2020-10-25

Comparing Variable Importance in Prediction of Silence Behaviours between Random Forest and Conditional Inference Forest Models.

Stephen Barrett Dr
Technological University Dublin, stephenbarrett1850@hotmail.com

Geraldine Gray Dr
Technological University Dublin, geraldine.gray@tudublin.ie

Colm McGuinness Dr
Technological University Dublin, colm.mcguinness@tudublin.ie

Recommended Citation

Barrett, S., Gray, G., & McGuinness, C. (2020). Comparing Variable Importance in Prediction of Silence Behaviours Between Random Forest and Conditional Inference Forest Models. *DATA ANALYTICS 2020: 9th International Conference on Data Analytics* pg. 28-34. doi:10.21427/9rsv-0479

Authors
Stephen Barrett Dr, Geraldine Gray Dr, Colm McGuinness Dr, and Michael Knoll Dr.

Comparing Variable Importance in Prediction of Silence Behaviours between Random Forest and Conditional Inference Forest Models

Stephen Barrett, Geraldine Gray
School of Informatics
Technological University Dublin
Dublin, Ireland
S.Barrett@live.ie, Geraldine.Gray@tudublin.ie

Colm McGuinness
School of Business
Technological University Dublin
Dublin, Ireland
Colm.McGuinness@tudublin.ie

Michael Knoll
Dept. of Work & Organizational Psychology
University of Leipzig
Leipzig, Germany
Michael.Knoll@uni-leipzig.de

***Abstract*—This paper explores variable importance metrics of Conditional Inference Trees (CIT) and classical Classification And Regression Trees (CART) based Random Forests. The paper compares both algorithms variable importance rankings and highlights why CIT should be used when dealing with data with different levels of aggregation. The models analysed explored the role of cultural factors at individual and societal level when predicting Organisational Silence behaviours.**

***Index Terms*—Random Forest; Variable Importance; Bias; PDP; Survey Data; Culture; GLOBE; Organisational Silence.**

Research indicates that there are many individual reasons why people do not speak up when confronted with situations that may concern them within their working environment. This phenomenon is referred to as Organisational Silence, and is considered both an individual and collective level behaviour [1]. Employees do not call attention to problems that make their life uncomfortable within an organisation, resulting in self-censorship and trivialisation of problems [2]. The end result is that employees make a decision to stay silent [3].

Organisational Silence impacts both an organisation's ability to adapt to change and individuals who experience it [4]. Companies are now becoming more global in their outlook, necessitating research into how culture may play a role in developing bespoke feedback mechanisms. In 1999, foreign sales by multi-national businesses exceeded $7 trillion dollars a year and had a growth rate of over 20% more than traditional exports [5]. This highlights a need for managers with a global mindset, of which a shortage exists among the fortune 500 companies [6].

Organisational Silence had been explored previously with respect to societal culture in a tangential manner but not as the core focus of research papers. During the literature review for this paper, it was found that several papers focused on facets of *organisational* culture and how they predicted sub domains of Silence [7]. At the time of writing no research was found that modelled the probability of engaging in Silence based on *societal* differences across cultures. The previous studies did not focus on cultural and organisational attributes that contribute to classify if a person engaged in Silence behaviours. It has been hypothesised that a manager's leadership type could moderate the effect of Silence behaviours as a result of culture [8], however there was no analytic work applied to the topic.

The aim of this study is to explore patterns found by models that predict an employee's propensity to engage in Silence. Variable importance measures are examined to understand why two models differed in their rankings of importance. *Partial Dependency Plot*s (PDP) are used to explore the impact of changes in the predictors on Silence behaviours. This study involves the analysis of data collected from three countries representing three different cultures (Germany, Italy and Poland).

Section I describes the two ensemble models used in analysing the survey data and highlights two methods for examining the role of predictors in predicting Silence behaviours. Section II describes the survey instrument used to collect the data for this study. Section III describes the analysis done, and the results of data modelling. We conclude this work in section IV.

I. Ensemble Modelling

An ensemble mechanism takes more than one model and trains it on a particular problem. Each model - especially if they are highly variable models like Decision Trees - takes advantage of the model's tendency to overfit the data [9]. To use an analogy, each model is an expert in its particular area, for example a specific attribute in the data set. When all the models are fit to the data, their expert opinions are combined to make a final decision. In modelling, this can be implemented as the mode across all model outputs in a classification problem. These methods use simpler base models as their constituent parts, where voting or aggregation of the results can produce extremely accurate classifiers. It has been pointed out in the context of bioinformatics that ensembles of Data Mining classifiers have the ability to reduce model bias and model overfitting, especially in datasets with class imbalance problems [10]. This study utilises two ensemble techniques, discussed next.

