# The benefits of Shapley-value in key-driver analysis

Received (in revised form): 14th December, 2020

**Marco Vriens**
Chief Executive Officer, Kwantum, USA

Marco Vriens is CEO of Kwantum, an analytics firm, and is an assistant professor at the University of Wisconsin-La Crosse. He is the author/editor of such books as 'The Business of Marketing Research' (2020), 'From Data to Decision: Handbook for the Modern Business Analyst' (2018), 'The Insights Advantage: Knowing How to Win' (2012) and 'The Handbook of Marketing Research' (2006). His work has been published in both academic journals and industry publications.

University of Wisconsin-La Crosse, 1725 State Street, La Crosse, WI 54601, USA
E-mail: mvriens@uwlax.edu

**Chad Vidden**
Associate Professor of Mathematics and Statistics, University of Wisconsin-La Crosse, USA

Chad Vidden is an associate professor of mathematics and statistics at the University of Wisconsin-La Crosse. He earned his PhD in applied mathematics from Iowa State University. He specialises in machine learning, data science, numerical analysis and computational mathematics.

University of Wisconsin-La Crosse, 1725 State Street, La Crosse, WI 54601, USA
E-mail: cvidden@uwlax.edu

**Nathan Bosch**
Master's Student, KTH Royal Institute of Technology, Sweden

Nathan Bosch is a master's student in machine learning at the KTH Royal Institute of Technology in Stockholm. He has a bachelor's degree in artificial intelligence from the University of Groningen. His research interests include system log file analysis, predictive maintenance and business analytics. He has also worked in applied machine learning.

KTH Royal Institute of Technology, Stockholm, Sweden
E-mail: nathangerbenbosch@gmail.com

**Abstract** Linear (and other types of) regression are often used in what is referred to as 'driver modelling' in customer satisfaction studies. The goal of such research is often to determine the relative importance of various sub-components of the product or service in terms of predicting and explaining overall satisfaction. Driver modelling can also be used to determine the drivers of value, likelihood to recommend, etc. A common problem is that the independent variables are correlated, making it difficult to get a good estimate of the importance of the 'drivers'. This problem is well known under conditions of severe multicollinearity, and alternatives like the Shapley-value approach have been proposed to mitigate this issue. This paper shows that Shapley-value may even have benefits in conditions of mild collinearity. The study compares linear regression, random forests and gradient boosting with the Shapley-value approach to regression and shows that the results are more consistent with bivariate correlations. However, Shapley-value regression does result in a small decrease in k-fold validation results.

KEYWORDS: driver modelling, regression, Shapley-value, customer satisfaction, random forests, gradient boosting

## INTRODUCTION

Customer satisfaction has decreased, and your firm has decided to conduct a customer satisfaction survey. An overall satisfaction measure is included along with various questions that pertain to satisfaction with three sub-elements of the product or service: satisfaction with overall technical support (V1); satisfaction with the product's functionality (V2); and satisfaction with the instruction manual (V3). Data are collected for each of these measures, as well as overall satisfaction, using a ten-point rating scale ranging from 1 (totally not satisfied) to 10 (totally satisfied). To determine how the three variables are affecting overall satisfaction, a linear regression analysis is conducted using overall satisfaction as the dependent variable and the various sub-elements as predictors (ie independent variables). The analysis will yield a regression coefficient for each element. The results can be used to predict the overall satisfaction scores. So far so good: linear regression — often referred to as driver modelling (or key-driver analysis) — is frequently used in commercial satisfaction research in this manner. In practice, driver models are used to model overall satisfaction, but they can also model other variables such as overall value and likelihood to recommend. The objectives of such key-driver studies are all very similar, in that they seek to improve the performance of the dependent variable. A frequent challenge presents itself, however: it is not uncommon to have 20 or more predictors. This means management will have to prioritise, as it is not possible to improve this many areas simultaneously. To help with prioritising improvements, regression results can be used to determine the relative importance of each element. In other words, it is possible to select the most important variables and try to improve performance on these, using the predictive model to set overall satisfaction goals. This is a straightforward approach.

