MACHINE LEARNING AND PERSONALITY TRAITS:

A DISTURBING CONTRIBUTION FROM THE

ALGORITHMIC CULTURE TO

BEHAVIORAL SCIENCE

By

MARCIO LIMA

Bachelor of Science in Agronomy Engineering
Universidade Federal de Viçosa
Viçosa, Minas Gerais
2002

Master of Science in Metallurgical Engineering
Universidade Federal de Minas Gerais
Belo Horizonte, Minas Gerais
2004

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
December, 2018

MACHINE LEARNING AND PERSONALITY

TRAITS: A DISTURBING CONTRIBUTION

FROM THE ALGORITHMIC CULTURE

TO BEHAVIORAL SCIENCE

Dissertation Approved:

Dursun Delen, Ph.D.

Bryan Edwards, Ph.D.

J. Craig Wallace, Ph.D.

Toby Joplin, Ph.D.

Name: MARCIO LIMA

Date of Degree: JULY, 2019

Title of Study: MACHINE LEARNING AND PERSONALITY TRAITS: A DISTURBING CONTRIBUTION FROM THE ALGORITHMIC CULTURE TO BEHAVIORAL SCIENCE

Major Field: TYPE MAJOR

Abstract: The time has come for behavioral scholars to benefit from the superior prediction accuracy of modern data mining practices over traditional data modeling. The current investigative study uses machine learning techniques to explore the prediction power of the HEXACO traits on work performance. To obtain reliable results, I employ the most prevailing machine learning algorithms in addition to logistic and multiple regression. The concomitant use of multiple prediction models that are grounded solidly in specific literature is applied to reveal the most accurate model for prediction purposes. One relevant methodological contribution of the present study is the employment a Random Forest-based heuristic method that computes the ratio of actual splits on a certain variable to the number of times that particular variable was selected as a candidate to split within the forest. By computing the order of importance of traits as job performance predictors, the current research illuminates the field with relevant and accurate information regarding the crucial role of humility, the strongest job performance predictor. Also, from a novel perspective and in light of Liebig's Law of the Minimum, the present study reveals a strong influence of relative proportions of traits (i.e., ratio between scores of traits) on job performance ratings. Interestingly the second most important predictor was found to be the ratio between scores on emotional stability and conscientiousness, followed by the ratio between scores on extraversion and openness to experience. In certain conditions, these results reveal that proportions between two different traits may be stronger predictors of job performance than individual traits. Taken all together, this research is the first academic study to use machine learning techniques on HEXACO personality scores to reveal job performance-related predictors with seamless high predictive accuracy. Indeed, the methodological and theoretical contributions obtained in the current study should be carefully examined by practitioners and scientists in order to pragmatically leverage the management research field.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Despite the fact that management researchers often deal with human behavior that is somewhat subjective, machine-learning techniques arise with a great potential to offer high-level insights to social scientists and practitioners. The current work intends to expand knowledge and "soft science" to hard science through the use of modern data analytics methods and by drawing a parallel between traditional research assumptions and Liebig's Law of Minimum (here after LLM), which is largely applied by organic chemistry scholars. Justus von Liebig proposed LLM in 1840, along with the process to understand the collective effect of chemical elements on plant growth. In general, LLM posits that in environments where important nutrients are in low concentrations, the most limiting of these will influence the outcome regardless of the levels of other nutrients (Novais, 2007). LLM explains that in nature several chemical elements interacting together will affect biological relationships and that the amount of these elements should be analyzed collectively, in addition to individually (Paris, 1992). Since the advent of personality theory in the management field, researchers have evolved and developed methods of assessing individuals' tendencies, characteristics, and probable behavior. Most of these methods have focused on revealing significant and nonsignificant effects between traits and management results. The vast majority of data interpretation in the social and behavioral sciences relies on techniques based on Null-Hypothesis Significant Testing (NHST). In this line, Loftus (1996) points to the fact that a chaotic phenomenon occurs in a number of social studies to the extent that similar results (e.g., $p = 0.049$, $p = 0.050$, and $p = 0.051$) can yield entirely different conclusions. This becomes critical since research relies on previous theoretical findings to build knowledge. If previous findings are not supported by solid and reliable conclusions, conflicting results in the management field are not astonishing. In contrast with what happens in natural science (e.g., the conjunct effect of elements and nutrients in plants), our field tackles human traits one by one or at most mediation and moderation effects involving a few traits and its effect on work-related outcomes (Shaffer & Postlethwaite, 2013; Do & Minbashian, 2014; Harari, Rudolph, & Laginess, 2015). There is a clear gap in the management literature that hasn't yet approached the collective influence of traits on work-related outcomes through the lens of machine learning techniques. Targeting an increased predictive accuracy with the use of powerful algorithms, this present work attempts to illuminate management science from a novel angle by uncovering how "collective amounts" of human traits may affect work-related outcomes. From a social science perspective, drawing on LLM and contrasting the large amount of academic work that builds on specific influences of personality traits on social interactions, the current work explores how different human traits collectively affect performance at the individual level.

Traditional statistical procedures have been used by social science scholars for decades. Interestingly, despite the recurring calls for the development of novel and reliable methods to help improve the fields' insufficient capacity to understand human behavior in the business context, not much has changed. In his outstanding article, Loftus (1996), elaborates on the intriguing power of psychological studies for diverse fields of science. He consistently argues that NHST reduces data analyses into a series of effect/no-effect decisions, which leads the field away from a correct decision support path in several ways. It suggests an illusion of certainty on a realm that is naturally ambiguous and subjective. The use of modern algorithms and its increased predictive power causes mixed feelings of excitement, anxiety, and angst about future directions of management science. Relying on inappropriate statistical methods to achieve conclusions and inferences related to personality data analysis may delay years of priceless advancements within the management field. Echoing Loftus' thoughts and speaking to the statistical community, Breiman (2001a) suggests that if the goal is to use data to solve problems, scholars should consider not relying exclusively on traditional data modeling but instead embrace a more elaborated set of tools.

According to Sharda, Delen, and Turban (2016) the analytics approach has become one of the most important decision-making drivers of this decade. The field of business analytics has evolved rapidly, achieving impressive predictive accuracy levels. Machine learning techniques' ability to extract knowledge has far surpassed humans' ability to do so, even for routine events. Data mining has been successfully predicting outcomes in several domain areas such as healthcare, medicine, entertainment, and homeland security. Tackling one of the most challenging areas of knowledge discovery, Delen (2009) developed prediction models to assess survivability of patients with prostate cancer, which helps health specialists to save lives. Algorithms have been employed to accurately predict individuals' traits from pictorial representation of faces, for instance, using deep learning algorithms and assessing Facebook likes across over 80,000 participants from different genders and ethnic groups. Youyou, Kosinski, and Stillwell (2015) argue that computers can predict husbands' and wives' traits more accurately than the couple can itself (Youyou et al., 2015).[1] It is clear that machine learning enlarges human possibilities, which is constrained by brain limitations and serves as a unique opportunity for management scholars to leverage the insights and assumptions generated by our community.

Previous work solely focusing on worker's traits or surface traits as predictors for service performance may also benefit from my results. Although there is a considerable amount of academic work supporting personality trait as a functional predictor for job performance (Christiansen, Sliter, & Frost, 2014; Cogliser, Gardner, Gavin, & Broberg, 2012; Do & Minbashian, 2014; Harari et al., 2015), scholars like Murphy (2005); Morgeson et al. (2007); and Sitser, Van der Linden, and Born (2013) argue that when selecting working personnel, personality traits may have limited use. As suggested by Morgeson, Campion, Dipbove, Hollenbeck, Murphy, and Schmitt (2007), one of the major problems related to eventual personality prediction power could be related to the fact that self-reported questionnaires are subject to individuals' bias. From a different angle, marketing scholars like Brown, Mowen, Donavan, and Licata (2002) find that emotional stability, agreeability, and need for activity explain 39% of the variance in customers' orientation measures. Brown et al. (2002) imply that surface traits may in fact be associated with work-related outcomes. The management literature provides vast evidence pointing to diverse directions when assessing the relevance of human traits and their association with job effectiveness. At this point, addressing the influence of collective levels of personality traits instead of individual ones appears to be a reasonable alternative to help conciliate the contrasts in the literature.

Addressing the impact of groups of human traits on performance with a focus on predictive accuracy is extremely relevant for organizations. When companies are unable to

[1] Publications by Kosinski available at http://www.michalkosinski.com/.

properly hire the right people for a certain position, they indirectly sabotage one of their main sources of competitiveness. Managers and directors are generally aware of the huge amount of effort that needs to be employed to construct a cohesive, skilful, and trustful work force within their firms. Many entrepreneurs consider human capital to be the secret for success. In this sense, screening and managing personality features that match specific work requirements seems to be a shrewd management strategy. To help organizations and scholars to better understand how people's traits impact work performance, modern predictive methods proposed by this present work may set the basis for a large avenue within the management field.

Personality features and their association with work outcomes have been exhaustively studied by management scholars. Addressing occupational, health, and safety issues, Wallace, Edwards, Paul, Burke, Christian, and Eissa (2016) suggest that personnel selection and proper training are two of the most important practices for organizations aiming at achieving low accident rates. They point out that consciousness, as one of the Big Five personality traits, has been repeatedly associated with safety outcomes. Christiansen et al. (2014) emphasize the relevance of addressing personality-based job fit for several work situations. They examine the relationship between job satisfaction, distress, and personality traits. Individuals scoring low on Agreeableness and Conscientiousness are perceived to be more distressed when the task is associated with these traits. Participants high in Neuroticism report more distress when facing tasks related to Extraversion. It appears that efforts towards a better understanding of workers' profiles with respect to personality traits should at least diminish the level of stressors at workplaces. Harari et al. (2015) suggest that Agreeableness, Extraversion, and Emotional Stability are positively related to high performance ratings. On the other hand, the same authors point out that the Big Five justified only between 6% and

22% of the performance ratings. In their meta-analyses, these authors elaborate on how complex is the association of personality traits with performance. They call attention to the need for deeper investigations about performance appraisals with respect to personality traits. Again, ambiguous results about the relationship between personality traits and work-related outcomes raise questions about whether scholars from this field should continue to aim at the individual effect of traits or broaden the conversation by investigating the collective effect and the relative quantity of these traits.

Traditional management research has not yet looked at the collective effect nor the effect of relative scores of workers' personality traits on work-related outcomes with the use of machine learning techniques. In this sense, I propose to assess the influence of the collective effect of individuals' personality traits (e.g., scores on HEXACO) on work-related outcomes and to compare predictive accuracy levels between traditional statistical methods (e.g., multiple regression models) and modern machine learning techniques (e.g., Artificial Neural Network — multi-layer perceptron). For such, I will work with a database gathered and provided by AOE Science that contains data from HEXACO assessment (Big Five plus Humility) and work-related outcomes from 682 participants. More than 20 years ago, Loftus (1996) pointed out that psychological theory runs the risk of becoming linear-model theory and consequently many game changing discoveries will be overlooked. Further, the current work aims to clarify possible illusion of insights that are far worse than no insights at all and that may delay an even steeper advancement of management science.

CHAPTER II

REVIEW OF LITERATURE

Decision Support, Business Intelligence, and Analytics

A brief historical overview of machine learning

Kononenko (2001) points out that at about the same time that computers actually became working tools, algorithms appeared to help with data processing. Three main branches of machine learning emerged. Symbolic learning was explored by Hunt, Marin, and Stone (1966), while statistical methods and primary versions of neural networks was tackled by Nilsson, Sejnowski, and White (1965). Frank Rosenblatt (1962) became known as the father of neurocomputing (Kononenko, 2001) as a consequence of the development of the first single-layer perceptron. During the following years, all three branches developed sophisticated methods such as pattern recognition, K-nearest neighbors, discriminant analysis, and Bayesian classifiers (Fulkerson, 1995; Kononenko, 2001).

The process of using information system support for decision making focuses on reporting structured data to facilitate managers' and leaders' work. Routine reports provide useful information about past events within the work place. Users develop a need for more details with diverse levels of granularity so that they are able to efficiently tackle the evolving challenges of the business. According to Sharda et al. (2016), during the 1970s, Scott Morton defined decision support systems (DSS) as a combination of individuals' intellectual skills with computers' capacity to process data to achieve and improve decision quality (Keen & Morton, 1978). The 1980s and 1990s were stages of substantial changes in the way data was processed and used for decision purposes. The advent of software systems like enterprise resource planning (ERP) and relational database management systems (RDBM) were then associated with procedures that made it possible to efficiently capture and process data. Concomitantly, the large amount of data that was being generated and the need to maintain its integrity led to the creation of data storage mechanisms known as data warehouses (DW).

According to Breiman (2001a), in the mid-1980s two powerful new algorithms became available for fitting data: neural networks and decision trees. Concomitantly an emerging research community that applied these tools with a focus on prediction accuracy began to grow. Composed by young scholars, experts in physics and engineering, and a few aging statisticians, this emerging community started exploring complex prediction challenges where it was crystal clear that traditional data modeling had no applicability. Some examples were rudimentary image recognition, nonlinear time series prediction, and predicting finance-related outcomes (Breiman, 2001a). The emerging "prediction-focused community" generated a large range of interesting and real-word insights in several domains that were once undeveloped by traditional statistics (Booker, 1988; Langley, 1989; Grefenstette, 1988; Quinlan, 1986).

Along with the globalization of the economy and the rapidly evolving need for efficient decision-making processes, the term "business intelligence" (BI) was coined in conjunction with DW and DSS. To leverage the potential of large data sets to generate meaningful insights to managers, software corporations developed data mining tools to process and extract knowledge fast and efficiently. Later in the 2010s, the emergence of the internet and widespread access to data fostered the maturity of the field. Recently, the explosion of Big Data along with new ways to collect and process data coupled with machine learning techniques has expanded human knowledge in a number of fields (Sharda et al., 2016).

