather, temperature or driving speed) which can influence (confound) the root cause; T is an binary indicator for whether the treatment is applied ($T = 1$ if so); and z is a binary response indicating if the vehicle has quality issue under consideration ($z = 1$ if so). For an overview of our
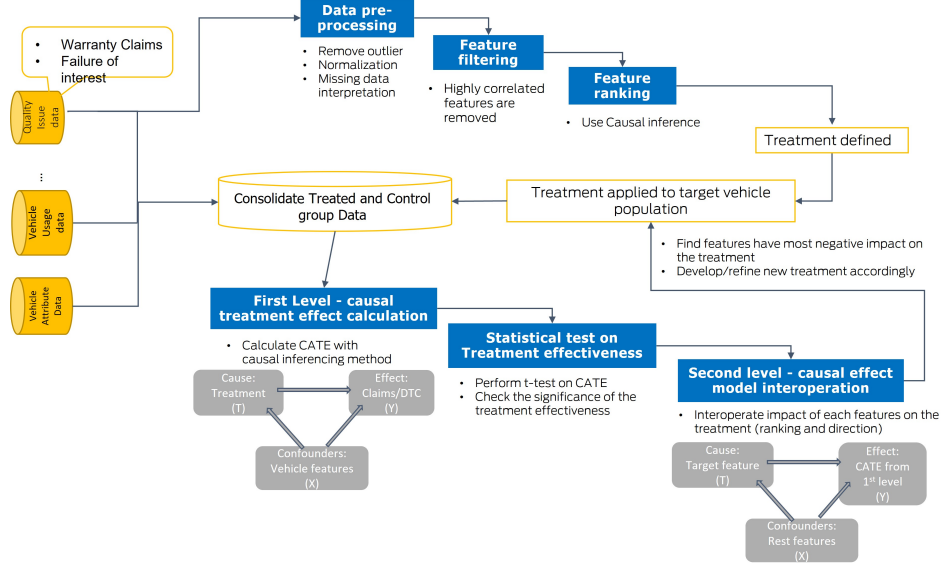
Figure 1: Flowchart of the proposed causal inference framework

proposed method, please see the flowchart attached (Figure 1). The corresponding pseudocode can be found in the Supplementary Material.

2.1 Feature filtering method for elimination of highly correlated binary features

Highly correlated features usually exist in industrial datasets, and they influence the performance of the subsequent root cause analysis. This is because correlated features will lead to the split of the same feature importance into each of the highly correlated features thereby removing significance from all features. This calls for the elimination of highly correlated features before further causal analysis.

This paper develops an empirical method to collect representative features *rep* from the set of all features and build a mapping to refer back from representatives to original features. Details can be found in the Supplementary Material. For the rest of this paper, input features used for analysis are all representative features, *rep* (i.e. highly correlated features are eliminated).

2.2 Root cause analysis based on causal inference

This section introduces the use of causal inference for the analysis of the potential root causes of a quality issue of interest. At this step, we have not yet identified the corresponding treatments, so data is only of the form $(x^{(k)}, y^{(k)}, z^{(k)})$.

First, we train the model $z = f(x, y)$ using XGBoost and remember to balance the class weight of z to compensate for the minority class $z = 1$ i.e. classes with quality issues reported/flagged. The quality outcome z varies with different vehicle features (x, y) and the features that drive the quality outcome closest to 1 are considered as strong influencers of the quality issue (since having quality issues means $z = 1$). These are the potential root causes that a treatment should focus on.

Thus, the strategy is to learn a first level model $z = f(x, y)$ using XGBoost to estimate the influence of the vehicle features $(x_1, x_2, .., x_n, y_1, y_2, ..., y_m)$ on the quality outcome z. After training f, the estimation for the effect of vehicle feature i on the quality outcome z is evaluated as follows:

$$\bar{e}_i = \frac{1}{N} \sum_k (f_{i1}(x^{(k)}, y^{(k)}) - f_{i0}(x^{(k)}, y^{(k)}))$$

where

$$f_{i1}(x, y) = f(x_1, .., x_i = 1, ..., x_n, y_1, ..., y_m)$$
$$f_{i0}(x, y) = f(x_1, .., x_i = 0, ..., x_n, y_1, ..., y_m)$$

3

Subsequently, the vehicle features with top "m" largest positive $\bar{e}_i$ are suggested as potential root causes leading to positive quality outcome $z = 1$. These "m" features are then used as heuristics to develop treatment T accordingly.