The application of linear (or logistic) regression models in such driver-modelling studies, however, has several shortcomings.[1,2] First, a regression model is optimised for prediction not relative importance. That is, if two highly correlated variables both explain variance in the dependent variable in some similar way, then once one of them has entered the model (ie contributed its explaining power) there is not much to be added by the second independent variable. This leads to the first variable being large, and likely statistically significant, whereas for the second variable the coefficient will be much smaller, may have a negative sign (counterintuitive effects), or may simply be non-significant. Its regression-derived importance will thus be low or not usable. Here, one would conclude that the first variable is important and the second is not, even though both variables have similar positive correlations with the dependent variable. This is not so good.

Secondly, customer satisfaction and other survey data can sometimes display collinearity and multicollinearity. Collinearity is said to exist when the independent variables are related to each other in a bivariate way, while multicollinearity exists when some independent variables can be explained as a linear combination of other independent variables. Without experimental data, there will always be some degree of (multi) collinearity, especially when independent variables are measured using the same scale. Multicollinearity can be further exacerbated by response style biases such as halo[3,4] response style — a cognitive tendency to evaluate all features positively based on a positive evaluation of an important (halo) feature or the reverse (ie to evaluate everything negatively) based on a negative impression of one feature.

Multicollinearity is a well-documented issue in regression analysis[5] and results in increased variance around the regression estimates potentially leading to sign reversal (ie counterintuitive effects), and coefficient instability, which can result in non-significant coefficients. The latter means that, in a commercial application, one must advise one's client that the topic with a non-significant coefficient has zero importance. This is hard to sell to stakeholders as it undermines the credibility of the results.[6]

Many different approaches have been proposed to mitigate multicollinearity. For example, principal components regression (PCR) derives orthogonal components (ie linear combinations of variables with mutually exclusive information) from predictor variables, and regresses those onto a dependent variable. The main disadvantage of this approach is that it results in the loss of the original variables' meaning, which is to say that the components are linear functions of the original variables. This can potentially cloud interpretability. Another method sometimes suggested for handling multicollinearity is ridge regression, which is seeks to reduce the variance around regression estimates by adding a ridge parameter (k) to the regression equation. Finding a value of k that minimises the variance around coefficients is not a simple task, however, and is subject to researcher bias.[7] Neither of these methods solve the challenge of multicollinearity fully.

In the context of linear regression, the Shapley-value approach has been proposed as an alternative method to mitigate the negative effects of multicollinearity and hence get better derived relative importance estimates.[8–10] Using the Shapley-value approach, the relative importance of a variable is defined as the average importance of that variable across all possible linear models that are feasible with the set of independent variables (main effects only).

In addition to linear (logistic) regression, one may choose various machine-learning approaches, such as random forests (RF) and gradient boosting (GB) methods.[11–13] These alternative models maintain interpretability, require few subjective decisions,[14] and may offer some protection against multicollinearity.[15] In a linear model, a good fit will only emerge when the relationships between the independent variables and the dependent variable are sufficiently linearly related. It may be reasonable to assume a linear relationship between a customer's overall satisfaction and their satisfaction with the price of the product. However, this assumption may be less reasonable when identifying the relationship between the user characteristics of the customer and their overall satisfaction (for example, the number of years the customer has been in the market for the product may affect their satisfaction in a non-linear way). RF and GB methods do not have such a strict linear relationship constraint and can also simply ignore irrelevant features and are invariant under scaling,[14] both of which make the implementation of such models much easier than other parametric approaches. In such cases, identifying the feature importance values through a non-linear model may be beneficial.

