COMPARING THE ACCURACY OF DECISION TREES AND LOGISTIC REGRESSION IN PERSONNEL SELECTION

by
Kyle Marks

A Thesis Submitted in Partial Fulfillment

of the Requirements for the Degree of

Master of Arts in Industrial and Organizational Psychology

Middle Tennessee State University
May, 2018

Committee Chair:

Dr. Alexander T. Jackson

Thesis Committee:

Dr. Ying Jin

Dr. Judith Van Hein

ABSTRACT

Being able to make better personnel decisions is a problem that many organizations consider. Actuarial methods have been shown to make more accurate decisions than human decision making. This study examines the performance of two actuarial methods. (1) The decision tree method, a fast and frugal approach to decision making that has been shown to be equally as accurate as other actuarial models in making decisions. As well as, (2) logistic regression a decision aid that has been often used in selection assisting in making selection decisions. Study one investigates the accuracy of each method using a simulated data set where performance is known. Study two examined how these methods performed when predicting acceptance to a graduate school program. Study one and two found that the decision tree method was equally as accurate as logistic regression in both scenarios.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE:

INTRODUCTION

Decisions are an everyday part of human life and have wide ranging consequences, both good and bad. Finding ways to improve the decision-making process can be useful in procuring a more favorable outcome. There are numerous different tools that can be used in aiding the decision-making process, such as taking advice (Hütter & Ache, 2016) quantitative aids (Diab, Pui, Yankelevich, & Highhouse, 2011; Highhouse, 2008), decision aids (Diab et al., 2011; Highhouse, 2008) and heuristics. (Artinger, Petersen, Gigerenzer, & Weibler, 2015). Each of these tools differs in how they improve the decision-making process. For instance, quantitative aids use data to support decision making, decision aids provide information on available choices, and heuristics are mental shortcuts that reduce effort in decision making (Artinger et al., 2015; Dawes, 1979). Multiple regression, logistic regression, neural networking, and decision trees are some of the quantitative approaches that improve decision-making processes by increasing the accuracy of decisions (Coussement, Van den Bossche, & De Bock, 2014; Stauffer & Ree, 1996). Of these approaches, decision trees are gaining interest by researchers (Coussement et al., 2014; Liu et al., 2011; Sinha & May, 2004; Stewart, Tuerk, Metzger, Davidson, & Young, 2016).

Purpose

The current study aims to compare the accuracy of decision trees to logistic regression in two personnel selection contexts. First, a simulated environment will be created to mimic applicants to an organization. For each applicant, there will be a cognitive ability score, conscientiousness rating, and a structured interview score. Job performance will be simulated as a function of the simulated scores. Additionally, different selection ratios will be applied to the simulated data to mimic how organizations select applicants and to determine whether the selection ratio has an impact on the accuracy of each analytic approach.

A second purpose of this study is to examine whether the decision strategies used by decision makers in a real selection context (graduate school admission decisions) reflect the strategies that the decision makers *should* be using. For each graduate school applicant, the predictors of undergraduate grade point average (GPA) and graduate record examination (GRE) scores will be used. To measure performance, final graduate GPA will be used. The performance data will be used to determine the accuracy of the decision tool in a real selection context. Additionally, there will be information on the decision of whether the applicant was admitted to the program. Whether an applicant was admitted will be matched to the prediction made by the regression analysis and the decision tree analysis to determine if what was done matches what the regression and decision tree models. This is meant to determine which analysis more closely matches how individuals *actually* make decisions.

Personnel Selection

Personnel selection is one of the primary responsibilities of human resources management. The Society for Human Resource Management (SHRM) and O*Net list selection as a primary duty for human resource managers and human resource professionals. Selection is the process employers use to make decisions about choosing the best candidate to hire from a group of applicants (Farr & Tippins, 2010). Currently, there are several methods that are commonly used in organizations to make selection decisions. For instance, intuition, heuristics, and regression have been used as selection techniques (Farr & Tippins, 2010; Miles & Sadler-Smith, 2011; Stauffer & Ree, 1996). Intuition refers to the automatic, associative, holistic nonverbal, and rapid process of decision making (Betsch, 2008). Additionally, intuition is described as a gut reaction that is based upon experience and is commonly used by hiring managers (Highhouse, 2008; Miles & Sadler-Smith, 2011). Conceptually similar to intuition is clinical judgment, which refers to an individual's judgment of an applicant using his or her own experiences and knowledge to inform the decision (Miles & Sadler-Smith, 2011). Despite the frequency with which intuition and clinical judgments are used, they are less reliable and accurate than quantitative methods (Dawes, 1971).

Dawes (1971) provided evidence that supports the notion that actuarial methods of selection, such as regression, outperform clinical judgments in nearly every situation. Additionally, Dawes (1979) found that improper linear models also outperform clinical judgments. In other words, even when the decision weights for each predictor are obtained through a sub-optimal method, such as unit weighting (equally distributing weights across predictors), the model produces better judgments and decisions than clinical judgments. Several reasons have been suggested for why clinical judgment is inferior to modeling techniques. When clinical judgments are used, weights are often assigned to predictors based on the individual's experiences, biases, and even moods (Dawes, 1979). This means that clinical judgments are prone to error and inconsistency. Furthermore, people fail to consistently apply the same weights to each applicant every time. For example, an individual might weigh the interview as the most important predictor for one applicant and a personality test as the top predictor for another applicant (Dawes, 1971). This leads to increased error. In contrast, statistical models apply the same weighting of the predictors for every candidate every time. This consistent application of weights to each candidate is mathematically designed to minimize error (Dawes, 1971). Thus, research evidence demonstrates that actuarial (quantitative) methods of decision making represent a far more reliable and accurate method for selection than simply using clinical judgments or intuition.

Despite the clear evidence for the superiority of actuarial methods, intuition and clinical judgments remain as a common method in many organizations (Miles & Sadler-Smith, 2011). One possible reason for this could be the common notion that a manager may think, “I will know it when I see it” (Highhouse, 2008). In other words, managers will rely on experience and their gut to make selection decisions. Another possible reason for the prevalence of clinical judgment and intuition is that actuarial methods have been said to dehumanize applicants (Dawes, 1979). This insinuates that applicants are merely viewed as numbers rather than individuals. Dawes (1979) argues that this statement is misguided. Rather, the argument for whether numbers are dehumanizing comes down to fairness. For example, selection predictors for graduate school such as GPA and GRE are meant to predict student performance. GPA is a collection of behavior representing approximately four (sometimes more) years of an applicant’s life. This numerical value tells a story of the applicant’s behavior better than any interview (Dawes, 1979). Additionally, Dawes (1979) argues that it is unethical to think that an individual’s judgment can be fairer than a test that has been shown to be both reliable and valid.

Although intuition represents an issue in selection, actuarial methods provide a viable solution.

Two examples of actuarial methods superior to intuition in selection are logistic regression and regression (Raju, Pappas, & Williams, 1989; Stauffer & Ree, 1996). Both methods can aid in the decision-making process of organizations. These methods differ in that linear regression predicts a continuous outcome, such as job performance, and logistic regression predicts categorical outcomes, such as hire or do not hire. These methods operate through the use of models that utilize weighing different predictors accurately and consistently across applicants. These quantitative methods are resistant to extraneous factors that influence the predictors such as bias, therefore reducing error in decision-making. However, not all predictors are equal, and some represent more reliable and valid options when predicting performance (Barrick, Mount, & Judge, 2001; Schmidt & Hunter, 1998; Schmidt & Hunter, 2004).

Predictors in Personnel Selection Predictors of performance are essential in selection regardless of the method used. Predictors are used to identify applicants who will perform best within the organization (Dawes, 1979). Schmidt and Hunter (2004) presented evidence that general mental ability (GMA) is the strongest predictor of performance on the job. GMA (i.e., cognitive ability or intelligence) is the aptitude to solve problems, learn quickly, and think abstractly (Gottfredson, 1998). The ability to solve problems and learn quickly are two integral skills in most occupations. As such, GMA is the most highly correlated predictor of performance on the job. Additionally, GMA is applicable across a variety of different positions and industries (Schmidt & Hunter, 2004). For instance, GRE scores are used as a measure of GMA (Littlepage, Bragg, & Rust, 1978) to predict performance in graduate school and make admission decisions (Kuncel, Ones, & Hezlett, 2001).

In addition to cognitive ability, specific personality characteristics have been shown to predict job performance for a wide variety of occupations. For instance, Barrick et. al (2001) examined the relationships between the big five personality traits (openness, conscientiousness, extraversion, agreeableness, and emotional stability) and job performance. The authors found that among all big five-personality traits, conscientiousness had the highest operational validity when predicting job performance (ρ=.31). Other personality traits also presented operational validity in predicting performance, but none predicted job performance as well as conscientiousness (.07 for openness to experience, .13 for extraversion, .13 for agreeableness, and .13 for emotional stability). Similarly, Barrick and Zimmerman (2009) examined the use of bio data, personality traits and attitudinal scales as predictors of performance. The authors found that conscientiousness and emotional stability (the opposite of neuroticism) were among the best predictors of performance and that the remaining big five personality traits added little incremental validity.