In this step, other root cause identification methods, as discussed in the introduction, can be used as well.

2.3 Statistical testing on treatment effectiveness

This section discusses how to leverage the data collected before and after treatment for the analysis of treatment effectiveness. The analysis has two steps: first we calculate the treatment effect for each vehicle; second, we build statistical tests for treatment effects at both the population level and at individual vehicle level.

Here we would like to explain the data collection before introducing the analysis methods. To test if the treatment is effective, we form two groups, the treatment group and the control group. The treatment group contains vehicles that have received the treatment, while the control group contains vehicles that have not yet received the treatment at the time of assessment. In both treatment and control groups, data is collected before and after the treatment application. Each group includes data from both problematic and healthy vehicles. After applying the treatment, the data has the form $(x^{(k)}, y^{(k)}, T^{(k)}, z^{(k)})$.

To analyze the treatment effectiveness, standard causal inference method like meta-learners (this paper uses X-learner) is applied to the data and the Conditional Average Treatment Effect (CATE) of each vehicle is calculated as $e_T^{(k)}$. During the training of base learners, since the quality outcome of vehicles may be biased, classes should be balanced. In this study, more weight is put on the minority class $z = 1$ i.e. the problematic ones.

To develop statistical tests to check the significance of the treatment effectiveness on the vehicle population, we assume CATE over the population to have Gaussian distribution $N\left(\mu, \sigma^2\right)$ with Average Treatment Effect (ATE) as the population mean μ. For a treatment to be effective at population level is equivalent to $ATE < 0$, and the hypothesis testing can be set up as

$$H_0 : \mu \geq 0; \quad H_1 : \mu < 0$$

T test is the standard way to test the sign of the population mean μ. Let $\bar{d}$: sample mean, s_d: sample standard deviation, N: number of samples, then

$$t = \sqrt{N}\bar{d}/s_d$$

We can declare ATE $= \mu < 0$ i.e. the treatment is significantly effective over the population at level α if

$$t < t_{N-1,\alpha}$$

However, it is trickier to assess if the treatment is effective for a specific vehicle. This can be informative in diagnosing the root cause. We apply bootstrapping to achieve a finer-level information of treatment effectiveness. A treatment being effective at the individual vehicle level $(x^{(k)}, y^{(k)})$ is equivalent to testing for $CATE\left(x^{(k)}, y^{(k)}\right) < 0$. We can apply X-learner on 1000 bags of bootstrapped data to calculate a two-sided $1 - 2\alpha$ confidence interval of $\mu^{(k)} := CATE\left(x^{(k)}, y^{(k)}\right)$, say $[lb, ub]$, which means

$$\mathrm{P}(lb \leq \mu^{(k)} \leq ub) = 1 - 2\alpha \Rightarrow \mathrm{P}(ub < \mu^{(k)}) = \alpha$$

We can declare $\mu^{(k)} = CATE\left(x^{(k)}, y^{(k)}\right) < 0$ i.e. the treatment is significantly effective on the individual vehicle $(x^{(k)}, y^{(k)})$ at level α if

$$ub < 0$$

2.4 Improvement analysis based on causal inference

After the hypothesis testing, a second level causal inference is employed to understand the effect of different vehicle features on the treatment, which can help the investigation on - which vehicle feature should be targeted next for further treatment improvement, if needed.

The treatment effect $e_T^{(k)}$ varies with different vehicles. The vehicle features that influence e_T in the negative direction the most should be the one that the current treatment has addressed, as this tends to move z most negatively from 1 to 0. On the other hand, the features that affected e_T in the most positive direction should be the next candidate for improvement via treatment, as it tends to move z most positively from 0 to 1.

Thus, a second level causal effect model $e_T = h(x, y)$ is built using XGBoost to estimate the influence of different vehicle features $(x_1, x_2, .., x_n, y_1, y_2, ..., y_m)$ on the treatment effect CATE e_T, which is calculated at the first level for each vehicle. After h is trained, the estimation formula for the effect of vehicle features i on the treatment effect e_T is

$$\bar{e}_i^{2nd} = \frac{1}{N} \sum_k (h_{i1}(x^{(k)}, y^{(k)}) - h_{i0}(x^{(k)}, y^{(k)}))$$

where

$$h_{i1}(x, y) = h(x_1, .., x_i = 1, ..., x_n, y_1, ..., y_m)$$
$$h_{i0}(x, y) = h(x_1, .., x_i = 0, ..., x_n, y_1, ..., y_m)$$

Subsequently, the vehicle features with top "m" largest positive $\bar{e}_i^{2nd}$ are selected. Since these are the major features that lead to the ineffectiveness of the current treatment T, we should use them as heuristics to develop a new treatment T'. This cycle can be iterated until a satisfactory treatment for the quality issues is developed. This concludes all analytical discussions of the proposed framework.