Two different cultures

From his distinguished position at the University of Berkeley, Leo Breiman elaborated on the gap between traditional statistics and machine learning. His seminal work, entitled "Statistical Modeling: The Two Cultures" explores the stochastic data and the algorithms models as ways to extract knowledge from data. According to Breiman (2001a), the data modeling approach assumes that data are generated by independent draws from response factors as functions of predictor variables and random noise. This vein relies on mechanisms such as goodness of fit, residual analysis, and *p*-values. The culture of algorithmic modeling considers the link between independent and dependent variables unknown and complex. Thus, it attempts to find an algorithm that manages to predict with the highest possible accuracy as its most important goal.

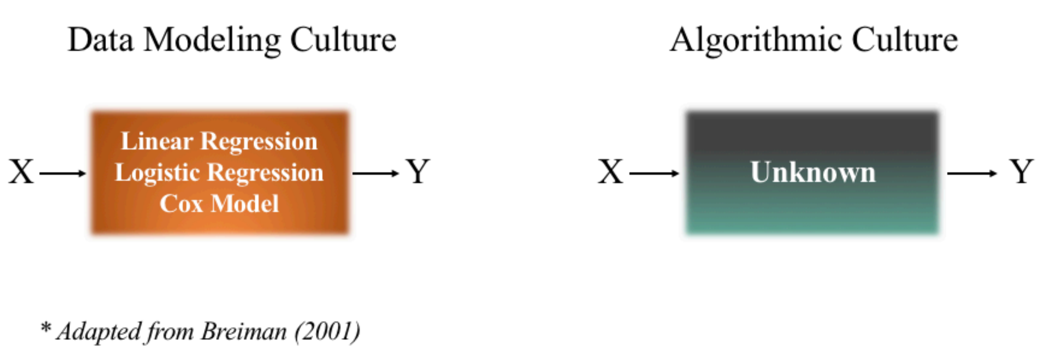


Figure 1. Data Modeling and Algorithmic Culture

Since the beginning of the algorithm culture, data models were seldom used. It relies on the assumption that in nature, data is generated in complex ways or ways that are at least partly inexplicable (i.e., black box). The challenge is in a test set to find a machine learning mechanism that will accurately predict future Y as a function of X. In contrast with traditional modeling where theory targets data models, the algorithm culture explores algorithms' properties, strengths, and ultimately their predictive accuracy. According to Breiman (2001a), the one assumption for the algorithm culture is that data generated by natural processes follows an unknown multivariate distribution. These separate both cultures by the essence of their basic assumptions.

Loftus (1996) points out that the majority of data assessments in the behavioral sciences relies on the traditional data modeling culture and utilizes the NHST. This matches Breiman's (2001a) argument that around 98% of all statisticians build conclusions and draw inferences based on goodness of fit and residual analyses. In his article, Breiman (2001a) suggests that by focusing knowledge building solely on traditional data practices, the statistical community may have been led to questionable scientific conclusions and also kept statisticians from working on exciting new problems. Following in the same vein, Bickel, Ritov, and Stoker (2006) argue that goodness-of-fit tests have very limited power except when the exact direction of the relationship is known. It seems that when the relationship between the predictor and the dependent variable is nonlinear, the traditional data modeling approach is highly subject to flawed conclusions.

Evidence has continuously pointed out the risk for the management field to keep generating excessive theory-oriented results at the expense of lacking practical real-work implications. Relying exclusively on the traditional data modeling approach to draw effective conclusions has channeled the field to, at best, achieving limited conclusions and at worst, building ambiguous knowledge (Bickel et al., 2006; Breiman, 2001a; Loftus, 1996).

In line with eminent scholars from the modern sciences and drawing on LLM, the current work represents a unique opportunity for the management field to enlarge its angle of vision when the goal is predicting within the workplace. It is of common sense that for any science to progress reasonably, its data analysis method must lead to valid genuine comprehension of whatever problematic issue is set out to be explored. The management field deals with humans as its principal subject matter (i.e., soft science), and therefore it contains a large number of uncontrolled variables and error variances. Thus, the algorithm culture with its predictive accuracy power may be a novel and genuine way of refreshing management science conclusions. Traditional data modeling, which banks its inferences on statistically significant levels (e.g., 0.05) indeed provides the appearance of objectivity as it creates rules for relationships between variables. However, as Loftus (1996) put it objectivity, is not, alas, sufficient for valuable insight. “Traditional modeling rules provide only the illusion of insight, which is worse than providing no insight at all” (Loftus, 1996, p.169).

Liebig’s Law of Minimum

Liebig’s Law of Minimum (LLM) has its origins in 1840. Justus Von Liebig provided evidence that in environments where one or more important inputs are in low concentration, the outcome will be affected regardless of optimum levels of others (Novais, 2007). In other words, it explains that the level of inputs and their effect on outcomes should be evaluated as a group (Paris, 1992). A classical application of the LLM on hard science relates to the assumption that when important nutrients are in relatively low concentration, the scarcest among them will restrain ultimate plant production levels. The LLM to this date is still one of the most important mechanisms for predictive inferences on ecological systems and has also been successfully applied to different areas of knowledge discovery. Previous research employs LLM's principals in ensembles of systems under different loads of factors that affect environments such as physiology, economics, and engineering (De Baar, 1994; Gorban, Pokidysheva, Smirnova, & Tyukina, 2011; Grimm, Paris, & Williams, 1987; Novais, 2007).

Gorban et al. (2011) approach LLM from the opposite end by exploring the assumption that when in fixed environments adaptation may represent a violation to LLM. The Law of the Minimum paradox proposes that adaptation may equalize the pressure of the absence of essential factors. That is, in well-adapted environments, we would have to expect violations of LLM (Gorban et al., 2011). However, the Law of the Minimum paradox was further contrasted by the Law of the Minimum Inverse paradox. Generally, it states that if many factors are equally important and they amplify each other, then after adaptation, a smaller amount of these factors is still fundamental to the respective outcome (Gorban, Smirnova, & Tyukina, 2010; Gorban et al., 2011). The image of a barrel has become popular as a representation of LLM and as a friendly way to explain the limiting effect of important factors (Figure 2). In a barrel built with staves of unequal length, the shortest stave will limit the maximum capacity of the barrel. Similarly, in plant crops the essential element with a too short or too large supply will limit maximum production. One way to apply LLM to management science would be to treat personality traits as environmental inputs and work-related outcomes as environmental outcomes.

In the context of evaluating the collective effect of factors and building on LLM, the limiting factor is approached based on its' relative quantity.

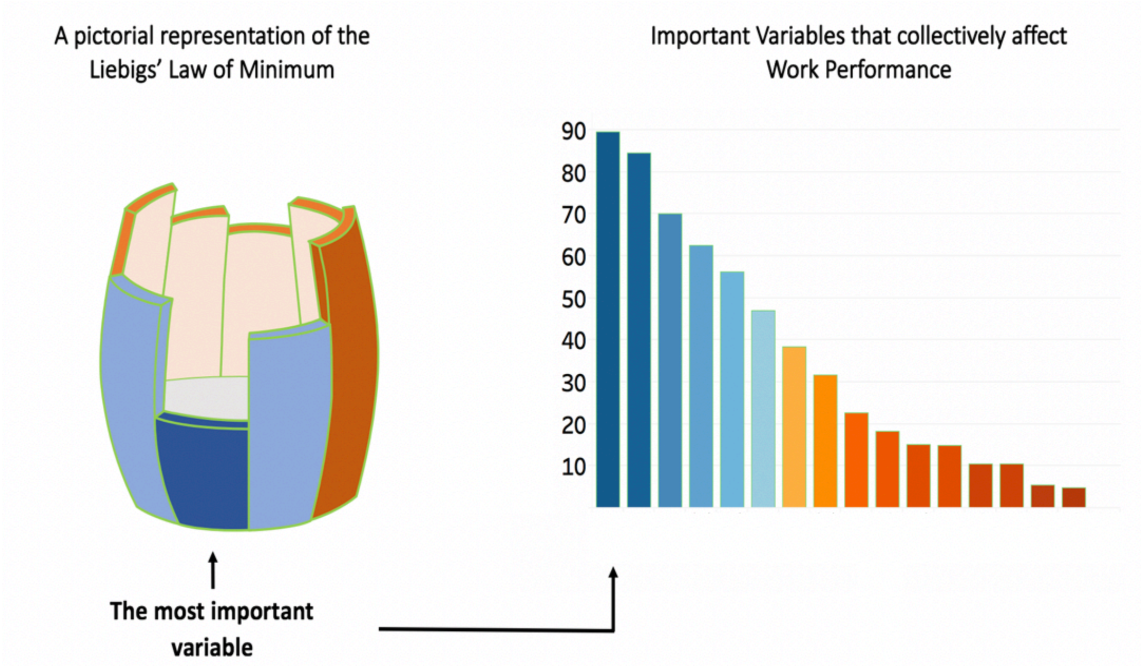


Figure 2. Pictorial Representation of Limiting Factors

Collective and Relative Influence of Elements in Nature

The collective effect of factors through their relative quantity is largely observable in nature. Research has found evidence of factors collectively influencing plants and social interactions in a number of studies (De Baar, 1994; Gorban et al., 2011; Grimm et al., 1987; Novais, 2007; Paris, 1992). In a similar fashion, the management research arena may benefit from the mechanics of this apparent widely applied law of nature (e.g., LLM) to scrutinize the predictive power of human traits on work-related outcomes. The following paragraphs will draw a parallel between the potential collective predictive power of several types of input variables and personality traits.

One of the key inferences of Liebig's Law relates to the assumption that essential factors that influence outcomes may not substitute for each other; instead they complement each other and exert a collective influence on the outcome as long as certain adequate levels of these factors are in place. In this sense, any increase in the scarcest element of a certain environment will affect the outcome until another element becomes the new limiting factor. For example, there are macro-nutrients in soil (e.g., phosphorous, potassium, magnesium, nitrogen) and micro-nutrients (e.g., iron, copper, zinc, and molybdenum). All of these elements play a crucial role on plant development concerning plant physiology. Some elements must be present at higher levels than others, which doesn't mean that in order to achieve high standards of plant growth or production all elements need to be present in high amounts. In other words, there is an optimal relative quantity for each element in the soil. This relative quantity affects plant growth and production because the environment lacks adequate levels of important factors.

Drawing on LLM, Warsi and Dykhuizen (2017) conduct a lab experiment with bacteria in which concentrations of nitrogen and magnesium on populations of *Escherichia coli* are tested. Results suggest that, "in low-nutrient environments, adaptation to the growth-limiting nutrient results in other nutrients at low concentrations to play a role in the evolutionary dynamics of the population" (Warsi & Dykhuizen, 2017, p. 1). This seems to be a good example of the above-mentioned Liebig's Law of the Minimum Inverse Paradox. In organic chemistry, not only the collective effect of nutrients is observed, but also their relative quantity. Take the example of the effect of the relative proportion between carbon and nitrogen on soil. The relative proportion of these elements affects the soil's organic matter quality and plant absorption rates of several important nutrients. According to Novais (2007), an ideal proportion of carbon in relation to nitrogen would be around 10:1 so that soil micro-organisms are able to transform organic matter into useful material. In a similar fashion, scholars who are considered pioneers of marine ecology propose an optimum ratio between

N:P as approximately 15:1 for marine environments (Cooper, 1937; De Baar, 1994; Harvey, Cooper, Lebour, & Russell, 1935).

In nature, there are several examples of the collective effect and the relative quantity of chemical elements on a number of biological or structural outcomes. The World Health Organization outlet published in 1996 mentions 17 trace elements that are considered essential to humans (Chapman, World Health Organization, Unesco, & United Nations Environment Programme, 1996). Based on estimates of their bioavailability, the committee report recommended ranges of intakes of adequate amounts of basic chemical elements to maintain or improve health. Similarly, several chemical elements exert collective influence on steel mechanical features. Silicon, for instance, improves oxidation resistance and fluidity of molten metal. In iron alloys, the element carbon when present at less than 2% helps to produce steel; but when its amount is superior to 2%, it is more likely to originate cast iron, which will have a whole different set of mechanical characteristics. Steel usually contains a minimum of 0.30% manganese because it assists in promoting greater strength by increasing the hardenability of the steel. Also, different types of cast iron such as Grey, Nodule, and Compact cast iron are composed by minimum amounts of certain essential elements to produce different mechanical features (De Campos, Lopes, Magina, Tavares, Kunioshi, & Golderstein, 2005). It seems that the collective effect and relative quantity of factors may indeed influence outcomes in several ways.

The extant management research provides diverse evidence regarding the influence of traits and types of behavior on work performance levels. Currently, the majority of studies approach the issue addressing predictors individually as functions of effect size, moderating, or mediating effect. LLM gives a hint for expanding management science by proposing a multicomponent and multifactor system model that assesses traits as performance predictors.

Modern machine learning techniques may bolster the robustness of the findings in a novel and powerful way by accurately predicting the effect of human traits on work performance.

Personality Theory

According to James and Mazerolle (2001), personality is described as an active system of cognitive structures and mental activities that governs one's emotions and behavior. It has been studied for many years from different approaches. Human behavior generates several theories and draws the attention of scholars from a diverse range of fields. Personality theory suggests that there is a recurring tendency in each individual's psychology, such as the individual's way of distinguishing and elaborating thoughts (Allport, 1961). According to Personality Theory, people are inherently biased and tend to consistently interact in the same manner across different situations (Allport, 1961; James & Mazerolle, 2001).