Conscientiousness is also used to predict performance in graduate admissions. Undergraduate GPA is often used as a partial predictor of conscientiousness. This is because it represents a students work ethic over the course of the individual's undergraduate study (Cheng & Ickes, 2009). Overall those who had a higher GPA in college displayed greater levels of conscientiousness or self-motivation (Cheng & Ickes, 2009). GPA is often standardized on a four-point scale making it easy to compare between students. Additionally, GPA is a universal measure of performance in undergraduate studies making it easy to collect for selection use. GRE is also a standardized measure where all who take the exam are scored on the same scale. This allows graduate admissions to easily measure both conscientiousness through GPA and GMA through the GRE (Kuncel et al., 2001).

The work sample is another strong predictor of performance in organizations (Schmidt & Hunter, 1998). Work sample tests are assessments modeled after actual work performed on the position. For work sample tests to be reliable and valid, it is necessary to assess tasks that are essential to the main function of the position (Bobko, Roth, & Buster, 2005). For a work sample test to be used, the applicant must already know how to do the job or have been trained in the occupation (Schmidt & Hunter, 1998). An example of a work sample test is having a mechanic work on an engine or having an applicant respond to different emails. Schmidt and Hunter (1998) conclude that the work sample test has a predictive validity of .54, slightly higher than the predictive validity of GMA.

The job interview is one of the most widely used selection tools, second only to the application blank. Interestingly, the validity of the interview depends largely on the structure of the interview (Hough & Oswald, 2000). Unstructured interviews have less predictive validity (ρ=.38) than structured interviews (ρ=.51) (Schmidt & Hunter, 1998). This is in part because structured interviews are standardized, meaning that each applicant is asked the same questions. This allows employers to more consistently measure desired constructs such as conscientiousness (Schmidt & Rader 1999; van der Zee, Bakker, & Bakker, 2002). Additionally, because structured interviews are standardized, the interview can be scored consistently across applicants (van der Zee et al., 2002). Unstructured interviews can either follow a loose outline with room to ask additional questions or can follow no format and change from applicant to applicant. This can be ineffective in measuring desired constructs across different applicants. Part of the reason for the ineffectiveness of unstructured interviews is the lack of consistency in the questions asked. This inconsistency leads to difficulty in scoring, measuring appropriate constructs, and comparing applicants (Schmidt et al., 1999). Constructs commonly measured by the interview are GMA, work motivation, and interpersonal skills (Huffcutt & Arthur, 1994). The structured interview is developed around essential functions of the position. Once these functions are identified, questions are formed to assess whether the applicant has experience in the tasks required by the position (Schmidt et al., 1999).

Several meta-analyses have been conducted examining the relationships between various predictors and job performance. Across these meta-analyses, the findings clearly indicate that GMA, conscientiousness, work samples, and structured interviews are the best predictors of performance (Barrick et al., 2001; Barrick & Zimmerman, 2009; Schmidt & Hunter, 1998; Schmidt & Hunter, 2004). Furthermore, the validity of these predictors generalizes to a variety of different positions and industries.

Selection Techniques

A common selection technique is the multiple hurdle approach. This process has several stages of assessing applicants on different predictors. For example, the first stage may be a cognitive ability test. For applicants to move on to the next stage they must score above a set cutoff score, if the cutoff score is met then the applicant would move to another stage such as an interview. Applicants who do not reach the cutoff score are dropped from the application process. The number of stages or hurdles depends on the organization but the goal remains the same: identifying the top candidates. For these candidates, there will be overall scores based on the predictors associated with the different stages (Finch, Edwards, & Wallace, 2009). Next, the organization can choose the highest overall score of the candidates or use discretion and choose the applicant they think best fits the organization. This approach is mostly objective with some room for judgment at the final selection.

As mentioned previously, logistic regression and linear regression are two common quantitative methods used in personnel selection. Logistic regression and regression operate similarly, the difference being that the outcome of logistic regression is a dichotomous outcome, such as the choice to hire or not hire an applicant (James, Witten, Hastie, & Tibishirani, 2013). Linear regression analysis uses continuous criteria. In selection, linear regression is meant to predict future performance. From the predicted performance, an organization can decide on a cutoff score for predicted performance that must be reached to hire applicants. This means that the applicant must score high enough on the predicted outcome to be hired, and those under the cutoff score will not be hired. As mentioned above, logistic regression is used to produce a dichotomous outcome (Stauffer & Ree, 1996). The logistic regression will be based upon the predictors that were chosen by the organization through job analysis, a process in which the key responsibilities, knowledge, skills, and abilities needed to perform the job are identified. Then according to the scores that the applicant provides, the logistic regression will predict an outcome of either “hire” or “don’t hire.” These methods have been shown to be more effective in selection of applicants to graduate school than clinical methods

(Dawes, 1971). Additionally, Staufer and Ree (1996) provide evidence that regression and logistic regression perform identically in the selection of individuals into pilot school.

Logistic regression and regression perform similarly as techniques for selection. Another common selection process is having applicants go through multiple hurdles. This necessitates that an applicant passes one test in order to be considered for the next. This is different from regression techniques in that these methods combine scores across all selection tests. This is considered a compensatory method whereas multiple hurdles is not a compensatory method (Farr & Tippins, 2010). Both methods operate similarly, transforming predictor scores into either predicted performance in the case of regression or “hire”/ “do not hire” in logistic regression. Both can be robust to error in data and changes in selection rate (Stauffer & Ree, 1996). However, there are several weaknesses of regression and logistic regression. First, these techniques can be difficult to explain because the weights for the predictors may seem random and have little meaning to someone without a statistical background. Additionally, neither regression technique allows for the use of qualitative predictors (predictors must be quantitative). An alternative actuarial method that compensates for some of these weaknesses is the decision tree.

Decision Trees

Decision trees are considered a fast and frugal approach to decision making (Raab & Gigerenzer, 2015), meaning that they are quick to use with few steps involved in interpretation. The basic structure of a decision tree begins with the root node. This is the best predictor of an outcome. For example, GMA as a predictor of job performance in personnel selection. All applicants enter the decision tree model through the root node. From the root node, the tree splits into at least two or more child nodes. The splits in a decision tree are made from the classification error rate. The classification error rate is "The fraction of the training observation in that region that do not belong to the most common class"(James et al., 2013). This is a measure of purity for the node. This means that the purest node would be when all observations fit into a class. The least pure is if half of the cases fit and half do not. This ratio is what determines the splits in the decision tree.

Equation 1 for creating decision tree.

$$\mathrm{E} = 1 - \max(\widehat{\mathrm{P}}\mathrm{mk})$$

The decision maker follows along the proceeding nodes based upon the applicant's value for the preceding independent variable (cognitive ability). The tree makes splits to subsequent nodes based on which predictor can best separate applicants into who should be hired and who should not be hired. For example, the node could split based into those who score above 50% on a GMA test and those who score 50% or below on a GMA test. The decision tree then branches to the next best predictors of job performance based upon the applicant's cognitive ability score. There can be one branch from cognitive ability or many, with the different branches representing different predictors of job performance. For example, for those with high cognitive ability, the next best predictor of job performance may be conscientiousness. For those with lower scores on cognitive ability, work experience may be the next predictor. Therefore, cognitive ability has two branches that lead to two different predictors (nodes) work experience and conscientiousness. The decision tree continues to branch from the previous predictors to the next best predictors. How one follows the decision tree is dependent upon where the applicant scores on the previous predictor (high cognitive ability to conscientiousness). This will continue until a terminal node is reached indicating that the decision should either be hire *or* do not hire. These final nodes are called terminal nodes. Figure 1 displays an example tree with the identifying characteristics labeled. Figure 2 displays a decision tree using a personnel selection example.

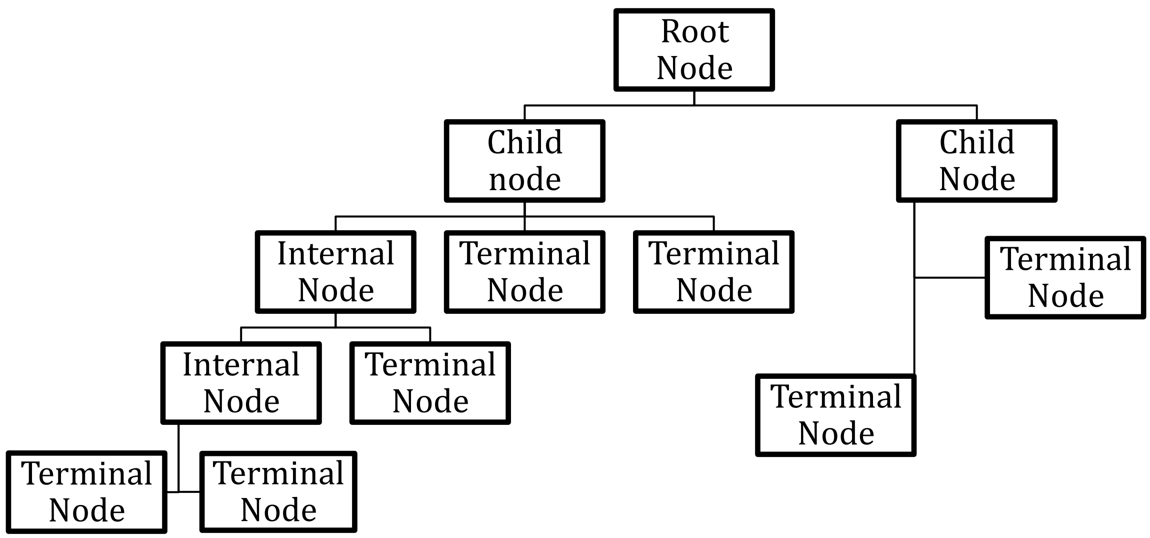

Figure 1. Decision tree with sections labeled.