 ISBN: 978-1-61208-816-7

A. Random Forest

Random Forest can be used with any modelling technique to attempt to improve its accuracy. However it is generally associated with Decision Trees or Regression Trees. For this study, one base learner was the CART model [11]. The user can specify the number of trees to be included in the Random Forest. Each tree is allowed to grow fully without being pruned back, producing many overfit trees. However, the algorithm introduces randomisation into the process by applying bootstrap sampling with replacement to the dataset. A second randomisation step only allows the model to split on a random selection of attributes at each node for each tree, decorelating the trees [12]. By necessitating that only a random sample of predictors is used at each split, the trees can focus on less predictive predictors that would have been overshadowed by more powerful predictors in the dataset. The two randomisation steps result in different trees overfitting different sections of the dataset. Classification is based on a majority vote amongst all trees.

CART uses the Gini Index as the impurity measure, as it forces the splits to be binary. The Gini index score is calculated for the data pre-split and post-split, with the lowest Gini score deciding the split point. However CART has several disadvantages, which are exacerbated by using it as a base learner for Random Forest. Trees are likely to select attributes with more variation in the predictor space [13]. This is especially prevalent in categorical data and variables with high levels of missing values. Consequently, this distorts variable importance measures. Some of CARTs disadvantages have been addressed by introducing CITs [14].

CITs use a generalised statistical test of independence to combat against overfitting and the aforementioned variable bias tendency. The algorithm operates in two steps, the first is attribute selection where an association test between the attribute and the outcome of interest is calculated [15]. The null hypothesis is that the attribute X_i has no association with the outcome variable Y. Due to the multiple comparisons, a Bonferroni corrected p-value threshold can be used. If the attribute and the outcome are both numeric, then the test statistic is a correlation test. If both attributes are categorical in nature, dummy variables are created and a χ^2 test of association is performed. If one of the attributes is numeric and the other categorical then an ANOVA is performed [14]. Once the attribute has been identified, the second step involves selecting a split point in the attribute, which can be determined using normal splitting procedures for Random Forest (see 16 for more details). Pruning is not used by default, as a stopping criteria can be set based on a cutoff (1 - p-value) pre-specified, which should produce an optimal predictive tree and can be tuned using cross validation [12]. CITs remove the bias that is inherent in CART, providing splits that are more reliable in interpretation of variable importance.

B. Variable Importance

In this study, permutation was used to determine a variable's importance. The predictor in question is shuffled so that the values in the dataset are basically random and any links with that predictor to the patterns in the rest of the dataset are broken [17]. The difference in accuracy is recorded per tree and aggregated across all the trees in the forest [12]. The importance is scaled per predictor based on the accuracy drop, and ranked. The method is referred to as Permutation Importance.

Permutation importance can be problematic when used with Random Forest if variables are correlated [18]. It has been shown that highly correlated data have a tendency to inflate the importance of non informative predictors as long as those predictors were correlated with predictive attributes [19]. Permutation should be done for groups of items that are highly correlated with each other. In the case of the silence attributes, the mean correlation was extremely high, which would mean that all silence attributes should be permuted together. This would produce a variable importance score where all the silence constructs in theory would be very important in comparison to other non correlated variables. However this method results in the loss of nuance on how the variables in isolation help in the prediction. Therefore a variation of this method was utilised in this study via the `party` package (version 1.3-1) in R. Variable importance ranking involved permutation of attributes within a group of attributes where the correlation among the variables was at a minimum of 0.2 [19]. The method is conceptually similar in spirit to partial correlation [18]. A conditioning grid is created based on the partition of feature space by individual trees within the Random Forest framework resulting in a discretised feature space. Then the variable is shuffled within this newly created grid and the Out Of Bag (OOB) error is recorded. The difference is then taken between the non-permutated and the permutated Random Forest OOB. Research suggested that using CITs as a basis for Random Forest to produce variable importance scores "*appear to strike a good balance between identification of significant variables and avoiding unnecessary flagging of correlated variables*" [20]. Interested readers are directed to [18] for an accessible version or [19] for a more in-depth treatment.

While the procedures described above tend to converge on their recommendations for what the most important variables are for predictions, they generally do not show if the variables positively or negatively impact the probability of the predictions.

C. Partial Dependancy Plots

Ensembles of models are more difficult to interpret due to the multiple models used in their construction. One method of model interpretation for classification models is the use of PDPs. On a conceptual level a PDP is used in conjunction with a model to plot model predictions when one or more of the independent variables is varied [21]. The average of the predicted value across all participants is taken and plotted

 ISBN: 978-1-61208-816-7

against the varying probability of engaging in Silence as the predictor changes. All other variables are held either at their mean or their median [22]. For example, take a hypothetical dataset with 20 predictors ($x_1..x_{20}$) and 2000 rows, with each row representing a participant in a survey and each predictor representing a factor that has been measured. A model M is generated to predict y_i. If the range of values that needs to be tested is 100 for x_i, then 100 datasets are created with the value of x_i being the only change for each copy. The PDP value is then calculated by using each of the datasets to generate a value y_i from the model M. The mean value y_i is then taken to give an average value for the model at that value x_i. This becomes computationally expensive when the number of copies becomes unmanageable so the median or mean for the values of x_i not being varied is a computational short cut. The result can then be plotted to show the changes in X producing a change in Y [9].