The main domain of Trait Activation is related to understanding in which type of situation an individual's trait is likely to emerge. Assessments centers, performance appraisals, and interview procedures have special interests in these approaches (Lievens, Chasteen, Day, & Christiansen, 2006). Employees who promptly adapt their behavior according to perceived situations may benefit by fostering a better quality of relationship. Traits that emerge in a certain situation for one person may not be the same to emerge in the same situation for another person. Surroundings may impact individuals' trait activation behaviors. Lievens et al. (2006) point out that it wouldn't be appropriate to evaluate one's trait for hostility in the course of a religious service because there would be rare cues to trigger the expected trait. Similarly, individuals who demonstrate an ability to manage their behavior in a frontline sales situation may not have the same trait-activation performance in a different work position. Van Hoye and Turban (2015) state that employee's traits are relevant and have different effects on attractiveness for organizations. They suggest that when an employee is aware of an ideal work profile, the fit of attractiveness based on personality is mediated. Although Trait-Activation Theory scholars report important progress (Lievens et al., 2006; Mussel & Spengler, 2015; Van Hoye & Turban, 2015), no study addresses the collective effect of those traits and how they may impact performance.

The Big Five Personality Traits

In the management literature, a number of studies point out that personality traits may affect performance (Shaffer & Postlethwaite, 2013; Sitser, Van der Linden, & Born, 2013; Wille, De Fruyt, & Feys, 2013). However, the majority of these studies assess the effect of traits individually. Barrick and Mount (1991) reveal associations between Conscientiousness and Job-Performance in several occupational groups. Extraversion was found to be a predictor for Social Interaction, while Openness to Experience and Extraversion are consistent predictors for training skills. These authors confirm that assessing the factors of the Big Five may be useful for numerous applications in the workplace, especially for personnel selection and training purposes. Shaffer and Postlethwaite (2013), suggest that Conscientiousness is predictive of performance when tasks are highly routinized. However, when Cognitive Ability is required, Conscientiousness would fail as a predictor for job performance. In line with this logic and in contrast with previous findings, Sitser et al. (2013) state that assessing the Big Five factors may not be the best way to predict sales performance. In a meta-analytic examination, Do and Minbashian (2014) explore previous findings about the correlation between extraversion and effective leadership. They find that a specific subset of extraversion has a positive impact on leadership, not the trait itself. The literature seems to be confused about which traits impact work-related outcomes. This could, to some extent, be due to the fact that organizations are different and have different ways of perceiving effective collaborators. A player within the financial sector may see effective employees as the ones who score high in Conscientiousness. On the other hand, advertisement firms may need collaborators high in Openness to New Experience or Extraversion, for instance. However, similar to what happens in nature, I attempt here to approach the issue from a novel angle. A minimum level of all important traits is required for a worker to achieve high standards of job performance. For example, a certain director of a certain company scores high in Conscientiousness, which — according to the literature — is one of the most common traits associated with job performance (Christiansen et al., 2014; Shaffer & Postlethwaite, 2013; Sitser et al., 2013; Wille et al., 2013). However, that same director may not have a minimum level of Humility, Openness to Experience, Agreeableness, Emotional Stability, or Extraversion that would allow him to be eligible to become the company's next CEO. On the other hand, another director who does not have that same high level of Conscientiousness, but rather meets the minimum required collective levels of other important traits may be the one who will become the next CEO. Machine learning techniques in conjunction with LLM seem to be an effective way to tackle the issue of whether minimum levels of certain traits may affect performance at some extent, even if they are not statistically significant in the eyes of the traditional data modeling culture. Approaching the collective effect of personality features may shed new light and help to illuminate the still-obscure role of traits on accurately predicting work performance outcomes.

HEXACO

In addition to the Big Five personality traits, many human factors have the potential to affect work performance. Humility is one strong candidate. George (2016) mentions that a precise sense of ones' skills, low self-focus, and the capacity to acknowledge limitations are sample features of humble people. When evaluating performance scores, individuals low in humility tend to rate themselves higher than is warranted by objective assessments of their performance (Hambrick & Chatterjee, 2007). Grijalva and Harms (2014) suggest that family boards with two or more less humble individuals who have an exaggerated sense of self and who strive to draw attention to themselves are likely to cause elevated levels of conflict. On the other hand, it is reasonable to assume that extreme high levels of humility may decrease self- esteem and thus affect performance. Grijalva and Harms (2014) bring to light significant differences between self and observed reports of narcissism, which is often perceived as opposed to humility. Leadership effectiveness is found to be positively related to high self-reported scores of narcissisms, but not with observed reports of this behavior. Hambrick and Chatterjee (2007) find interesting evidence pointing out that less narcissistic CEOs may perform better in dynamic business environments. In this sense, the authors infer that humbleness is a crucial trait when evaluating workers' performance. HEXACO embraces six personality factors: Honesty-Humility (H), Emotionality (E), Extraversion (X), Agreeableness (A), Conscientiousness (C), and Openness to Experience (O) and has been extensively explored by behavioral scholars (Ashton & Lee, 2007; Lee & Ashton, 2004). According to Ashton and Lee (2007), the HEXACO model represents a valid and reliable alternative to the largely employed Big Five-Factor Model. The description of the factors and primary dimensions are shown in Table 1 (adapted from George, 2016). The first column presents the primary dimensions, the second column presents the factors related to each dimension, and the third column presents a description of each factor.

Table 1. HEXACO Dimensions and Factors (George, 2016)

Honesty-Humility	Sincerity	Low scores use flattery and are often seen as "fake," whereas high scorers are viewed as being sincere and do not manipulate others
	Fairness	Low scorers might cheat or steal; high scorers are unlikely to take advantage of others
	Greed Avoidance	Low scorers want to enjoy and to display wealth and privilege; high scorers are not overly concerned with material possessions and social status
	Modesty	Low scorers consider themselves superior and entitled; high scorers see themselves as ordinary people
Emotional Control	Fearlessness	Low scorers are extremely fearful of physical harm; high scorers are relatively tough, brave, and not overly sensitive to physical injury
	Composure	Low scorers worry excessively, even with minor issues; high scorers remain calm, even about major issues
	Independence	Low scorers want encouragement and/or comfort from others; high scorers are self-assured and able to deal effectively with problems
	Stoical	Low scorers show strong emotions and have strong emotional attachments; high scorers show little emotion and have weak emotional attachments
Extraversion	Social Self-Esteem	High scorers have self-respect and see themselves as likeable; low scorers tend to feel worthless and unpopular
	Social Boldness	Low scorers are typically shy or awkward, particularly in leadership positions or large settings; high scores are comfortable leading groups and communicating with a variety of people.
	Sociability	Low scorers prefer solitary activities and work tasks; high scorers enjoy talking, visiting, and interacting with others
	Liveliness	Low scorers are generally not overly cheerful or dynamic; high scorers are generally enthusiastic and in high spirits
Agreeableness	Forgiveness	Low scorers might "hold a grudge" against those who have wronged them; high scorers can forgive and are willing to work towards re-establishing friendly relations
	Gentleness	Low scorers are generally critical of others; high scorers tend not to be judgmental of others
	Flexibility	Low scorers are viewed as stubborn and likely argumentative; high scorers tend not to be judgmental of others
	Patience	Low scorers tend to get angry or upset easily; high scorers generally are more tolerant before possibly getting angry or upset
Conscientiousness	Organization	Low scorers are generally sloppy or haphazard; high scorers are generally well-organized and prefer a structured approach to tasks
	Achievement	Low scorers lack self-discipline and are not strongly motivated to achieve; high scorers are strongly motivated to achieve due to a strong "work ethic"
	Detailed	Low scorers are tolerant of errors in their work; high scorers carefully check for mistakes and potential improvements
	Prudence	Low scorers follow impulses and do not consider consequences; high scorers consider multiple options and are generally careful and self-controlled

Openness	Aesthetic Appreciation	Tendency to see and enjoy beauty in art, physical surroundings, and nature. Low scorers don't care for art, aesthetics, or natural wonders; high scorers have a deep appreciation for a variety of art forms (e.g., nature, physical space)
	Inquisitiveness	Low scorers are generally not curious; high scorers are curious and prefer to know how things work or came to be
	Creativity	Low scorers have little inclination for original thought; high scorers actively seek new solutions to problems and express themselves in art
	Unconventionality	Low scorers avoid things that are out of the ordinary; high scorers are open to strange or out of the ordinary ideas

Although narcissism and humility are intuitively opposite sides of the same coin, Owens, Wallace, and Waldman (2015) demonstrate that they are not and provide sound arguments for such. Would Steve Jobs have developed Humility by his second incursion as the head of Apple, or was he simply a narcissist so high in Conscientiousness or Emotional Intelligence that he was able to activate (Trait Activation Theory) Humility, aiming to adjust to his new context? According to Owens et al. (2015), narcissists are persistent in their pursuit of goals despite adversity. Thus, depending on how hard a narcissist wants to achieve a certain goal, it is plausible to assume that he or she could activate a specific trait to comply with certain situations, thereby reaching the ultimate goal. Given the mixed and recurrent ambiguous results when addressing traits and their association with work-related outcomes, machine learning with its enhanced predictive capability may help to illuminate the field by uncovering what traits are actually important predictors of work performance in a novel and powerful way.

Performance

Focusing on understanding how students' demographics and background features may affect student performance, Cortez and Silva (2008) successfully apply machine learning techniques to select the most important features associated with students' performance. Using a binary classification for the dependent variable (i.e., pass, fail), the authors employ

Decision Trees (DT), Random Forests (RF), Neural Networks (NN), and Support Vector Machines (SVM). Sensitivity analysis shows that previous grade score was the most important factor when predicting students' performance. Minaei-Bidgoli, Kashy, Kotemeyer, and Punch (2003) suggest that genetic algorithm (GA) improves prediction accuracy by between 10% and 12% when compared to non-GA classifier. The authors use a combination of multiple classifiers (CMC) method and GA to improve classification rates and discover that "total number of correct answers" and "total number of tries" are the most important variables in the model. To assess variable importance, Minaei-Bidgoli et al. (2003) use entropy, which is a statistical property known as information gain. Generally, it measures how effectively a certain feature separates the training examples associated with the target classes. An experiment conducted by Kotsiantis, Pierrakeas, and Pintelas (2004) finds that the Naıve Bayes algorithm is most appropriate for predicting new undergrad students' performance. The potential predictors are demographics attributes and tutors' academic assessments. Demographics alone predict with 62% accuracy, but when adding tutor assessments as predictors, it is possible to predict with 82% accuracy which students passed or failed the class. Similarly, there are a number of academic works within the algorithm culture assessing how well demographics, math question levels of complexity, or student background features predict students' performance (Ramesh, Parkavi, & Ramar, 2013; Saarela, Yener, Zaki, & Kärkkäinen, 2016; Xu, Moon, & Van Der Schaar, 2017).

To measure work-related performance, the traditional statistical culture frequently employs task-related performance measures as a valid and reliable tool. For instance, Wallace, Edwards, Arnold, Frazier, and Finch (2009) use the Welbourne, Johnson, and Erez (1998) performance scale to investigate the relationship between work stressors, organizational support, and work-performance. Although there is a considerable amount of academic work supporting personality trait as a functional predictor for job performance (Christiansen et al., 2014; Harari et al. 2015), scholars like Morgeson et al. (2007) and Sitser et al. (2013) argue that when selecting working personnel, personality traits may have limited use. As suggested by Morgeson et al. (2007), one of the major problems related to an eventual personality prediction power could be related to the fact that self-reported questionnaires are subject to individuals' bias. Again, the high focus on predictive accuracy of modern machine learning algorithms may help shed light on this important management problem. Arguably, the management research community would collect scalable benefits by embracing and absorbing the algorithm culture into its academic outlets.

The Predictive Power of Machine Learning

According to Quinlan (1986), since artificial intelligence (AI) first achieved recognition as a discipline in the mid 1950s, machine learning has been approached as a fundamental research area. He gives reasons for this. Learning skills are a trademark of effective behavior; therefore, any attempt to identify or quantify intelligence as a phenomenon must necessarily understanding the process of learning behind it. Therefore, the process of learning provides a powerful methodology for building high performance systems to build valuable knowledge.

Perhaps "knowledge discovery" would a more suitable way to describe data mining techniques. Delen and Al-Hawamdesh (2009) elaborate on an interesting framework for the management community to optimize the knowledge discovery process. These authors suggest that high standards of knowledge discovery require a certain level of harmonization between individuals, technology, and information (Delen & Al-Hawamdeh, 2009). In this sense, the relatively recent explosion of Big Data and consequently intuitive data mining (i.e., knowledge discovery) tools represents a colossal opportunity to management scholars to extract knowledge for our field. Eichstaedt (2017) generates relevant data-driven insights by finding that tweets can be used to categorize counties according to their prevalence of heart disease. Tweets associated with hostility, aggression, and boredom predict counties where heart disease is high. He compares his predictions to actual Center for Disease Control (CDC) incidence ratings and finds that Twitter alone is a better predictor than all of the most common demographic risk factors combined (Eichstaedt, Schwartz, Kern, Park, Labarthe, Merchant, & Weeg, 2015; Eichstaedt, 2017). Interestingly, Sharda and Delen (2006) use neural networks to predict the financial performance of movies at the box office before their launch. Artificial neural networks perform significantly better than logistic regression and discriminant analysis, reinforcing once more the superior predictive power of machine learning techniques over traditional data modeling.