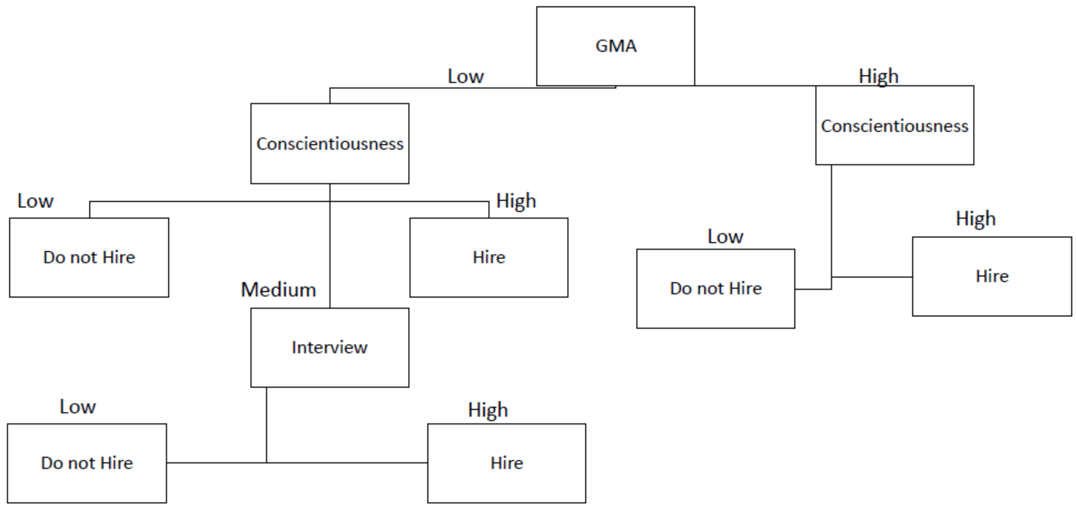

Figure 2. Example decision tree for personnel selection.

Decision trees can be applied to numerous types of decisions. Various industries use decision trees for making decisions. For instance, decision trees have been used in identifying illnesses in medicine (Liu et al., 2011; Nakayama et al., 2012; Raab & Gigerenzer, 2015), determining financial risk in the finance industry (Sinha & May, 2004), and more quickly and accurately diagnosing post-traumatic stress disorder (PTSD) (Stewart et al., 2016). Additionally, decision trees have emerged as a useful tool for businesses. Coussement, Van den Bossche, and De Bock (2014) compared decision trees to logistic regression and the recency, frequency, and monetary value analysis (RFM) commonly used in marketing. They found that decision trees were the most accurate method in categorizing customers into groups based on their purchasing behavior. Additionally, they found that decision trees were more robust against errors (missing data that was then randomly replaced) in the data. Research spanning multiple industries clearly demonstrates that decision trees can be a useful tool in aiding decision-making.

The success of decision trees in other fields suggests that the method can be applied successfully in other decision-making situations. The increased use and research interest in decision trees is in part due to several advantages of decision trees over other quantitative methods. First, decision trees are more easily understood by users of the trees (James et al., 2013). Other quantitative methods, such as regression, require some statistical knowledge to interpret, while one only needs to be able to follow a diagram to use decision trees (James et al., 2013; Stewart et al., 2016). This contrast in ease of use and interpretation is in large part due to the visual representation of the decision process that the tree provides. Second, decision trees can predict both qualitative and quantitative responses and more easily incorporate qualitative variables as predictors (James et al., 2013). For instance, a score on a cognitive ability test (quantitative) or what school an applicant graduated from (qualitative) can both be used as predictors in a decision tree. In contrast, regression only predicts quantitative outcomes and requires extra steps in order to properly handle qualitative variables, such as creating dummy variables. In some cases, regression uses qualitative variables, but they must first be assigned a numerical value. Decision trees do not have these quantitative restrictions. Third, decision trees better mirror actual human decision making than other methods (Artinger et al., 2015). Decision trees are based upon contingencies set in place by the nodes of the tree. Subsequently moving from node to node based upon how the situation at hand meets the criteria of the various predictor nodes. This method is like a searching rule in human decision-making in which one identifies factors that must be satisfied to select a decision (Artinger et al., 2015). After it has been determined whether the situation meets the series of predictors, the decision tree ends in terminal nodes that identify a decision.

This is similar to a stopping rule in human decision making. A stopping rule is when an individual has reached a conclusion based on identified factors, thus making a decision (Artinger et al., 2015). Because decision trees offer advantages in terms of interpretability and ease of use without sacrificing accuracy, decision trees may be a promising tool for more accurate decision-making in other applied contexts, such as human resource management and personnel selection.

Hypothesis 1: Decision tree analysis will be as accurate as logistic regression in predicting performance outcomes in the simulated environment. Accuracy of decision trees will not be significantly different from that of regression.

Hypothesis 2: Decision tree analysis will be as accurate as logistic regression in predicting acceptance into a graduate school program. Accuracy of decision trees will not be significantly different from that of regression

Selection Ratio

Predictor based selection occurs when weights are assigned to variables that are meant to predict the success of an applicant on the job based on some criterion, usually job performance (De Corte, 1999). The success ratio or hit rate is a ratio of the number of applicants hired that meet an organizationally set cutoff score for performance (i.e., good performers) to the total number of applicants who were hired (De Corte, 1999). The selection ratio an organization sets can have an impact on the hit rate of the selection process. The selection ratio is a ratio of the number of applicants hired to the total number of applicants (De Corte, 1999).

There are three commonly used selection ratios: fixed quota, threshold selection, and mixed quota/threshold decision (De Corte, 1999, 2002). Fixed quota refers to having a set number of positions available and only selecting that number from the available applicants. Threshold selection requires that anyone who scores above a cut score on the selected predictor variables will be hired into the organization. This could potentially mean that all applicants are selected or that none of the applicants are selected. Lastly, mixed quota/threshold decision means that there is a set number of positions available and that applicants are only selected for the available positions if the applicant meets a predetermined cut off score (De Corte, 1999, 2002). Depending on the selection ratio method chosen, the hit ratio can be overstated. This is due to the formula for success ratio not accounting for various selection ratios (De Corte, 1999, 2002). Based upon the material presented above the following hypothesis is presented.

Hypothesis 3: More stringent selection ratios will reduce the ability to accurately identify the top performing applicants for all methods.

Hypothesis 4: Selection ratios will interact with decision technique such that decision trees will be more robust to changes in selection ratio than logistic regression.

CHAPTER TWO:

STUDY ONE

Method

The purpose of the first study was to compare the prediction accuracy of logistic regression and the decision tree method using cross validation techniques in a simulated dataset. First, a simulated population was created containing 500,000 job applicants with predictor scores for: conscientiousness, GMA, and structured interview. Job performance was used as an outcome variable. Job performance scores were generated based on Equation 1 displayed below. Three predictors (GMA, conscientiousness, and structured interview scores) were included in Equation 1 to predict job performance scores. Performance scores will be normally distributed. According to (Barrick et al., 2001; Schmidt & Hunter, 1998), these predictors were frequently observed in previous studies as significant and valid predictors of job performance. Regression coefficients of the predictors in Equation 1 were extracted from previous empirical studies (Cortina, Goldstein, Payne, Davison, & Gilliland, 2000). Scores of the three predictors were generated from the standard normal distribution which means that the population had a mean of 0 and standard deviation of 1. Data were generated using R (Therneau, Atkinson, Ripley, & Ripley, 2015). See Appendix A for the R code used to set up the simulated population.

$$Perf\widehat{\widehat{orm}}ance = \beta_0 + .43 * GMA + .24 * Consientiousness + .58 * structuredinteview + Error \tag{1}$$

A sample size of 250 applicants will be taken from the simulated population. This sample size was chosen because it represents the average number of applicants to corporate positions (Joyce, 2016). The first sample of 250 is used to create both a decision tree model and logistic regression model. The models that result from this first sample can then be applied to other samples to assess classification accuracy. This process is known as training and allows for models to be created on one data set then tested on other samples of data. Next, the selection ratio will be manipulated to determine the number of applicants that will be selected from the testing sample. The first selection ratio, fixed quota will have a .10 selection ratio meaning that the top ten percent of the applicants with the highest predicted job performance will be selected. Next, fixed threshold will set a cutoff score for performance of .69 (third quartile of the population) and any applicants above the predicted performance score are hired. Lastly, quota/threshold where of those above .69 only those in the top 10 percent of applicants are selected. These selection ratios are similar to what would be used in an organization (De Corte, 1999, 2002). See Appendix B for the R code for selection ratios. This creates three separate conditions. In each condition, both logistic regression and decision tree models created from the training data are applied to samples under the different selection ratio conditions. The decision tree method and logistic regression will then classify applicants as either 1 (*hire*) or 0 (*do not hire*). This allows comparison of the classification accuracy between the two approaches. Appendix C displays the R code used for the decision tree, and Appendix D displays the R code used for the logistic regression.