3 Use case study

This section discusses the application of the proposed framework on a real-world use case for a vehicle fleet. The problem at hand was related to high battery drain after vehicle key-off [How]. This issue lead to dead batteries, sometimes requiring early battery replacement under warranty. To fix the battery drain issue and avoid additional warranty cost, engineers leveraged connected vehicles data and the proposed framework to identify the root causes, take action, track the treatment effectiveness, and refine the treatments as necessary.

Overall, the data consists of three parts. First, input X, Y includes binary features for vehicle attribute and continuous features for vehicle usage. The total number of features are greater than 1500. Second, the output Z indicates the existence of vehicle battery drain issue. The number of problem cases was less than 10% of the total fleet population. The last component of data set is a treatment label, T, indicating whether a vehicle had a successful treatment applied. Treatments were made available via over the air updates (OTA). Since successful OTA updates can vary by vehicle and availability of connectivity, this provides an opportunity for a natural clustering of a treatment group and a control group. Hence, the control group forms with vehicles that have not yet had the treatment at the time of assessment.

The root cause for the battery drain issue was found to be related to a particular vehicle attribute (Feature A) in a particular model year vehicle. Therefore, the entire population is targeted at this model year vehicles. The treatment was then applied via OTA. Further sampling showed that approximately 75% of the impacted vehicles had successfully received this treatment. The remaining vehicles naturally formed the control group. Even though we tried to collect data for both before and after treatment, there still remained less than 10% of vehicles that lacked either before-treatment or after-treatment data due to inconsistencies in connectivity authorization by users. It is noted that the binary features - vehicle attributes are static, while the continuous features - vehicle usage might change overtime, i.e. for a same vehicle, before and after treatment, the distribution of continuous features might be changed.

All codes are built in Python, with major packages from CausalML [Che+20] and EconML [Bat+19]. To clean and filter the data, all missing data is interpreted by filling with median values, and representative features are selected, which reduces the dimension of feature space from 1500+ to around 300. The root cause analysis using the causal ML approach shows that the Feature A is within top 10 risk factors.

For treatment analysis, first, to evaluate the effect of mix types features, we evaluated the proposed framework with two different scenarios. One is with only binary features x, and the other with both binary x and continuous features y.

Table 1: Comparison results between four different causal models under two scenarios

CATE model	Binary only				Binary + Continuous			
	t-test stat	feature interpretation			t-test stat	feature interpretation		
		SHAP	Meta S	Self		SHAP	Meta S	Self
Meta X-LR+XGBoost	-466	1	3	n/a	-266	9	8	n/a
Meta X-XGBoost only	-466	2	2	n/a	-239	25	23	n/a
CF - DML	-390	7 (slow)	1	4	-356	slow	8	55
CF - GRF	-375	slow	1	1	-366	slow	5	2

Second, to calculate the causal treatment effect (CATE), four different causal effect models are built based on different methods - 1) Meta X-learner with Logistic Regression (LR) for propensity score calculation and XGBoost for CATE, 2) Meta X-learner with XGBoost for both propensity score and CATE calculation, 3) Causal Forest - Double Machine Learning (CF-DML) [Che+18] and 4) Causal Forest - Generalized Random Forest (CF-GRF) [ATW19] for CATE.

Then the proposed statistical test is conducted on the calculated CATE. As shown in Table1, given the critical value $t_{N-1,\alpha} = -1.645$ at level $\alpha = 0.05$, all test statistics are much smaller than the critical value, which shows the current treatment T is significantly effective for eliminating the battery drain issue under all models.

Finally, another three approaches are stacked with each learnt first-level causal effect model to interpret the influence direction and significance of each vehicle feature on the treatment effect. The three methods are 1) SHAP [LL17], 2) the proposed 2nd-level Meta S-learner (Meta S), 3) self-interpretation of tree based methods (Self, only for Causal Forest). The first two methods can provide both direction and significance information, while the third method can only provide the significance. For a given vehicle binary feature *A* as the known root cause and knowledge of the target of the current treatment, we evaluated the performance of the proposed framework by comparing its capability for interpreting the effect of this particular feature on the treatment. In other words, results should show that this feature has a strong impact on the treatment effectiveness. During the evaluation, we combine the direction and significance information into *Ranking in the good direction* as the evaluation metric for the first two methods, while only use significance information for ranking for the third. Feature *A* should be ranked on top for all three methods. The comparison results are shown in Table 1.