Medical research is a fields that has benefitted from machine learning techniques. Delen, Walker, and Kadam (2005) use a comparative study involving two popular data mining algorithms, neural networks and decision trees, in addition to the commonly used statistical method of logistic regression to predict breast cancer survivability. To avoid biased estimation the authors apply 10-fold cross-validation. Results highlight Decision Tree and its 93.6% accuracy as the best predictive method for that case. Using sensitivity analysis, which provides information about the relative importance of the input variables, the authors are able to detect the degree of differentiation of the tumor and the stage at which the cancer has spread, respectively, as the first and the second most important predictors. Bayat, Cuggia, Rossille, Kessler, and Frimat (2009) use Bayesian Network, Decision Tree, and Sensitivity Analysis to address the relative importance of factors such as age, diabetes, cardiovascular disease, and albumin on predicting access to renal transplantation waiting list. Age and cardiovascular disease were the first and second most important variables, respectively. The authors found that both Bayesian Network and Decision Tree algorithms predict with above

90% accuracy. Results are suggested as practical help for optimizing healthcare processes in that field. More recently, neuropsychiatry coupled with electrophysiology scholars and used machine learning techniques to support electroencephalographic features analysis as reliable predictors of working memory in schizophrenic and healthy adults. Support Vector Machine algorithms, which use different kernel functions and varying degrees of nonlinearity, predict with 84% accuracy (Johannesen, Bi, Jiang, Kenney, & Chen, 2016). Delen, Oztekin, and Kong (2010) employ a Cox Regression Model on three sets of variables to determine survival time after organ transplantation. Variables are selected from published literature, machine learning algorithms, and domain experts. For the machine learning data set, Support Vector Machine provides the best fit, which is correspondent to an R^2 value of 0.879. By using machine learning on the integrated Cox Regression Model, the authors are able to innovate creating a robust and effective way of assessing thoracic transplantation prognosis with substantial practical implications. Thus, whether nature produces outcomes through complex relationships between variables (i.e., black box for human brains) or through linear relationships, we reach a time where the management research field must embrace machine learning as one of its most powerful allies for revealing meaningful knowledge.

For organizations and scholars, it appears that the enhanced predictive power of machine learning techniques not only help reveal meaningful knowledge but also bolster managers' confidence in a number of decision making situations. Importantly, because behavioral science is somewhat subjective, data quality is a huge chapter of the story. In this sense, along the years management scholars have developed numerous ways of collecting reliable data that can be analyzed through the lens of machine learning. Behavioral scientists use an array of techniques to assess survey item reliabilities and construct validities. In Chapter III, I describe the good practices that ensure data quality for the purpose of this current work.

Assuming that data is reliable, data-driven decisions derived from machine learning research regarding human factors has been shown to be a flourishing alternative to traditional data modeling. For instance, Carnahan, Meyer, and Kuntz (2003) use curriculum scores and commercial driver license exam performances to provide evidence of superior predictive accuracy of machine learning classification models over discriminant analysis and logistic regression. The genetic algorithm employed, which is inspired by biological evolution, and the C4.5 algorithm predicted correctly four out of five test cases (80%). Similarly, assessing the applicability of LLM through machine learning techniques and reliable data on personality traits may be a novel way of revealing the actual predictive power of work performance traits.

Data mining far exceeds humans' brain ability to process knowledge in a number of fields. An interesting work conducted by Wang and Kosinski (2017) uses deep neural networks to analyzed features from 35,326 images of participants' faces. These facial features were processed in such a way that by evaluating one facial image, the algorithm could accurately predict the sexual orientation of the participants. The predictive accuracies were 81 % when distinguishing gay and heterosexual men and 74% when distinguishing gay and heterosexual women. The sports field is also an interesting area explored by data mining scientists. Delen, Cogdell, and Kasap (2012) developed regression and classification-type models to predict bowl outcomes. The data was originated from eight seasons of college football bowl games embracing a total of 244 games. Twenty-eight input variables involving game outcomes, team composition, and score differences when playing at home and away were used in the model. Their work reveals that Decision Trees as a classification technique using the 10-fold cross-validation produced the highest predictive accuracy of 86% (Delen et al., 2012; Sharda et al., 2016).

Diverse research areas have explored data mining techniques. Addressing the education field, Delen (2011) explores the causes behind freshman student's attrition. The author uses Artificial Neural Network (ANN), DT, and Logistic Regression as model types. Neural Network achieved the highest performance accuracy, which was 81% on the hold-out sample. The ANN architecture employed is known as Multi-Layer Perceptron (MLP) with back-propagation; it is a supervised learning algorithm. This ANN architecture is one of most commonly used by data mining scholars to learn arbitrarily complex nonlinear functions (Delen, 2011). Sharda et al. (2016) cited a number of successful cases where algorithms surpassed traditional data modeling predictive accuracy in fields like economics, politics, sports, medicine, and business. Given the relevance of data analytics to advancing science in a number of fields, it is surprising that management research has not yet extensively addressed human behavior through data mining tools. The current research allies management theory (i.e., Personality Theory) with modern predictive techniques (i.e., machine learning) to offer meaningful knowledge to both theorists and practitioners within the business domain.

Interval Versus Ordinal Scales

Although algorithms are capable of efficiently processing interval and ordinal measures, the nature of the data that will be assessed in this current work is a subject of concern. The dilemma concerning types of measurements applied to behavioral science is not new and still permeates debates among high-level scholars from the field. In the 1950s, ordinal scales were simply described as measures in which events are ordered in the same way as the arithmetic order of the numbers assigned to them. On the other hand, interval scales would be characterized by having equality of unit over different parts of the measure (Stevens, 1951). This led to the assumption that in ordinal scales the only transformation permissible is monotonic since it maintains rank order unchanged, while in interval scales linear transformations may be applied because they preserve relative distance unchanged. For example, when individuals respond to survey questions inquiring about the extent to which they agree or disagree with a certain statement, it is reasonable to assume that there is an implicit rank order assigned to these events. In this case, changing the arithmetic order of the numbers assigned to each possible answer would completely change the survey results. Following this rational, and in line with Anderson (1961), consider a measurement scale that assigns numbers to certain class events; this scale would be an ordinal scale and not necessarily an interval scale. Looking to clarify whether experimental subjects perceive Likert-type scales as ordinal or interval answers, Parker, McDaniel, and Crumpton-Young (2002) investigated distances and distributions of responses. Their study elaborates that when using a five-level ordinal scale, the normality assumption associated with parametric hypothesis testing such as analysis of variance (ANOVA) and *t*-test is likely to be violated. That is, if a certain Likert scale produces ordinal measures, the distances between the events cannot be said to be constant. So, employing traditional parametric tests would be subject to flawed conclusions. However, Parker et al. (2002) infer that the way questions are stated and displayed may influence whether respondents perceive questions as ordinal or interval.

The issue related to understanding whether respondents' answers are ordinal or interval brings up another interesting point: the process of dichotomization. Researchers have been using mean split as the most common method of dichotomization; it converts continuous variables into two groups of categorical variables. There are scholars pointing to benefits but also to statistical losses when using dichotomization on continuous variables. Farrington and Loeber (2000) provide evidence that working with dichotomization produces meaningful insights that have clear practical use and are easy to understand. Cohen (1983) comments that although dichotomization may result in significant reduction of statistical power, this loss may not be substantial when working with real data. On the other hand, traditional data modeling researchers argue that dichotomization is hardly justifiable and may yield ambiguous results (Aguinis, Gottfredson, & Wright, 2011; MacCallum, Zhang, Preacher, & Rucker, 2002). A revealing piece of research produced by DeCoster, Iselin, and Gallucci (2009), who are scholars from the social science and medical research fields, interviewed 118 scientists and scrutinized their justifications for using dichotomization on psychology fields. The majority of the scholars from diverse fields, including behavioral science, offered solid justifications for using dichotomization. Interestingly, the assumption that the relationship between the latent and outcome variable is nonlinear is one of the most cited reasons. Other reasons were pointed out as important (DeCoster et al., 2009), such as: “the latent variable has an irregular distribution” and “Results from analyses with dichotomized variables typically lead to the same conclusions as those with continuous variables.” These authors conclude that dichotomization is justifiable when the goal of the research is to evaluate how observed variables relate to the dichotomized measure being tested. It seems that in cases where it is assumed that nature produces data in nonlinear relationships and possibly not normally distributed, dichotomization practices align with machine learning’s high emphasis on predictive power and focus on meaningful practical implications.

CHAPTER III

METHODOLOGY

Respectable science involves both sound methods and rigorous theory (Gray, 2017). In the same fashion, Greenwald (2012) points out that meticulous theory when coupled with reliable methods are complementary in generating meaningful knowledge. As a controversial subject among scientists, theory has been challenged and considered secondary by many. Mischel (2008), for example, elaborates on the role of theory by commenting that within the psychology field, theories are often approached like a toothbrush problem. That is, “Everyone wants their own and no one wants to use others” (Gray, 2017; Mischel, 2008, p.737). Conversely, Greenwald (2012) advocates that new methods can uncover new knowledge that in turn shapes theory. On one hand, machine learning explores data with the major goal of improving models’ ability to predict outcomes. On the other hand, the dataset used in the current study is strongly based in rigorous theoretical studies (HEXACO factors). Therefore, the current work represents a unique opportunity to align science and theory with the potential to advance the management field to a whole new frontier of knowledge discovery.

Basically, there are two reasons for analyzing the data. Prediction, which relates to predicting outputs as responses to future input variables and information (i.e., data description), has to do with extracting useful information about how nature relates certain response variables to certain input variables. As stated in the previous section, there are two options for approaching these tasks, traditional data modeling and machine learning techniques. To merge science with theoretical assumptions about the HEXACO traits, which currently relies on stochastic data modeling types, I follow a widely used data mining procedure known as Cross Industry Standard Process for Data Mining (CRISP-DM) (Shearer, 2000). This six-step process embraces (1) exploring, understanding the research domain, and clarifying the study goals; (2) accessing and making sense of relevant related data sources; (3) working the data such that any required cleaning, preprocessing, and transformation are conducted; (4) studying and assessing different models through comparable analytical techniques; (5) analyzing the validity and implications of using the models and how well they attain the study goals; and (6) implementing the results for use in decision making processes (Delen et al., 2012). Figure 3 shows an illustration of the CRISP-DM procedure.

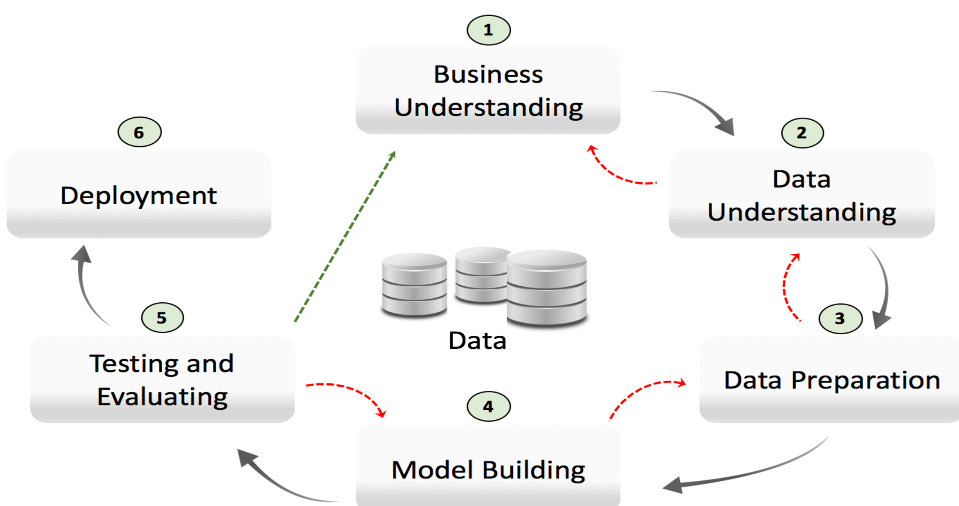


Figure 3. CRISP-DM Model (Adapted from Sharda et al., 2016)

CRISP-DM

Business understanding

Business understanding relates to clearly knowing what the study goals are. As was repeatedly stated in the previous sections, the management research arena has not yet employed machine learning techniques to predict work-related outcomes as function of scores on personality traits. Machine learning and its inherent focus on predictive accuracy calls several research domains to our attention. The main goal of the current study is to enrich the management literature with meaningful insights about the actual predictive power of traits when assessing work-related outcomes. Further, it aims to compare the predictive accuracy of stochastic data modeling techniques with machine learning algorithms.

Data understanding

According to Sharda et al. (2016), one of the most important steps in the data mining process is to identify relevant data. As for data understanding, a thorough literature review embracing the most important findings of the theoretical management domain is conducted. The data collection method as well as the items employed to measure the input and output variables are described in this chapter. To compare the predictive power of both algorithm and the traditional data modeling culture, I use the same final dataset across all models.

The present study uses secondary data from George (2016). Health care professionals from a large medical center completed an online Qualtrics survey to provide the dataset used here. All respondent answers were matched to their immediate supervisors. The electronic data collection platform also collected demographic information such as gender, age, race/ethnicity, work status, job function, job level, and tenure.

Performance

The current study approaches role-based performance embracing task performance, citizenship performance, and customer service performance. Generally, it addresses workers' ability to perform well in their job positions. Supervisors' performance ratings and employees' ePerformance scores computed from the firm's PeopleSoft system were matched and collapsed to form a role-based performance measure as the output variable. According to Welbourne et al. (1998), task performance relates to workers' ability to follow their job description while citizenship performance measures workers' concern for the organization. Customer service performance measures the extent to which employees excel in their relationship with customers/patients (Chen & Klimoski, 2003; George, 2016; Wallace et al, 2009).

Task Performance

The organizations' internal measure of task performance involves employees' performances on individual goals and competencies. Task performance was measured on a five-point Likert scale (1 = Far below expectations, 5 = Far exceeds expectations) (α = .92)

Citizenship Performance

To measure citizenship performance, George (2016) used a four-item questionnaire developed by Welbourne et al. (1998). Supervisors rate employees' citizenship behavior using a five-point Likert scale (1 = Needs much improvement, 5 = Excellent). Sample items are "The employee does things that help others when it's not part of his/her job" and "the employee volunteers for additional work," (α = 84).

Customer Service Performance

The four-item scale developed by Chen and Klimoski (2003) was employed to assess customer service performance as an additional component of the role-based performance measure. Supervisors rated employees on a five-point Likert scale (1 = Needs much improvement, 5 = Excellent). Sample items are “the employee interacts professionally with customers/patients” and “the employee establishes excellent relationships with customers/ patients,” (α = .92).

