ABSTRACT

ANDREWS, JOSHUA SCOTT. Random Forest for Scale and Item Level Prediction Analysis in the Social Sciences: An Application Using Organizational Deviance Data. (Under the direction of Dr. Samuel Pond III).

This study examines the utility of machine learning (ML) algorithms, specifically random forest (RF), to outperform traditional predictive modeling approaches in social science. Specifically, this study compares RF to Ordinary Least Squares Regression (OLS) and Least Absolute Shrinkage and Selection Operator (LASSO) regression for their ability to interpret item level validity. Using an archival data set that includes Maladaptive Traits and Beliefs (MTBs) and Counterproductive Work Behaviors (CWBs), I illustrate the process of conducting OLS, LASSO, and RF using a traditional approach (aggregate scales) and an item level approach (scale items). I also discuss the differences in variable importance ratings across models, which I used to rank order the importance of items. Findings from this study support the utility of RF, which had the best overall performance compared to OLS and LASSO using both scale and item level predictors. Furthermore, using the RF variable importance metrics, I was able to reduce the MTB item RF model to only two items and consequently achieve the highest fit indices of all study models. This study also presents a generalizable (i.e. soft-coded) version of the R code that I used to analyze the data, which researchers can easily apply to their own data sets.
Random Forest for Scale and Item Level Prediction Analysis in the Social Sciences: An Application Using Organizational Deviance Data

by
Joshua S. Andrews

A Dissertation submitted to the Graduate Faculty of North Carolina State University in partial fulfillment of the requirements for the degree of Doctorate of Philosophy

Psychology

Raleigh, North Carolina

2019

APPROVED BY:

______________________________  ________________________________
Dr. Samuel B. Pond, III              Dr. S. Bartholomew Craig
Chair of Advisory Committee

______________________________  ________________________________
Dr. Sarah L. Desmarais              Dr. Seong Hee Cho
DEDICATION

To Charlotte, we keep moving forward.

To Dr. Samuel Pond III, thanks for turning me in the right direction.

“That boy’s got real spark, lots of spirit. Throws himself, heart and soul, into everything he does. And that’s really worth something. If it could only be turned in the right direction.”

-Merlin (Disney’s Sword in the Stone)
BIOGRAPHY

Originally from Raleigh, North Carolina, Joshua Andrews earned his 2nd Dan Black Belt in Tae Kwon Do and Hapkido from the World Tae Kwon Do Federation in 2003 and received his Eagle Scout award from Troop 214 in 2004. He attended North Carolina State University where he received his Bachelors of Arts in Psychology in 2013 and his Masters of Science in Industrial and Organizational Psychology in 2018. Josh has an interest in cross-disciplinary research and enjoys research on machine learning and computer science topics. He enjoys playing guitar, with his favorite genres being Blues, Rock and Roll, and Heavy Metal. He also loves musicals and Disney movies, both of which he watches often. He lives with his spouse Charlotte and their three cats.
# TABLE OF CONTENTS

LIST OF TABLES ........................................................................................................... v  
Deviant Behaviors at Work .......................................................................................... 2  
Correlates of Deviant Behavior .................................................................................. 3  
  The Dark Triad .......................................................................................................... 4  
  Impulsivity ............................................................................................................... 4  
  Entitlement ............................................................................................................... 5  
  Moral Disengagement .............................................................................................. 5  
Approaches to Item-level Criterion Analysis ............................................................. 6  
  Linear Regression .................................................................................................... 6  
  Machine Learning Algorithms .................................................................................. 7  
  LASSO ...................................................................................................................... 7  
  Decision Trees ......................................................................................................... 7  
  Random Forests ....................................................................................................... 8  
Methods ....................................................................................................................... 13  
  Sample ..................................................................................................................... 13  
  Materials and Procedures ....................................................................................... 13  
  Measures .................................................................................................................. 14  
  Presentation of Measures ....................................................................................... 15  
Results ......................................................................................................................... 16  
  Validation and Hyperparameter Optimization ....................................................... 16  
  Maladaptive Traits/Beliefs Scale Analyses ............................................................ 19  
  Maladaptive Traits/Beliefs Item Level Analysis .................................................... 22  
  Differences in Variable Importance Rankings ....................................................... 25  
Discussion ................................................................................................................... 25  
  Limitations .............................................................................................................. 30  
  Concluding Remarks .............................................................................................. 32  
References .................................................................................................................. 33  
APPENDICES ............................................................................................................. 45  
  Appendix A ............................................................................................................. 46  
  Appendix B ............................................................................................................. 47  
  Appendix C ............................................................................................................. 48  
  Appendix D ............................................................................................................. 49  
  Appendix E ............................................................................................................. 50  
  Appendix F ............................................................................................................. 51  
  Appendix G ............................................................................................................. 52
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Fit Statistics for OLS, LASSO, and RF Models Using Maladaptive Traits/Beliefs Scales and Items to Predict CWBs, (n=99)</th>
<th>38</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 2</td>
<td>Beta weights from Linear and LASSO Regression Models Using Anti-Social Beliefs/Traits Scales to Predict CWBs, (n=402)</td>
<td>39</td>
</tr>
<tr>
<td>Table 3</td>
<td>Variable Importance Metrics from Random Forest Model Using Anti-Social Beliefs/Traits Scales to Predict CWBs, (n=402)</td>
<td>40</td>
</tr>
<tr>
<td>Table 4</td>
<td>Beta weights from Linear Regression Model Using Anti-Social Beliefs/Traits Items to Predict CWBs, (n=402)</td>
<td>41</td>
</tr>
<tr>
<td>Table 5</td>
<td>Beta weights from LASSO Regression Model Using Anti-Social Beliefs/Traits Items to Predict CWBs, (n=402)</td>
<td>42</td>
</tr>
<tr>
<td>Table 6</td>
<td>Variable Importance Metrics from Random Forest Model Using Anti-Social Beliefs/Traits Items to Predict CWBs, (n=402)</td>
<td>43</td>
</tr>
<tr>
<td>Table 7</td>
<td>OLS, LASSO, and RF Item Importance Ranks for Maladaptive Traits and Beliefs Items Across Variable Importance Metrics</td>
<td>44</td>
</tr>
</tbody>
</table>
Random Forest for Scale and Item Level Prediction Analysis in the Social Sciences: An Application Using Organizational Deviance Data

Interdisciplinary interest in Big Data continues to grow, resulting in spurred advancement of analytic techniques that can model large applied data sets. This advancement in analytic techniques now has many names: “machine learning, data mining, statistical learning, applied mathematics, and data science” (Hindman, 2015, p 48-49). Regardless of the label, the approaches all seek to reduce uncertainty and increase prediction accuracy over previously established techniques. Specifically, machine learning (ML) is a subfield of Artificial Intelligence (AI) that focuses on creating programs that can access data and use it to learn automatically without human intervention. Most ML models applicable to social scientists are supervised learning models, meaning the model attempts to predict a set outcome variable. This process typically incorporates splitting the data into training and test sets. Researchers can apply ML techniques to social science, where these techniques can outperform the utility of standard regression models for purposes of prediction and variable importance analysis. When deciding which statistical model offers the best explanation, out-of-sample prediction and parsimony are the two most important metrics (Hindman, 2015). The majority of social science research has relied heavily on linear regression models to examine these metrics (Taagepera, 2008; Hindman, 2015), but ML models now offer several valuable alternatives.

Random Forest (RF) regression is one ML algorithm that can increase predictive power and accuracy over standard linear and parametric approaches. The current study proposes to characterize RF in the context of Counterproductive Work Behaviors (CWBs), which are a typical outcome of importance in Industrial and Organizational (I/O) Psychology literature. Within I/O literature, CWBs represent a collection of behaviors in opposition to an
organization’s best interests (Spector, Fox, Penney, Bruursema, Goh, & Kessler, 2006). However, predictive models for deviant behavior are a topic of interdisciplinary scientific concern. Findings from this study are likely of particular interest to disciplines other than I/O, such as Forensic Psychology and Criminology. Consideration over the validity and utility of both measures and prediction models is a topic of growing attention in these fields, culminating in a recent court case (Loomis v. Wisconsin, 2017) over the ethical nature of such algorithms. Findings from the current study provide insight to researchers regarding when, how, and why ML algorithms can outperform standard linear regression techniques while also providing equal or better information about variable importance in predicting social science outcomes. While general fondness for ML algorithms, such as RF, has grown in recent years, few studies seem to address the pragmatic application of ML techniques, which can overcome many of the limitations in current approaches and potentially provide a better fit to applied data sets. The purpose of the present study is to illustrate and compare several methods for analyzing the utility of scale items with a focus on examining advances in ML algorithms. Specifically, this study assesses several previously established measures and compares the utility of standard linear regression techniques with that of ML models for assessing scale level and item level predictors, in this instance predicting CWBs.

**Deviant Behaviors at Work**

CWBs encompass a range of deviant or anti-social behaviors and represent a major topic of research attention within I/O literature. Such behaviors include abuse towards others, production deviance, sabotage of equipment, theft, and work withdrawal (Spector et al., 2006). Many published articles have examined the relationship between CWBs and variables such as integrity (Van Iddekinge, Roth, Raymark, & Olde-Dusseau, 2012) and personality (Ones,
Viswesvaran, & Schmidt, 2003). When authors refer to the criterion-related validity between CWBs and various predictors, they are typically referring to a standard linear regression’s $R^2$ coefficient. However, some researchers have addressed that CWBs characteristically have highly skewed, non-normal distributions (Andrews, 2018; Van Zyl & De Bruin, 2018). CWBs, like other deviant outcomes, do not have well behaved distributions. CWB research is one area that could immediately benefit from an exploration of ML alternatives to typical OLS approaches. It is therefore pragmatic for the present study to investigate the utility of RF for modeling the relationship between predictors and CWB data.

**Correlates of Deviant Behavior**

Regarding variables associated with deviant behaviors, researchers often refer to integrity as a major predictor of CWBs (Schmidt & Hunter, 1998; Van Iddekinge, Roth, Raymark, & Odle-Dusseau, 2012). The majority of integrity research seems to focus on Maladaptive Traits and Beliefs (MTBs). Researchers have established that MTBs, such as the Dark Triad (narcissism, Machiavellianism, psychopathy), are important predictors of various deviant behaviors (Sherer, Baysinger, Zolynsky, & LeBreton, 2013). Such variables include traits and beliefs about the acceptability of behavior, which researchers claim are significant predictors of deviant behavior (Knight, Garner, Simpson, Morey, & Flynn, 2006). The current study measures a multitude of variables that fall within the content domain of MTBs, including narcissism, Machiavellianism, psychopathy, impulsivity, entitlement, and moral disengagement. These variables are not necessarily unique indicators; however, in isolation from one another, measures for these constructs have established criterion-related validity with CWBs.
The Dark Triad

The dark triad is a term for the combination of narcissism, Machiavellianism, and psychopathic traits as three highly related indicators of delinquent behavior. Previous research suggests that the three traits are similar enough to form a higher-order construct (Jonason & Webster, 2010). Social scientists have established links between the dark triad traits and delinquency in both prison inmate samples (Mandell, 2006) and worker populations (Ragatz, Fremouw, & Baker, 2012). Meta-analysis results also indicate that all three components of the dark triad are highly associated with CWBs (O’Boyle, Forsyth, Banks, & McDaniel, 2012). The dark triad is likely the most well established interdisciplinary deviance predictor.