The present paper aims to add to previous research by making the following contributions. First, it is well known that multicollinearity is not always a huge problem.[16] This paper illustrates how relative importance estimates are affected in a mild multicollinearity situation. The study compares the relative importance estimates derived from linear regression, the Shapley-value approach applied to linear regression models, to RF, GB trees, and the Shapley-value method applied to the RF and GB modelling approaches.

In the literature, it has been argued that Shapley-value results in better estimates of relative importance than estimates based on the OLS coefficients.[8,9,17]

Shapley-value regression estimates are more in line with Pearson correlation

**BOX 1: THE SHAPLEY-VALUE METHOD**

1. Explained variance of a model with only V1
2. Explained variance of a model with V1 and V2 minus the explained variance of a model with only V2
3. Explained variance of a model with V1 and V3 minus the explained variance of a model with only V3
4. Explained variance of a model with V1 and V4 minus the explained variance of a model with V4
5. Explained variance of a model with V1, V2 and V3 minus a model with V2 and V3.

The process is repeated until one reaches:

6. Explained variance of a model with V1, V2, V3 and V4 minus the explained variance of a model with V2, V3 and V4.

Then, the mean explained variance is calculated for each variable (V1, V2, V3, V4), and these numbers are re-scaled to 100.

coefficients and the results are more stable; for example, there will be less variation across coefficients across, say, multiple waves in a tracking study.[18] This is why some claim such estimates are better. If they truly are better, then this would imply they should provide a better representation of the data-generating mechanism, which in turn would imply the Shapley-value adjusted coefficients should yield superior results in a k-fold evaluation.

A method has been derived to adjust the linear regression coefficient using Shapley regression results; it has also been suggested that those coefficients should be used for predictions in place of ordinary least squares (OLS)-derived coefficients.[17] However, little is known about how using Shapley-value affects the predictive power of the model. The k-fold predictive accuracy metric[19] is an indication of how strong the model is and, as a result, how confident one can be that the results are representative of the data-generating mechanism. With this in mind, the present study evaluates Shapley-value adjusted results in terms of k-fold predictive accuracy. (Note that no such adjustments currently exist for RF and GB). This paper uses a blinded commercial dataset to evaluate the Shapley-value method.

## THE SHAPLEY-VALUE APPROACH

The Shapley-value approach is a fully recursive modelling approach that determines the relative importance of predictor variables by running and comparing every possible model among a set of predictors.[15,17] As an example, say one wished to understand the effects of four predictors (v1, v2, v3, v4) on a single dependent variable, Shapley regression calculates all possible models involving those variables, compares them (as shown in Box 1), and computes the average across them to find the unique contribution of v1 to $R^2$ of the whole model. It then repeats that process for v2, v3 and v4.

Shapley-value is a general approach to partition the variance explained by the various predictors and has several appealing features. First, Shapley-value is in essence a bagging approach, which is to say it uses the mean explained variance from across a number of models (see Box 1). This means that it captures the effect of a given variable without the dominating effect of other independent variables. It also means that it is highly unlikely that any given variable will obtain a value of zero. Secondly, it is a general framework that can be applied to different types of analyses, such as logistic regression and other machine-learning approaches. Of course, the exact implementation must be adjusted to fit the modelling approach at hand.

As an example, say one wished to understand the effects of four predictors (v1, v2, v3, v4) on a single dependent variable. Shapley regression calculates all possible models involving those variables, compares them as shown in Box 1, and computes averages across them to find the unique contribution of v1 to $R^2$ of the whole model. It then repeats that process for v2, v3 and v4.

There are several appealing features to this variance partitioning approach. By averaging some statistics across all possible models involving a given predictor, one can capture the effect of that predictor independently from the effects of other related predictors.

One disadvantage of Shapley-value regression is that it is computationally intensive. As per Box 1, 15 separate regression models must be run to find the unique contribution of each of the predictors. (The same 15 regression models can be used to find unique contributions of any of the four variables, the results just need to be compared in different ways.) Running 15 models given four predictors is not that hard computationally, but the size of the problem increases exponentially as predictors are added. The construct for determining the number of models, given p predictors is $2^p - 1$. Models with 10, 20 and 30 predictors require computing roughly 1,000, 1 million and 1 billion separate regression models, respectively. For larger problems, one soon runs out of the computing resources required for Shapley regression to be a practical approach.