HEXACO

As previously stated, the six dimensions of the HEXACO personality inventory are Honesty-Humility (H), Emotionality (E), Extraversion (X), Agreeableness (A), Conscientiousness (C), and Openness to Experience (O). To assess these personality dimensions, participants answered the 96-item HEXACO questionnaire by Wallace and Edwards (2015). Using a five-point Likert scale (1 = Strongly Disagree, 5 = Strongly Agree), participants were asked to rate the extent to which they agreed or disagreed with the HEXACO-related statements (e.g., “I am a sincere person to those that I work with (H),” “I am deeply moved when others are upset (E),” “I am deeply moved when others are upset (X),” “I am generally a mild-mannered person when dealing with other people (A),” “I push myself hard to complete tasks successfully (C),” “I am a very curious person (O),” .

Demographic Information

The demographic characteristics of the 682 participants are presented in Figure 4 and are described as follows. A large majority of participants were females (*n* = 575, 84.3%). The age groups were separated as follows: participants in the age groups of 30 to 39 years of age (*n* = 183, 26.8%), 40 to 49 years of age (*n* = 170, 24.9%), and 50 to 59 years of age (*n* = 150, 22.0%). African American participants accounted for the majority of responses (*n* = 365, 53.5%) as opposed to Caucasian (*n* = 214, 31.4%). Full-time workers were *n* = 427, 73.4%, as compared to part time workers (*n* = 109, 18.8%). Nurses represented the largest number of participants (*n* = 291, 42.7%). Work tenure groups were divided as follows: from 0 to 4 years (*n* = 256, 37.5%), 5 to 9 years (*n* = 136, 19.9%), and 10 to 14 years in current job (*n* = 140, 20.5%).

Reliability measures and descriptive statistics are reported to better describe and understand the data (as shown in Table 2). The data set is provided in a numeric fashion in order to allow proper assessment of the predictive accuracy of the different modeling techniques; some data transformation will be conducted.

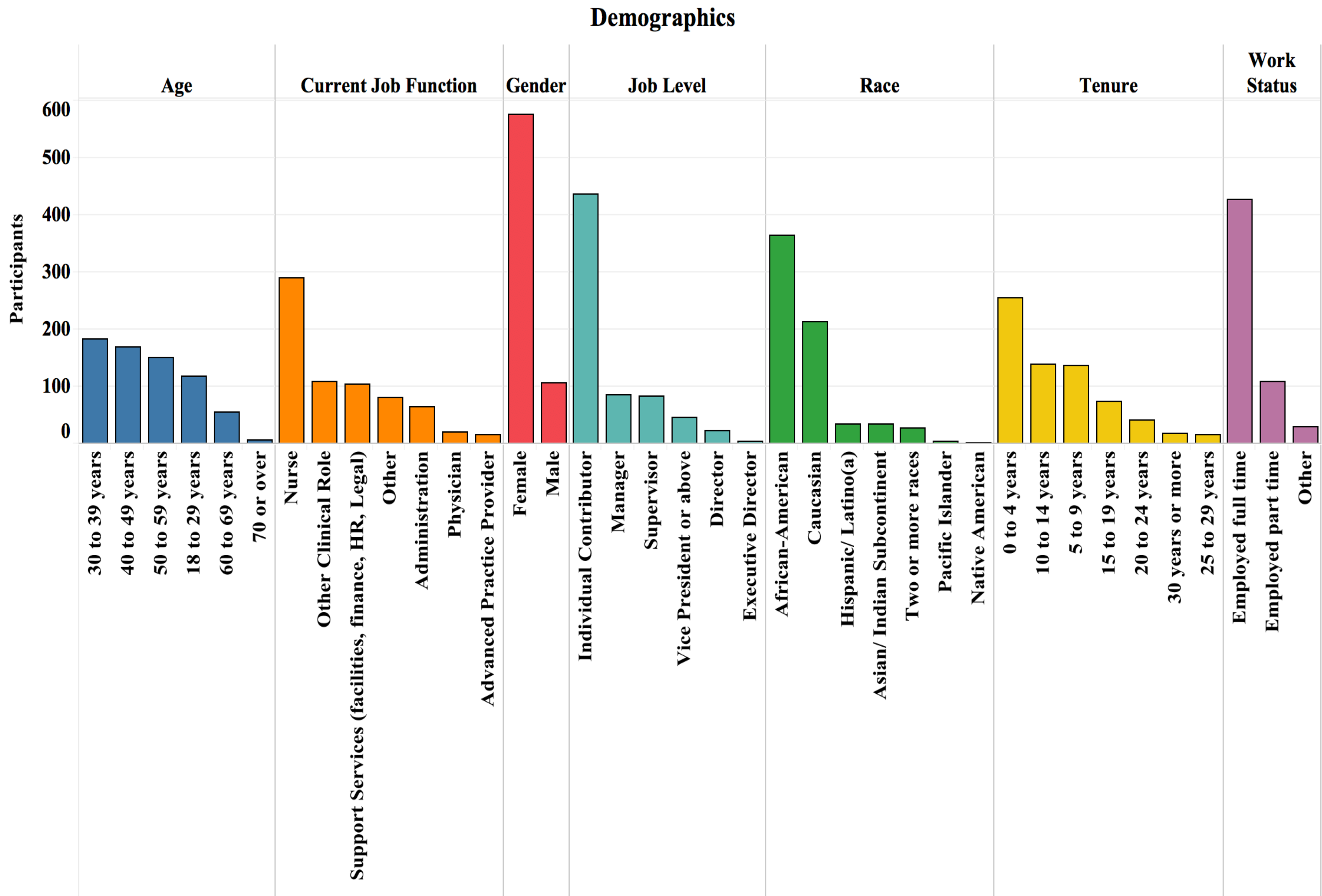


Figure 4. Demographic Information for the Dataset

Table 2. Descriptive Statistics and Bivariate Correlations for Composite Scores

Variable	M	SD	H	E	X	A	C	O	RBP
H	3.65	0.42	**-0.719**						
E	2.79	0.44	0.160**	**-0.70**					
X	3.51	0.57	0.150**	0.15**	**-0.70**				
A	3.42	0.74	0.300**	-0.09*	0.23**	**-0.787**			
C	4.03	0.46	0.380**	0.08*	0.50**	0.230**	**-0.864**		
O	3.53	0.49	-0.020	-0.04	0.52**	0.200**	0.410**	**-0.836**	
RBP	3.23	0.71	0.210**	0.03	0.22**	0.090*	0.210**	0.300	**-0.923**

N = 682; Coefficient a in parentheses.;**Significant at the <0.01 level (2-tailed); *Significant at the <0.05 level (2-tailed)

Data preparation

Frequently, the data preparation or data preprocessing step is the one that consumes the most time. According to Sharda et al. (2016), it can account for roughly 80% of the time dedicated to a project.

Although data imputation methods are commonly applied by researchers in general, Delen et al. (2006) point out that even when using sophisticated imputation techniques, imputed values may create biased results because they are not real. Using the same dataset, George (2016) justified not including the missing values on the final dataset by making sure that there was enough data to properly conduct the statistical analyses and by not disrupting the original distribution of the variables (Delen et al., 2006; George, 2016). In line with these researchers, I use no imputation technique on the current dataset. Therefore, from 723 participants who took the survey, 41 answers were removed because of missing data points. The final number of participants was 682.

To test algorithms and statistical techniques such as ANN, RF, and Logistic Regression, data transformation is conducted. For example, in some cases I normalize the data by reducing the range of score values to a standard range between 0 and 1 across all the input variables. Also, in some cases, I conducted dichotomization (e.g., mean split) of the output variables with two primary goals. The first goal is to produce meaningful insights that have clear practical use and are easy to understand (Farrington, & Loeber, 2000); the second is to properly assess an assumed nonlinear relationship between the input variables and the outcome (DeCoster et al., 2009).

Neural Networks

Neural networks (NN) come from a family of machine learning techniques that are based on the biological neural network functioning process (e.g., human brain). This technique has its roots in the 1960s. Since then it has been improved to become one of the most largely used machine learning techniques. NN is often used to explore complex and nonlinear relationships between predictors and outcomes in diverse research realms such as medicine, sports, finance, and manufacturing (Haykin, 2008; Delen et al. 2012). In NN the models predict results of new observations by "learning" how patterns of pre-existing events led to the outcome. In the current study, I use a popular NN mechanism called multi-layer perceptron (MLP) with back-propagation type in a supervised-learning algorithm. Concerning the different types of NN, the feed-forward back-propagation is the first and still most popular structural mechanism of NN (Wu, Jennings, Terpenny, Gao, & Kumara, 2017). Delen et al. (2012) used MPL as a powerful NN architecture to produce classification and regression prediction models having the outcome variable labeled as both nominal and numeric. As a type of NN, MLP consists of several processing elements (nonlinear neurons called perceptrons) arranged in layers that are connected in a feed-forward, multi-layer process. In a single hidden layer structure, the input layer transmits the data/signal to one hidden layer that then passes it to the output layer. Next, the final signal in the output layer is compared to the original observation and the noise (error) is then fed back to the network so that a continuous correction of parameters and weights is carried on in an ongoing process called "learning" (Delen, Tomak, Topuz, & Eryarsoy, 2017).

Decision Tree

The literature points out that Decision Trees have been developed since 1930s. According to Delen et al. (2017), in its early stages, Decision Trees relied heavily on expert knowledge (deductive approach) instead of using data (inductive approach). With the explosion of the internet and the huge amount of data generated and stored for decision-making processes in several fields, Decision Trees emerged as a popular and important complementary tool for data mining purposes. The inner structure of the Decision Trees explains how the predictions are achieved. This is the biggest advantage of Decision Trees and Random Forests (i.e., ensemble models of several Decision Trees) over more complex machine learning tools such as NN and SVM.

Generally, Decision Trees recurrently split the training set of the data until each one of the divisions contain a pure representation with members of the same class or until it reaches a predetermined stopping condition. The split points (nonleaf nodes) test attributes and determine whether the data will be split. To determine the splitting point, evaluating the goodness of the split, the information gain, and the Gini index are the most popular splitting indices (Sharda et al., 2016). In this present study, I employ the C4.5 Decision Tree algorithm developed by Quinlan (1996) that uses information gain as a form of evaluating the goodness of split at a nonleaf node level.

Random Forest

Basically, ensembles refer to aggregating records from two or more information sources. Similarly, in machine learning, ensemble models combine information from two or more models (e.g., Decision Trees, Neural Network, Logistic Regression) to generate robust and reliable prediction information (Sharda et al., 2016). Although ensemble models are usually more accurate than composing models (Seni & Elder, 2010), they may also increase the model complexity, which can make the process of understanding the underlying mechanism that generated the predictive accuracy a difficult task.

Random Forest is an ensemble model largely used by the machine learning community. Basically, it develops a number of small trees from which information is then computed and aggregated. According to Breiman (2001b), Random Forest combines tree predictors in a way that each tree relates to a randomly sampled vector with the same distribution for all trees in the model. Breiman points out that “internal estimates monitor error, strength, and correlation and these are used to show the response to increasing the number of features used in the splitting. Internal estimates are also used to measure variable importance” (2001, p.1). This current study uses Random Forest, among the other prediction models, to reveal the predictive power of traits when targeting work performance.

Logistic Regression

Since its inception in the 1940s, logistic regression has become one of the most popular statistical techniques for predicting dichotomous classifications. It is based on probabilistic assumptions and employs a supervised-based expectation maximization algorithm (Delen et al., 2012). Logistic Regression uses an exploratory technique by which instead of predicting data points, it estimates the odd ratio of the potential occurrence of this point. Two limitations of this predictive technique are restrictive assumptions of independence between input variables and a normal distribution of the data set. The present work uses logistic regression representing the traditional data modeling approach to evaluate the predictive efficacy of the models.

General Linear Regression

Basically, the General Linear Model or Multivariate Regression Model may be written as

$Y = \beta_0 + \beta_1 X_i + U$, where Y represents a matrix of outcome measurements, X represents a matrix of the input variables, β (i.e,. coefficient) represents a matrix of the parameters that will be estimated, and U represents the random error (Christensen, 2011). Two basic assumptions of a General Linear Model are a multivariate normal distribution of the data and a situation in which the measurement errors are not correlated. Examples of General Linear approaches are ANOVA, multivariate analysis of the variance (MANOVA), multivariate analysis of covariance (MANCOVA), and regression analyses. My work here uses a general linear model to test the predictive accuracy of stochastic models against machine learning techniques.

Testing and evaluating − Cross-validation

At a high level, the cross-validation methodology splits the data into two mutually exclusive subsets. The training subset is used to build the model, while the test subset assesses the predictive power of the model. It is possible that one single split of the dataset may incur uneven representations of the training and test subsets (Delen et al., 2012). To avoid the nonhomogeneity of the subsets and a potential bias in the trained model, I will employ multirounds of cross-validation. This procedure is called *K*-fold cross-validation where *K* represents the number of splits that will be performed on the dataset so that *K* number of equal-sized subsets is obtained. Several *K* rounds of training and testing the model will be conducted. On each round, the model is trained in all but one ($K - 1$) fold and tested in the excluded fold, which is the testing subset for that round. The average of the test outcomes from all *K* times that the process is run is then compiled for analysis. According to Delen et al. (2012), because the cross-validation method relies on the random assignment of single samples to *K* folds, it is convenient to stratify the folds to reduce bias. My work employs a stratified *K*-fold cross-validation so that the folds have approximately the same proportion of class variables as the original dataset. Olson and Delen (2008) point out that using stratified cross-validation tends to reduce bias when compared to regular cross-validation. In line with Delen et al. (2012), my study will set the value of *K* to 10. Equation 1 shows that the overall accuracy from the cross-validation procedure is calculated as a function of the average of the *K* single accuracy measures.

$$CV = \frac{1}{k}\sum_{i=1}^{k} A_i \tag{1}$$

In the above equation, *CV* represents the cross-validation accuracy, *A* represents the accuracy measure of the *K* folds, and *K* is the number of folds that were generated in the *K* fold cross-validation settings (Delen et al., 2012).