Impulsivity

Impulsive individuals are prone to rapid, unplanned reactions to stimuli with diminished concerns for the consequences of their actions (Morean et al., 2014). The general theory of crime suggests that impulsivity is involved in all forms of crime because humans naturally desire pleasure and self-gratification (Benda, 2003). Previous research has established impulsivity as a predictor of a wide range of undesirable behaviors including heavy substance abuse (Leeman & Potenza, 2012), drunk driving (Moan, Nordstrom, & Storvoll, 2013), and intimate partner violence (Shorey, Brasfield, Eeabres, & Stuart, 2011). Impulsivity is also a known antecedent of self-reported CWBs (Cohen, Panter, Turan, Morse, & Kim, 2014). Impulsive individuals seem to lack the ability to impede behaviors and are consequently less likely to engage in ethical reasoning before acting. Given its prominence in deviant behavior literature, impulsivity is an important variable under examination in the current study.
Entitlement

The majority of literature on entitlement stems from researchers’ focus on narcissism (Campbell, Bonacci, Shelton, Exline, & Bushman, 2004). Although researchers note that entitlement is likely a component of narcissism (Campbell et al., 2004), there may be added utility in the inclusion of entitlement as a unique construct. Given that current measures of narcissism are highly related to other dark triad traits, entitlement scales may capture content outside the range of current narcissism measures. The current study will examine empirically the utility of entitlement measures to predict CWBs in addition to dark triad measures.

Moral Disengagement

Moral disengagement (Bandura, 1990; Bandura, 2002) represents a cognitive process involved in ethical decision-making. This process involves cognitively restructuring unethical decisions so that they appear less harmful. By framing decisions as unrelated to typical moral guidelines, moral disengagement serves to permit an individual to engage in unethical behavior without feeling distress or cognitive dissonance. Once an individual accepts a morally disengaged thought, that thought becomes an attitude about the acceptability of certain behaviors. A propensity to morally disengage represents the existence of attitudes reflective of morally disengaged beliefs (Moore, Detert, Trevino, Baker, & Mayer, 2012). Researchers have established a link between moral disengagement and various unethical behaviors including self-reported lying/cheating/stealing, responses to an ethically charged hypothetical situation, and self-interest-based decision-making (Moore et al., 2012). Researchers also found that moral disengagement was significantly associated with the acceptance of unethical negotiation tactics (Cohen et al., 2014) and self-reported CWBs (Andrews, Thompson, & Williams, 2016). Moral disengagement further samples the content domain of MTB characteristics; its inclusion in the
study will provide unique information about item importance in the presence of other dark personality measures.

**Approaches to Item-level Criterion Analysis**

**Linear Regression**

Ordinary Least Squares (OLS) regression is a foundational technique for all psychologists. In some cases, linear regression is a sufficient, or even the best, technique for investigating research hypotheses. Nevertheless, there are also instances in social science where data sets encompass variables that do not meet the assumptions of OLS. In fact, some researchers have suggested widespread overdependence on linear regression in social sciences, stating, “these [linear regression] methods have been overused and misused in social sciences” (Taagepera, 2008, p. 91). In the current study, CWBs have a non-normal distribution and the relationship between predictors and CWBs is likely non-linear to some degree. A non-linear model would better capture the strength and nature of these relationships and at worst would perform equally well to OLS models. In addition, the high correlations between predictors are another major issue in the current study. When predictor variables have high correlations with each other, beta weight estimates can become unreliable. Specifically, when researchers add more correlated variables to a model the standard error estimates become increasingly large, which in turn increases confidence intervals for estimated beta weights. Scale items designed to measure the same underlying construct should be highly correlated. Furthermore, OLS is non-parsimonious; any variable in the model will produce a non-zero coefficient (Hindman, 2015). ML models can likely increase predictive power over OLS and can provide variable importance metrics that are more accurate than traditional beta weights. Researchers can use these reliable
ML importance metrics to reduce the number of predictor items and create a more parsimonious model while still maintaining predictive accuracy.

**Machine Learning Algorithms**

**LASSO.** Least Absolute Shrinkage and Selection Operator (LASSO) regression (Tibshirani, 1996) is one technique that can assess variable importance and therefore reduce items for OLS regression. LASSO can address the high correlations between items and offer more reliable beta weights for determining item importance. LASSO, a type of penalized regression, is relatively equivalent to running OLS with some additional constraints on the coefficients (Hindman, 2015). LASSO constrains (penalizes) the sum of the absolute value of all regression coefficients. Lambda is the denotation for the penalty term in LASSO models. Researchers use cross-validation to choose the optimal lambda (constraint) value. When lambda is large, more variables will have zero coefficient values. When the LASSO procedure reduces lambda to zero, the model yields the same results as OLS. While LASSO does provide an advanced approach for adapting OLS regression, it is still limited by the OLS assumption of a linear relationship between x and y. In the present study, it is simply an automated means of selecting the best performing variables within a linear regression model, a process researchers call regularization (Hindman, 2015).

**Decision Trees.** Decision trees are ML methods that work by repeatedly splitting the data in a binary fashion, creating branches of decisions that resemble a tree. These algorithms select a value for the best predicting variable at each branching point to produce the binary partitioning (Breiman, Friedman, Olshen, & Stone, 1984). Researchers refer to the family of decision tree methods as Classification and Regression Tree (CART) methods (Hindman, 2015). Classifier trees apply to discrete outcomes while regression trees apply to continuous outcomes, in which
case the model assigns the end nodes an average value for the estimated outcome. Plotting packages in R can graphically represent decision trees like flow charts, providing high levels of interpretability. One major issue with decision trees is that, similar to OLS regression, extreme cases in a data set heavily influence the procedure and makes it prone to over-fitting (Hindman, 2015).

**Random Forests.** Random forest (Brieman, 2001) is a modification of a decision tree algorithm that produces multiple trees (i.e. a forest) based on bootstrapped samples of the dataset, which serves primarily to mitigate the issue of over-fitting. Random Forest (RF) is a type of ensemble learning, where a method generates many classifiers and then aggregates the results. Specifically, RF generates a number of decision trees based on bootstrapped samples of the data. This process uses a bagging method, as opposed to a boosting method, where each tree has the same weight in deciding the final algorithm, similar to a majority vote (Liaw & Wiener, 2002). The unique approach of RF is that the algorithm generates each decision tree using only a subset of the variables as well as a subset of cases (observations). In a standard tree, the algorithm splits each node using the best from all possible variables, but RF splits each node using a subset of predictors randomly chosen for that node. Hence, the decision tree forest is comprised of randomized trees. While this strategy may seem counterintuitive, researchers have asserted that it makes RF models very robust against over-fitting, which is always a concern with ML algorithms (Liaw & Wiener, 2002). Furthermore, RF models also provide an additional benefit in that they are more robust against miscalculations due to extreme observations (Hindman, 2015), which is a typical characteristic of deviance data. Linear regression models are particularly prone to over-fitting. Decision tree methods, such as RF, typically outperform standard regression (Hindman, 2015) and offer a relatively simple ML alternative to OLS. RF offers multiple
methods for assessing variable importance, including Gini impurity, percent increase to Mean Standard Error (MSE), and Permutation Importance (PIMP), making it particularly well suited for the current study in comparison to other ML options.

**Variable Importance Metrics for Random Forest.** In ML literature, researchers use the term feature to describe any quantitative input in a ML model. In social sciences, this term is synonymous with independent variables or predictor variables. Therefore, feature importance is synonymous with variable importance in the present context. For examining variable importance in RF models, researchers have several metrics to examine. First is the Gini impurity coefficient, which indicates a variable’s relative importance in the model. For RF regression models, Gini-based importance assesses a variable’s increase to node purity, which is the reduction to sum of squared errors whenever the algorithm chooses that variable to represent a node. The actual unit of measurement for the Gini coefficient is arbitrary, only the relative magnitude of the coefficient between predictors is important. The second variable importance indicator is percent increase in MSE. This calculation uses a permutation process. In the present study, a permutation process redistributes the values of a variable across observations (participants), i.e. shuffles values within a column, creating a random (noise) variable. The percent increase in MSE metric assesses the increase to model MSE when using permuted (noise) values in place of the normal variable. This indicator provides information about a variable’s global importance, i.e. its contribution to model accuracy. The final variable importance metric is permutation importance (PIMP), which researchers introduced as a corrected measure of feature importance (Altmann, Tolosi, Sander, & Lengauer, 2010). The PIMP method is another permutation-based approach; however, in order to preserve the relationship between predictor variables, this method permutes the outcome variable instead of predictor variables (Altmann et al., 2010). The PIMP algorithm is a method for
correcting bias in other importance metrics, specifically the Gini importance metric, which occurs when predictors are highly correlated. The algorithm permutes the outcome variable a set number of times by randomly shuffling the values of the outcome across participants. This process creates a set of permuted versions of the outcome variable. For each permutation of the outcome variable, the algorithm assesses the importance of each predictor variable, yielding a number of importance estimates equal to the number of outcome permutations. The resulting distribution of importance estimates represents a null distribution, which researchers call the null importances (Altmann et al., 2010). Using the null importances distribution, the algorithm then calculates the probability (p-value) of observing the non-permuted variable importance value in the null distribution. This process generates p-values for the non-permuted variable importance metric, i.e. Gini coefficient.

To understand better the utility of RF for item level analysis compared to traditional social science approaches, this study will examine the ability of ML algorithms to predict CWBs using scale items as inputs. Typically, however, researchers aggregate scale items into a singular scale level variable. This process assumes equality in item function, i.e., that scales consist of items that contribute equally to a scale-level score. While aggregation is a useful psychometric tool, it also limits researchers’ ability to examine item level function and to determine low functioning items with regard to predictive validity. To better illustrate the utility of RF for item level analysis, I will first create models using the aggregated MTB scales as predictor variables for comparison. Hypothesis 1 concerns the ability of RF models to explain additional variance beyond that of standard linear regression models when using the typical approach of aggregating scale items.
**Hypotheses 1.** A random forest model that uses scale level maladaptive traits/beliefs as predictors of CWBs will provide more variance explained than either an OLS or LASSO regression using the same method.

The previous scale-level models offer a reference point for comparing the utility of RF as a technique that can provide a reliable item importance analysis. Aggregating items into scales reduces the total information available (i.e. variance), but researchers often apply this technique for psychometric reasons. Both LASSO regression and RF provide an alternative to scale aggregation because they are better for analyzing correlated predictors. This alternative allows both approaches to make use of the total information available by analyzing the predictive validity of an entire pool of items, but LASSO regression is still constrained by assumptions of linearity and parametric distributions. Again, ML algorithms, such as RF models, should outmatch standard linear regression, with the exception of cases where the relationship under examination is truly linear. Although researchers typically assume linearity, they seldom find support for the assumption in applied datasets, particularly those involving deviant behavior data.

Hypothesis 2a concerns the ability of RF models to explain additional variance beyond that of LASSO models when using scale items as predictors.

**Hypotheses 2a.** A random forest model that uses item level maladaptive traits/beliefs as predictors of CWBs will provide more variance explained than an OLS or a LASSO regression using the same method.

Hypothesis 2b concerns the ability of RF to create better predictive models when using scale items as input rather than aggregated scale scores. Using items as model features provides more information to the RF model, which allows for greater diversity in the creation of the
random forest trees. If scale items are not truly equal in their function, the RF model using scale item input should explain more variance in the data.

**Hypothesis 2b.** A random forest model that uses item level maladaptive traits/beliefs as predictors of CWBs will provide more variance explained than a random forest model using scale level maladaptive traits/beliefs as predictors of CWBs.

The primary goal of the current study is to illustrate the utility of RF for creating better predictive algorithms compared to OLS and LASSO regression that also allow for reliable estimates of item-level importance. As mentioned, high correlations among predictor items may bias the OLS estimates of item importance. The research question concerns the pragmatic nature of applying ML models to social science data, specifically delinquency data. While some research has explored the use of RF classifiers in predicting outcomes such as mindfulness (Sauer, Lemke, Zinn, Buettner, & Kohls, 2014), researchers have largely ignored the use of RF for analyzing item importance in social science data. As previously mentioned, RF algorithms have several metrics for analyzing variable importance including Gini impurity, permuted increase to Mean Standard Error (MSE), and Altmann et al.’s (2010) PIMP. This range of importance metrics allows researchers to assess item importance easily in RF frameworks. The current study uses these metrics to assess item importance results for comparison to linear approaches (OLS and LASSO regression) and to illustrate variable reduction approaches in RF frameworks.