Previous research has evaluated the Shapley-value method in the context of identifying customer satisfaction drivers. Tang and Weiner[18] raised the issue of multicollinearity in customer satisfaction studies and offered the Shapley-value method as a possible solution. Part of their reasoning was that customer satisfaction studies are often tracking studies (ie they are repeated over time, say once or twice a year). This creates the following challenge: say there are four predictors (v1, v2, v3 and v4) and that all of these are highly correlated with each other and with overall satisfaction. In wave 1, the effect of v1 turns out to be statistically significant and has a relatively large regression coefficient, while the effect of v2 is minimal and has a regression coefficient near zero. One would conclude from this that v1 is important and v2 is not, even though both are substantially correlated with the dependent variable. This faulty conclusion is driven by multicollinearity among the variables in the model. Wave 2 of the same research might easily reverse the relationship between v1 and v2 because of very small changes in correlations, which would reverse the prior conclusion about which variable is most important. From a 'communication of insights' perspective, this situation is clearly undesirable as, following Wave 1, management will be recommended to prioritise V1, but following Wave 2, management will be told to forget V1 and go improve V2.

The Tang and Weiner study had two additional interesting results. First, using data from a commercial customer satisfaction study, they found that the mean gap between Wave 1 and Wave 2 is much smaller when using Shapley-value instead of relative importance values derived from linear regression. Secondly, they found that as sample size decreases, these gaps grow much faster for OLS than for Shapley-value relative importance values.

The Shapley-value approach has traditionally been applied to linear regression. Recently, SHAP has been developed, which can help find feature importance estimates for many machine-learning models, and uses game theory to be more efficient than standard Shapley-value approaches.[20] Several variants have been proposed; for example, TreeSHAP is an approach to perform Shapley-value regression on tree-based models that is more efficient than standard applications.[21]

## AN EMPIRICAL COMPARISON AND ILLUSTRATION

A disguised commercial dataset is used to illustrate the above arguments. The data were collected in 2020 and pertain to a business-to-business service. The sample size is $n = 497$, and there is one variable measuring overall value, and seven independent variables. The Pearson inter-correlations ranged from roughly 0.40 to 0.70. The variance-inflation factors (VIF) ranged from about 1.4 to about 2.4 — well below the threshold beyond which collinearity is said to be an issue (sometimes VIF > 5, sometimes VIF > 10).[16,22] So, this represents a situation of mild multicollinearity.

A linear regression model was run on the datatset, and the coefficients used to determine the relative importance. The authors also ran a Shapley-value analysis for RF and GB and used Python code along with scikit-learn and shap modules. There are several ways Shapley-value regression can be implemented. The present analysis used the approach described by Mishra.[23]

The feature importance estimation in random forests and gradient boosting methods is generally done the same way as decision trees, but instead averaged over each tree. The feature importance estimates may suffer from some bias.[24,25] Here too, one can employ a Shapley-value approach to get a better approximation of the feature importance values.

Seven sets of relative importance values were calculated: (1) derived from the regression coefficients, the squared standardised Betas are used to re-scale the relative importance values; (2) using Shapley-value regression; (3) using RF; (4) using Shapley-value RF; (5) using GB; (6) using Shapley-value GB; and (7) using Pearson correlation coefficients between dependent variable and the independent variables. The results are shown in Table 1.

All three methods of determining relative importance produced comparable rank orders among the predictors, especially among the top-ranked predictors. This was mainly due to these predictors being very close to one another in relative importance across the methods. As Table 1 shows, the difference between the maximum relative importance and the minimum relative importance varies dramatically across methods. The Shapley-value regression shows the smallest difference apart from the Pearson correlation coefficient.