One of the advantages of the technique is the ability to see the relationship type (linear, non linear) between the independent variable and the object being predicted [23, Section 5.1]. A number of studies utilised PDPs to tease out the relationships in models [24]–[26]. The technique does have some problems when used in conjunction with correlated data where some combinations of correlated values are not reasonable [23, Section 5.1]. PDPs were used in this study to interrogate constructs where such a pattern was permissible and interpretable. It is recommended to use rug plots as part of any PDP plot to limit interpretation to within the range of the training set, thus avoiding models extrapolating beyond the data range [27].

II. Survey Instrument

The survey consisted of 136 questions and was designed by Organisational Silence researchers under the administration of the fourth author of this paper. All questions pertaining to this survey were taken from previously published research papers or added by the researchers based on their expertise in the area of organisational silence research. All scales were translated into their local languages and then translated back to English to confirm there was nothing lost in the translation. Inconsistencies were resolved by communication with the project administrator. Construct's names start with a capital letter.

The survey included questions on demographic information such as age, gender, country, industry worked and type of contract the participant was on. Cultural aspects of silence were measured using 13 constructs from the *Global Leadership and Organisational Behaviour Effectiveness* (GLOBE) questionnaire [28]. The constructs measured both societal and organisational practices related to: (1) Power Distance (collective response to power; acronym ends with "_pd"); (2) Uncertainty Avoidance Practices (effort undertaken by the collective to avoid uncertainty in their lives; acronym ends with "_ua"); (3) Future Orientation Practices (the process by which a group plans and is rewarded for future orientated behaviour; acronym ends with "_fo"); (4) Institutional Collectivism (propensity of people to act as a collective; acronym ends with "_c"); (5) Humane Orientation (propensity to promote and reward humane behaviour; acronym ends with "_ho"); (6) Performance Orientation (attitude to high standards and performance improvement; acronym ends with "_po") and finally (7) Gender Egalitarianism (collectives attempts to maximise or minimise the differences between men and women; acronym ends with "_g"). The constructs were generated from questions where the participants were queried about both their organisation and society. The constructs were aggregated to individual and societal level for the GLOBE societal constructs and individual and industry level for the GLOBE organisational constructs. Satisfaction With Life was measured by adjusting *"The Satisfaction with Life Scale"*. It originally consisted of 5 statements where the respondents answered on a 7 point Likert scale [29]. This study adapted the structure of the original questions to ask if respondents were satisfied (Satisfaction) with their health, their jobs, their life and their ability to do their jobs. The four questions were treated as separate constructs.

Additional constructs relating to the individual were included in the survey. These included Organisational Citizenship Behaviour (7 questions inspired by [30]); Mental Health (5 questions taken from [31]); and Health (16 questions taken from [32]). Individual perceptions of Climate For Authenticity (indv_cfa_calc) was covered using a 6 question construct inspired by [33]. It measured if participants could be true to themselves and express themselves in a manner that was consistent with their feelings irrespective of outside influences. Psychological Safety Climate (ind_psc_calc, 7 questions taken from [34]) measured if the climate within an organisation was amenable to employees taking personal risks. They were included to measure perceived office environments. Expectation of remaining in the same job (from [35]) was extended to expectation of remaining in the same organisation, and profession, until retirement.

Six silence constructs were added to the survey including: (1) Acquiescent (employees feel their opinion does not matter and it will not change anything; indv_sil_as); (2) Quiescent (fear of the consequences either from their management or from their co-workers because they do not agree with the group; indv_sil_qs); (3) Prosocial (to protect co-workers or the company; indv_sil_ps); (4) Opportunistic (to gain advantages for themselves; indv_sil_os); Diffident (silence due to lack of confidence; indv_sil_di) and Disengaged Silence (due to the individual being disengaged from their role within the organisation; indv_sil_de) [36], [37]. Relationship To Organisation was measured using three questions asking how much a participant identified with their colleagues, line manager and company [38]. Finally, the question *how often did you express concerns or opinions to someone who is able to change the situation* comprised of four possible answers but was binned into a binary identifying those participants who engaged in Silence. This was the label predicted by the two Random Forest models used in the study.

III. Data Analysis

A. Data Manipulation

All variables were standardised. Dummy binary variables were created for all character variables. Validity was determined using Confirmatory Factor Analysis. Reliability analysis was carried out using McDonald's Omega as described in [39]. Only valid and reliable constructs were interpreted. In total, there were 774 (Italy = 191, Germany = 450, Poland = 133) records with 91 columns. There was a class imbalance in the dataset for those that engaged in Silence (yes = 596, no = 178). Research suggests upsampling/downsampling methods could be used to balance the data leading to improved predictive accuracy. However, this idea was discarded as some research suggests it could result in loss of information or overfitting the model [40]. The *Receiver Operator Curve* (ROC) metric is resilient against the effects of unbalanced datasets [12].