**Research Question 1.** What are the differences in relative item importance for predicting CWBs using OLS and LASSO regression compared to random forest algorithms?
Methods

Sample

The sample for this study is an archival data set collected for a previous study that included MTB and CWB data (Andrews, 2018). A total of 734 individuals from Amazon’s Mechanical Turk (MTurk) attempted the study. The study required participants to be 18 or older and to have worked full time (29 + hours per week) for at least one year. I removed 1 participant for being under 18 years old and 71 participants because they worked less than the minimum 29 hours for participation in the study. Additionally, I removed 1 participant who did not provide consent and 64 participants who indicated that they lived outside of the U.S. I removed 50 participants for failing to respond accurately to attention check items. Finally, 46 participants failed to complete the survey instrument fully. The final participant pool included 501 U.S. workers who sufficiently completed the survey and received compensation.

The final data set (n=501) comprises data from 239 male participants and 262 female participants. The mean age is 36 years (median 33) with a minimum age of 19 years and a maximum age of 71. The distribution of race is 72% white (363), 12% Asian (62), 6% black (34), 4% Native American (24), <1% Native Hawaiian/Pacific Islander (3), and 2% “other” (15). Overall, participants represented all twenty-three job sectors established by the Bureau of Labor Statistics, with the most participants holding occupations in Business (N=53), Sales (N=55) and Education (N=56).

Materials and Procedures

Participants responded to an anonymous online survey via Qualtrics. They accessed the survey through an invite-only link by way of an MTurk HIT (Human Intelligence Task). All participants received compensation of one dollar for successfully completing the study.
Measures

The archival data set included the following previously published scales that purport to measure their corresponding constructs. These scales are the property of their individual authors and they are available for academic and research purposes. Participants responded to all CWB items using a Likert scale (1=Never, 2=Once or twice, 3=Once or twice per month, 4=Once or twice per week, 5=Everyday). For all other items, participants responded to the prompt “Please indicate how much you agree with the following statements” with a Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree).

**Counterproductive Work Behaviors Checklist (CWB-C).** The data set includes a measure of CWBs collected using Spector et al.’s (2006) 33-item measure (α = .97). Participants responded to the prompt “How often have you done each of the following things on your job(s) within the past year?” I altered the original scale prompt from “…on your present job?” to “…on your job(s) within the past year?” The rationale for this alteration was two-fold. First, an individual’s ability to recall events accurately beyond one year may be limited. Second, I wanted to sample individuals’ tendencies across organizational settings if they happened to have more than one job. Example CWB events included “purposely wasted your employer’s materials/supplies” and “came to work late without permission.” A list of all items is available in Appendix A.

**The Dirty Dozen.** The data set includes a measure of the dark triad traits collected using Jonason and Webster’s (2010) 12-item dark triad measure (α = .86). Example items include “I tend to want others to admire me” and “I have used deceit or lied to get my way.” A list of all items is available in Appendix B.
Barrett Impulsivity Index (BIS). The data set includes a measure of impulsivity collected using an abbreviated 8-item version (Morean et al., 2014) of the original Barratt Impulsiveness scale (Patton, Stanford, & Barratt, 1995; α = .82). Example items include “I say things without thinking” and “I act on the spur of the moment.” A list of all items is available in Appendix C.

Psychological Entitlement. The data set includes a measure of entitlement collected using Campbell et al.’s (2004) 9-item Personal Entitlement measure (α = .90). Example items include “I honestly feel I’m just more deserving than others” and “I demand the best because I am worth it.” A list of all items is available in Appendix D.

Propensity to Morally Disengage. The data set includes a measure of moral disengagement collected using Moore et al.’s (2012) 8-item Propensity to Morally Disengage measure (α = .85). Example items include “people can’t be blamed for doing things that are technically wrong when all their friends are doing it too” and “I think it is okay to spread rumors to defend those you care about.” A list of all items is available in Appendix E.

Attention Checks. The data collection included several attention check items to ensure that participants were thoroughly reading each item prior to responding. These items asked participants to select a certain response option in order to proceed with the survey. An example of an attention check item is “for this question, select the option, “agree””. If a participant failed an attention check item, the survey platform immediately prevented further participation in the study.

Presentation of Measures

The electronic survey first presented respondents with demographic items. A list of all demographic items is available in Appendix F. Next, the survey presented CWB and
Organizational Citizenship Behavior items to participants in random order. An overall prompt preceded this section: “Please answer the following questions honestly; remember that your answers are anonymous.” I intended this prompt to encourage participants to answer the CWB items honestly, even though these items describe socially undesirable behaviors.

**Results**

I conducted all statistical analyses using R 3.5.2 (R Core Team, 2018). Appendix G presents the full reproducible and annotated R code. To begin data analysis, I first split the total sample (N=501) into a training set and test set. I split the data into 80% for the training sample (n=402) and 20% for the test sample (n=99), which is a common splitting practice in ML literature. The purpose of splitting the data into a test set and training set was to allow for model tuning on the training set while using the test set as a holdout sample to later produce generalized performance estimates for each model (Zheng, 2015). I partitioned the model using the createDataPartition function within the caret package (Kuhn, 2018). This function takes a stratified random sample of the data to create training and tests sets while preserving the proportion of categories in y, i.e. the samples should have similar distributions. Given that CWBs do not have a normal distribution in the data set (Andrews, 2018), using random sampling could create training and test sets that differ dramatically in distribution because of sampling error.

**Validation and Hyperparameter Optimization**

Generally, in ML applications, analysts split the training set into a training and validation set so that the total data splits into three groups: training, validation, and testing. Researchers use validation sets to tune model hyperparameters. Predictive models estimate parameter values (e.g. Beta weights) directly from a data set, but hyperparameters represent preset model configurations for which the model cannot estimate values directly from the data. An algorithm
must search for an optimal hyperparameter through trial and error by testing an array of models with different hyperparameter values (i.e. configurations). ML researchers call this process hyperparameter tuning (Zheng, 2015). For example, OLS estimates beta weights as a parameter. The LASSO algorithm adds a regularization hyperparameter to OLS, which is the lambda constraint coefficient. Researchers must search through a large set of possible lambda values in order to determine its optimal value. This process of testing competing preset configurations uses the validation set as a testing set for the different configurations. When creating a ML model, the model should never have access to the holdout test cases during the training or validation processes. In other words, model validation is a separate step aside from model testing (Zheng, 2015). Instead of manually splitting the training set into training and validation sets, I used a process called k-fold cross-validation. K-fold is a method for generating training and validation splits for the purposes of hyperparameter tuning.

The k-fold method optimizes the use of a data set for training and validating ML models by splitting the data set into k (number) folds with equal sample sizes. K-fold validation uses each fold as a validation set while training the model using the other folds (k - 1). This validation process uses each observation in the training data exactly one time. The number of tested models on each fold is equal to the number of hyperparameter combinations under assessment. For example, if a model has two hyperparameters with three and four options each, the number of tested models (i.e. possible configurations) would equal twelve. In the end, the model with the best mean performance across folds represents the best hyperparameter configuration. The k-fold method of validation is simply a means to select the best hyperparameter configuration. Researchers then use the best configuration to train a new model on the entire training set using the optimal hyperparameter values. ML researchers refer to this process as hyperparameter
optimization (Zheng, 2015). The important point here is that training, validation, and testing sets are not interchangeable. ML researchers use the term data leakage when referring to mixing of training, validation, and/or testing data sets and they considered this a serious error that will decrease generalizability of ML models. Data leakages can easily occur if researchers do not take care to prevent accidental exposure of the model to test set data. For example, the LASSO approach requires standardization of predictor variables. Researchers must calculate the standardization using only the information in the training data set. If standardization occurs before the data split, the calculation for the standardized scores will include instances from the test set, which is a form of data leakage (Nawara, 2019).

I trained the LASSO models and RF models using k-fold (k = 10) cross-validation. For the linear regression models, cross-validation was not necessary because OLS has no hyperparameters to select. LASSO regression has one hyperparameter to optimize, which is the previously mentioned lambda constraint value. In the current study, the RF model also has one hyperparameter, which is the number of variables the algorithm randomly selects for assessment at each decision tree node. After training via k-fold, I retrained the models using the entire training set and then estimated the predictive accuracy of each model using the test data. To generate model fit statistics, I compared each model’s predicted CWB values for cases in the test set with their observed CWB values. To do this, I used the postResample function included in the caret package. Table 1 presents fit estimates, Root Mean Squared Error (RMSE), R squared, and Mean Absolute Error (MAE), for each of the study models using the test set. Here, I have rank ordered models from best to worst fit. I use this information to address the research hypotheses. I explain how I acquired each model’s estimates in the following sections. For the scale level analyses, I created the models by using each of the MTB scales (item aggregates) as predictors
(features) and an aggregated CWB scale as the continuous outcome variable. For the item level analyses, I created the models by using the individual MTB items as predictors (features) and an aggregated CWB scale as the continuous outcome variable. The following sections detail the model creation process for each of the six models.

**Maladaptive Traits/Beliefs Scale Analyses**

**OLS.** As previously noted, OLS regression has no hyperparameters. The use of cross-validation is unnecessary with OLS. I simply trained the model on the entire training set using linear modeling (i.e., the `lm` function) in base R. I then used the trained model to predict CWBs for cases in the test set. Fit statistics for the OLS model are in Table 1. Table 2 presents the beta weights and p-values for each of the MTB scales in the model.

**LASSO.** I used the `glmnet` (Friedman, Hastie, & Tibshirani, 2010) package in R to compute results for the LASSO model. For the LASSO model, I standardized the predictor variables within the `glmnet` function. When conducting LASSO regression, it is common practice to standardize predictor variables, as suggested by the original author (Tibshirani, 1996). LASSO places constraints on the size of coefficients for each variable and the value of the constraint is dependent upon the magnitude of each variable. Without standardization, the constraints placed on each predictor are uneven. I used the caret package to cross-validate the LASSO model in order to select the best lambda constraint value. Using a grid search method, I tested all possible values of lambda between 0.001 and 1.0 using increments of .001, i.e. 1,000 possible values. I searched an atypically large range of lambda values to aid in the generalizability of the included R code, but this does have the consequence of increasing computation time. Researchers may choose to reduce the range of lambda values to meet the needs of their own data set.
The optimal lambda value from the cross-validated LASSO model using MTB scales was 0.022. After selecting the lambda value, I then trained a new LASSO model on the entire training set using glmnet. The final training model has the same beta weights as the cross-validated model. In this sense, while it does have a hyperparameter, LASSO is unique due to the inherent connection between its model parameters and its hyperparameter. Specifically, LASSO is a regularization process that should help prevent over-fitting. Finally, I used the trained LASSO model to predict CWBs for cases in the test set. Fit statistics for the LASSO model using MTB scales are in Table 1. Table 2 presents the constrained beta weights from the OLS model using MTB scales. Predictors with coefficients greater than zero are significant by nature of having beta weights greater than zero (i.e. not suppressed). The model does not calculate p-values for the beta weights because the weights themselves are constrained values.

**Random Forest.** I used the randomForest package (Liaw & Wiener, 2002) in R to compute results for the random forest model. I created the RF model using 100 individual trees. ML literature argues that researchers should not tune the number of trees in a RF model and that the first 100 trees in a RF model provide the biggest performance gains (Probst & Boulesteix, 2017). The addition of more trees should not significantly influence results on a final model but would increase computational load; therefore, a setting of 100 trees offers a balanced threshold between optimal performance and unnecessary computation. I used the caret package to cross-validate the RF model in order to select the best hyperparameter. For the randomForest package, the available hyperparameter, called mtry, is the number of variables randomly selected for evaluation at each decision tree node. The optimized number of variables evaluated at each node was two. After selecting the hyperparameter value using the cross-validated RF model, I then trained a new RF model on the entire training set using the randomForest package. Finally, I
used the trained RF model to predict CWBs for cases in the test set. Model fit indices for the RF model using MTB scales are in Table 1. The RF model using anti-social beliefs/traits scales as predictors had a lower MSE, a lower MAE, and a higher $R^2$ than the OLS and LASSO models using the same methods. Results support Hypothesis 1.