To Compare Model Performances

I will employ the same commonly used methods to evaluate and compare the predictive accuracy of the models on both algorithmic and traditional data modeling cultures. In addition to looking at differences in the predictive power between machine learning techniques and stochastic data modeling, I seek to uncover which traits likely exert stronger influence on the outcome variable. I will apply Accuracy, Sensitivity, and Specificity as the performance criteria to assess the models. Equation 2 shows how True Positives (TP), True Negatives (TN), False Positives (FP) and False Negatives (FN) interact together to generate the measures of performances. For my purposes, TP refers to accurate predictions of high performance, TN refers to accurate predictions of low performance, FP refers to inaccurate predictions of low performance as high, and FN refers to inaccurate predictions of high performance as low. In this sense, Accuracy measures how well the model predictions work to indicate the overall probability of accurate predicting performance. Sensitivity and Specificity address how precise the model is when predicting high and low performance respectively and individually.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \tag{2}$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{3}$$

$$\text{Specificity} = \frac{TN}{TN+FP} \tag{4}$$

To compare the general linear model prediction outcomes that will provide numerical values for predicted performance, I will use dichotomization to assign class variables to both actual and predicted work performance. This post-hoc dichotomization procedure and the subsequent confusion matrix will allow for a direct comparison using the same performance criteria across all models. As stated in previous sections, dichotomization has been reported as a useful technique when assessing real data and when there is an easy-to-understand focus on the predictive power of models (Farrington & Loeber, 2000).

In addition to the direct comparison of the models by scores on Accuracy, Sensitivity, and Specificity, I will apply a pairwise two-tail *t*-test on the errors generated by the *K*- fold cross validation. Each time the training set is split and tested, an error is generated. The *K* error measures generated by each model will be compared using 1% and 5% levels of significance.

Variable Importance

To assess the relevance order of variables, I will employ the actual splitting rate (ASR) using data from the attribute statistics of the Random Forest model. Basically, this Random Forest-based heuristic method assesses variable importance by computing the ratio of the number of actual splits on a certain variable to the number of times that particular variable is selected as a candidate to split within the forest. Random Forest as an extension of the Classification and Regression Tree (CART) models randomly select input variables at each node for each tree within the forest with no pruning rule (i.e., stopping rule) (Breiman, 2001a.). This method builds on the notion of reduced entropy (lack of predictability) at each time a certain variable is used to split a node (Minaei-Bidgoli et al., 2003). Basically, the decrease in the Gini impurity criterion is computed and then the average of all decreases for that particular variable Gini impurity; each time it generates a split in the forest determines the splitting hierarchy (Archer & Kimes, 2008). Generally, the number of times that each input variable is chosen to be split (i.e., candidate to split) as well as the number of times that particular variable was actually split (i.e., split) are computed and informed by most data mining software providers. To eliminate potential bias on the Random Forest algorithm, my model computes 1,000 trees on the forest learner node in substitution for the standard 100 trees. This procedure ensures that all variables are properly selected as candidates to split. Thus, given that most of the machine learning algorithms' prediction mechanisms are difficult for human brains to grasp (i.e., black box), I choose to apply ASR as a Random Forest-based heuristic method to assess variable importance.

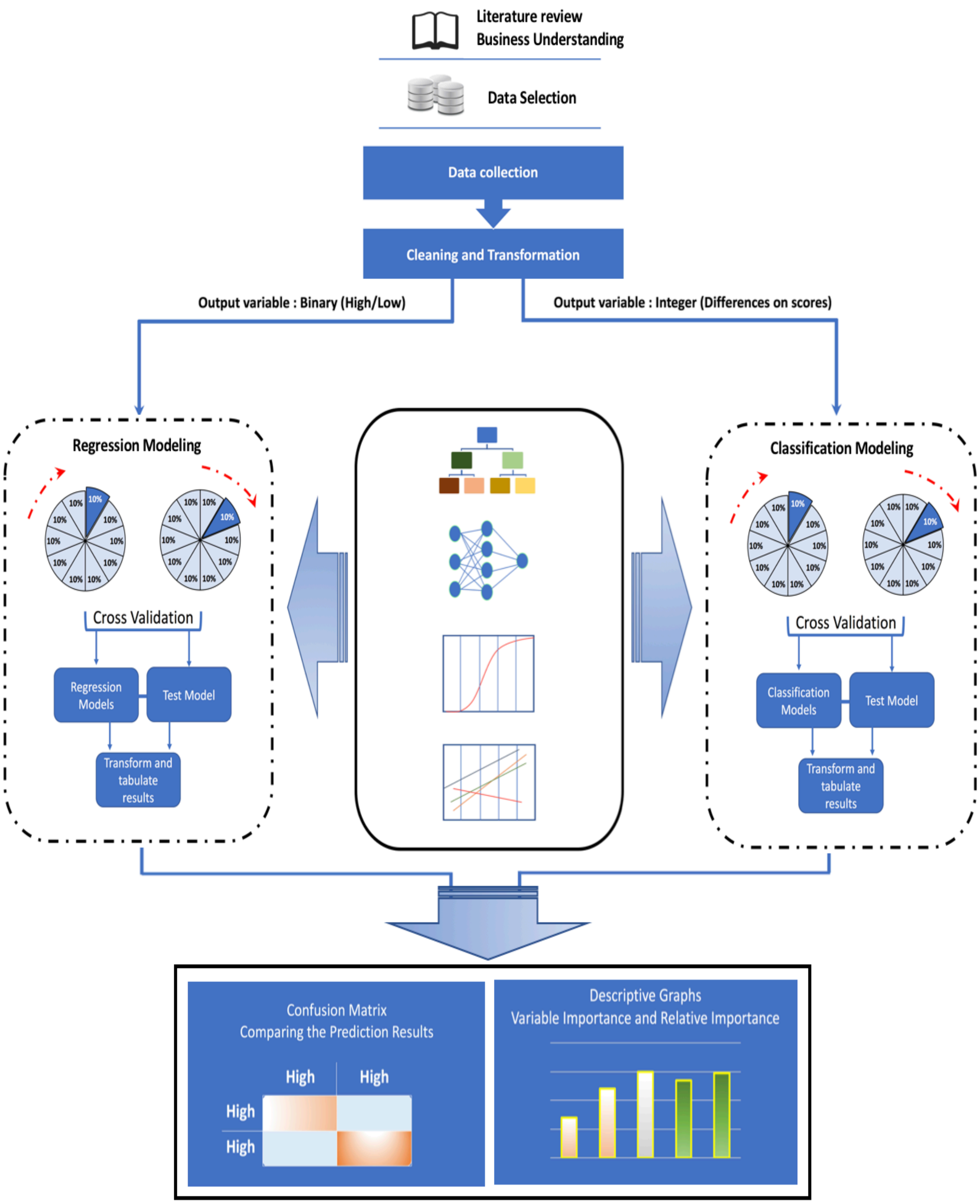
Assessing a Potential "Ideal Proportion" of Traits

My study will assess the ratio of the most important traits as potential predictors of work performance. In nature, several examples of the influence of the ratio (i.e., relative proportion) between relevant factors to produce certain outcomes exist. For example, as previously stated, the ideal proportion of carbon relative to nitrogen in the soil is around 10:1 so that plants can grow proficiently (Novais, 2007). Similarly, a ratio of nitrogen and phosphorous around 15:1 is considered optimal for marine ecosystems (Cooper, 1937). In the same vein but in a different realm, Delen, Kuzey, and Uyar (2013) pointed out sets of financial ratios as relevant predictors for firm performance. The Earnings Before Tax-to-Equity Ratio was the leading predictive variable for that particular study. In line with LLM and with what nature presents concerning the relative and collective effect of input factors on a number of outcomes, I will assess the predictive performance of individual traits as well as how well the ratio between the most important traits predict work performance.

It is inherently crucial for several research domains to determine which variables are the most relevant predictors to explore the study goal. Also, the removal of redundant or unnecessary input variables may reduce overfitting (i.e., memorizing the data set instead of identifying the underlying causal effect and the distribution) and help to achieve accurate models. Importantly, asking the right questions with respect to what predictors are actual predictors to a certain outcome may save money, time, and effort (Dreiseitl & Ohno-Machado, 2002). This work relies on the extant management theories and findings to investigate the HEXACO subfactors individually and the ratio between them as potential predictors for work performance.

Deployment

To deploy the model results for use in decision making, I use Tableau, which is a commercial analytics software with enhanced descriptive analytics features. Measures of model accuracy, sensitivity, specificity, and variable importance are elaborated on Tableau charts as well as on MS Excel documents. According to Sharda et al. (2016), the deployment phase does not constitute the end of the data mining project as it needs to be constantly revisited so that new and perhaps more effective machine learning tools may be applied. The evolving and changing nature of human behavior patterns across the years represents a solid reason for recursive efforts toward understanding the predictive power of traits when predicting work performance. In a novel way, drawing on LLM and in line with what nature shows regarding the collective influence of relevant elements on several outcomes, this current work addresses human traits, their relative importance, and their collective effect on work performance. Figure 5 shows a graphical representation of the methodology used (adapted from Delen et al., 2012).



* Adapted from Delen et. al. (2012)

Figure 5. Current Study Methodology

CHAPTER IV

RESULTS

The prediction results of the five modeling techniques are shown in Table 3. The binary output variable containing Low Performance or High Performance classes reflect the median split from the role-based performance (RBP) measure. As stated in previous sections, for RF, ANN, DT, and Logistic Regression, I conducted ad-hoc dichotomization; for Multiple Regression, I used post-hoc dichotomization. The median value for RBP was 3.17.

The confusion matrixes from the 10-fold cross-validation results demonstrate the superior overall accuracy (79.17%) of the RF model over the other four models. As shown, the classification-type prediction models were more effective when predicting work performance than the regression-based models. DT was the second best prediction method with 76.09% accuracy, followed by ANN with 65.54%. Multiple Regression and Logistic Regression were, respectively, the fourth and fifth best methods, achieving 61.73% and 58.35% accuracy. Examination of the sensitivity and specificity measures showed once again that machine learning algorithms outperformed traditional regression-based statistical analysis. I employed a pair-wise *t*-test on the error rates from the *K*-fold cross-validation analysis to explore whether differences on overall accuracy measures

Table 3. Tabulation of Prediction Results Based on the Ten-Fold Cross Validation Methodology

Model Type		Confusion Matrices		Accuracy (%)	Sensitivity (%)	Specificity (%)
		Low Performance	High Performance			
Artificial Neural Networks (ANN)	Low Performance	262	117	65.54	62.25	68.94
	High Performance	118	185			
Random Forest	Low Performance	299	80	**79.17**	73.33	78.27
	High Performance	83	220			
Decision Tree	Low Performance	213	166	76.09	73.33	78.27
	High Performance	118	185			
Logistic Regression	Low Performance	318	61	58.35	52.70	64.35
	High Performance	81	222			
Multiple Regression	Low Performance	202	177	61.73	55.30	70.62
	High Performance	84	219			

In bold = Highest predictive accuracy

between the classification models were significant. In Table 4, I show that the RF model accuracy is significantly higher than DT, ANN, and Logistic Regression. DT accuracy is significantly higher than Logistic Regression, but not higher than ANN. Importantly, Logistic Regression's accuracy is significantly lower than RF, DT, and ANN at 1% significance.

Table 4. Tabulation of the *t*-Test (*p*-Values) for Accuracy Measures of the Four Classification Prediction Methods

	Artificial Neural Networks (ANN)	Random Forest	Decision Tree	Logistic Regression
Artificial Neural Networks (ANN)	—	0.0034^{***}	0.1155	0.0046^{***}
Random Forest	0.0034^{***}	—	0.0196^{***}	0.0002^{***}
Decision Tree	0.1155	0.0196^{***}	—	0.0003^{***}
Logistic Regression	0.0046^{***}	0.0001^{***}	0.0003^{***}	—

$^{*}p$-value < 0.05, $^{***}p$-value < 0.01

Because the Multiple Regression model used post-hoc dichotomization, I compare its prediction accuracy using exclusively the measures of overall accuracy, sensitivity, and specificity. The summary of model fit for the regression model is shown in Table 5.

Table 5. Summary of Model Fit of the Multiple Regression Model

R^2	0.205203
R^2 Adj.	0.171123
Root Mean Square Error	0.644482
Mean of Response	3.231305
Observations (or Sum Weights)	682.000000
Prob > F	< 0.000100

In addition to providing more understandable results that are easier to grasp, RF and DT algorithms allow for a more comprehensive analysis of variable importance. Figure 6 shows the importance order of factors that may affect work performance according to the ASR analysis from the RF attribute statistics.