First, we found that the ranking results from the models built with only binary features outperform those with both binary and continuous features. It it noted that we only give ranks for binary features. This indicates that mix types of input features may challenge the base-model learning.

Second, self-interpretation fails to give direction information, and this makes it incapable of identifying features for treatment improvement as we cannot distinguish "good" features from "bad" features. Also, it is only available for Causal Forest (tree based) type of methods and not generally applicable.

When stacking SHAP with Causal Forest methods, it takes significantly longer to interpret the model. Hence SHAP is not suitable for tree based methods when time matters. But when stacked with Meta Learners, it performs well and gave top rankings for the target feature. Meta Learner with LR+XGBoost performs better than Meta learner with XGBoost only.

Finally, our proposed second level S-learner is generally stackable with any first-level causal model and can interpret the results in a reasonable amount of time (minutes) and gave top rankings with direction information. Moreover, the second level S-learner results in higher ranking when stacked with Causal Forest models than Meta Learners.

4 Conclusion

This paper proposed a two-level Causal Inference framework for analyzing root cause and treatment effectiveness for on-road vehicle applications. The proposed method enables a full cycle of managing in-use vehicle quality issues by leveraging connected vehicle data and advanced causal Machine Learning methods. Results demonstrated that the proposed framework provides an efficient and interpretable approach for investigating root causes of a quality issues, for track the effectiveness of applied treatment, and for refinement of treatments as needed. In this paper, we showed that the

proposed method can also provide information on potential risk factors that impede the effectiveness of treatments, however, the validation process of this step is still under investigation. Future work will also refine our method for multiple root causes and/or multi-level treatment effectiveness analysis.

References

[RR83]   Paul R Rosenbaum and Donald B Rubin. "The central role of the propensity score in observational studies for causal effects". In: *Biometrika* 70.1 (1983), pp. 41–55.

[RRZ94]   James M Robins, Andrea Rotnitzky, and Lue Ping Zhao. "Estimation of regression coefficients when some regressors are not always observed". In: *Journal of the American statistical Association* 89.427 (1994), pp. 846–866.

[UW95]   Antony Ugoni and Bruce F Walker. "The Chi square test: an introduction". In: *COMSIG review* 4.3 (1995), p. 61.

[Hal99]   Mark A Hall. "Correlation-based feature selection for machine learning". PhD thesis. The University of Waikato, 1999.

[Bre01]   Leo Breiman. "Random forests". In: *Machine learning* 45.1 (2001), pp. 5–32.

[BR05]   Heejung Bang and James M Robins. "Doubly robust estimation in missing data and causal inference models". In: *Biometrics* 61.4 (2005), pp. 962–973.

[TJ06]   MTCAJ Thomas and A Thomas Joy. *Elements of information theory*. Wiley-Interscience, 2006.

[WFH11]   Ian H. Witten, Eibe Frank, and Mark A. Hall. *Data Mining: Practical Machine Learning Tools and Techniques*. 3rd. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2011. ISBN: 0123748569.

[GLH12]   Quanquan Gu, Zhenhui Li, and Jiawei Han. "Generalized fisher score for feature selection". In: *arXiv preprint arXiv:1202.3725* (2012).

[HLS13]   David W Hosmer Jr, Stanley Lemeshow, and Rodney X Sturdivant. *Applied logistic regression*. Vol. 398. John Wiley & Sons, 2013.

[CG16]   Tianqi Chen and Carlos Guestrin. "Xgboost: A scalable tree boosting system". In: *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. 2016, pp. 785–794.

[LL17]   Scott M Lundberg and Su-In Lee. "A unified approach to interpreting model predictions". In: *Advances in neural information processing systems* 30 (2017).

[Che+18]   Victor Chernozhukov et al. *Double/debiased machine learning for treatment and structural parameters*. 2018.

[WA18]   Stefan Wager and Susan Athey. "Estimation and inference of heterogeneous treatment effects using random forests". In: *Journal of the American Statistical Association* 113.523 (2018), pp. 1228–1242.

[ATW19]   Susan Athey, Julie Tibshirani, and Stefan Wager. "Generalized random forests". In: *The Annals of Statistics* 47.2 (2019), pp. 1148–1178.