If Shapley-value regression is truly better (ie it is more likely that this approach reveals the true data-generating mechanism) then this should translate into a model with better predictive accuracy than the linear regression model. Shapley-value numbers can be used to adjust the original linear regression numbers and these adjusted coefficients can be used to make predictions. A comparison is made of both linear regression, Shapley-value regression, RF and GB on their performance in k-fold predictive validation. Currently, no procedures exist to Shapley-value adjust RF and GB modelling results. The R package Carat is used for this. The results are shown in Table 2.

Fit statistics in predicting holdouts were similar across the various methods, with the OLS method performing only slightly better than the Shapley-adjusted regression coefficients. RF had the same in-sample fit as OLS but performed worse in the k-fold validation. GB also had better in-sample fit and had the same k-fold validation than OLS. So, Shapley-value does perform worse than base regression methods. To shed

Table 1: The relative importance of variables: Regression vs Shapley-value vs correlation

| | **OLS Regression** | | | **Shapley value** | | **Random Forests** | | **Shapley value** | | **Gradient Boosting** | | **Shapley value** | | **Pearson correlation coefficients** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Coefficients** | **Relative Importance** | **Rank** | **Relative importance** | **Rank** | **Relative Importance** | **Rank** | **Relative importance** | **Rank** | **Relative Importance** | **Rank** | **Relative importance** | **Rank** | **Correlations** | **Relative Importance** | **Rank** |
| Predictors | | | | | | | | | | | | | | | | |
| R1 | 0.17 | 12.7% | 3 | 15.2% | 3 | 7.9% | 3 | 11.7% | 3 | 10.3% | 3 | 17.2% | 3 | 0.54 | 14.6% | 3 |
| R2 | 0.32 | 45.5% | 2 | 25.1% | 2 | 36.2% | 2 | 28.4% | 2 | 37.7% | 2 | 30.6% | 2 | 0.65 | 21.3% | 1 |
| R3 | 0.06 | 1.4% | 5 | 9.6% | 5 | 2.0% | 5 | 3.8% | 4 | 2.2% | 5 | 6.3% | 4 | 0.48 | 11.6% | 5 |
| R4 | 0.04 | 0.8% | 7 | 7.5% | 6 | 2.0% | 5 | 3.2% | 6 | 0.4% | 7 | 1.0% | 7 | 0.43 | 9.4% | 6 |
| R5 | 0.29 | 36.8% | 1 | 24.8% | 1 | 48.1% | 1 | 47.5% | 1 | 43.8% | 1 | 35.9% | 1 | 0.65 | 21.2% | 2 |
| R6 | 0.06 | 1.7% | 4 | 10.9% | 4 | 2.3% | 4 | 3.6% | 5 | 3.3% | 4 | 4.9% | 5 | 0.51 | 12.9% | 4 |
| R7 | 0 (N.S.) | 1.1% | 6 | 6.9% | 7 | 1.1% | 6 | 1.5% | 7 | 2.0% | 6 | 3.7% | 6 | 0.42 | 9.0% | 7 |
| DIFFERENCE | | 44.4 | | 18.2 | | 47.0 | | 46.0 | | 41.8 | | 32.2 | | 12.2 | 12.3 | |
| MAXIMUM - | | | | | | | | | | | | | | | | |
| MINIMUM | | | | | | | | | | | | | | | | |

Table 2: K-fold validation for linear regression, Shapley-value regression, RF and GB (N = 497; 7 independent variables)

| OLS | | | |
|---|---|---|---|
| In-sample | | K-fold validation | |
| RMSE | MAE | RMSE | MAE |
| 0.64 | 0.49 | 0.65 | 0.5 |
| Shapley-value (linear regression) | | | |
| In-sample | | K-fold validation | |
| RSME | MAE | RSME | MAE |
| 0.66 | 0.51 | 0.66 | 0.52 |
| Random forest | | | |
| In-sample | | K-fold validation | |
| RMSE | MAE | RMSE | MAE |
| 0.6 | 0.47 | 0.66 | 0.51 |
| Gradient boosting | | | |
| In-sample | | K-fold validation | |
| RSME | MAE | RSME | MAE |
| 0.59 | 0.48 | 0.65 | 0.5 |

further light on this, Figure 1 plots the actual versus predicted y-values.