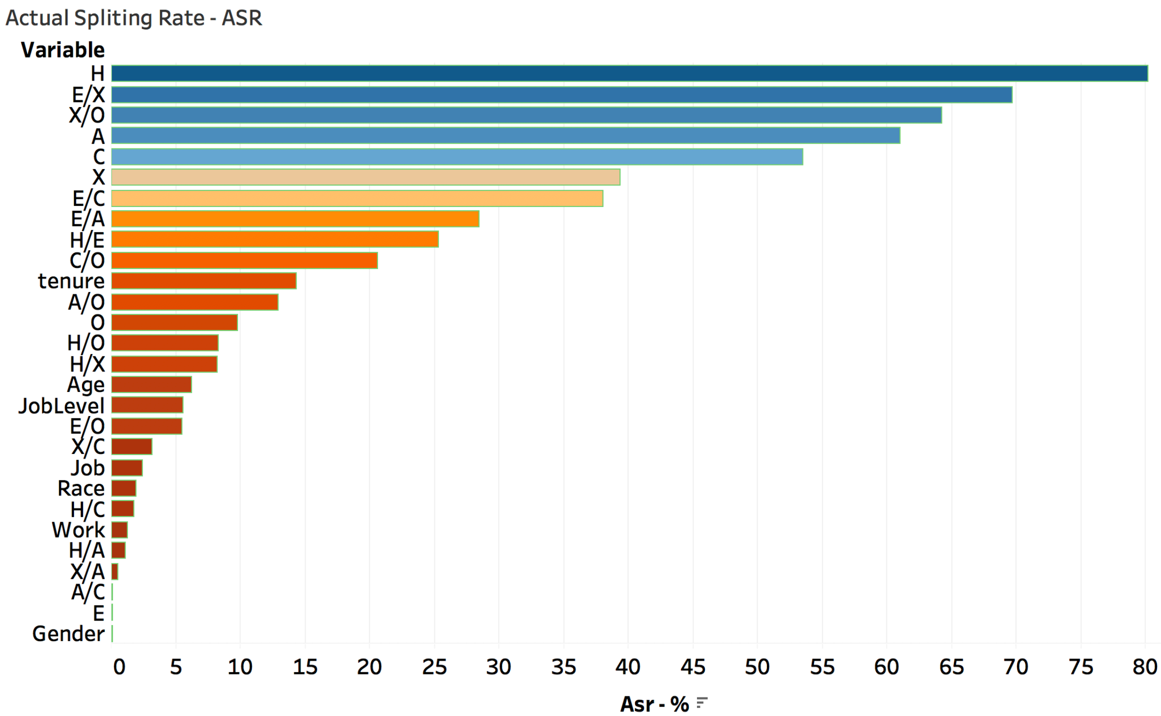


Figure 6. Variable Importance Order

Within the RF algorithm, the variable Humility was split 80.11% of the times it was chosen as a candidate to become a splitting point. The second most important variable was the ratio between Emotional Stability and Conscientiousness with 69.64% ASR, followed by the ratio between Extraversion and Openness to Experience with 64.20%. The fourth and fifth important variables were Agreeableness and Conscientiousness with 61.02% and 53.44% ASR, respectively. The first group of important variables embracing the factors mentioned above are consistent with existing regression-based studies. For example, Grijalva and Harms (2014) suggest the trait Humility as one of the strongest candidates to exert positive influence on Performance. These authors imply that family businesses' boards with two or more less humble individuals who have an exaggerated sense of self and who strive to draw attention to themselves are likely to experience elevated levels of conflict. As humble people tend to show a low self-focus and a precise sense of their own skills, they are more prone to strive hard and overcome their limitations, which in turn increases Performance. In this line, Hambrick and Chatterjee (2007) find scientific support for the assumption that humble CEOs may perform better in dynamic business environments. Although there are not many studies pointing to a potential strong positive relationship between Agreeableness and Performance, the regression-based specific literature finds consistently that Conscientiousness may be a strong predictor for Performance. Barrick and Mount (1991) suggest a positive relationship between Conscientiousness and Performance in several occupational groups. Interestingly, Shaffer and Postlethwaite (2013) call attention the notion that Conscientiousness may be a valid predictor of Performance when tasks are highly routinized. That said, results reveal the ratio between Emotional Stability and Conscientiousness and the ratio between Extraversion and Openness to Experience as strong predictors for Performance with over 80% prediction accuracy. In short, this means that consistently balanced levels of those traits are likely to produce high Performance ratings. It is worth noting that a certain proportion between Emotional Stability and Extraversion is found to be more important than the scores on these traits when they are evaluated separately. According to most the accurate algorithm, the same type of analogy can be developed for certain proportions of Extraversion and Openness to Experience. The ratio between scores on these traits may be said to be a better predictor for Performance than when scores on these traits are evaluated separately.

The second group of predictors include Extraversion, the ratios between Emotional Stability and Conscientiousness, Emotional Stability and Agreeableness, Humility and Emotional Stability, and Conscientiousness and Openness to Experience. These variables are found to be moderately important when the goal is predicting Performance. Although Extraversion has been suggested as a potential predictor for better social interactions, training skills, and performance (Barrick & Mount, 1991; Sitser et al., 2013; Do & Minbashian, 2014), in the present study this trait was not among the most important Performance predictors.

According to the ASR computed from the RF attribute statistics, the third group with the least important variables include all demographics, Openness to Experience, Emotional Stability, and the remaining possible ratios between the HEXACO traits. The contribution of this group of variables was rather marginal.

For the purpose of illustration, Figure 6 shows a pictorial representation of a DT from the RF learner node.

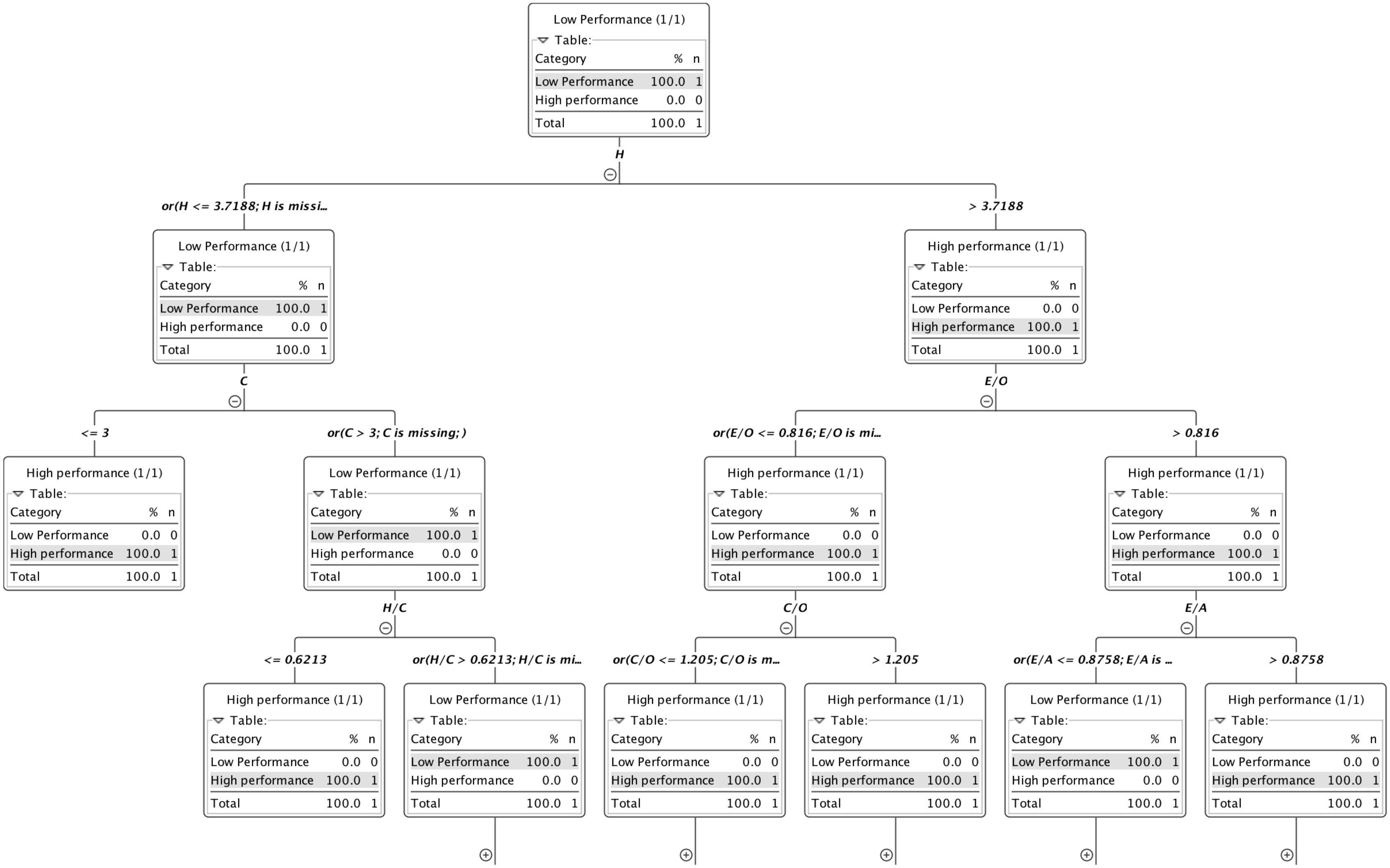


Figure 7. A Decision Tree View From the Random Forest Learner Node

Although results point out Humility as the most important variable, it doesn't mean that scores on this particular variable need to be as high as possible to achieve high work performance. Rather, it informs us that when all other factors are collectively considered, Humility exerts the strongest influence on the outcome as the most important predictor. Figure 7 illustrates the order of important factors by drawing a parallel between the LLM and the variable importance order from the RF algorithm. Figure 8 displays the average scores of high Performance employees. An examination of Figures 8 and 9 makes it easier to visualize that although scores on Conscientiousness and Agreeableness (i.e., C and A) are the highest ones, for high performers their relative importance order are fifth and forth, respectively. It is also worth noting that even though the ratio between Emotional Stability and Extraversion is pointed out as the second most important variable, its absolute value is the third lowest for high Performance. This is in line with the algorithm culture assumption that in nature, data is generated in complex ways that are not necessarily normally distributed nor linearly correlated. As results from this study indicate, the pattern recognition capability with impressive predictive accuracy of modern machine learning techniques has been shown to be an existing and reliable way to unearth human behavior-related knowledge for the management field.

Working Predictive Analytics Results with Descriptive Analytics

Following most organizations' tendencies to replace their traditional flat reports with modern and more interactive data visualization tools. The present research employs descriptive analytics with aiming to produce valuable insights with regard to potential associations between human traits and job performance.

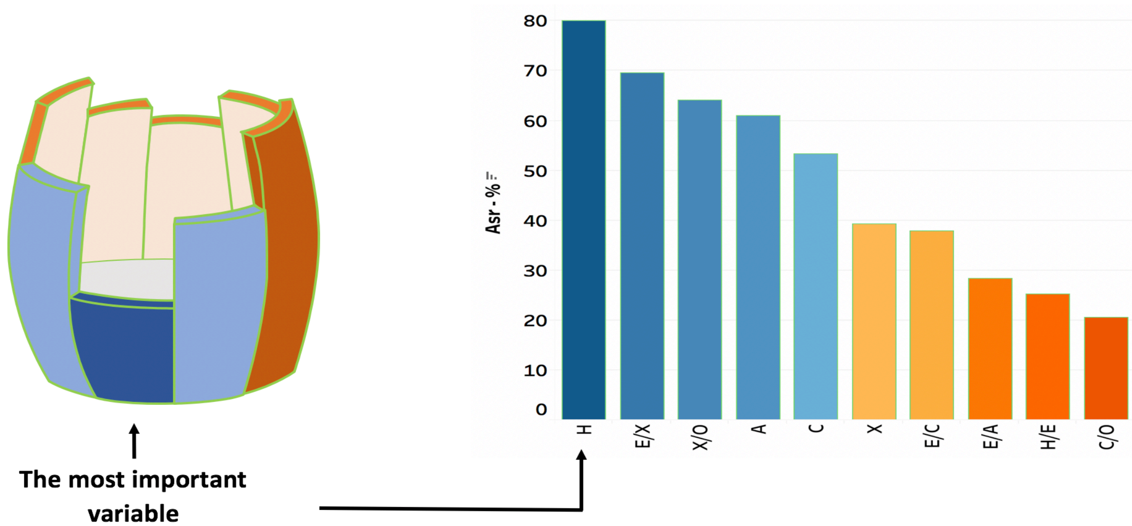


Figure 8. A Pictorial Representation of the Factors that Collectively Affect Work Performance

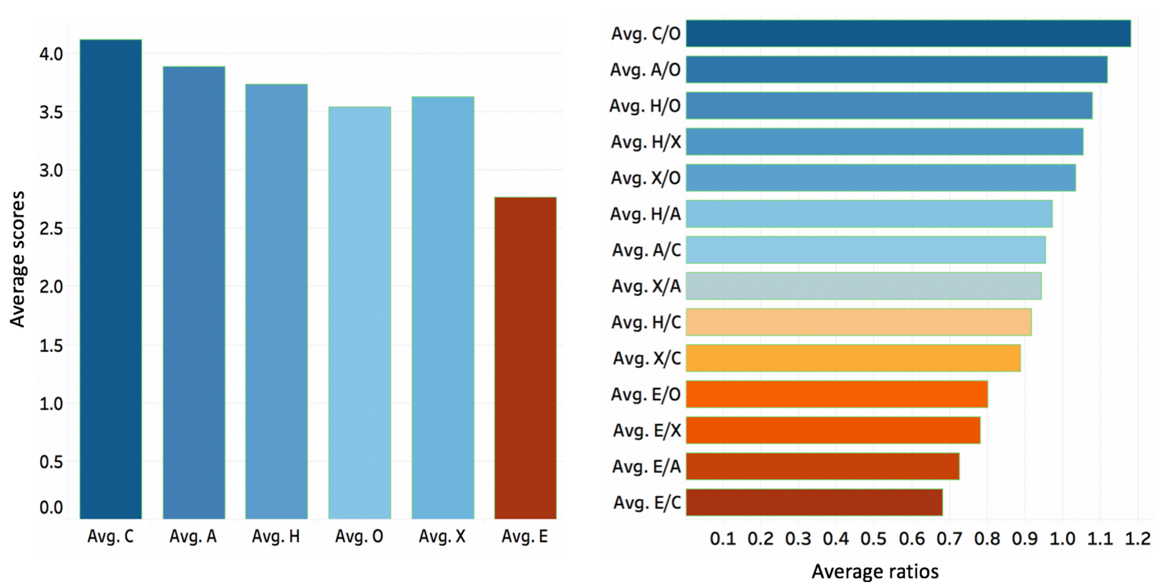


Figure 9. A Pictorial Illustration of Average Scores and Ratios Between Scores on Traits Related to High Work Performance

According to Sharda et al. (2016), efficient descriptive analytics relates to properly acknowledging what is happening with use of visual analysis that enables powerful insights. When coupled with predictive analytics that involve machine learning algorithms, visualization can bolster the decision-making process and generate powerful information. Figures 10 and 11 are generated with the visualization tool Tableau and provide an interesting pictorial representation of a box plot with the scores and ratios between scores related to high performance workers.