The mean was taken for each participant for every question related to each construct. For example the construct Acquiescent Silence was generated by the equation $\bar{AS} = (sil_9 + sil_{10} + sil_{12})/n$. All constructs aggregated to individual level are prefixed with "indv_". The GLOBE country level scores were generated by taking the average score for all the individuals per country, producing one score per construct per country. They are pre-fixed with "grp_". Similarly the organisational constructs were aggregated to industry level by taking the average score for all individuals per industry and producing a single score per industry. These are prefixed in the data with "ind_"

B. Modelling

The Random Forest using CART as a base learner was tuned by varying *mtry* from 2 to 80 and minimum node size from 2 to 20 in increments of five. The optimal values used in the final models were 17 for *mtry*, and 2 for node size. The Conditional Inference Forest model was tuned across the same *mtry* range where the optimal *mtry* value was found to be 57. All other tuning parameters were left at their defaults.

All tuned models were run on a ten fold cross validated dataset and repeated ten times, to get an average *Area Under The Curve* (AUC) performance. The CART based Random Forest model had an AUC_μ = 0.65 and a standard deviation of AUC_{sd} = 0.08 ($Kappa_\mu$ = 0.1, $Kappa_{sd}$ = 0.09). Conditional Inference Forests produced a result of AUC_μ = 0.65 with a standard deviation of AUC_{sd} = 0.08 ($Kappa_\mu$ = 0.13, $Kappa_{sd}$ = 0.11). Figure 1 shows confusion matrices for both models where opacity indicates number of classifications, with blue indicating the majority of classifications. Both models had problems with identifying people who did not engage in Silence. The high precision for both models in predicting those that engage in Silence contrasts a low precision for those that did not engage in Silence.

Variable importance results can be seen in Table I, along with the number of distinct values per attribute. Uncertainty Avoidance Societal Practices (grp_gls_ua) was the most important predictor for Conditional Inference Forests but the

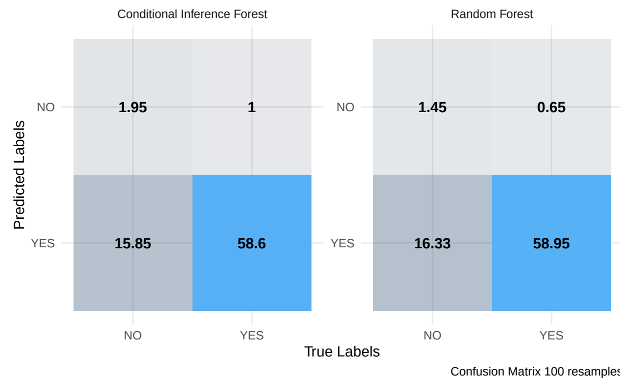

Fig. 1. Average values for Confusion Matrices for all models across 100 samples.

Random Forest suggested it was the 36th most important predictor. Models were in relative agreement when ranking the importance of the next four predictors (within 3 places), each had a relatively high number of distinct values. In contrast, ind_calc_education, qual_level_second_level and qual_calc_degree all have large disparities between rankings illustrating CART Random Forest's bias towards attributes with a larger number of distinct values.

TABLE I
Top 10 Variable Importance Ranking of Conditional Inference Forest With corresponding CART Random Forest Ranking

Feature Code	Distinct Values	CIF Rank	CRF Rank
grp_gls_ua	3	1	36
indv_sil_as	19	2	2
indv_sil_qs	19	3	3
indv_gls_pd	29	4	1
indv_cfa_calc	35	5	6
ind_calc_education	2	6	65
indv_gls_ho	24	7	5
indv_sil_de	19	8	14
indv_glo_ua	19	9	7
indv_sil_os	18	10	22
indv_sil_di	19	11	15
present	5	12	40
qual_level_second_level	2	13	59
indv_sil_ps	19	14	8
qual_calc_degree	2	15	56

[a] CIF = Conditional Inference Forest; CRF = CART Random Forest

PDPs were generated in Figure 2 for the top 5 predictors for Conditional Inference Forests. Jitter was added to the rug plot to show where the models may be extrapolating beyond their bounds. Figure 2 shows that the silence predictors Acquiescent Silence (indv_sil_as) and Quiescent Silence (indv_sil_qs) had an expected positive relationship to the probability of someone engaging in Silence. Panel D highlights that as Individual Power Distance Societal Practices increases the probability of engaging in Silence increases. In several studies, high power distance societies tended not to openly express their anger or dissatisfaction with their superiors compared with low power