Results

This study sought to evaluate the classification accuracy of the decision tree method and logistic regression in correctly identifying applicant's performance across various selection ratios. This was performed repeatedly through a loop created in R (see Appendices C and D). One hundred random samples of 250 applicants were drawn from the simulated population with replacement. For each sample, a cutoff score was applied based on the selection ratio (See Appendix B for the R code for cutoff scores). The cutoff was used to categorize applicants as either "hire" or "do not hire" based on the simulated performance of that applicant. In order to apply the decision tree method and logistic regression, equations were created using a training sample from the population. This training sample was used to create the statistical models to be tested. The developed models were then applied to the 100 different samples. For each sample, the classification accuracies were averaged to create an overall classification across each condition. A 2 (decision tree vs. logistic regression) x 3 (ratio, threshold, mixed model) repeated measures ANOVA was conducted to determine whether the decision tree method and the logistic regression method significantly differed on ability to accurately classify applicants across the different selection ratios. A familywise alpha of .05 was used for analyses. Method used (decision tree and logistic regression) and selection ratio (fixed quota, threshold, and mixed) were used to predict classification accuracy (hire, do not hire).

The mean classification accuracy of each condition across the 100 trials was used for analysis (see Table1). Hypothesis 1 stated that the decision tree method will *not* be significantly different than logistic regression in accurately predicting performance of applicants. The average overall classification accuracy for the decision tree method (M = 74%, SD = 0.02) was slightly lower than the average overall classification accuracy for the logistic regression method (M = 76%, SD = 0.03). See Figure 3 for the decision tree.

In this tree, the root node is the interview score, if the score is above 0.068 then the user would move to GMA scores. If above 0.41 for GMA the user would move to look at the interview score again, if below then the user would examine the applicants conscientiousness score. This would continue until the end of the tree which would indicate "hire" or "do not hire". However, there was not a significant main effect of analytic method ($F(1, 5) = 2.95$, $p = .23$, $\eta^2 = .05$), meaning that although there was a slight difference, it was not enough to state that one method outperformed the other (See Figure 3 for the decision tree that resulted). Therefore, hypothesis 1 was supported (see Table 2).

Table 1: Means for Classification Percent Accuracy Based on Method and Selection Ratio

	Fixed Quota	**SD**	**Threshold**	**SD**	**Mixed**	**SD**	**Average**
Decision Tree	82%	0.02	73%	0.02	66%	0.03	74%
Logistic Regression	84%	0.02	77%	0.03	66%	0.03	76%
Average	83%		75%		66%		

Table 2: Pairwise Comparison Between Selection Ratios

	Fixed Quota	**Threshold**	**Mixed**
Fixed Quota	-	.02	.005
Threshold	.02	-	
Mixed	.005		-

**p* Values significant at .05.

Hypothesis 3 stated that more stringent selection ratios will reduce the ability to accurately identify the top performing applicants. There was a significant main effect of selection ratio on classification accuracy ($F(1, 5) = 179.10$, $p < .01$, *partial* $\eta^2 = .18$). Pairwise comparisons revealed that the average overall classification for fixed quota selection ratio (M = 0.83, SD = 0.03) had significantly higher classification accuracy than the threshold selection ratio (M = 0.75, SD = 0.03, $t(185) = 29.30$, p >.01) and the mixed model selection ratio (M = 0.66, SD = 0.09, $t(185) = 57.54$, $p > .01$). Furthermore, the threshold selection ratio had significantly higher classification accuracy than the mixed model, $t(197) = 26.11$, $p < .01$. This indicates that selection ratio method is a significant predictor of classification accuracy and that more selective types of selection ratios (among these three ratios) diminish overall accuracy. Therefore, hypothesis 3 was supported.

Hypothesis 4 states that selection ratio will interact with decision technique such that decision trees will be more robust to changes in selection ratio than logistic regression. This hypothesis was tested using a 2x3 repeated measures ANOVA with classification accuracy as the dependent variable (see Table 3). Table 1 displays the means and standard deviations for each of the conditions. There was not a significant interaction between decision technique and selection ratio method on classification accuracy, ($F(2, 4) = 0.11$, $p = .77$, η^2 ~~=~~ $\leq$ .01). This indicates that changes in selection ratio have little to no impact on the overall accuracy of either the decision tree method or logistic regression. Hypothesis 4 was *not* supported by the data.

Table 3: ANOVA for Study One

Source	*df*	*F*	*p*	η^2
Selection	5	179.10	.01	.18
Method	5	2.95	.22	.05
Selection*Method	4	0.11	.77	<.01

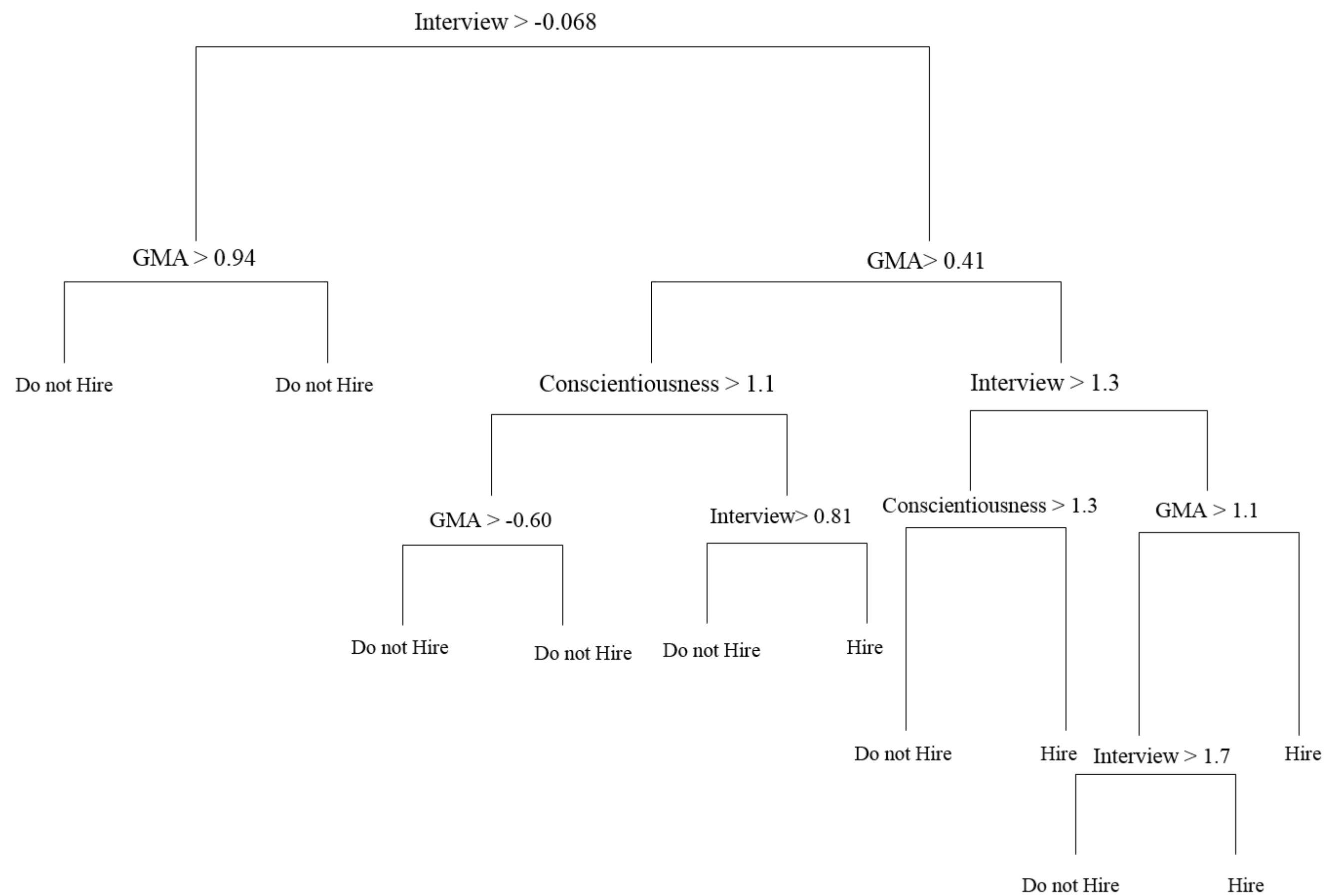

Figure 3. Example Decision tree from study one. “0” represents do not hire where as “1” represents hire.

Discussion

The first part of the analysis was meant to determine whether the decision tree method was significantly different from the logistic regression method. The analysis indicated that there were no significant differences between the methods. This supports the initial hypothesis that both methods are equally accurate in classifying applicant's performance. These findings suggest that the decision tree method may be used in place of logistic regression. This has several benefits for organizations, the most important of which is interpretability. Decision trees can be much easier for hiring managers to use and understand due to the visual representation of the decision tree (James et. al, 2013). The ease of use creates the potential for a more helpful selection tool for hiring managers who lack a statistical background. If the decision tree method is favored by managers above methods like regression and logistic regression, this would be important for organizations because it would move decision makers away from using clinical judgements when making hiring decisions and toward using an actuarial method.