I calculated the RF variable importance metrics on the trained model. The randomForest package provides estimates of the Gini impurity and percent increase to MSE metrics but does not generate p-values (i.e. significance values) for these metrics. I used a permutation approach to generate p-values for these metrics using the rfPermute package (Archer, 2018). This approach permutes (shuffles) each predictor’s values across observations (participants) a set number of times, creating a set of permuted versions of the original variables. The algorithm then estimates variable importance metrics for each of these permuted variables in order to create a distribution of null importances for each variable. This method then generates probability values (p-values) for observing a variable’s non-permuted importance statistic within the null importance distribution for that variable. The typical social science cut-off value of $p<.05$ applies to these values, suggesting that the chance of observing the non-permuted metric in the null distribution is less than 5%. I used the vita (Celik, 2015) package in R to compute the PIMP statistic. Table 3 presents results from the variable importance analysis on the cross-validated RF model using MTB scales. The importance metrics from the RF model using MTB scales yields roughly the same information as that of the OLS and LASSO beta weights. Specifically, that the moral disengagement scales had the highest importance ratings and that the impulsivity and entitlement scales offered little additional prediction power.
Maladaptive Traits/Beliefs Item Level Analysis

**OLS.** I trained the OLS model using MTB items as predictors on the entire training set using the lm function in base R. I then used the trained model to predict CWBs for cases in the test set. Model fit results from the OLS regression using MTB items are in Table 1. For answering research question 1, Table 4 contains linear regression beta weights for each of the MTB items.

**LASSO.** I used the glmnet package in R to compute results for the LASSO model using MTB items as predictors and again standardized the predictor variables within the glmnet function. I used the caret package to cross-validate the LASSO model, using the same grid search method previously noted, to select the best lambda constraint value, which was .017. After selecting the optimal lambda value, I then trained a new LASSO model on the entire training set using glmnet. Results from LASSO model using MTB items are in Table 1. For answering research question 1, Table 5 contains the LASSO beta weights for each of the MTB items.

**Random Forest.** I used the randomForest package (Liaw & Wiener, 2002) in R to compute results for the random forest model using MTB items as predictors. I again created the RF model using 100 individual trees. I used the caret package to cross-validate the RF model in order to select the best hyperparameter. The optimized number of variables evaluated at each node was twelve. Next, I computed the Gini impurity, percent increase in MSE, and PIMP metrics using the cross-validated model. For answering research question 1, Table 6 contains results from the variable importance analysis on the cross-validated RF model using MTB items. After selecting the hyperparameter value and assessing variable importance using the cross-validated RF model, I then trained a new RF model using all 37 items on the entire training set.
Finally, I used the trained RF model to predict CWBs for cases in the test set. Model fit indices for the RF model using all MTB items are in Table 1. The RF model created using MTB items as predictors had a lower RMSE, a lower MAE, and a higher R² than the OLS and LASSO models created using MTB items. Results support Hypothesis 2a.

Hypothesis 2b stated that the RF model using MTB items as predictors would have higher predictive power than the RF model using MTB scales as predictors. The RF model using MTB items as predictors had a lower RMSE than the RF model using MTB scales, but the two models had roughly equivalent MAE values with a difference of .001. The former model does have a higher R² value, but this estimate is likely biased when comparing models using different predictor sets. In general, making comparisons between non-nested models that use different predictor sets is generally difficult because both out-of-sample prediction and parsimony are important concepts when comparing models. The purpose for comparing RF models in the current study is to generate a wider sense of how RF models will function using traditional psychology approaches (scale aggregation) compared to item level analysis. It seems that the RF model using all MTB items functioned only marginally better than the RF model using MTB scales as predictors. To assess hypothesis 2b further, I conducted a variable reduction approach on the RF model using MTB items.

**Variable Reduction Approach.** The research question for this study addressed the comparison of relative item importance for predicting CWBs across the OLS, LASSO, and RF models using MTB items as predictors. The variable importance metric for OLS and LASSO are beta weights. Specifically, LASSO is a regularization technique that can optimize OLS by suppressing predictors by using the lambda constraint. In that regard, researchers use LASSO to determine which predictors are truly important and then adapt the model accordingly by simply
using the optimized lambda. RF uses very different variable importance metrics for determining which variables are most important for prediction. Two of the three RF variable importance metrics use permutations to assess the importance of items; the other metric (Gini) assesses the rate at which the RF model uses a variable to make splits.

Importantly, the pragmatic reason for assessing variable importance is so researchers can remove unimportant variables, in this case scale items. Researchers can potentially use these importance metrics to decrease the number of items included in psychological instruments. LASSO performs a kind of variable removal process by constraining beta weights to zero. RF has a different procedure for variable reduction. The RF model using MTB items as predictors has a much greater number of features than the RF model using MTB scales as predictors, so I illustrated the variable reduction approach using the former. Specifically, I used the model training method illustrated by Altmann and colleagues (2010) to remove items with non-significant PIMP variable importance metrics before training a new RF model after the hyperparameter cross-validation process. To assess the utility of RF variable importance metrics for item reduction, I tested a second RF model using MTB items as predictors, but this time I retained only the items with significant PIMP p-values from cross-validation (see Table 6). Only two items, moral disengagement six and two, had significant PIMP p-values (p<.05). Next, I used the trained reduced variable RF model to predict CWBs for cases in the test set. Results from the reduced RF model analysis suggest that the two remaining predictor items explained 69.66% of variance in the test set (\(R^2=.69\)), with a RMSE of .340 and a MAE of .249. The reduced RF model containing only two predictors has better fit-statistics compared to the RF model containing all 37 predictors (see Table 1) and is more parsimonious. If we consider that RF model using MTB items functions better after item reduction, then these results yield further
support for Hypothesis 2b. In addition, these results clearly illustrate the utility of RF models relative to LASSO models for variable importance analysis and item reduction.

**Differences in Variable Importance Rankings**

The research question for this study concerned the differences in variable importance ratings across different approaches (models) when the predictors were MTB items. Table 7 presents the MTB items’ variable importance rankings across each of the three models using MTB items as predictors of CWBs. The table shows that item rank ordering across the metrics is similar but does fluctuate across the different metrics. Notably, the two moral disengagement items selected for the final reduced RF model have top rankings across all metrics. This would suggest that the PIMP method is effective in selecting predictors that show consistent significant links to the outcome variable. Furthermore, the PIMP method appears well suited for item level analysis in social science due to its performance in selecting significant items even among highly correlated items.

**Discussion**

In general, results from this study support the utility of RF models for predicting important outcomes in social science, and more specifically deviant behavior. Findings from this study support Hindman’s (2015) assertion that ML algorithms can outperform traditional regression approaches such as OLS (linear regression). Furthermore, this study demonstrates the appropriate steps for conducting RF and provides the computer code (R code) to run the illustrated analyses. The provided R code is soft-coded, which means that it automatically generalizes to new data sets that have different numbers of predictors without needing to change the code. The major roadblock that prevents more social scientists from applying ML techniques is a lack of access to interpretable resources. Psychologists are not computer scientists, but we
have use for algorithms and products of computer science. It is likely difficult for social scientists to implement ML due to lack of concrete examples of ML algorithms applied to social science data and a lack of computer programming knowledge necessary to perform these analyses. This study helps bridge the gap between two fields by providing an example application of ML to social science data as well as illustrative and generalizable programming code to perform these techniques. This can aid other social scientists and help them to make use of valuable new analytic resources.

Hypothesis 1 addressed the ability of RF models to outperform traditional linear models when using the typical social science approach of aggregating items into scales. The RF model using MTB scales as predictors explained an additional ten percent of variance over the OLS model using scales as predictors and an additional eight percent of variance over the LASSO model using scales as predictors. This study demonstrates the ability of ML algorithms to outperform traditional regression approaches using the typical predictors of social science data, i.e. psychometric scales.

This study specifically explored the utility of random forest algorithms for interpreting item level prediction, which is a more atypical approach in social science. Hypothesis 2a claimed that RF models would outperform traditional linear models when using MTB items as predictors of CWBs. The RF model using MTB items as predictors explained an additional sixteen percent of variance over the OLS model using MTB items as predictors and an additional fifteen percent of variance over the LASSO model using MTB items as predictors. Clearly, RF models are more adept at utilizing items as predictor variables compared to traditional regression approaches in social sciences. In general, RF models performed better than OLS and LASSO models in the current study. One reason for this is that OLS models tend to over-fit on the training data, which
lowers their predictive power when applied to new data (i.e. test data). The design of RF models reduces the issue of over-fitting, which in turn helps to maintain predictive accuracy when applying these models to new data. Another likely reason the RF models performed better is that there is some degree of non-linearity in the relationship between predictor variables and the target outcome (CWBs). RF models can account for non-linearity in the relationship between predictors and CWBs while OLS and LASSO cannot.

Hypothesis 2b claimed that an RF model would function better when using individual scale items as predictors rather than using the typical approach of aggregating items into scale scores as predictors. While the RF item model did have a lower RMSE than the RF scale model, the two models had roughly equivalent MAEs. The more important finding here, though, is that after variable reduction, the final reduced item RF model used only 6% of the original predictors (i.e., two items) and improved across fit indices compared to the RF model that included all 37 items. Therefore, I contend that researchers should chose the reduced item RF model as the best model for prediction due to its out-of-sample performance and parsimony. This finding greatly supports the utility of RF variable importance metrics and underlines the fact that using RF modeling on social science data has clear pragmatic applications.

The research question for this study addressed the difference in variable importance metrics for MTB items across algorithms. The RF model provides a wider range of item importance indicators, which researchers can use to decide which items to remove while maintaining the best predictive power. The current study provided three RF variable importance indices for determining which items have truly low importance in the model. First, the Gini coefficient provides a computationally inexpensive calculation of variable importance. This index denotes the mean decrease in sum of squared errors when the model selects an item for a
node, relative to other items under consideration within the model. Second, the permuted percent increase in MSE metric shuffles a predictor variable’s values across observations and then runs the RF model using the permuted (shuffled) version of the variable. The calculation then compares the MSE of the RF model using the un-permuted variable to that of the model using the permuted variable. The difference indicates the loss in accuracy the model suffers because of using the permuted version of a variable. In this sense, an item’s importance is equivalent to its impact on the model’s fit indices. Finally, the Altman (2010) permutation importance (PIMP) is the most computationally heavy index and appears to be the most stringent of the item importance indices. This method permutes the outcome variable rather than the predictor variables, thus preserving the relationships between predictors. Using multiple permutations, this method creates a null distribution of variable importances, which the model uses to generate \( p \)-values for each variable’s importance (Gini) metric. For this method, a variable’s importance represents its ability to predict the target outcome while considering its relationship to other predictors in the model. While it is common for researchers to refer to ML algorithms as black box phenomena (Card, 2017), RF models actually provide more information about the functionality of predictors (features) than traditional regression approaches. A wider range of variable importance indicators provides researchers with more information to leverage when deciding which items to remove. This can aid researchers in reducing the number of item in psychometric instruments while still maintaining confidence that the remaining items will retain previous levels of predictive capability.

To illustrate the value of RF variable importance metrics, I conducted a variable reduction procedure suggested by Altmann et al. (2010). I used a PIMP \( p \)-value of <.05 from the cross-validated RF model as cut-offs for significant variable importance when creating a reduced
variable RF model. This reduced RF model had the best predictive power compared to all other models and used only two of the original thirty-seven scale items. This magnitude of variable reduction displayed in the current study is promising; researchers have not previously demonstrated such stark results using OLS approaches. This final reduced model represents the best fitting model and the most parsimonious of all study models. This study has fully demonstrated the utility of RF using the current data set.