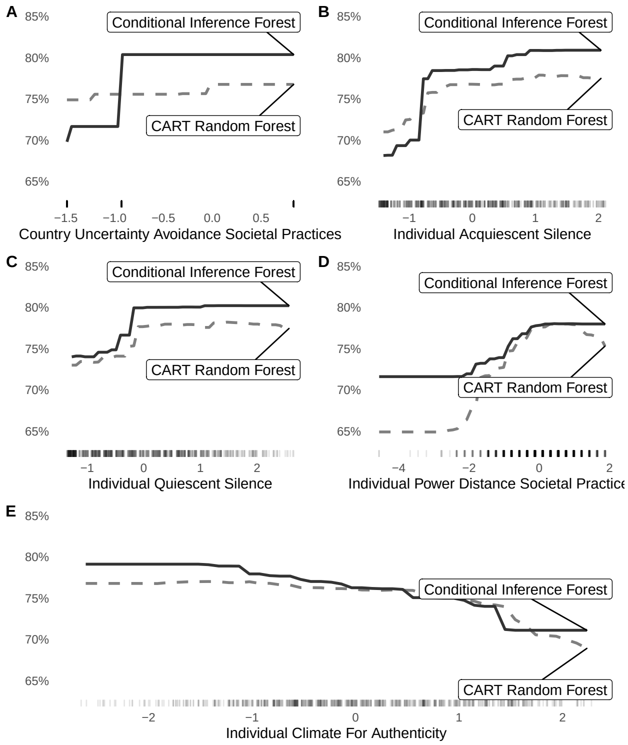

Fig. 2. Partial Dependancy Plots for top Predictors for Conditional Inference Forests.

distance nations [41]. Panel E indicates that as a Climate for Authenticity increases, the probability of participants engaging in Silence decreases. In previous research, it was found that Climate for Authenticity was found to have a positive relationship with voice, while a negative relationship to Quiescent Silence, Pro-Social Silence, Opportunistic Silence, and Quiescent Silence [42]. Panels B, C, D and E modelled expected patterns based on existing research, reinforcing the external validity of the model [36].

It is apparent from both Conditional Inference Forest and Random Forest, that as Uncertainty Avoidance Societal Practices (grp_gls_ua) increased, the probability of engaging in Silence also increased. This was a previously unknown pattern, although several components of Uncertainty Avoidance that may promote Silence behaviours had been documented previously. For example, societies having a highly formalised management structure, an inclination towards hierarchical structures and exhibit a strong resistance to change [43] [44]. This suggests that people in high Uncertainty Avoidance societies may engage in Silence behaviours because they feel any feedback they give would result in no changes in the status quo, in essence Acquiescent Silence.

IV. Conclusions and future work

This paper shows how Random Forests in conjunction with PDPs can be used with variable importance measures to highlight non linear relationships between predictors and target variables. However, a CART based random forest showed a bias for predictors with more values. A CIT based forest did not have the same bias. For example, Uncertainty Avoidance Societal Practices was the number one predictor for the Conditional Inference Forest, but was not even in the top 20 predictors for the CART based Random Forest. It is also worth noting that while the pattern plotted for the construct in Panel A of Figure 2 highlights that both Random Forests are in agreement with the relationship between the predictor and the outcome, the pattern is far more pronounced in the Conditional Inference Forest.

Based on these findings, it is suggested where the predictor space has varying number of distinct values per predictor, and model interpretation is the goal of the analysis, that Conditional Inference Forest is better than Random Forest for exploring variable importance. This finding is particularly pertinent for researchers who wish to use tree based modelling for survey data where the questions pertaining to the constructs have a different number of available options.

A weakness in the study was average predictive accuracy of the models. However, moderate AUC scores are common when analysing psychometric survey data. For example, a study that applied a Generalized Additive Model to predict the frequency participants would take cocaine reported an AUC of 0.567 [45]. Another example used an online questionnaire to record several constructs to identify features that would highlight individuals social support needs for "*Online Health Social Networks*" . The resulting mean AUC performance was 0.8. Finally a third study ran several machine learning algorithms to try to predict major depressive disorders from self reported surveys. The study attempted to predict five such disorders with an average AUC of 0.71 (0.71, 0.63, 0.73, 0.74, 0.76) [46].

Modelling in this study highlighted how culture plays a role in whether someone will engage in Silence or not. The models appeared to suggest that people in environments with high Authenticity are less likely to engage in Silence behaviours. It also highlights that the higher the power distance within a society, the more likely someone will engage in Silence. Both of these findings have already been previously explored in the Silence literature. However, the patterns related to uncertainty avoidance were not previously known, and indicate that the higher the value in this construct, the more inclined someone is to engage in Silence behaviours. The results of this analysis suggests that, within this data, culture plays a role in silence engagement both at local and country level. This is demonstrated in Figure 2 for both Societal level uncertainty avoidance and the local Climate for Authenticity. Panels A-D show that the four main constructs level off, suggesting interactions may play a role in mediating the role of each construct in influencing Silence. Data from more countries is needed to confirm the pattern is not an artifice of having too little data.

Partial dependency plots suffer when applied to highly correlated data because they condition on the marginal distribution (See [23, Section 2.1.1]). Other methods exist that are not so readily impacted by correlation such as *Accumulated*

Local Effects (ALE) plots which condition on the conditional distribution while averaging over the differences in the predictions, as opposed to the average of the predictions [47]. ALE plots could be investigated to see if the same patterns highlighted in Figure 2 remain consistent with a larger dataset.