As one can see in Figure 1, the predicted performance across the various methods is very similar, except for low value of the dependent variable, where the difference between Shapley-value seems to deteriorate worse relative to the other methods.

## DISCUSSION

Survey-based customer satisfaction key-driver analysis research is one of the most pervasive applications in marketing research, and quite often the customer satisfaction line-item is the single biggest budget item in the marketing research budget.[6] This is especially true when the customer satisfaction study is executed as a tracking study. When survey questions capture multiple potential sub-components of customer satisfaction, management must identify the key drivers. In other words, management must determine which of

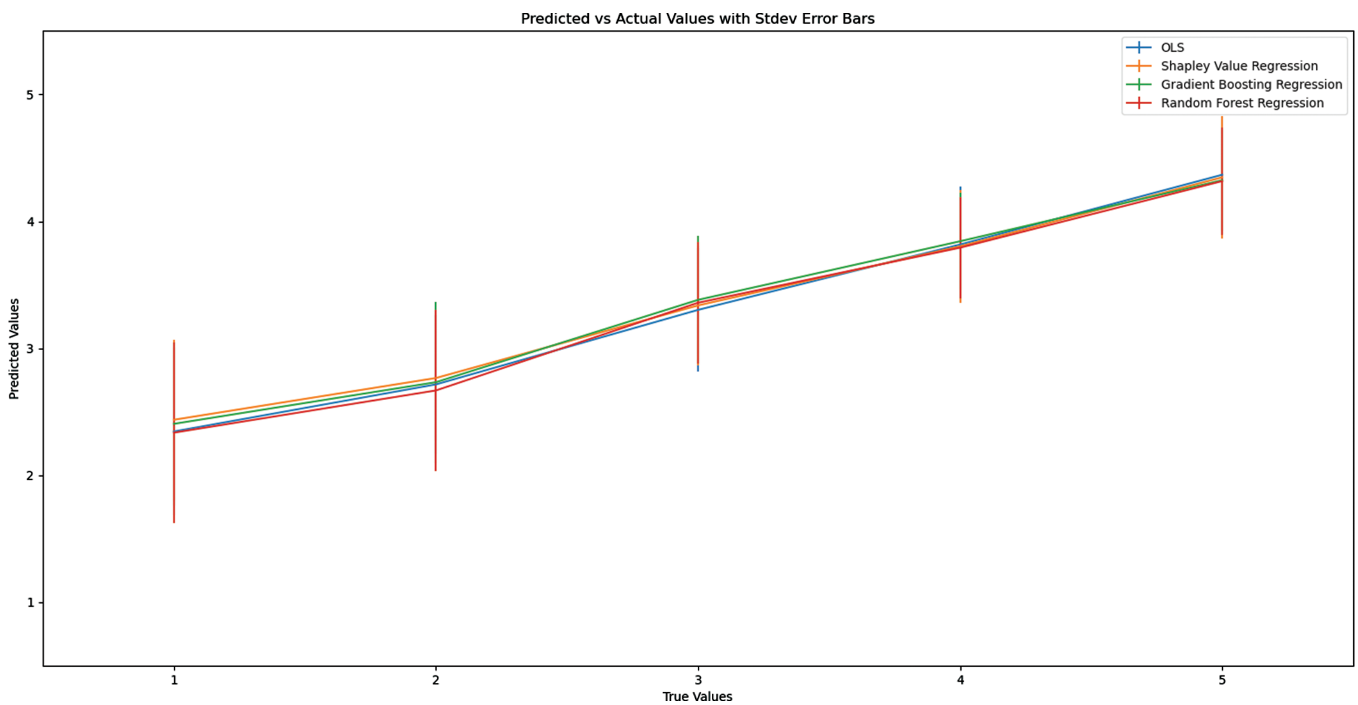


**Figure 1:** Actual versus predicted values from the various methods

the many sub-components it should focus on improving as there is unlikely to be sufficient budget or time to improve them all. To provide management with insight into the relative importance of the sub-components, researchers often rely on linear or logistic regression techniques. As this paper has noted, however, this exercise is not without its challenges as one may encounter collinearity and multicollinearity. As multicollinearity increases (ie as indicated by VIF > 5, and in some cases even sooner),[16,22] the risk of misleading relative importance estimates increases as well. Tang and Weiner[11] also showed that this can be an even bigger problem in tracking studies, as it can flip the relative importances across waves, thus compromising credibility.

This paper presented the results of a commercial study, and found that the Shapley-value method does a better job of evening out the relative importance of multiple variables versus the relative importance of estimates extracted directly from regression coefficients and machine-learning approaches such as RF and GB. Regression and machine-learning approaches can be very sensitive to what one might refer to as the 'first come, first served' problem: In a regression model, the first variable able to explain a decent chunk of the variance will deny other variables the opportunity to show their explanatory power. The Shapley-value method provides insurance against this phenomenon.