This study also applied different selection ratios to the data in order to mimic different organizational settings. For example, in a hospital there are many different nurse positions, while there may be far fewer accountants. Additionally, different positions may require separate cutoff scores; some positions may choose more strict thresholds while others much more lenient. Selection ratio depends on both the organization and the position. It was hypothesized that more selective types of selection ratios would decrease the classification accuracy of both logistic regression and the decision tree model. The pairwise comparisons for the main effect of selection ratio showed that classification accuracy was greater under the quota (top ten percent) and threshold selection ratios (applicants above .69) than under the mixed model (a combination of the other two selection ratios). This supports the hypothesis because the mixed model had the most selective selection ratio. This means that the more lenient type of selection ratio of the three produced the most accurate prediction for both models. However, this does not mean that lowering the selection ratio will result in better candidates. Although the least selective of the selection ratios in this study produced the best accuracy, continually lowering selection ratio does not increase performance. Too selective of a selection ratio type will result in missing potential good employees, while selection ratios that are less restrictive will result in some poor performing employees being hired. This means that it is important to find a selection ratio that maximizes accuracy while not being too selective or lenient to result in poor hires. In conclusion, in this study the most lenient selection ratio produced the most accurate results from each model. However, in practice lowering the selection ratio or changing selection ratio type to one that is less selective does not translate to better hiring but instead increases the amount of poor performers that will be considered for employment.

This study has several limitations. The data used was simulated, and although this gives an idea of how these models perform under similar conditions, the data lacks the fidelity that real applicant data provides. For example, a limitation of the dataset is a lack of error. This deficiency makes it difficult to estimate how these models would function given errors in the data, such as missing data and incorrect values (negative numbers when there should only be positive). This limitation will be addressed in study two in which real applicant data from a master's in industrial/organizational psychology graduate program is used to compare the decision tree and logistic regression techniques.

CHAPTER THREE:

STUDY TWO

In study one, cross validation techniques were applied to simulated data to compare the decision tree method to logistic regression. This allows for the identification of what decision makers *should* do when making decisions. In study two, cross validation techniques will be used on archival data from Middle Tennessee State University's Industrial Organizational Psychology master's program in order to compare the results of the simulation with real applicant selection data. This will also allow for a comparison of the decision makers' selection strategies to those made based on actuarial methods. This study tested hypothesis two by comparing the actuarial methods in classifying applicant acceptance into a graduate school program.

Method

The data consisted of 435 applicants from the years of 2011 to 2016 (accepted and not accepted applicants included). Applicants who had missing data from GRE, GPA, or a decision made by faculty, were deleted from the analysis. This resulted in 300 total applicants. Any GRE scores that were recorded using past formats were converted to the current scoring measures using the GRE score concordance tables (ETS, 2017). When applicants apply to the program, undergraduate grade point average (GPA), the Graduate Record Exam (GRE) verbal scores, GRE quantitative scores, and GRE written scores are required material. These were used as predictors of acceptance. These predictors are similar to the predictors used in study one with GPA being comparable to conscientiousness and GRE being a measure of intelligence (Cheng & Ickes, 2009;

Kuncel et al., 2001). Whether or not the applicant was selected to the program was used as the outcome variable.

Training the Data: The data was randomly split into thirds using code in R that sampled from the 300 applicants (See Appendix B). One third of the data was used for training data. Similar to study one both a decision tree model and a logistic regression model were created based on the training data. The remaining two thirds was used to test the logistic regression model and the decision tree model. This was done using a loop in R that sampled 100 applicants from the remaining 200. This loop created 1,000 different samples resulting in 1,000 different classification accuracies for the decision tree and the logistic regression. The models predicted selection into the program classifying applicants as either 1 (admit) or 0 (do not admit). This was then compared to the decision that was actually made for each applicant. This allowed for comparison between which actuarial method most accurately represents the admissions decisions made.

Results

Table 3 displays the variables used and their descriptive statistics for the study. Hypothesis 2 states that decision tree analysis will be as accurate as logistic regression in predicting acceptance into a graduate school program. Accuracy of decision trees will not be significantly different from that of regression. This in part will validate the results in study one. Hypothesis 2 was tested using a two-sample independent *t*-test. The decision tree classification accuracy was (M = 53%, SD = 0.05). See Figure 4 for the classification tree. In the classification tree created, root node is GPA. If GPA is less than 3.325 then the tree advises not to select. If it is greater than 3.325 then the user would proceed to the next branch which is GRE written. If the written score is greater than 3.29 the user would proceed to the next branch until a decision of either admit or do not admit is reached. The logistic regression classification accuracy was (*M* = 52%, *SD* = 0.05). This means that the decision tree method was able to accurately classify applicant admission based on human judges decision making 53% of the time will logistic regression was able to accurately identify applicants 52% of the time. The two-sample proportion test was not significant ($t(1998) = 0.56$, $p = .57$). Hypothesis two was supported.

Table 4: Descriptive Statistics for Predictors in Study Two

	N	**Mean**	**SD**
GRE Verbal	300	152.51	5.71
GRE Quantitative	300	149.93	5.75
GRE Written	300	3.88	0.61
Undergraduate GPA	300	3.49	0.33

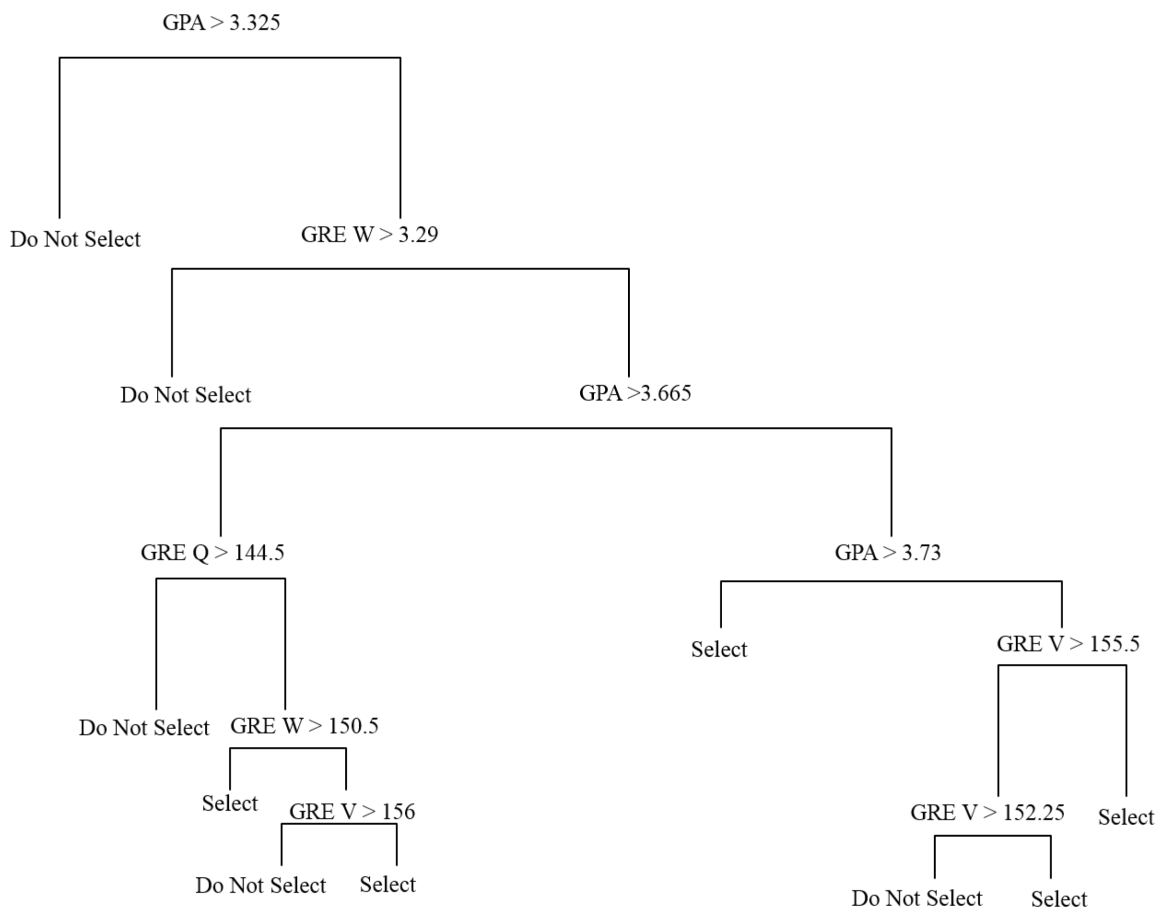

Figure 4. Decision tree from study two. “0” is do not admit and “1” is admit

Discussion

Study two sought to examine how different actuarial methods most accurately modeled human decision making in admission to graduate school. I found that both decision trees and logistic regression performed similarly in modeling human decision making. Both methods accurately predicted human decision making at a rate of 53% for decision tree method and 52% for the logistic regression method. This shows that there is a small albeit non-significant difference between the decisions made by actuarial methods in predicting what human decision makers did. Part of the reason for the relatively low accuracies of both of these models could be contributed to the fact that the models were created to predict human decision making. Human decision making is complex and takes into account many other variables than are represented here which likely led to the low accuracy for both methods. James et. al. (2013) stated that decision tree models can better encapsulate human decision making than other actuarial methods. In this study, that did not hold true. Although, the decision tree method did not produce more similar results to human decision makers, the interpretation of the decision tree method is more similar to human decision making (Artinger et al., 2015) which could lead to greater use by decision makers.