ACKNOWLEDGEMENT

The Italian data sample was collected by Paola Gatti and Chiara Ghislieri (Universit degli Studi di Torino, Italia), the Polish sample by Sylwiusz Retowski (SWPS University of Social Sciences and Humanities) and the German sample by Rolf van Dick (Goethe Universitt Frankfurt, Germany).

All analysis and exploration of data was carried out in `R` [48]. This paper was produced using `bookdown` (version 0.16) and `rticles` (version 0.14) packages [49], [50]. Scale validity and reliability was undertaken using the `lavaan` package (version 0.6-5) [51]. Data manipulation and plotting of PDPs were undertaken using `tidyverse` (version 1.2.1), `pdp` (version 0.7.0), `ggrepel` (version 0.8.0) and `cowplot` (version 0.9.3) packages [27], [52]–[54]. Modelling was applied using `tidymodels` (version 0.02) and `caret` (version 6.0-80) packages [55], [56]. The CART Random Forest model as well as permutated variable importance was generated via the `randomForest` package (version 4.6-14) [57]. The Conditional Inference Forest Model and variable importance of the same model was applied using the `party` package (version 1.3-1) [58].

REFERENCES

[1] E. W. Morrison, "Employee Voice and Silence," *Annual Review of Organizational Psychology and Organizational Behavior*, vol. 1, no. 1, pp. 173–197, 2014, doi: https://doi.org/10.1146/annurev-orgpsych-031413-091328.

[2] M. L. DeVault, *Liberating method: Feminism and social research*. Temple University Press, 1999.

[3] F. Tahmasebi, S. M. Sobhanipour, and M. Aghaziarati, "Burnout ; Explaining the Role of Organizational Silence and Its Influence (Case study : Selected Executive Organizations of Qom Province)," vol. 3, no. 8, pp. 272–282, 2013.

[4] A. Edmondson, "Psychological safety and learning behavior in work teams," *Administrative science quarterly*, vol. 44, no. 2, pp. 350–383, 1999, doi: https://doi.org/10/b9rmv3.

[5] R. J. House, P. J. Hanges, S. A. Ruiz-Quintanilla, P. W. Dorfman, M. Javidan, and M. V. Dickson, "Cultural influences on leadership: Project GLOBE," *Advances in Global Leadership*, vol. 1, pp. 171–233, 1999, doi: https://doi.org/8042818.

[6] M. M. Javidan, P. W. Dorfman, M. S. de Luque, R. J. House, M. S. De Luque, and R. J. House, "In the Eye of the Beholder: Cross Cultural Lessons in Leadership from Project GLOBE." *Academy of Management Perspectives*, vol. 20, no. 1, pp. 67–90, 2006, doi: https://doi.org/10/b9ds7b.

[7] J. Wynen, B. Kleizen, K. Verhoest, P. Lgreid, and V. Rolland, "Just keep silent… Defensive silence as a reaction to successive structural reforms," *Public Management Review*, pp. 1–29, 2019.

[8] R. Bogosian, "The intersection of national cultural values and organizational cultures of silence and voice, and the moderating effect of leadership," *AIB Insights*, vol. 18, no. 2, pp. 16–20, 2018.

[9] T. Hastie, R. Tibshirani, and J. Friedman, "The Elements of Statistical Learning," *Bayesian Forecasting and Dynamic Models*, vol. 1, pp. 1–694, 2009, doi: https://doi.org/10/cw7rjn.

[10] T. M. Khoshgoftaar, D. J. Dittman, R. Wald, and W. Awada, "A review of ensemble classification for dna microarrays data," in *2013 IEEE 25th International Conference on Tools with Artificial Intelligence*, 2013, pp. 381–389, doi: https://doi.org/10/gf5hbg.

[11] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*. CRC press, 1984.

[12] M. Kuhn and K. Johnson, *Applied Predictive Modeling*. Springer, 2013.

[13] A. P. White and W. Z. Liu, "Technical Note: Bias in Information-Based Measures in Decision Tree Induction," *Machine Learning*, vol. 15, no. 3, pp. 321–329, 1994, doi: https://doi.org/10/dbhwp2.

[14] T. Hothorn, K. Hornik, and A. Zeileis, "Unbiased recursive partitioning: A conditional inference framework," *Journal of Computational and Graphical Statistics*, vol. 15, no. 3, pp. 651–674, 2006, doi: https://doi.org/10/d47hbq.

[15] A. Venkatasubramaniam, J. Wolfson, N. Mitchell, T. Barnes, M. Jaka, and S. French, "Decision trees in epidemiological research," *Emerging Themes in Epidemiology*, vol. 14, no. 1, pp. 1–12, 2017, doi: https://doi.org/10/gf5g96.

[16] T. Hothorn, K. Hornik, and A. Zeileis, "Ctree: Conditional inference trees," *The Comprehensive R Archive Network*, vol. 8, 2015.