In predictive models, including linear regression, RF and GB, the estimates are identified in such a way that they optimise predictive accuracy (ie explained variance). So, this study tested the degree to which in-sample and out-of-sample (k-fold validation) accuracy compared when based on the original regression coefficients vis-à-vis the Shapley-value adjusted coefficients. The results show that the accuracy indeed decreases a little bit for the Shapley-value approach, albeit not by much. This is consistent with what others have recommended: ie not to adjust the original regression coefficients using the Shapley-value numbers.[26] It is worth noting that Shapley-value performs worst in the lower range of the dependent variable.

Based on the results of the present study, and the results of previous studies, it is recommended to use the results of a Shapley-value regression alongside the results from standard OLS regression, Random Forest, or other base regression approach when determining the relative importance of attributes is a primary objective. Looking at Shapley-value holds value in cases of severe multicollinearity, but also has value in situations with mild multicollinearity.

In sum, the work by Tang and Weiner,[16] considered in tandem with the results described herein, make a strong case for the use of the Shapley-value approach to determine relative importance of potential drivers in key-driver analysis research.

## References


1. Stratmann, W.C., Zastowny, T.R., Bayer, L.R., Adams, E.G., Black, G.S. and Fry, P.A. (1994) 'Patient satisfaction surveys and multicollinearity', *Quality Management in Healthcare*, Vol. 2, No. 2, pp. 1–12.
2. Kalnins, A. (2016) 'Multicollinearity: How common factors cause Type 1 errors in multivariate regression', *Strategic Management Journal*, Vol. 39, No. 8, pp. 2362–2385.
3. Thorndike, E.L. (1920) 'A constant error on psychological rating', *Journal of Applied Psychology*, Vol. 4, No. 1, pp. 25–29.
4. Bueschken, J. , Otter, T. and Allenby, G. M. (2010) 'Dimensionality of customer satisfaction surveys and implications for driver analysis', *Marketing Science*, Vol. 32, No. 4, pp. 533–553.
5. Ofir, C. and Khuri, A. (1986) 'Multicollinearity in marketing models: Diagnostics and remedial measures', *International Journal of Research in Marketing*, Vol. 3, No. 3, pp. 181–205.
6. Vriens, M., Rademaker, D. and Verhulst, R. (2020) 'The Business of Marketing Research', Cognella Academic Publishing, San Diego, CA.
7. Dorugade, A.V. and Kashid, D.N. (2010) 'Alternative method for choosing ridge parameter for regression', *Applied Mathematical Sciences*, Vol. 4, No. 9, pp. 447–456.
8. Budescu, D.V. (1993) 'Dominance analysis: A new approach to the problem of relative importance of predictors in multiple regression', *Psychological Bulletin*, Vol. 114, No. 3, pp. 542–551.