Another major contribution of this study is the R code that I produced and attached to this document. Without doubt, ML topics have gained popularity in social science, but social scientists may encounter difficulty when attempting to apply ML to prediction efforts in their respective fields. The vast majority of ML research exists in journals well outside of social science. Furthermore, researchers with limited coding experience will certainly encounter difficulty when attempting to set up ML algorithms. This study, and the annotated R code, documents the process of implementing LASSO and RF regression from the perspective of a social science researcher. Furthermore, the R code is generalizable (i.e. soft-coded), meaning researchers can use the code for their own studies by following the short instructions at the beginning of the R document (Appendix G). For example, a personality researcher could repurpose the prediction analyses from this study to use on a data set that includes the Big-Five personality index as indicators and any continuous outcome of interest. No alteration to the fundamental analysis code is necessary.

Lastly, results from this study are uniquely interesting for researchers of anti-social or deviant behavior. Particularly, the application of RF provides a major insight into item importance for many common associates of deviant behavior. Using the steps suggested by Altmann and colleagues, (2010), the RF procedure was able to reduce the total number of
relevant predictor items greatly. Perhaps even more important to I/O psychologists, forensic psychologists, and criminologists is the nature of the two remaining items. Both items with significant variable importance (PIMP) belonged to the moral disengagement scale. The first item, moral disengagement six, reads “Taking personal credit for ideas that were not your own is no big deal.” The second item, moral disengagement two, reads “Taking something without the owner’s permission is okay as long as you’re just borrowing it.” The content of these two items converges around the notion of wrongfully taking something of value, either social credit or physical objects. It seems logical that these two items would predict workplace deviance well.

**Limitations**

Researchers should take some caution when generalizing findings from the current study to new scenarios. The context of the target behavior is likely important. The RF item reduction process created a narrow predictor set that is specifically suited to predict CWBs. That is, had the target outcome been criminal behavior or arrest rates, different MTB items may have performed better. This is an obvious topic for future research, and the code provided in this document would make the application to new data sets straightforward. Additionally, the randomForest package will automatically adapt for categorical or continuous outcomes, such that the provided R code can generalize to categorical outcomes with only few alterations to the code generating fit statistics. Researchers can use the provided code to generalize the analyses in this study and produce results that are relevant to different target behaviors.

The RF approach reduced the pool of predictor items down to two relevant predictors, but this finding has caveats. I would not recommend that a study use a set of only two predictor items. Item reduction serves to reduce survey length, which helps to avoid participant fatigue. Reducing a survey to only two predictor items, however, would be overcompensation. In
addition, responding to only two predictor items would almost certainly influence the way that participants respond to the survey items. For example, survey respondents may be more likely to alter or distort their responses if they know that only two items will represent their characteristics on a survey. In an applied setting, such as employee selection, using only two indicator items may influence survey respondents’ perceptions of face validity or their perceptions of justness regarding the selection instrument. In other words, there may be many unforeseen consequences of such a drastic reduction in survey length. Instead, researchers should consider these results as an opportunity to search for new relevant predictors. Future studies on CWBs should examine new measures in addition to the measures/items that were significant in the present study.

The RF item reduction process outlined in this study is not a replacement for psychometric design principles, but it can add to current psychometric approaches. For example, Andrews (2018) examined the psychometric properties of the scales and items in the current data set. This process established content validity for the current measures; however, the present study suggests that the vast majority of items are not relevant in predicting the intended outcome. Current findings question the pragmatic nature of including a large set of predictor items if the majority of items have limited criterion-related validity. The approach outlined in this study can help researchers narrow down a small set of highly predictive scale items after first conducting the typical scale development procedures. Researchers may then attempt to improve or add scale items if necessary. Over repeated survey administrations, RF models can likely identify a small pool of items that will consistently predict a target outcome. Again, the target behavior is important and the pool of significant items may change depending on the specific outcome of interest.
Concluding Remarks

Interest in ML topics is on the rise in social sciences, with the Society for Industrial and Organizational Psychology (SIOP) preparing to host its 2nd annual ML competition. Nonetheless, social scientists may encounter roadblocks when trying to apply ML to their respective fields. One roadblock is the lack of information on ML within social science literature, forcing researchers to consult and extrapolate meaning from literature pertaining to unfamiliar subjects such as bioinformatics (e.g. Altmann et al., 2010). A second roadblock is difficulty in setting up computer code to run ML models. The current study helps bridge both of these gaps between ML and social science literature. Firstly, I provide an illustration of RF’s utility and application in the context of social science data. Secondly, I provide researchers the tools (R code) to reenact the analyses from this study on their own data, even without substantial knowledge of computer languages. Furthermore, having information presented in a familiar format can help social scientists transfer knowledge about ML to their respective fields. While this study focused specifically on CWB prediction, the applications and methods in this study are widely applicable to other topics of social science research. As the social sciences continue their courtship with ML topics, the current study offers a guide for application of ML techniques to research questions in specific fields.
References


Loomis v. Wisconsin, 137 S.Ct. 2290 (June, 2017).


Table 1

*Fit Statistics for OLS, LASSO, and RF Models Using Maladaptive Traits/Beliefs Scales and Items to Predict CWBs, (n=99).*

<table>
<thead>
<tr>
<th>Model</th>
<th>RMSE</th>
<th>$R^2$</th>
<th>MAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random Forest (Items)</td>
<td>0.396</td>
<td>0.64</td>
<td>0.256</td>
</tr>
<tr>
<td>Random Forest (Scales)</td>
<td>0.419</td>
<td>0.58</td>
<td>0.257</td>
</tr>
<tr>
<td>LASSO (Scales)</td>
<td>0.455</td>
<td>0.50</td>
<td>0.321</td>
</tr>
<tr>
<td>LASSO (Items)</td>
<td>0.459</td>
<td>0.49</td>
<td>0.339</td>
</tr>
<tr>
<td>OLS (Scales)</td>
<td>0.461</td>
<td>0.48</td>
<td>0.331</td>
</tr>
<tr>
<td>OLS (Items)</td>
<td>0.466</td>
<td>0.48</td>
<td>0.348</td>
</tr>
</tbody>
</table>

*Note. CWBs = Counterproductive Work Behavior.*
<table>
<thead>
<tr>
<th>Variable</th>
<th>OLS β</th>
<th>OLS p-value</th>
<th>LASSO β</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moral Disengagement</td>
<td>0.35</td>
<td>0.00**</td>
<td>0.33</td>
</tr>
<tr>
<td>Dark Triad</td>
<td>0.23</td>
<td>0.00**</td>
<td>0.18</td>
</tr>
<tr>
<td>Impulsivity</td>
<td>0.04</td>
<td>0.17</td>
<td>0.05</td>
</tr>
<tr>
<td>Entitlement</td>
<td>-0.05</td>
<td>0.15</td>
<td>0.00</td>
</tr>
</tbody>
</table>

*Note.* *p* < .05, **p** < .01, CWBs = Counterproductive Work Behaviors
Table 3

*Variable Importance Metrics from Random Forest Model Using Anti-Social Beliefs/Traits Scales to Predict CWBs, (n=402).*

<table>
<thead>
<tr>
<th>Variable</th>
<th>Gini Impurity</th>
<th>% MSE</th>
<th>Gini p-value</th>
<th>% MSE p-value</th>
<th>PIMP p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moral Disengagement</td>
<td>58.64</td>
<td>16.54</td>
<td>0.00**</td>
<td>0.00**</td>
<td>0.01*</td>
</tr>
<tr>
<td>Dark Triad</td>
<td>47.00</td>
<td>9.94</td>
<td>0.00**</td>
<td>0.11</td>
<td>0.48</td>
</tr>
<tr>
<td>Impulsivity</td>
<td>19.60</td>
<td>5.95</td>
<td>1.00</td>
<td>0.58</td>
<td>0.33</td>
</tr>
<tr>
<td>Entitlement</td>
<td>16.79</td>
<td>2.07</td>
<td>1.00</td>
<td>0.58</td>
<td>0.82</td>
</tr>
</tbody>
</table>

*Note.* *p < .05, **p < .01, CWBs = Counterproductive Work Behaviors.
Table 4  
*Beta weights from Linear Regression Model Using Anti-Social Beliefs/Traits Items to Predict CWBs, (n=402).*

<table>
<thead>
<tr>
<th>Item</th>
<th>$\beta$</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Impulsivity 1(R)</td>
<td>0.358</td>
<td>0.62</td>
</tr>
<tr>
<td>Impulsivity 2(R)</td>
<td>-0.018</td>
<td>0.19</td>
</tr>
<tr>
<td>Impulsivity 3(R)</td>
<td>-0.041</td>
<td>0.32</td>
</tr>
<tr>
<td>Impulsivity 4(R)</td>
<td>0.027</td>
<td>0.01*</td>
</tr>
<tr>
<td>Impulsivity 5</td>
<td>-0.099</td>
<td>0.40</td>
</tr>
<tr>
<td>Impulsivity 6</td>
<td>0.024</td>
<td>0.05</td>
</tr>
<tr>
<td>Impulsivity 7</td>
<td>0.061</td>
<td>0.00**</td>
</tr>
<tr>
<td>Impulsivity 8</td>
<td>0.072</td>
<td>0.95</td>
</tr>
<tr>
<td>Dark Triad 1</td>
<td>0.002</td>
<td>0.75</td>
</tr>
<tr>
<td>Dark Triad 2</td>
<td>0.009</td>
<td>0.07</td>
</tr>
<tr>
<td>Dark Triad 3</td>
<td>0.044</td>
<td>0.47</td>
</tr>
<tr>
<td>Dark Triad 4</td>
<td>0.017</td>
<td>0.21</td>
</tr>
<tr>
<td>Dark Triad 5</td>
<td>0.042</td>
<td>0.00**</td>
</tr>
<tr>
<td>Dark Triad 6</td>
<td>0.089</td>
<td>0.17</td>
</tr>
<tr>
<td>Dark Triad 7</td>
<td>-0.041</td>
<td>0.42</td>
</tr>
<tr>
<td>Dark Triad 8</td>
<td>0.024</td>
<td>0.90</td>
</tr>
<tr>
<td>Dark Triad 9</td>
<td>0.003</td>
<td>0.11</td>
</tr>
<tr>
<td>Dark Triad 10</td>
<td>-0.040</td>
<td>0.10</td>
</tr>
<tr>
<td>Dark Triad 11</td>
<td>0.040</td>
<td>0.80</td>
</tr>
<tr>
<td>Dark Triad 12</td>
<td>0.006</td>
<td>0.46</td>
</tr>
<tr>
<td>Entitlement 1</td>
<td>-0.021</td>
<td>0.20</td>
</tr>
<tr>
<td>Entitlement 2</td>
<td>-0.044</td>
<td>0.50</td>
</tr>
<tr>
<td>Entitlement 3</td>
<td>-0.023</td>
<td>0.00**</td>
</tr>
<tr>
<td>Entitlement 4</td>
<td>0.089</td>
<td>0.41</td>
</tr>
<tr>
<td>Entitlement 5(R)</td>
<td>0.025</td>
<td>0.07</td>
</tr>
<tr>
<td>Entitlement 6</td>
<td>-0.050</td>
<td>0.79</td>
</tr>
<tr>
<td>Entitlement 7</td>
<td>-0.007</td>
<td>0.94</td>
</tr>
<tr>
<td>Entitlement 8</td>
<td>0.002</td>
<td>0.86</td>
</tr>
<tr>
<td>Entitlement 9</td>
<td>0.005</td>
<td>0.70</td>
</tr>
<tr>
<td>Moral Disengagement 1</td>
<td>0.011</td>
<td>0.23</td>
</tr>
<tr>
<td>Moral Disengagement 2</td>
<td>0.034</td>
<td>0.00**</td>
</tr>
<tr>
<td>Moral Disengagement 3</td>
<td>0.122</td>
<td>0.55</td>
</tr>
<tr>
<td>Moral Disengagement 4</td>
<td>-0.016</td>
<td>0.49</td>
</tr>
<tr>
<td>Moral Disengagement 5</td>
<td>0.016</td>
<td>0.00**</td>
</tr>
<tr>
<td>Moral Disengagement 6</td>
<td>0.089</td>
<td>0.00**</td>
</tr>
<tr>
<td>Moral Disengagement 7</td>
<td>0.086</td>
<td>0.92</td>
</tr>
<tr>
<td>Moral Disengagement 8</td>
<td>0.002</td>
<td>0.20</td>
</tr>
</tbody>
</table>

*Note. *$p < .05, **p < .01, CWBs= Counterproductive Work Behaviors.*
Table 5
Beta weights from LASSO Regression Model Using Anti-Social Beliefs/Traits Items to Predict CWBs, \( n=402 \).