[Bat+19]   Keith Battocchi et al. *EconML: A Python Package for ML-Based Heterogeneous Treatment Effects Estimation*. https://github.com/microsoft/EconML. Version 0.x. 2019.

[Kün+19]   Sören R Künzel et al. "Metalearners for estimating heterogeneous treatment effects using machine learning". In: *Proceedings of the national academy of sciences* 116.10 (2019), pp. 4156–4165.

[Che+20]   Huigang Chen et al. *CausalML: Python Package for Causal Machine Learning*. 2020. arXiv: 2002.11631 [cs.CY].

[How]   Phoebe Howard. *2021 Ford F-150 batteries going dead due to phantom battery drain*. https://www.freep.com/story/money/cars/ford/2021/02/15/2021-ford-f-150-pickups-battery-drain/6753464002/. Accessed: 2022-09-29.

A Supplementary Material

A.1 Feature filtering method for elimination of highly correlated features

Highly correlated features usually exist in the industrial dataset, and they will influence the performance of the subsequent root cause analysis. This is because they will lead to the split of the same feature importance into each highly correlated feature, causing none of them being significant. This calls for the elimination of those highly correlated features before further causal analysis.

We developed an empirical method to collect the representative features *rep* from all the features and build a mapping m from each feature to its representative. All the features in the same representative group share the same feature importance as that of the representative. This allows for better interpretation of the results in the later stages.

The detailed example of the elimination process for binary features is as follows.

Algorithm 1 The elimination of highly correlated features

Require: Vehicle data with features $(x_1, x_2, ..., x_N, z)$

1: Calculated the mutual information of the response z with each of $(x_1, x_2, ..., x_N)$ and got results $(I_1, I_2, ...I_N)$
2: Reorder $(x_1, x_2, ..., x_N)$ by the magnitude of I_i in descending order
3: Initialize representatives list *rep* = $[x_1]$
4: Assign x_1 as the representative of x_1
5: **for** $i \leftarrow 2$ to N **do**
6:     Calculate the absolute values of correlation coefficients of x_i with each feature in *rep*
7:     **if** all the absolute correlation coefficients $\leq$ threshold **then**
8:         Add x_i to the *rep*
9:         Assign x_i as the representative of x_i
10:     **else**
11:         Find the feature $x_{m(i)}$ in *rep* that has highest absolute correlation coefficient with x_i
12:         Assign $x_{m(i)}$ as the representative of x_i
13: **Return** the representative list *rep* and the representative mapping m

---

A.2 Summary of the proposed Causal Inference framework

In this section, the detailed pseudo code for the two level causal framework is listed.

Algorithm 2 Two level causal inference framework for vehicle quality issues

---

Require: Vehicle data of the form $(x^{(k)}, y^{(k)}, z^{(k)}), k = 1, 2, \cdots, N$

1: Use Algorithm 1 to eliminate highly correlated features and get representatives
2: Train the model $z = f(x, y)$ using XGBoost
3: **for** $i \leftarrow 1$ **to** n **do**
4:     Estimate the effect of vehicle attribute feature i on the quality outcome: $\bar{e}_i = \frac{1}{N} \sum_k (f_{i1}(x^{(k)}, y^{(k)}) - f_{i0}(x^{(k)}, y^{(k)}))$
5: Pick the features with top largest positive $\bar{e}_i$ as suggested root causes and figure out treatment T accordingly
6: **while** treatment T is not promising **do**
7:     Apply the treatment to vehicles in the treatment/control groups and get new data of the form $(x^{(k)}, y^{(k)}, T^{(k)}, z^{(k)})$
8:     Estimate the treatment effect $CATE = e_T^{(k)}$ for each vehicle using the X-learner with base learners as XGBoost
9:     Test the significance of treatment effectiveness at population/vehicle levels as in section 2.3
10:     **if** treatment T is promising **then**
11:         **break**
12:     **else**
13:         Use the estimated $e_T^{(k)}$ to train the new model $e_T = h(x, y)$ using LASSO or XGBoost
14:         **for** $i \leftarrow 1$ **to** n **do**
15:             Estimate the effect of vehicle attribute feature i on the treatment: $\bar{e}_i^{2nd} = \frac{1}{N} \sum_k (h_{i1}(x^{(k)}, y^{(k)}) - h_{i0}(x^{(k)}, y^{(k)}))$
16:             Pick the vehicle attribute feature with top largest positive $\bar{e}_i^{2nd}$ and use them as heuristics to develop new treatment T
17: **Return** treatment T