[17] S. Nembrini, I. R. Knig, and M. N. Wright, "The revival of the Gini importance?" *Bioinformatics*, vol. 34, no. 21, pp. 3711–3718, 2018, doi: https://doi.org/10/gdkz3q.

[18] C. Strobl, T. Hothorn, and A. Zeileis, "Party on! A new, conditional variable importance measure available in the party package," 2009, Accessed: Aug. 20, 2020. [Online]. Available: https://cran.r-project.org/web/packages/party/index.html.

[19] C. Strobl, A.-L. Boulesteix, T. Kneib, T. Augustin, and A. Zeileis, "Conditional Variable Importance for Random Forests," *BMC Bioinformatics*, vol. 9, no. 307, 2008, Accessed: Aug. 20, 2020. [Online]. Available: http://www.biomedcentral.com/1471-2105/9/307.

[20] L. Auret and C. Aldrich, "Empirical comparison of tree ensemble variable importance measures," *Chemometrics and Intelligent Laboratory Systems*, 2011, doi: https://doi.org/10/chrsv7.

[21] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Annals of statistics*, pp. 1189–1232, 2001, doi: https://doi.org/10/fbgj35.

[22] A. Natekin and A. Knoll, "Gradient boosting machines, a tutorial," *Frontiers in Neurorobotics*, vol. 7, no. DEC, 2013, doi: https://doi.org/10/gfts5q.

[23] C. Molnar, *Interpretable machine learning*. lulu.com, 2020.

[24] D. P. Green and H. L. Kern, "Modeling heterogeneous treatment effects in large-scale experiments using Bayesian Additive Regression Trees," in *The annual summer meeting of the society of political methodology*, 2010.

[25] R. A. Berk and J. Bleich, "Statistical procedures for forecasting criminal behavior: A comparative assessment," *Criminology & Pub. Pol'y*, vol. 12, p. 513, 2013.

[26] J. Elith, J. R. Leathwick, and T. Hastie, "A working guide to boosted regression trees," *Journal of Animal Ecology*, vol. 77, no. 4, pp. 802–813, 2008, doi: https://doi.org/10/fn6m6v.

[27] B. M. Greenwell, "Pdp: An R Package for constructing partial dependence plots," *R J*, vol. 9, no. 1, p. 421, 2017, Accessed: Jul. 01, 2020. [Online]. Available: https://cran.r-project.org/web/packages/pdp/index.html.

[28] R. J. House, P. W. Dorfman, M. Javidan, and P. J. Hanges, *Strategic Leadership across Cultures: The GLOBE Study of CEO Leadership Behavior and Effectiveness in 24 Countries*. 55 City Road, London, 2019.

[29] E. Diener, R. A. Emmons, R. J. Larsen, and S. Griffin, "The Satisfaction with Life Scale," *Journal of Personality Assessment*, vol. 49, no. 1, pp. 71–75, 1985, doi: https://doi.org/10/fqqbmr.

[30] R. Van Dick, M. W. Grojean, O. Christ, and J. Wieseke, "Identity and the extra mile: Relationships between organizational identification and organizational citizenship behaviour," *British Journal of Management*, vol. 17, no. 4, pp. 283–301, 2006, doi: https://doi.org/10/fjv4dk.

[31] J. E. Ware, "SF-36 health survey update." *Spine*, vol. 25, no. 24, pp. 3130–3139, 2000, doi: https://doi.org/10/fv3fbd.

[32] L. R. Derogatis, R. S. Lipman, K. Rickels, E. H. Uhlenhuth, and L. Covi, "The Hopkins Symptom Checklist (HSCL): A self?report symptom inventory," *Behavioral Science*, vol. 19, no. 1, pp. 1–15, 1974, doi: https://doi.org/10/fn39gn.

[33] A. Grandey, S. C. Foo, M. Groth, and R. E. Goodwin, "Free to be you and me: A climate of authenticity alleviates burnout from emotional labor." *Journal of Occupational Health Psychology*, vol. 17, no. 1, pp. 1–14, 2012, doi: https://doi.org/10/fpps98.

[34] M. Baer and M. Frese, "Innovation is not enough: Climates for initiative and psychological safety, process innovations, and firm performance,"

Journal of Organizational Behavior, vol. 24, no. 1, pp. 45–68, 2003, doi: https://doi.org/10/fvkggk.

[35] S. C. Liebermann, J. Wegge, and A. Mller, "Drivers of the Expectation of Remaining in the Same Job until Retirement Age: A Working Life Span Demands-Resources Model," *European Journal of Work and Organizational Psychology*, vol. 22, no. 3, pp. 347–361, 2013, doi: https://doi.org/10.1080/1359432X.2012.753878.

[36] M. Knoll and R. van Dick, "Do I Hear the Whistle...? A First Attempt to Measure Four Forms of Employee Silence and Their Correlates," *Journal of Business Ethics*, vol. 113, no. 2, pp. 349–362, Apr. 2013, doi: https://doi.org/10/r67.