<table>
<thead>
<tr>
<th>Item</th>
<th>( \beta )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Impulsivity 1(R)</td>
<td>0.000</td>
</tr>
<tr>
<td>Impulsivity 2(R)</td>
<td>-0.016</td>
</tr>
<tr>
<td>Impulsivity 3(R)</td>
<td>0.000</td>
</tr>
<tr>
<td>Impulsivity 4(R)</td>
<td>-0.047</td>
</tr>
<tr>
<td>Impulsivity 5</td>
<td>0.013</td>
</tr>
<tr>
<td>Impulsivity 6</td>
<td>0.048</td>
</tr>
<tr>
<td>Impulsivity 7</td>
<td>0.058</td>
</tr>
<tr>
<td>Impulsivity 8</td>
<td>0.000</td>
</tr>
<tr>
<td>Dark Triad 1</td>
<td>0.010</td>
</tr>
<tr>
<td>Dark Triad 2</td>
<td>0.039</td>
</tr>
<tr>
<td>Dark Triad 3</td>
<td>0.003</td>
</tr>
<tr>
<td>Dark Triad 4</td>
<td>0.017</td>
</tr>
<tr>
<td>Dark Triad 5</td>
<td>0.081</td>
</tr>
<tr>
<td>Dark Triad 6</td>
<td>0.000</td>
</tr>
<tr>
<td>Dark Triad 7</td>
<td>0.012</td>
</tr>
<tr>
<td>Dark Triad 8</td>
<td>0.000</td>
</tr>
<tr>
<td>Dark Triad 9</td>
<td>0.000</td>
</tr>
<tr>
<td>Dark Triad 10</td>
<td>0.010</td>
</tr>
<tr>
<td>Dark Triad 11</td>
<td>0.000</td>
</tr>
<tr>
<td>Dark Triad 12</td>
<td>0.000</td>
</tr>
<tr>
<td>Entitlement 1</td>
<td>0.000</td>
</tr>
<tr>
<td>Entitlement 2</td>
<td>0.000</td>
</tr>
<tr>
<td>Entitlement 3</td>
<td>0.045</td>
</tr>
<tr>
<td>Entitlement 4</td>
<td>0.000</td>
</tr>
<tr>
<td>Entitlement 5(R)</td>
<td>-0.028</td>
</tr>
<tr>
<td>Entitlement 6</td>
<td>0.000</td>
</tr>
<tr>
<td>Entitlement 7</td>
<td>0.000</td>
</tr>
<tr>
<td>Entitlement 8</td>
<td>0.000</td>
</tr>
<tr>
<td>Entitlement 9</td>
<td>0.000</td>
</tr>
<tr>
<td>Moral Disengagement 1</td>
<td>0.023</td>
</tr>
<tr>
<td>Moral Disengagement 2</td>
<td>0.107</td>
</tr>
<tr>
<td>Moral Disengagement 3</td>
<td>0.000</td>
</tr>
<tr>
<td>Moral Disengagement 4</td>
<td>0.000</td>
</tr>
<tr>
<td>Moral Disengagement 5</td>
<td>0.075</td>
</tr>
<tr>
<td>Moral Disengagement 6</td>
<td>0.084</td>
</tr>
<tr>
<td>Moral Disengagement 7</td>
<td>0.000</td>
</tr>
<tr>
<td>Moral Disengagement 8</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Note. CWBs = Counterproductive Work Behaviors.
Table 6
Variable Importance Metrics from Random Forest Model Using Anti-Social Beliefs/Traits Items to Predict CWBs, (n=402).

<table>
<thead>
<tr>
<th>Item</th>
<th>Gini Impurity</th>
<th>% MSE</th>
<th>Gini p-value</th>
<th>% MSE p-value</th>
<th>PIMP p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Impulsivity 1(R)</td>
<td>1.62</td>
<td>1.53</td>
<td>0.99</td>
<td>0.65</td>
<td>0.53</td>
</tr>
<tr>
<td>Impulsivity 2(R)</td>
<td>1.16</td>
<td>1.07</td>
<td>1.00</td>
<td>0.37</td>
<td>0.66</td>
</tr>
<tr>
<td>Impulsivity 3(R)</td>
<td>1.12</td>
<td>-0.11</td>
<td>1.00</td>
<td>0.55</td>
<td>0.84</td>
</tr>
<tr>
<td>Impulsivity 4(R)</td>
<td>0.99</td>
<td>-0.49</td>
<td>1.00</td>
<td>0.23</td>
<td>0.85</td>
</tr>
<tr>
<td>Impulsivity 5</td>
<td>2.82</td>
<td>1.84</td>
<td>0.29</td>
<td>0.51</td>
<td>0.61</td>
</tr>
<tr>
<td>Impulsivity 6</td>
<td>3.84</td>
<td>3.15</td>
<td>0.22</td>
<td>0.08</td>
<td>0.37</td>
</tr>
<tr>
<td>Impulsivity 7</td>
<td>4.37</td>
<td>2.49</td>
<td>0.32</td>
<td>0.35</td>
<td>0.43</td>
</tr>
<tr>
<td>Impulsivity 8</td>
<td>3.91</td>
<td>0.82</td>
<td>0.80</td>
<td>0.42</td>
<td>0.72</td>
</tr>
<tr>
<td>Dark Triad 1</td>
<td>7.27</td>
<td>4.67</td>
<td>0.01*</td>
<td>0.01*</td>
<td>0.09</td>
</tr>
<tr>
<td>Dark Triad 2</td>
<td>4.34</td>
<td>1.41</td>
<td>0.16</td>
<td>0.27</td>
<td>0.74</td>
</tr>
<tr>
<td>Dark Triad 3</td>
<td>1.77</td>
<td>1.13</td>
<td>1.00</td>
<td>0.50</td>
<td>0.66</td>
</tr>
<tr>
<td>Dark Triad 4</td>
<td>6.50</td>
<td>3.71</td>
<td>0.01*</td>
<td>0.07</td>
<td>0.32</td>
</tr>
<tr>
<td>Dark Triad 5</td>
<td>10.15</td>
<td>3.59</td>
<td>0.01*</td>
<td>0.12</td>
<td>0.15</td>
</tr>
<tr>
<td>Dark Triad 6</td>
<td>2.23</td>
<td>1.70</td>
<td>0.73</td>
<td>0.19</td>
<td>0.62</td>
</tr>
<tr>
<td>Dark Triad 7</td>
<td>8.42</td>
<td>3.36</td>
<td>0.20</td>
<td>0.08</td>
<td>0.19</td>
</tr>
<tr>
<td>Dark Triad 8</td>
<td>2.06</td>
<td>0.86</td>
<td>1.00</td>
<td>0.37</td>
<td>0.71</td>
</tr>
<tr>
<td>Dark Triad 9</td>
<td>1.00</td>
<td>1.18</td>
<td>1.00</td>
<td>0.59</td>
<td>0.84</td>
</tr>
<tr>
<td>Dark Triad 10</td>
<td>1.53</td>
<td>1.89</td>
<td>1.00</td>
<td>0.27</td>
<td>0.65</td>
</tr>
<tr>
<td>Dark Triad 11</td>
<td>1.35</td>
<td>1.43</td>
<td>1.00</td>
<td>0.49</td>
<td>0.77</td>
</tr>
<tr>
<td>Dark Triad 12</td>
<td>2.43</td>
<td>0.90</td>
<td>0.49</td>
<td>0.18</td>
<td>0.87</td>
</tr>
<tr>
<td>Entitlement 1</td>
<td>1.63</td>
<td>3.07</td>
<td>1.00</td>
<td>0.65</td>
<td>0.55</td>
</tr>
<tr>
<td>Entitlement 2</td>
<td>1.45</td>
<td>0.57</td>
<td>1.00</td>
<td>0.74</td>
<td>0.79</td>
</tr>
<tr>
<td>Entitlement 3</td>
<td>1.93</td>
<td>2.19</td>
<td>0.93</td>
<td>0.49</td>
<td>0.60</td>
</tr>
<tr>
<td>Entitlement 4</td>
<td>0.82</td>
<td>1.05</td>
<td>1.00</td>
<td>0.96</td>
<td>0.85</td>
</tr>
<tr>
<td>Entitlement 5(R)</td>
<td>2.00</td>
<td>1.12</td>
<td>0.97</td>
<td>0.52</td>
<td>0.77</td>
</tr>
<tr>
<td>Entitlement 6</td>
<td>1.16</td>
<td>1.51</td>
<td>1.00</td>
<td>0.37</td>
<td>0.82</td>
</tr>
<tr>
<td>Entitlement 7</td>
<td>2.13</td>
<td>1.69</td>
<td>1.00</td>
<td>0.84</td>
<td>0.67</td>
</tr>
<tr>
<td>Entitlement 8</td>
<td>1.04</td>
<td>1.12</td>
<td>1.00</td>
<td>0.47</td>
<td>0.77</td>
</tr>
<tr>
<td>Entitlement 9</td>
<td>1.08</td>
<td>0.36</td>
<td>1.00</td>
<td>0.46</td>
<td>0.89</td>
</tr>
<tr>
<td>Moral Disengagement 1</td>
<td>8.45</td>
<td>1.46</td>
<td>0.01*</td>
<td>0.51</td>
<td>0.59</td>
</tr>
<tr>
<td>Moral Disengagement 2</td>
<td>16.24</td>
<td>5.05</td>
<td>0.01*</td>
<td>0.02*</td>
<td>0.02*</td>
</tr>
<tr>
<td>Moral Disengagement 3</td>
<td>2.67</td>
<td>2.96</td>
<td>0.99</td>
<td>0.08</td>
<td>0.72</td>
</tr>
<tr>
<td>Moral Disengagement 4</td>
<td>1.63</td>
<td>-0.56</td>
<td>0.98</td>
<td>0.84</td>
<td>0.86</td>
</tr>
<tr>
<td>Moral Disengagement 5</td>
<td>10.73</td>
<td>4.11</td>
<td>0.01*</td>
<td>0.02*</td>
<td>0.12</td>
</tr>
<tr>
<td>Moral Disengagement 6</td>
<td>16.64</td>
<td>5.31</td>
<td>0.01*</td>
<td>0.01*</td>
<td>0.02*</td>
</tr>
<tr>
<td>Moral Disengagement 7</td>
<td>1.89</td>
<td>1.32</td>
<td>0.98</td>
<td>0.80</td>
<td>0.87</td>
</tr>
<tr>
<td>Moral Disengagement 8</td>
<td>1.94</td>
<td>0.17</td>
<td>0.99</td>
<td>0.93</td>
<td>0.82</td>
</tr>
</tbody>
</table>

Note. *p < .05, **p < .01, P-values for % increase in MSE and Gini impurity are the scaled p-values from the rfPermute package in R. CWBs = Counterproductive Work Behaviors.
Table 7
OLS, LASSO, and RF Item Importance Ranks for Maladaptive Traits and Beliefs Items Across Variable Importance Metrics.