Study two examined the classification accuracy of the decision tree method and logistic regression method. The findings from this study also help support the results of study one. In study one, I found that the decision tree method was not significantly different from logistic regression in ability to accurately predict performance. A weakness of the first study is that the data was simulated. This study used real data and came to the same conclusion. This provides initial evidence that decision trees can be used to make selection decisions as accurately as other actuarial methods. As mentioned in study one, the implication of this is that because decision trees are quick and easy to interpret more managers may be willing to use this method, which will in turn result in better decisions.

There are several limitations to this study. First, in order to be able to show the difference between human decision making and each actuarial method, performance data would be needed. Although this could have been gathered for some applicants, it could not be for those who either did not attend the program or dropped out of the process. This is a problem that many organizations face when attempting to validate their selection measures. It is difficult to be able to state that selection measures truly differentiate between applicants. For this reason, both models predicted what decision was actually made by decision makers. This does not allow for the study to determine whether the actuarial methods better predict performance. However, an abundance of research supports the notion that actuarial methods do predict performance better than clinical judgments (Dawes, 1971, 1979). Another limitation of this study is due to the small sample size. Missing data limited the number of different applicants that were used for analysis. Actuarial methods perform better when they have more data points available to them (Bartlett, Kotrlik, & Higgins, 2001). Part of the reason that the models accurately predicted less than half of applicants could be due to a relatively small sample size used to train the models, although, this limitation is not uncommon in most organizations (Bartlett et al., 2001).

CHAPTER FOUR:

GENERAL DISCUSSION

These studies provide initial evidence that the decision tree method can be used effectively in selection scenarios within human resources. In study one, it was found that the decision tree method and logistic regression performed equally well in classifying applicant performance. Findings from other fields have shown that the decision tree method is as accurate as other actuarial methods, and the findings from this study support that conclusion (Coussement et al., 2014; Liu et al., 2011; Stewart et al., 2016). Additionally, selection ratio type had a significant impact on the accuracy of the decision tree and logistic regression such that less restrictive selection ratios increased accuracy of both methods. Although both models performed better under less restrictive selection ratio types, this does not mean that better applicants will be selected. A final finding of study one was that decision trees were not more (or less) robust to change in selection ratios than logistic regression. Previous research demonstrated that decision trees were more robust to changes in data, however, these results were not reproduced in this study (Coussement et al., 2014). Both methods performed equally across changes. Finally, study two found that the decision tree technique and logistic regression were equally as accurate in predicting acceptance to graduate school. This in conjunction with the results of study one demonstrates that the decision tree method could be used as an equally accurate replacement to other actuarial methods.

Implications

Other industries have already begun to take use of the decision tree method due to how quickly it can be used and how easy it is to interpret. For instance, the decision tree method has been used to reduce the time of diagnosis and to more easily make decisions (Artinger et al., 2015; Stewart et al., 2016). These studies show that there is utility in using decision trees provided that they are as accurate as other actuarial methods. Study one and two demonstrated that the decision tree method performed equally well to logistic regression. A potential benefit to this finding is that decision trees could be used in place of regression when making selection decisions. Due to the ease of interpretation, this could be a strong choice for managers when making decisions.

Different types of selection ratios were used to mimic different positions within organizations. For example, some positions have few openings while others have many open positions. Study one, found that less restrictive selection ratios increased the accuracy of both methods. As discussed previously, in this study more lenient selection ratio types resulted in increased accuracy. However, this does not necessarily result in higher performing candidates. If the selection ratio type is too lenient or strict, then this increases the chances of hiring poor employees or missing high performing employees. For example, if the selection ratio is low then more applicants will be hired thus increasing the chance that poor performers are selected. Therefore, it is essential to find a selection ratio that limits poor performers and is not too strict to exclude good performers. To get the most out of the models, there must be a balance between desired accuracy and appropriate selection ratio. However, because both models performed equally under each selection ratios, managers have the option of using the decision tree method for a variety of different positions. A final finding was that neither the decision tree method nor logistic regression responded differently to changes in data due to selection ratios. An implication of this is that both methods perform quite similarly across different situations. Due to both methods performing equally well, either method could be used for selection decisions with similar results. Additionally, these results have reproduced the findings of other studies that have reported that the decision tree technique is as accurate as other methods across different situations and industries (Coussement et al., 2014; Liu et al., 2011; Stewart et al., 2016).

Limitations

There are several limitations of these studies. In study one, simulated data was used. Therefore, it could be difficult to generalize the findings to a real-world scenario. A similar limitation in study two is that the data was from a graduate school program. This data may lack characteristics that data in applied organizations possess, such as attrition from applicants dropping out of the process, missing data, and human bias in judgements for predictors, such as structured interview scores. Additionally, because performance data was not available for all applicants, it was not possible to compare true accuracy of the models. Both models' accuracies were judged based upon the decision made by the selection committee. A more accurate way to judge each model would be to compare available performance data for each applicant with the decision that each model produces. If this was available, then one could compare human decision making to both models based on a quantitative measure of performance.

Another limitation is that the null hypothesis was predicted in both studies. In study one it was predicted that there would be no difference between decision tree and logistic regression. Study two also predicted that there would be no difference between the models in classification accuracy. In null hypothesis testing, a researcher either rejects the null or fails to reject the null. When the null is predicted, there is a higher threshold necessary to provide evidence that the null is correct. This evidence is often obtained by replication where support for the null is found repeatedly (Bernardo, Estadística, & Matemáticas, 2003). Since this study predicted the null, replication is necessary in order to confirm that the decision tree truly is as accurate as logistic regression.

Future Directions

These studies provide an initial foundation for more research on decision tree use for personnel selection decision-making. For instance, the present research leads to a question of whether managers may be more willing to rely on decision trees than logistic regression or even their own intuitive processes. Previous research has shown that hiring managers over rely on their own intuition (Highhouse, 2008; Miles & Sadler-Smith, 2011). A study that could answer whether managers would use the decision tree method is to survey current hiring managers and ask questions about current selection systems and willingness to use the decision tree method. Another, component of this would be to determine which method managers find easier to use. The purpose of this would be to determine managers’ acceptance of the decision tree method answering the question whether managers would use this technique. Future research should also examine how error (missing data, attrition, bias in judgment) effects both of these models in a selection context. Coussement et. al. (2014) demonstrated that the decision tree method performs slightly better than other actuarial methods when there is missing data. Missing data is common error in data and can adversely affect actuarial methods (Gelman & Hill, 2007). Due to the prevalence of missing data, future research should determine how missing data will affect the use of decision trees when compared to other selection methods, such as regression or human decision making. This could be done by randomly deleting data from a dataset then testing each of the selection methods for accuracy. Another potential research question is whether the decision tree could reduce the number of items on a predictor while maintaining accuracy. Stewart et. al. (2016) demonstrated how the decision tree method cut down the time it took to diagnose PTSD while maintaining similar accuracy. A future study could investigate whether lengthy selection devices could be cut down to be more frugal, saving the organization both time and money. For example, if an organization has a structured interview with many questions, a decision tree could be applied and potentially reduce questions that are less predictive, thus shortening administration time.

Conclusion

Actuarial methods such as logistic regression have been shown to be more accurate in predicting applicant's future performance when compared to clinical judgments (Dawes, 1971, 1979). Despite this, organizations continue to use suboptimal methods when making decisions (Highhouse, 2008). Decision trees have gained continued research interest as a fast and easy to interpret actuarial method (Raab & Grigerenzer, 2015). Decision trees are an attractive alternative due to the ease of interpretation. With methods like logistic regression, prior statistical knowledge is needed in order to interpret the model. Decision trees require no such prerequisites to interpret. This has led to numerous studies in fields ranging from finance to medicine that have demonstrated decision trees to be as accurate as other actuarial methods (Coussement et al., 2014; Liu et al., 2011; Sinha & May, 2004; Stewart et al., 2016). The two studies conducted provided initial evidence that a) decision trees are as accurate in classifying performance as logistic regression, b) decision trees can be effective in selection scenarios, c) of the three selection ratios the least stringent ratio increased the accuracy of both models over the other models, and d) selection ratios will interact with decision technique in predicting performance.