[37] C. T. Brinsfield, "Employee silence motives: Investigation of dimensionality and development of measures," *Journal of Organizational Behavior*, vol. 34, no. 5, pp. 671–697, 2013, doi: https://doi.org/10/r66.

[38] T. Postmes, S. A. Haslam, and L. Jans, "A single-item measure of social identification: Reliability, validity, and utility," *British Journal of Social Psychology*, vol. 52, no. 4, pp. 597–617, 2013, doi: https://doi.org/10/gddc6w.

[39] K. A. Bollen, "Issues in the comparative measurement of political democracy," *American Sociological Review*, pp. 370–390, 1980.

[40] G. M. Weiss, K. McCarthy, and B. Zabar, "Cost-sensitive learning vs. Sampling: Which is best for handling unbalanced classes with unequal error costs?" *Dmin*, vol. 7, nos. 35-41, p. 24, 2007.

[41] X. Huang, E. V. de Vliert, G. V. der Vegt, E. V. D. Vliert, and G. V. D. Vegt, "Breaking the silence culture: Stimulation of participation and employee opinion withholding cross-nationally," *Management and Organization Review*, vol. 1, no. 3, pp. 459–482, 2005, doi: https://doi.org/10/fn4hfr.

[42] M. Knoll and R. van Dick, "Authenticity, employee silence, prohibitive voice, and the moderating effect of organizational identification," *The Journal of Positive Psychology*, vol. 8, no. 4, pp. 346–360, 2013, doi: https://doi.org/10/gfbzbr.

[43] G. Oliver, *Organisational Culture for Information Managers*. Elsevier, 2011.

[44] C. N. Grove, "Introduction to the GLOBE research project on leadership worldwide," *Professional Knowledge Center, GROVEWELL LLC. Retrieved July*, 2005. https://www.grovewell.com/wp-content/uploads/pub-GLOBE-intro.pdf (accessed Nov. 29, 2017).

[45] R. Suchting, J. N. Vincent, S. D. Lane, C. E. Green, J. M. Schmitz, and M. C. Wardle, "Using a Data Science Approach to Predict Cocaine Use Frequency from Depressive Symptoms," *Drug and alcohol dependence*, vol. 194, pp. 310–317, 2019.

[46] R. C. Kessler *et al.*, "Testing a Machine-Learning Algorithm to Predict the Persistence and Severity of Major Depressive Disorder from Baseline Self-Reports," *Molecular psychiatry*, vol. 21, no. 10, p. 1366, 2016.

[47] D. W. Apley and J. Zhu, "Visualizing the effects of predictor variables in black box supervised learning models," *arXiv preprint arXiv:1612.08468*, 2016.

[48] R Core Team, *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing, 2018.

[49] Y. Xie, *Bookdown: Authoring Books and Technical Documents with R Markdown*. Chapman and Hall/CRC, 2016.

[50] J. Allaire *et al.*, "Rticles: Article formats for r markdown," 2020, Accessed: Aug. 15, 2020. [Online]. Available: https://CRAN.R-project.org/package=rticles.

[51] Y. Rosseel, "Lavaan: An R Package for Structural Equation Modeling," *Journal of Statistical Software*, vol. 48, no. 2, pp. 1–36, 2012, Accessed: Apr. 13, 2020. [Online]. Available: https://cran.r-project.org/web/packages/lavaan/index.html.

[52] C. O. Wilke, "Cowplot: Streamlined plot theme and plot annotations for 'ggplot2'," 2018, Accessed: Mar. 16, 2020. [Online]. Available: https://CRAN.R-project.org/package=cowplot.

[53] K. Slowikowski, "Ggrepel: Automatically Position Non-Overlapping Text Labels with 'ggplot2'," 2018, Accessed: May 19, 2020. [Online]. Available: https://cran.r-project.org/web/packages/ggrepel/index.html.

[54] H. Wickham, "Tidyverse: Easily install and load the 'tidyverse'," 2017, Accessed: Jul. 01, 2020. [Online]. Available: https://CRAN.R-project.org/package=tidyverse.html.

[55] K. Max and H. Wickham, "Tidymodels: Easily install and load the 'tidymodels' packages," 2018, Accessed: Apr. 20, 2020. [Online]. Available: https://CRAN.R-project.org/package=tidymodels.html.

[56] Max Kuhn *et al.*, "Caret: Classification and regression training," 2018, Accessed: Apr. 20, 2020. [Online]. Available: https://CRAN.R-project.org/package=caret.

[57] A. Liaw and M. Wiener, "Classification and regression by randomForest," *R News*, vol. 2, no. 3, pp. 18–22, 2002, Accessed: Jun. 11, 2020. [Online]. Available: https://cran.r-project.org/web/packages/randomForest/index.html.

[58] C. Strobl, A. L. Boulesteix, A. Zeileis, and T. Hothorn, "Bias in random forest variable importance measures: Illustrations, sources and a solution," *BMC Bioinformatics*, 2007, Accessed: Aug. 20, 2020. [Online]. Available: https://cran.r-project.org/web/packages/party/index.html.