<table>
<thead>
<tr>
<th>Item</th>
<th>OLS β</th>
<th>LASSO β</th>
<th>Gini Impurity</th>
<th>% MSE</th>
<th>PIMP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Impulsivity 1(R)</td>
<td>25</td>
<td>N/A</td>
<td>26</td>
<td>17</td>
<td>10</td>
</tr>
<tr>
<td>Impulsivity 2(R)</td>
<td>14</td>
<td>13</td>
<td>30</td>
<td>27</td>
<td>17</td>
</tr>
<tr>
<td>Impulsivity 3(R)</td>
<td>19</td>
<td>N/A</td>
<td>32</td>
<td>35</td>
<td>30</td>
</tr>
<tr>
<td>Impulsivity 4(R)</td>
<td>2</td>
<td>7</td>
<td>36</td>
<td>36</td>
<td>32</td>
</tr>
<tr>
<td>Impulsivity 5</td>
<td>21</td>
<td>14</td>
<td>13</td>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td>Impulsivity 6</td>
<td>8</td>
<td>6</td>
<td>12</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Impulsivity 7</td>
<td>7</td>
<td>5</td>
<td>9</td>
<td>11</td>
<td>9</td>
</tr>
<tr>
<td>Impulsivity 8</td>
<td>37</td>
<td>N/A</td>
<td>11</td>
<td>31</td>
<td>21</td>
</tr>
<tr>
<td>Dark Triad 1</td>
<td>30</td>
<td>16</td>
<td>7</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Dark Triad 2</td>
<td>10</td>
<td>9</td>
<td>10</td>
<td>21</td>
<td>23</td>
</tr>
<tr>
<td>Dark Triad 3</td>
<td>26</td>
<td>18</td>
<td>23</td>
<td>24</td>
<td>18</td>
</tr>
<tr>
<td>Dark Triad 4</td>
<td>12</td>
<td>12</td>
<td>8</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>Dark Triad 5</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>Dark Triad 6</td>
<td>13</td>
<td>N/A</td>
<td>16</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>Dark Triad 7</td>
<td>22</td>
<td>15</td>
<td>6</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>Dark Triad 8</td>
<td>34</td>
<td>N/A</td>
<td>18</td>
<td>30</td>
<td>20</td>
</tr>
<tr>
<td>Dark Triad 9</td>
<td>15</td>
<td>N/A</td>
<td>35</td>
<td>23</td>
<td>31</td>
</tr>
<tr>
<td>Dark Triad 10</td>
<td>16</td>
<td>17</td>
<td>27</td>
<td>13</td>
<td>16</td>
</tr>
<tr>
<td>Dark Triad 11</td>
<td>32</td>
<td>N/A</td>
<td>29</td>
<td>20</td>
<td>24</td>
</tr>
<tr>
<td>Dark Triad 12</td>
<td>24</td>
<td>N/A</td>
<td>15</td>
<td>29</td>
<td>35</td>
</tr>
<tr>
<td>Entitlement 1</td>
<td>11</td>
<td>N/A</td>
<td>25</td>
<td>9</td>
<td>11</td>
</tr>
<tr>
<td>Entitlement 2</td>
<td>23</td>
<td>N/A</td>
<td>28</td>
<td>32</td>
<td>27</td>
</tr>
<tr>
<td>Entitlement 3</td>
<td>5</td>
<td>8</td>
<td>21</td>
<td>12</td>
<td>13</td>
</tr>
<tr>
<td>Entitlement 4</td>
<td>20</td>
<td>N/A</td>
<td>37</td>
<td>28</td>
<td>33</td>
</tr>
<tr>
<td>Entitlement 5(R)</td>
<td>9</td>
<td>10</td>
<td>19</td>
<td>26</td>
<td>25</td>
</tr>
<tr>
<td>Entitlement 6</td>
<td>31</td>
<td>N/A</td>
<td>31</td>
<td>18</td>
<td>28</td>
</tr>
<tr>
<td>Entitlement 7</td>
<td>36</td>
<td>N/A</td>
<td>17</td>
<td>16</td>
<td>19</td>
</tr>
<tr>
<td>Entitlement 8</td>
<td>33</td>
<td>N/A</td>
<td>34</td>
<td>25</td>
<td>26</td>
</tr>
<tr>
<td>Entitlement 9</td>
<td>29</td>
<td>N/A</td>
<td>33</td>
<td>33</td>
<td>37</td>
</tr>
<tr>
<td>Moral Disengagement 1</td>
<td>17</td>
<td>11</td>
<td>5</td>
<td>19</td>
<td>12</td>
</tr>
<tr>
<td>Moral Disengagement 2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Moral Disengagement 3</td>
<td>28</td>
<td>N/A</td>
<td>14</td>
<td>10</td>
<td>22</td>
</tr>
<tr>
<td>Moral Disengagement 4</td>
<td>27</td>
<td>N/A</td>
<td>24</td>
<td>37</td>
<td>34</td>
</tr>
<tr>
<td>Moral Disengagement 5</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Moral Disengagement 6</td>
<td>6</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Moral Disengagement 7</td>
<td>35</td>
<td>N/A</td>
<td>22</td>
<td>22</td>
<td>36</td>
</tr>
<tr>
<td>Moral Disengagement 8</td>
<td>18</td>
<td>N/A</td>
<td>20</td>
<td>34</td>
<td>29</td>
</tr>
</tbody>
</table>

Note. N/A under LASSO results is for items with beta weights reduced to zero.
Appendix A
Counterproductive Work Behavior Scale items

Prompt: “How often have you done each of the following things on your job(s) within the past year?”
Scale Anchors: 1 (Never), 2 (Once or twice), 3 (Once or twice per month), 4 (Once or twice per week), 5 (Everyday).

1. Purposely wasted your employer’s materials/supplies
2. Purposely damaged a piece of equipment or property
3. Purposely dirtied or littered your place of work
4. Came to work late without permission
5. Stayed home from work and said you were sick when you were not
6. Taken a longer break than you were allowed to take
7. Left work earlier than you were allowed to
8. Purposely did your work incorrectly
9. Purposely worked slowly when things needed to get done
10. Purposely failed to follow instructions
11. Stolen something belonging to your employer
12. Took supplies or tools home without permission
13. Put in to be paid for more hours than you worked
14. Took money from your employer without permission
15. Stole something belonging to someone at work
16. Told people outside the job what a lousy place you work for
17. Started or continued a damaging or harmful rumor at work
18. Been nasty or rude to a client or customer
19. Insulted someone about their job performance
20. Made fun of someone’s personal life
21. Ignored someone at work
22. Blamed someone at work for error you made
23. Started an argument with someone at work
24. Verbally abused someone at work
25. Made an obscene gesture (the Finger) to someone at work
26. Threatened someone at work with violence
27. Threatened someone at work, but not physically
28. Said something obscene to someone at work to make them feel bad
29. Did something to make someone at work look bad
30. Played a mean prank to embarrass someone at work
31. Looked at someone at work’s private mail/property without permission
32. Hit or pushed someone at work
33. Insulted or made fun of someone at work
Appendix B
The Dirty Dozen (Dark Triad)
Jonason & Webster (2010)

**Prompt:** “Please indicate how much you agree with the following statements.”

**Anchors:** 1 (Strongly Disagree), 2 (Disagree), 3 (Neither Disagree nor Agree), 4 (Agree), 5 (Strongly Agree).

1. I tend to manipulate others to get my way.
2. I have used deceit or lied to get my way.
3. I have used flattery to get my way.
4. I tend to exploit others towards my own end.
5. I tend to lack remorse.
6. I tend to be unconcerned with the morality of my actions.
7. I tend to be callous or insensitive.
8. I tend to be cynical.
9. I tend to want others to admire me.
10. I tend to want others to pay attention to me.
11. I tend to seek prestige or status.
12. I tend to expect special favors from others.
Appendix C
Shortened Barratt Impulsiveness Scale (BIS) Items
Morean, DeMartini, Leeman, Pearlson, Anticevic, Krishnan, Krystal, & O’Malley (2014)

Prompt: “Please indicate how much you agree with the following statements.”
Anchors: 1 (Strongly Disagree), 2 (Disagree), 3 (Neither Disagree nor Agree), 4 (Agree), 5 (Strongly Agree).

1. I plan tasks carefully. (R)
2. I am self-controlled. (R)
3. I concentrate easily. (R)
4. I am a careful thinker. (R)
5. I do things without thinking
6. I don’t pay attention
7. I say things without thinking
8. I act on the spur of the moment
Appendix D
Psychological Entitlement
Campbell, Bonacci, Shelton, Exline, & Bushman (2004)

Prompt: “Please indicate how much you agree with the following statements.”
Anchors: 1 (Strongly Disagree), 2 (Disagree), 3 (Neither Disagree nor Agree), 4 (Agree), 5 (Strongly Agree).

1. I honestly feel I’m just more deserving than others.
2. Great things should come to me.
3. If I were on the titanic, I would deserve to be on the first lifeboat!
4. I demand the best because I’m worth it.
5. I don’t necessarily deserve special treatment. (R)
6. I deserve more things in my life.
7. People like me deserve an extra break now and then.
8. Things should go my way.
9. I feel entitles to more of everything.
Appendix E
Moral Disengagement
Moore, Detert, Trevino, Baker, & Mayer (2012)

Prompt: “Please indicate how much you agree with the following statements.”
Anchors: 1 (Strongly Disagree), 2 (Disagree), 3 (Neither Disagree nor Agree), 4 (Agree), 5 (Strongly Agree).

1. It is okay to spread rumors to defend those you care about.
2. Taking something without the owner’s permission is okay as long as you’re just borrowing it.
3. Considering the ways people grossly misrepresent themselves, it’s hardly a sin to inflate your own credentials a bit.
4. People shouldn’t be held accountable for doing questionable things when they were just doing what an authority figure told them to do.
5. People can’t be blamed for doing things that are technically wrong when all their friends are doing it too.
6. Taking personal credit for ideas that were not your own is no big deal.
7. Some people have to be treated roughly because they lack feelings that can be hurt.
8. People who get mistreated have usually done something to bring it on themselves.
Appendix F
Demographic Questions

Prompt: “Please respond to the following demographic questions.”

1. What is your age in years?
2. What is your biological sex? (Male/Female)
3. What gender do you identify with? (Male/Female/Neither)
4. What country do you reside in?
5. How many hours do you typically work in a week?
6. Are you currently employed part-time (28 hours per week or less) or full-time (29+ hours per week)?
7. Have you worked full-time for at least one year? (Yes/No)
8. Do you have more than one employer? (Yes/No)
   a. If Yes: How many different employers do you have?
9. How long have you been with your current employer?
   a. If yes to question 8: How long have you been with your current employers? (Spaces provided for multiple entries)
10. Which of the following occupations bests describes your work?
    a. Management Occupations
    b. Business and Financial Operations Occupations
    c. Computer and Mathematical Occupations
    d. Architecture and Engineering Occupations
    e. Life, Physical, and Social Science Occupations
    f. Community and Social Service Occupations
    g. Legal Occupations
    h. Education, Training, and Library Occupations
    i. Arts, Design, Entertainment, Sports, and Media Occupations
    j. Healthcare Practitioners and Technical Occupations
    k. Healthcare Support Occupations
    l. Protective Service Occupations
    m. Food Preparation and Serving Related Occupations
    n. Building and Grounds Cleaning and Maintenance Occupations
    o. Personal Care and Service Occupations
    p. Sales and Related Occupations
    q. Office and Administrative Support Occupations
    r. Farming, Fishing, and Forestry Occupations
    s. Construction and Extraction Occupations
    t. Installation, Maintenance, and Repair Occupations
    u. Production Occupations
    v. Transportation and Material Moving Occupations
Appendix G
R programing language Code

.userdetailsed
library(dplyr)
library(tidyverse)
library(randomForest)
library(glmnet)
library(vita)
library(caret)
library(glmpath)
library(rfPermute)
library(car)
library(caret)
library(elasticnet)
library(lars)

#Data cleaning for Andrews (2019;Dissertation)
###read in data
raw_complete <- raw_data[complete.cases(raw_data[, "MTurkCode"]),]