REFERENCES

Artinger, F., Petersen, M., Gigerenzer, G., & Weibler, J. (2015). Heuristics as adaptive decision strategies in management. *Journal of Organizational Behavior*, *36*(S1), S33–S52. https://doi.org/10.1002/job.1950

Barrick, M., Mount, M., & Judge, T. (2001). Personality and performance at the beginning of the new millennium: What do we know and where do we go next? *International Journal of Selection and Assessment*, *9*(June), 9–30. https://doi.org/10.1111/1468-2389.00160

Barrick, M. R., & Zimmerman, R. D. (2009). Hiring for retention and performance. *Human Resource Management*, *48*(2), 183–206. https://doi.org/10.1002/hrm.20275

Bartlett, J. E., Kotrlik, J. W., & Higgins, C. C. (2001). Organizational research: Determining appropriate sample size in survey research. *Information Technology, Learning, and Performance Journal*, *19*(1), 43–50. https://doi.org/10.1109/LPT.2009.2020494

Bernardo, J. M., Estadística, D. De, & Matemáticas, F. De. (2003). Bayesian statistics. *Encyclopedia of Life Support Systems (EOLSS)*, *Probabilit*, 1–46. https://doi.org/10.1002/sim.1982

Betsch, C. (2008). Chronic preferences for intuition and deliberation in decision making: Lessons learned about intuition from an individual differences approach. In H. Plessner, C. Betsch, & T. Betsch (Eds.), *Intuition in judgment and decision* (pp. 231–248). https://doi.org/10.4324/9780203838099

Bobko, P., Roth, P. L., & Buster, M. A. (2005). Work sample selection tests and expected reduction in adverse impact: A cautionary note. *International Journal of Selection* *and Assessment*, *13*(1), 1–10. https://doi.org/10.1111/j.0965-075X.2005.00295.x

Cheng, W., & Ickes, W. (2009). Conscientiousness and self-motivation as mutually compensatory predictors of university-level GPA. *Personality and Individual Differences*, *47*(8), 817–822. https://doi.org/10.1016/j.paid.2009.06.029

Cortina, J. M., Goldstein, N. B., Payne, S. C., Davison, K. H., & Gilliland, S. W. (2000). The incremental vaidity of interview scores over and above cognitive ability and conscientiousness scores. *Personnel Psychology*, *53*(2), 325–351. https://doi.org/10.1111/j.1744-6570.2000.tb00204.x

Coussement, K., Van den Bossche, F. A. M., & De Bock, K. W. (2014). Data accuracy's impact on segmentation performance: Benchmarking RFM analysis, logistic regression, and decision trees. *Journal of Business Research*, *67*(1), 2751–2758. https://doi.org/10.1016/j.jbusres.2012.09.024

Dawes, R. M. (1971). A case study of graduate admissions. *American Psychologist*, *26*, 180–188. https://doi.org/10.1037/h0030868

Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist*, *34*(7), 571–582. https://doi.org/10.1037/0003-066X.34.7.571

De Corte, W. (1999). A note on the success ratio and the utility of fixed hiring rate personnel selection decisions. *Journal of Applied Psychology*, *84*(6), 952–958. https://doi.org/10.1037/0021-9010.84.6.952

De Corte, W. (2002). Sampling variability of the success ratio in predictor-based selection. *The British Journal of Mathematical and Statistical Psychology*, *55*(Pt 1), 93–107. https://doi.org/10.1348/000711002159716

Diab, D. L., Pui, S. Y., Yankelevich, M., & Highhouse, S. (2011). Lay perceptions of selection decision aids in US and non-US samples. *International Journal of Selection and Assessment*, *19*(2), 209–216. https://doi.org/10.1111/j.1468-2389.2011.00548.x

ETS. (2017). GRE Verbal Reasoning and Quantitative Resaoning Concordance Table.

Farr, J. L., & Tippins, N. T. (2010). Handbook of employee selection: An introduction and overview. *Handbook of Employee Selection.* https://doi.org/10.4324/9780203809808

Finch, D. M., Edwards, B. D., & Wallace, J. C. (2009). Multistage selection strategies: simulating the effects on adverse impact and expected performance for various predictor combinations. *The Journal of Applied Psychology*, *94*(2), 318–340. https://doi.org/10.1037/a0013775

Gelman, A., & Hill, J. (2007). Missing-data imputation. In *Data Analysis Using Regression and Multilevel/Hierarchical Models* (pp. 529–543). https://doi.org/10.3758/s13428-011-0071-2

Gottfredson, L. S. (1998). The general intelligence factor. *Scientific American Presents*, *9*(4), 24–29. https://doi.org/Article

Highhouse, S. (2008). Stubborn reliance on intuition and subjectivity in employee selection. *Industrial & Organizational Psychology*, *1*(3), 333–342. https://doi.org/10.1111/j.1754-9434.2008.00058.x

Hough, L. M., & Oswald, F. L. (2000). Personnel selection: Looking toward the future - remembering the past. *Annual Review of Psychology*, *51*, 631–64. https://doi.org/10.1146/annurev.psych.51.1.631

Huffcutt, A. I., & Arthur, W. (1994). Hunter and Hunter (1984) revisited: Interview validity for entry-level jobs. *Journal of Applied Psychology*, *79*(2), 184–190. https://doi.org/10.1037/0021-9010.79.2.184

Hütter, M., & Ache, F. (2016). Seeking advice: A sampling approach to advice taking. *Judgment and Decision Making*, *11*(4), 401–415.

James, G., Witten, D., Hastie, T., & Tibishirani, R. (2013). *An Introduction to Statistical Learning*. *Springer Texts in Statistics*. https://doi.org/10.1007/978-1-4614-7138-7

Joyce, S. P. (2016). The internet revolution : Digital disruption in recruiting & job search. *Career Planning & Adult Development Journal*, *32*(2), 14–21.

Kuncel, N. R., Ones, D. S., & Hezlett, S. a. (2001). A comprehensive meta-analysis of the predictive validity of the graduate record examinations: implications for graduate student selection and performance. *Psychological Bulletin*, *127*(1), 162–181. https://doi.org/10.1037/0033-2909.127.1.162

Littlepage, G. E., Bragg, D. M., & Rust, J. O. (1978). Relations between admission criteria, academic performance, and professional performance. *Teaching of Psychology*, *5*(1), 16–20. https://doi.org/10.1207/s15328023top0501_5

Liu, Y., Lin, D., Xiao, T., Ma, Y., Hu, Z., Zheng, H., … Gao, Y. (2011). An immunohistochemical analysis-based decision tree model for estimating the risk of lymphatic metastasis in pN0 squamous cell carcinomas of the lung. *Histopathology*, *59*(5), 882–891. https://doi.org/10.1111/j.1365-2559.2011.04013.x

Miles, A., & Sadler-Smith, E. (2011). Personnel Review with recruitment I always feel I need to listen to my gut: The role of intuition in employee selection. *Personnel Review Iss Journal of Managerial Psychology Personnel Review Career*

Development International, *43*(6), 606–627. https://doi.org/10.1108/PR-04-2013-0065

Nakayama, N., Oketani, M., Kawamura, Y., Inao, M., Nagoshi, S., Fujiwara, K., … Mochida, S. (2012). Algorithm to determine the outcome of patients with acute liver failure: A data-mining analysis using decision trees. *Journal of Gastroenterology*, *47*(6), 664–677. https://doi.org/10.1007/s00535-012-0529-8

Raab, M., & Gigerenzer, G. (2015). The power of simplicity: A fast-and-frugal heuristics approach to performance science. *Frontiers in Psychology*, *6*(OCT), 1–6. https://doi.org/10.3389/fpsyg.2015.01672

Raju, N., Pappas, S., & Williams, C. P. (1989). An empirical monte carlo test of the accuracy of the correlation, covariance, and regression slope models for assessing validity generalization. *Journal of Applied Psychology*, *74*, 901–911. https://doi.org/10.1037/0021-9010.74.6.901

Schmidt, F., & Hunter, J. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, *124*(2), 262–274. https://doi.org/Doi 10.1037//0033-2909.124.2.262

Schmidt, F. L., & Hunter, J. (2004). General mental ability in the world of work: occupational attainment and job performance. *Journal of Personality and Social Psychology*, *86*(1), 162–173. https://doi.org/10.1037/0022-3514.86.1.162

Schmidt, F., & Rader, M. (1999). Exploring the boundary conditions for interview validity: Meta-analytic validity findings for a new interview type. *Personnel Psychology*, *52*(2), 446–464. https://doi.org/10.1111/j.1744-6570.1999.tb00169.x

Sinha, A. P., & May, J. H. (2004). Evaluating and tuning predictive data mining models using receiver operating characteristic curves. *Journal of Management Information Systems*, *21*(3), 249–280. https://doi.org/10.1080/07421222.2004.11045815

Stauffer, J., & Ree, M. J. (1996). Predicting with logistic or linear regression: Will it make a difference in who is selected for pilot training? *The International Journal of Aviation Psychology*, *6*(3), 233–240.