9. Johnson, J.W. (2000) 'A heuristic method for estimating the relative weight of predictor variables in multiple regression', *Multivariate Behavioral Research*, Vol. 35, No. 1, pp. 1–19.
10. Lipovetsky, S. and Conklin, M. (2001) 'Analysis of regression in game theory approach', *Applied Stochastic Models in Business and Industry*, Vol. 17, No. 4, pp. 319–330.
11. Breiman, L. (2001) 'Random forests', *Machine Learning*, Vol. 45 No. 1, pp. 5–32.
12. Friedman, J. H. (2001) 'Greedy function approximation: a gradient boosting machine', *Annals of Statistics*, Vol. 29, No. 5, pp. 1189–1232.
13. Friedman, J. H. (2002) 'Stochastic gradient boosting', *Computational Statistics and Data Analysis*, Vol. 38, No. 4, pp. 367–378.
14. Hastie, T., Tibshirani, R. and Friedman, J. (2009) 'The Elements of Statistical Learning: Data Mining, Inference, and Prediction', Springer Science and Business Media, New York, NY.
15. Kotsiantis, S. B. (2013) 'Decision trees: a recent overview', *Artificial Intelligence Review*, Vol. 39, No. 4, pp. 261–283.
16. Johnston, R., Jones, K. and Manley, D. (2018) 'Confounding and collinearity in regression analysis: A cautionary tale and an alternative procedure, illustrated by studies of British voting behaviour', *Quality and Quantity*, Vol. 52, No.4, pp. 1957–1976.
17. Conklin, M., Powaga, K. and Lipovetsky, S. (2004) 'Customer satisfaction analysis: Identification of key drivers', *European Journal of Operational Research*, Vol. 154, No. 3, pp. 819–827.
18. Tang, J. and Weiner J. (2005) 'Multicollinearity in CSAT studies', in: 'Proceedings of the 11th Sawtooth Software Conference, San Diego, CA, 6th–8th October, 2004', pp. 83–90.
19. Shao, J. (1993) 'Linear model selection by cross-validation', *Journal of the American Statistical Association*, Vol. 88, No. 422, pp. 486–494.
20. Lundberg, S.M. and Lee, S. (2017) 'A unified approach to interpreting model predictions', in: 'Proceedings of 31st International Conference on Neural Information Processing Systems', Longbeach, CA, pp. 4765–4774.
21. Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., and Lee, S.-I. (2020) 'From local explanations to global understanding with explainable AI for trees', *Nature Machine Intelligence*, Vol. 2, No. 1, pp. 2522–5839.
22. Chennamaneni, P.R., Echambadi, R., Hess, J.D. and Syam, N. (2015) 'Diagnosing harmful collinearity in moderated regressions: A roadmap', *International Journal of Research in Marketing*, Vol. 33, No. 1, pp. 172–182.
23. Mishra, S.K. (2016) 'Shapley-value regression and the resolution of multicollinearity', *Journal of Economics Bibliography*, Vol. 3, No. 3, pp. 498–515.
24. Deng, H., Runger, G. and Tuv, E. (2011) 'Bias of importance measures for multi-valued attributes and solutions', in 'Proceedings of the 21st International Conference on Artificial Neural Networks', Springer, Berlin and Heidelberg, pp. 293–300.
25. Toloşi, L. and Lengauer, T. (2011) 'Classification with correlated features: unreliability of feature ranking and solutions', *Bioinformatics*, Vol. 27, No. 14, pp. 1986–1994.
26. Grömping, U. and Landau, S. (2010) 'Do not adjust coefficients in Shapley value regression', *Applied Stochastic Models in Business and Industry*, Vol. 26, No. 2, pp. 194–202.