#Data Cleaning For Study##
#fix name error
names(raw_complete)[names(raw_complete) == "CWB.2"]<-
study_vars <-
names(raw_complete) %in% c("CWB_1", "CWB_2", "CWB_3", "CWB_4",
"CWB_5", "CWB_6", "CWB_7", "CWB_8",
"CWB_9", "CWB_10", "CWB_11", "CWB_12", "CWB_13",
"CWB_14", "CWB_15", "CWB_16",
"CWB_17", "CWB_18", "CWB_19", "CWB_20", "CWB_21",
"CWB_22", "CWB_23", "CWB_24",
"CWB_25", "CWB_26", "CWB_27", "CWB_28", "CWB_29",
"CWB_30", "CWB_31", "CWB_32",
"CWB_33",
"BIS_1R", "BIS_2R", "BIS_3R", "BIS_4R", "BIS_5", "BIS_6",
"BIS_7", "BIS_8",
"DD_1", "DD_2", "DD_3", "DD_4", "DD_5", "DD_6", "DD_7",
"DD_8", "DD_9", "DD_10", "DD_11", "DD_12",
"MD_1", "MD_2", "MD_3", "MD_4", "MD_5", "MD_6", "MD_7",
"MD_8",
"PE_1", "PE_2", "PE_3", "PE_4", "PE_5R", "PE_6", "PE_7", "PE_8",
"PE_9")
clean_data <- raw_complete[study_vars]

###Reverse score items
reverse_items_names <- names(dplyr::select(clean_data, ends_with("R")))
for(i in 1:length(reverse_items_names)) {
clean_data[ ,reverse_items_names[i]] <- car::recode(clean_data[ ,reverse_items_names[i]] ,
"1= 5; 2= 4; 3= 3; 4= 2; 5= 1")
}
## Scale Computation
clean_data$CWB_scale <- rowMeans(dplyr::select(clean_data, CWB_1:CWB_33), na.rm = TRUE)
clean_data$BIS_scale <- rowMeans(dplyr::select(clean_data, BIS_1R:BIS_8), na.rm = TRUE)
clean_data$DD_scale <- rowMeans(dplyr::select(clean_data, DD_1:DD_12), na.rm = TRUE)
clean_data$PE_scale <- rowMeans(dplyr::select(clean_data, PE_1:PE_9), na.rm = TRUE)
clean_data$MD_scale <- rowMeans(dplyr::select(clean_data, MD_1:MD_8), na.rm = TRUE)

##############################################################################
## End Data Cleaning Andrews (2019; Dissertation)

##############################################################################
#### Soft-Code Reproduction Instructions

#### For automatic generalizability of code, follow procedures below

#### Assumes no missing data (clean missing prior to administration)

##### STEP 1: upload data

#### Name your total data set clean_data by entering the name of your data to the right of the <-
#### first, delete the # in front of clean_data

# clean_data <- read.csv("PATH TO YOUR DATA HERE", na.strings = c("", "NA"))

##### STEP 2 Option 1: auto-predictor selection

#### Option one automatically selects names for scales and predictors, but requires you to upload
#### pre-cleaned data sets

#### Upload a file containing only predictor SCALES of interest
#### first, delete the # in front of scales_dat

dscales_dat <- read.csv("PATH TO YOUR SCALE DATA HERE", na.strings = c("", "NA"))

#### Upload a file containing only predictor ITEMS of interest
#### first, delete the # in front of items_dat

#items_dat <- read.csv("PATH TO YOUR ITEM DATA HERE", na.strings = c("", "NA"))

#### Automatically select names scales and items
#### first, delete the # in front of items_names and scale_names

#item_names <- names(items_dat)
#scale_names <- names(scales_dat)

##### STEP 2 Option 2: manual predictor selection

#### Do not use both options 1 and 2!!!

# Use the code below to select predictors (two sets of predictors are created for this study)
# replace variables inside c() with your study predictor names

scale_names <- names(clean_data) %in% c("DD_scale", "BIS_scale", "PE_scale", "MD_scale")
item_names <- names(clean_data) %in% c("BIS_1R", "BIS_2R", "BIS_3R", "BIS_4R", "BIS_5", "BIS_6", "BIS_7", "BIS_8", "BIS_9", "BIS_10", "BIS_11", "BIS_12")
"DD_1", "DD_2", "DD_3", "DD_4", "DD_5", "DD_6", "DD_7", 
"DD_8", "DD_9", "DD_10", "DD_11", "DD_12", 
"MD_1", "MD_2", "MD_3", "MD_4", "MD_5", "MD_6", "MD_7", 
"MD_8", 
"PE_1", "PE_2", "PE_3", "PE_4", "PE_5R", "PE_6", "PE_7", "PE_8", 
"PE_9")

BEGIN R

#STEP 3: Rename your outcome variable
rename your dependent variable, new name is outcome
replace CWB_scale in the code below with the name of your DV, keep the quotations
#NOTE: outcome must be a continuous variable

#STEP 4: split data into training and test sets
#createDataPartition splits the data while preserving the proportion of categories in y

train_row_numbers <- createDataPartition(clean_data$outcome, p = .8, list = FALSE)
train_data <- clean_data[train_row_numbers,]
test_data <- clean_data[-train_row_numbers,]
x_scales_train <- train_data[scale_names]
x_items_train <- train_data[item_names]
y_train <- train_data$outcome
x_scales_test <- test_data[scale_names]
x_items_test <- test_data[item_names]
y_test <- test_data$outcome

BEGIN Soft-Code Instructions

BEGIN Data Analysis

#Scale Level Analysis

#Linear Regression

train_scales <- train_data[scale_names]
lm_train_data <- cbind(y_train, train_scales)
ols_model <- lm(lm_train_data)
summary(ols_model)

ols_model_predict <- predict(ols_model, newdata = test_data)
ols_model_predict_val <- postResample(ols_model_predict, test_data$outcome)

#LASSO

lasso_model <- train(x_scales_train, y_train,
method = "glmnet",
trControl = trainControl(  
  method = "cv",
  number = 10,
  verboseIter = TRUE),
tuneGrid = expand.grid(    
  alpha = 1,
  lambda = seq(0.001, 1.0, by = 0.001)
 )
)

## you may get a warning "there were missing values in resmapled performance measures
# this warning indicates that some lambda values produced R-squared of 0, this is normal

c <- lasso_model$finalModel, s = lasso_model$bestTune$lambda)
# train final lasso model using selected hyperparameters
lasso_final <- glmnet(as.matrix(x_scales_train), as.matrix(y_train), lambda =
  lasso_model$bestTune$lambda)

### these coefficients will be that same as the cross-validated model. because lambda is set
c <- lasso_final

### Test Set Prediction
lasso_predict <- predict(lasso_model, newdata = test_data, lambda =
  lasso_model$bestTune$lambda)
lasso_predict_val <- postResample(lasso_predict, test_data$outcome)

### Random Forest
### Hyperparameter Optimization
set.seed(123)
rf_model <- train(x_scales_train, y_train,
  method = "rf",
  tuneLength = length(x_scales_train),
  ntree = 100,
  importance = TRUE,
  trControl = trainControl(    
    method = "cv",
    number = 10,
    verboseIter = TRUE)
)

### variable importance
randomForest::importance(rf_model$finalModel)

set.seed(123)
rf_model_permute <- rfPermute(x_scales_train, y_train, nrep = 100, ntree = 100, mtry =
  rf_model$finalModel$mtry)
rf_model_permute$pval
rf_model_pimp <- PIMP(x_scales_train, y_train, rf_model$finalModel, S = 100, seed = 123)
rp_model_pimp_test <- PimpTest(rf_model_pimp)
rp_model_pimp_test$pvalue
Train Final Model

```r
set.seed(123)
rf_final <- randomForest(x_scales_train, y_train, mtry = rf_model$finalModel$mtry, ntree = 100)
```

Test Set Prediction

```r
rf_predict <- predict(rf_final, newdata = test_data)
rf_predict_val <- postResample(rf_predict, test_data$outcome)
```

Item Level Analysis

```r
# auto-data wrangle
train_items <- train_data[item_names]
godzilla_train_data <- cbind(y_train, train_items)

Linear Model (Godzilla)
godzilla <- lm(godzilla_train_data)

Test Set Prediction
godzilla_predict <- predict(godzilla, newdata = test_data)
godzilla_predict_val <- postResample(godzilla_predict, test_data$outcome)

#coefficients
summary(godzilla)
```

LASSO

```r
set.seed(123)
mecha_godzilla <- train(x_items_train, y_train,
    method = "glmnet",
    trControl = trainControl(
        method = "cv",
        number = 10,
        verboseIter = TRUE),
    tuneGrid = expand.grid(
        alpha = 1,
        lambda = seq(0.001, 1.0, by = 0.001)
    )
)

#you may get a warning "there were missing values in resmapled performance measures
#this warning indicates that some lambda values produced R-squared of 0, this is normal

#Coefficients
coeff(mecha_godzilla$finalModel, s = mecha_godzilla$bestTune$lambda)

#train final lasso model using selected hyperparameters
mecha_godzilla_final <- glmnet(as.matrix(x_scales_train), as.matrix(y_train), lambda = mecha_godzilla$bestTune$lambda)

#these coefficients will be that same as the cross-validated model. because lambda is set
coeff(mecha_godzilla_final)

Test Set Prediction

mecha_godzilla_predict <- predict(mecha_godzilla, newdata = test_data, lambda = mecha_godzilla$bestTune$lambda)
mecha_godzilla_predict_val <- postResample(mecha_godzilla_predict, test_data$outcome)

###############################Random Forest
set.seed(123)
mecha_king_ghidorah <- train(x_items_train, y_train,
    method = "rf",
    tuneLength = length(x_items_train),
    ntree = 100,
    importance = TRUE,
    trControl = trainControl(
        method = "cv", number = 10,
        verboseIter = TRUE)
)

mecha_king_ghidorah$finalModel

###variable importance
imp <- randomForest::importance(mecha_king_ghidorah$finalModel)
write.csv(imp, "RF_Items_VarImp.csv")

###obtain p-values for variable importance statistics
set.seed(123)
mecha_king_ghidorah_permute <- rfPermute(x_items_train, y_train, nrep = 100, ntree = 100,
    mtry = mecha_king_ghidorah$finalModel$mtry)

mecha_king_ghidorah_permute$pval

##Run Altmann (2010) Permutation Importance (PIMP)
mecha_king_ghidorah_pimp <- PIMP(x_items_train, y_train,
    mecha_king_ghidorah$finalModel, S = 100, seed = 123)
mecha_king_ghidorah_pimp_test <- PimpTest(mecha_king_ghidorah_pimp)

mecha_king_ghidorah_pimp_test$pvalue

###############################Train Final Model
set.seed(123)
mecha_king_ghidorah_final <- randomForest(x_items_train, y_train, mtry =
    mecha_king_ghidorah$finalModel$mtry, ntree = 100)

#############################Test Set Prediction
mecha_king_ghidorah_predict <- predict(mecha_king_ghidorah_final, newdata = test_data)
mecha_king_ghidorah_predict_val <- postResample(mecha_king_ghidorah_predict, test_data$outcome)

############################RF Variable Reduction Model
#This code identifies items with significant Altmann (2010) p-values
#then creates new datasets with just significant items
sig_items <- pimp_pvals[pimp_pvals <= .05,]
sig_items_names <- names(sig_items)
x_items_train_reduced <- train_data[sig_items_names]
x_items_test_reduced <- test_data[sig_items_names]

###re-train hyperparameter
##In study example, model cannot be hyperparameterized because
#mtry is set to the only possible value of 2
```r
set.seed(123)
mecha_king_ghidorah_reduced <- train(x_items_train_reduced, y_train,
    method = "rf",
    tuneLength = length(x_items_train_reduced),
    ntree = 100,
    importance = TRUE,
    trControl = trainControl(
        method = "cv", number = 10,
        verboseIter = TRUE)
)
#####retrain
set.seed(123)
mecha_king_ghidorah_reduced_final <- randomForest(x_items_train, y_train, mtry =
    mecha_king_ghidorah_reduced$finalModel$mtry, ntree = 100)
############################Test Set Prediction
mecha_king_ghidorah_reduced_predict <- predict(mecha_king_ghidorah_reduced_final,
    newdata = test_data)
mecha_king_ghidorah_reduced_predict_val <-
    postResample(mecha_king_ghidorah_reduced_predict, test_data$outcome)

#End R Script
```