Stewart, R. W., Tuerk, P. W., Metzger, I. W., Davidson, T. M., & Young, J. (2016). A decision-tree approach to the assessment of posttraumatic stress disorder: Engineering empirically rigorous and ecologically valid assessment measures. *Psychological Services*, *13*(1), 1–9. https://doi.org/10.1037/ser0000069

Therneau, T., Atkinson, B., Ripley, B., & Ripley, M. B. (2015). rpart: recursive partitioning and regression trees. *R Package Version 4.1-10, https://CR*. Retrieved from http://cran.ma.ic.ac.uk/web/packages/rpart/rpart.pdf%5Cnhttps://cran.r-project.org/web/packages/rpart/rpart.pdf

van der Zee, K. I., Bakker, A. B., & Bakker, P. (2002). Why are structured interviews so rarely used in personnel selection? *The Journal of Applied Psychology*, *87*(1), 176–184. https://doi.org/10.1037/0021-9010.87.1.176

APPENDICES

APPENDIX A: R Code Study One

```
###Population Set Up###
n = 250
condition=function(b1,b2,b3,r.square,n)
{
### b1-b3 are the regression coefficients of the predictors (see Equation 1)
### r.square is R square found Cortina et al. (2000)
gma=rnorm(n,0,1)
cons=rnorm(n,0,1)
interview=rnorm(n,0,1)
integrity=rnorm(n,0,1)
error=rnorm(n,0,sqrt(1-r.square))
perform=b1*gma+b2*cons+b3*interview+error
data=cbind(perform,gma,cons,interview)
colnames(data)=c("performance","GMA","conscientiousness","interview")
return=data
}
test=condition(b1=.43,b2=.24,b3=.44,r.square=0.46,n=50)
test

### Replicating Data Sets ###
rep=1 ### number of replications
con=lapply(1:rep, function(x) condition(.43,.24,.44,.46,500000))
for (g in 1:rep){dataname=paste("data",g,sep="")
write.table(con[[g]],file=paste("C:/Users/.../data/",dataname,".txt",sep=""),col.names=FA
LSE,row.names=FALSE,quote=FALSE)}

###Selection Ratio 1###
cut.off=quantile(train[,1], .9)
Decision=as.factor(ifelse(train[,1] <=cut.off,0,1))

###Selection Ratio 2###
Decision=as.factor(ifelse(train[,1] <=.69,0,1))

###Selection Ratio 3###
cut.off=quantile(train[,1], .9)
Decision=as.factor(ifelse(train[,1] <=(cut.off&.69),0,1))
```

```
###Decision Tree Formation###
set.seed(123)
train=read.table("C:/Users/.txt")
cut.off=quantile(train[,1], .9)
Decision=as.factor(ifelse(train[,1] <=cut.off,0,1)) # 0 for not hire, 1 for hire
train=data.frame(train[,-1],Decision)
index=sample(500000,250,replace=FALSE)
train.data=train[index,]
tree.train=tree(train.data[,4]~train.data[,1]+train.data[,2]+train.data[,3])
m1=c()
for(i in 1:100)
{
all.test.data=train[-index,]
index=sample(499750,250,replace=TRUE)
test.data=all.test.data[index,]
tree.pred=predict(tree.train,test.data,type="class")
my=table(tree.pred,test.data[,4])
m1[i]=(my[1,1]+my[2,2])/(my[1,1]+my[1,2]+my[2,1]+my[2,2])
}
mean(m1)
m1

###Logistic Regression Model Formation###
set.seed(123)
train=read.table("C:.txt")
cut.off=quantile(train[,1], .9)
Decision=as.factor(ifelse(train[,1] <=cut.off,0,1)) # 0 for not hire, 1 for hire
train=data.frame(train[,-1],Decision)
index=sample(500000,250,replace=FALSE)
train.data=train[index,]
model.train=glm(train.data[,4]~train.data[,1]+train.data[,2]+train.data[,3],family=binomial
(logit))
m2=c()
for(i in 1:100)
{
all.test.data=train[-index,]
index=sample(499750,250,replace=TRUE)
test.data=all.test.data[index,]
model.pred=predict(model.train,test.data,type="response")
fitted.results=as.factor(ifelse(model.pred<=.5,0,1))
my1=table(fitted.results,test.data[,4])
m2[i]=(my1[1,1]+my1[2,2])/(my1[1,1]+my1[1,2]+my1[2,1]+my1[2,2])
}
mean(m2)
m2
```

```
###ANOVA Code###
data1=read.table(" ",header=FALSE)
colnames(data1)=c("selection","method","error")#1 is decision, #2 is logistic
data1
aov(error ~selection*method,data=data1)
pairwise.t.test(data1$error,data1$method)
pairwise.t.test(data1$error,data1$selection)
```

APPENDIX B: R Code Study Two

```
library(tree)
set.seed(110)
data=read.table("C:/.txt",sep="")
data[,2]=as.numeric(data[,2])
data[,3]=as.numeric(data[,3])
data[,4]=as.numeric(data[,4])
data[,5]=as.numeric(data[,5])
data[,6]=factor(as.numeric(data[,6]))
index=sample(300,100,replace=FALSE)
train.data=data[index,]
tree.train=tree(train.data[,6]~train.data[,2]+train.data[,3]+train.data[,4]+train.data[,5])

my=c()
for(i in 1:1000)
{
all.test.data=data[-index,]
index=sample(200,100,replace=TRUE)
test.data=all.test.data[index,]
tree.pred=predict(tree.train,test.data,type="class")
m1=table(tree.pred,test.data[,6])
my[i]=(m1[1,1]+m1[2,2])/(m1[1,1]+m1[1,2]+m1[2,1]+m1[2,2])
}
my

train.data=data[index,]
model.train=glm(train.data[,6]~train.data[,2]+train.data[,3]+train.data[,4]+train.data[,5],fa
mily=binomial(logit))

my2=c()
for(i in 1:1000)
{
all.test.data=data[-index,]
index=sample(200,100,replace=TRUE)
test.data=all.test.data[index,]
model.pred=predict(model.train,test.data,type="response")
fitted.results=as.factor(ifelse(model.pred<=.5,1,2))
m2=table(fitted.results,test.data[,6])
my2[i]=(m2[1,1]+m2[2,2])/(m2[1,1]+m2[1,2]+m2[2,1]+m2[2,2])
}
my2

t.test(my,my2,var.equal=TRUE)
```

APPENDIX C: IRB Acceptance Form

IRB

INSTITUTIONAL REVIEW BOARD
Office of Research Compliance,
010A Sam Ingram Building,
2269 Middle Tennessee Blvd
Murfreesboro, TN 37129

IRBN007 – EXEMPTION DETERMINATION NOTICE

Tuesday, January 30, 2018

Investigator(s): Kyle Marks; Alexander Jackson
Investigator(s') Email(s): Kcm3z@mtmail.mtsu.edu; alexander.jackson@mtsu.edu
Department: Psychology

Study Title: Comparing the accuracy of decision trees and logistic regression in personnel selection
Protocol ID: **18-1157**

Dear Investigator(s),

The above identified research proposal has been reviewed by the MTSU Institutional Review Board (IRB) through the **EXEMPT** review mechanism under 45 CFR 46.101(b)(2) within the research category *(4) Study involving existing data* A summary of the IRB action and other particulars in regard to this protocol application is tabulated as shown below:

IRB Action	EXEMPT from furhter IRB review***
Date of expiration	**NOT APPLICABLE**
Participant Size	Existing Data
Participant Pool	**De-identified exisiting data**

Mandatory Restrictions	Only de-identified data covered by the approved permission letter on file with the MTSU Office of Research Compliance may be accessed	
Additional Restrictions	**None at this time**	
Comments	None at this time	
Amendments	**Date**	**Post-Approval Amendments**
		None at this time

***This exemption determination only allows above defined protocol from further IRB review such as continuing review. However, the following post-approval requirements still apply:

- Addition/removal of subject population should not be implemented without IRB approval
- Change in investigators must be notified and approved
- Modifications to procedures must be clearly articulated in an addendum request and the proposed changes must not be incorporated without an approval
- Be advised that the proposed change must comply within the requirements for exemption
- Changes to the research location must be approved – appropriate permission letter(s) from external institutions must accompany the addendum request form
- Changes to funding source must be notified via email (irb_submissions@mtsu.edu)
- The exemption does not expire as long as the protocol is in good standing

- Project completion must be reported via email (irb_submissions@mtsu.edu)
- Research-related injuries to the participants and other events must be reported within 48 hours of such events to compliance@mtsu.edu

The current MTSU IRB policies allow the investigators to make the following types of changes to this protocol without the need to report to the Office of Compliance, as long as the proposed changes do not result in the cancellation of the protocols eligibility for exemption:

- Editorial and minor administrative revisions to the consent form or other study documents
- Increasing/decreasing the participant size

The investigator(s) indicated in this notification should read and abide by all applicable postapproval conditions imposed with this approval. Refer to the post-approval guidelines posted in the MTSU IRB's website. Any unanticipated harms to participants or adverse events must be reported to the Office of Compliance at (615) 494-8918 within 48 hours of the incident.

All of the research-related records, which include signed consent forms, current & past investigator information, training certificates, survey instruments and other documents related to the study, must be retained by the PI or the faculty advisor (if the PI is a student) at the sacure location mentioned in the protocol application. The data storage must be maintained for at least three (3) years after study completion. Subsequently, the researcher may destroy the data in a manner that maintains confidentiality and anonymity. IRB reserves the right to modify, change or cancel the terms of this letter without prior notice. Be advised that IRB also reserves the right to inspect or audit your records if needed.

Sincerely,

Institutional Review Board

Middle Tennessee State University

Quick Links:

Click here for a detailed list of the post-approval responsibilities.
More information on exmpt procedures can be found here.