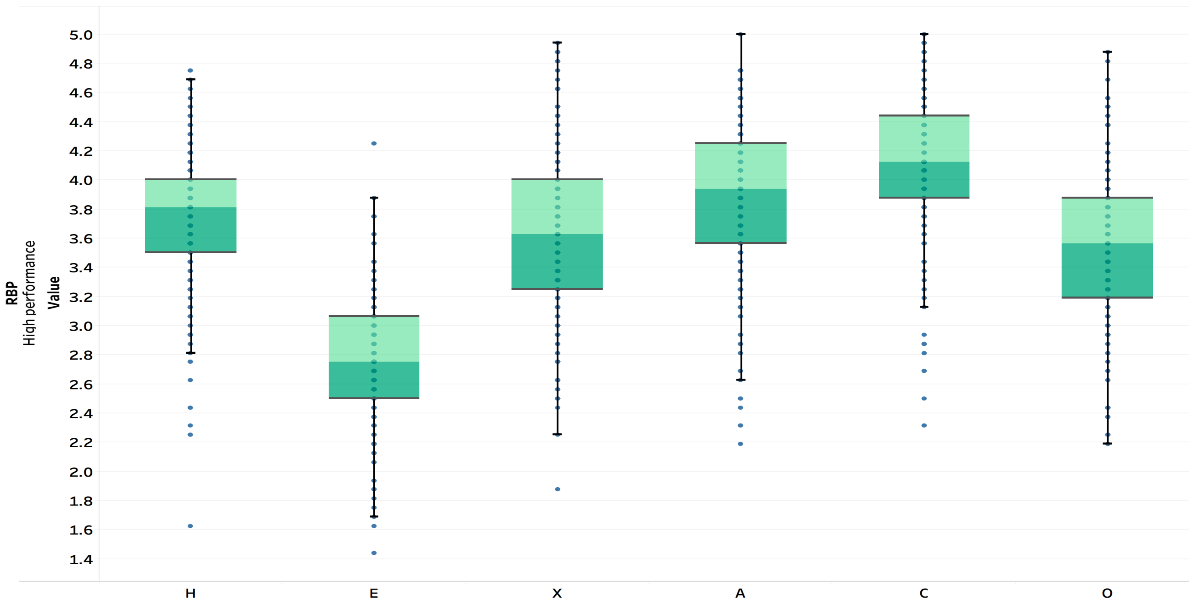


Figure 10. Average Scores of the HEXACO Traits for High Work Performance

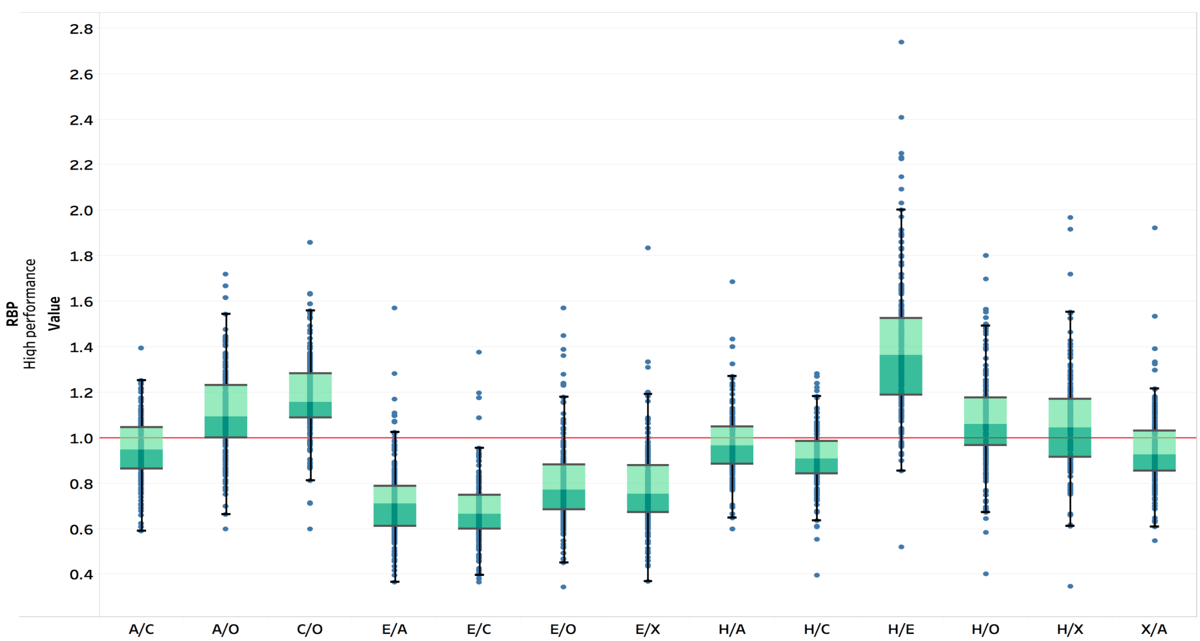


Figure 11. Average Ratios Between Scores on the HEXACO Traits for High Work Performance

Another example of how descriptive analytics can be applied in conjunction with predictive analytics within the context of the current research is expressed by Figure 12. A critical analysis of Figure 9 allows for a clear understanding of percental differences between the five most important predictors of job performance between high and low work performers.

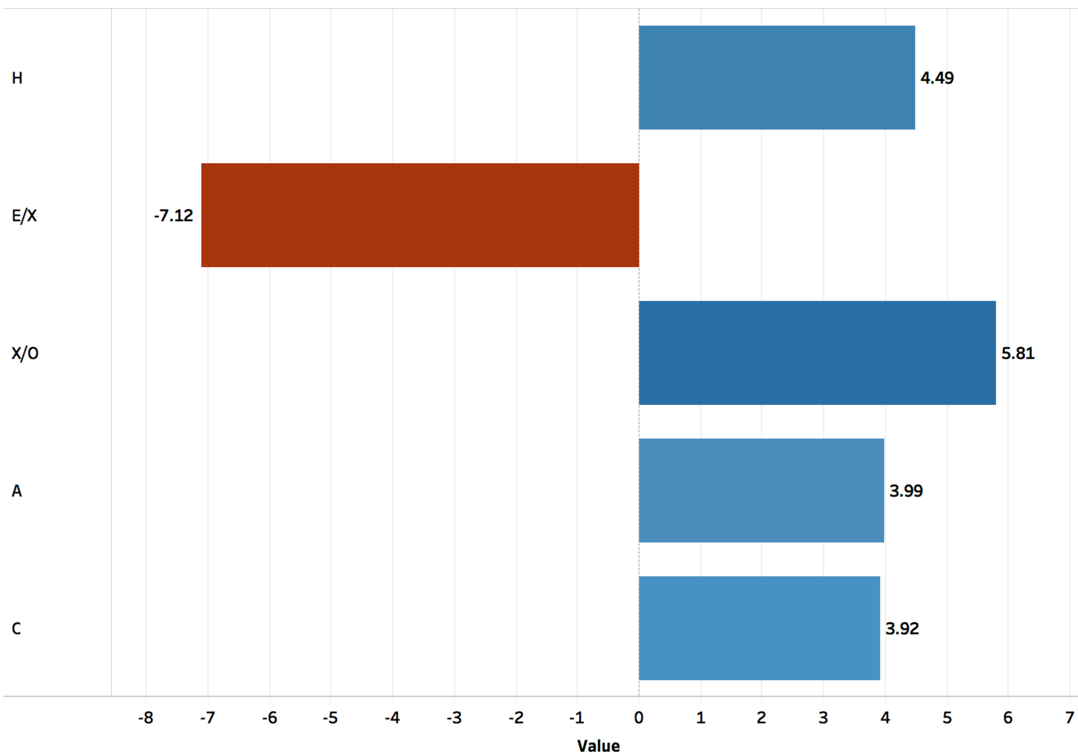


Figure 12. Percentile Differences on Average Scores Between High and Low Performers for the Five Most Important Predictors (According to the ASR from the Random Forest Algorithm)

Average scores on Humility, Agreeableness, and Conscientiousness are found to be, respectively, 4.49%, 3.99%, and 3.92% higher for high work performers when compared to low work performers. At this time, it is worthy to recall that in 2005 John Collins used the *Harvard Business Review* management magazine to elaborate on the concept of the Level Five Leader. According to Collins (2007), the Level Five Leader may be thought of as a professional where humbleness combines paradoxically with a vivid and passionate professional that will to achieve outstanding work performance standards. Interestingly, the ratio between average scores on Emotional Stability and Extraversion was found to be 7.12% lower for high work performers than low work performers. At a very high level and given that all important predictors are present in a certain measure, these results suggest that in a consistent fashion, employees who tend to behave with balanced levels of these traits are likely to achieve high work performance. According to the data and as is shown in Figure 8, the average value of the ratio between Emotional Stability and Extraversion is 0.78 for high work performers. Within the context of the present study, the 0.78 value indicates that employees who behave consistently more Extrovertly than Open to Experience are likely to perform better at work. Similarly, the ratio between average scores on Extraversion and Openness to Experience as the third more important predictor was found to be 5.81% higher for high work performers than low work performers. The average value of the ratio between Emotional Stability and Extraversion is 1.03 for high work performance. Notably, the fact that this ratio value is fairly close to 1.00 suggests that workers who get similar scores on Emotional Stability and Extraversion tend to perform well. Importantly and in line with LLM and the results from the RF output, all the above-mentioned inferences about traits and ratios between traits need to be considered in the presence of minimum levels of all other important predictors for job performance.

CHAPTER V

DISCUSSION AND CONCLUSION

This present investigative study was conducted with the purpose of expanding the management field knowledge discovery process beyond the extant methodological limitations. Numerous investigations and experiments with potential model types and modeling parameters — random forest, decision tree (C 4,5), neural networks, logistic regression, and multiple regression — were explored to evaluate five prediction models (5 * 10-fold cross validation = 50). To ensure reliability of the data and in accordance to CRISP, a thorough examination was conducted of a rich data set containing demographics, performance ratings, and scores on personality traits from 682 health care professionals from a large medical center. The dataset was meticulously preprocessed such that all five modeling types could be properly tested and compared. According to the cross-validation results, the best performing and most accurate model to predict job performance was the Random Forest algorithm followed by Decision Tree, Neural Networks, Multiple Regression, and Logistic Regression. Taken together, the data and results demonstrate that “state of art” machine learning algorithms are far more accurate than regression-based models to predict job performance. The best machine learning model predicted with 79.17% overall accuracy, while the best regression-based model predicted with 61.73% overall accuracy.

Responding to a clear gap in the management literature, this work approaches the collective influence of traits on work performance using powerful machine learning techniques. Supported by Liebig's Law of Minimum framework and the prediction accuracy results, the collective influence of certain amounts of traits were found to be important suggesting that group effects should be considered by management scholars. Even more importantly, the ratio between scores on some traits were found to represent better predictors of job performance than their related traits' scores when evaluated separately. Further, assuming that science is an ongoing and eternal learning process, it seems that the time has come for behavioral and management scholars to effectively embrace the opportunities of knowledge discovery that is fostered by modern prediction methodologies such as machine learning algorithms. Along this line, recent research conducted by Bleidorn and Hopwood (2018) points out that most current machine learning approaches to personality analysis focus on using social media data and other digital records to the neglect of more comprehensive construct validation frameworks. As reported in previous sections, my research answers this call by taking into consideration fundamental assumptions of construct validity employing reliable data that was previously tested in accordance with the most prevailing techniques for such.

Although disturbing, prediction-focused, and thought-provoking, these results can help build theoretical knowledge for social sciences. As discussed, the majority if not all academic papers within the management field rely mainly on regression-based analysis. Consequently, the process of knowledge discovery in the field is based on specific and limiting assumptions such as normal distribution of datasets, absence of multicollinearity issues, and insufficient capacity to compute nominal input variables on models. Most of the theories of the management field are based on effect size and moderating and mediation effects between constructs. Often, the theoretical relationships between independent and dependent variables involve roughly 3 to 10 constructs depending on the type of study. Conversely, machine learning algorithms are capable of extracting and computing patterns and relationships hidden deep in very large and complex datasets that can embrace all types of input variables. In this sense, the obvious superior predictive accuracy of machine learning impacts several management theories that solely rely its inferences on basic assumptions of traditional stochastic data modeling.

From a practical standpoint, asking the right questions with respect to predictors may help to save money, time, and effort (Dreiseitl & Ohno-Machado, 2002). One of the major goals of this study is to help managers to properly assess and manage their employees' personality traits with confidence to achieve elevated levels of work performance, reduced turnover, and increased profits. The machine learning techniques that are used in the current work not only have the potential to accurately predict performance but also to reveal the order of importance of the traits that may strongly affect work performance. By acknowledging what personality profiles are more likely to produce better work results, managers can focus efforts on effectively assigning employees to certain job positions to obtain greater work results. For example, if Humility, Conscientiousness, or even the ratio between them are found to be the most important predictors, managers may allocate workers with the proper score on these personality features to key positions within the company in an efficient manner to improve performance.

Despite the relevance of the current work, it is important to address its shortcomings and limitations. First, I am only assessing the influence of the six-dimensional model HEXACO and not any other trait that may have some sort of influence on work-related performance ratings. For example, it would be fruitful to include and examine other related constructs if available such as Working Memory, Bottom Line Mentality or even Proactivity as input variables and potential predictors of job performance. The management literature points to the influence of several traits when it comes to predicting work performance. Although machine learning algorithms can compute thousands of input variables, using only the HEXACO factors and subfactors as input variables may be a limitation of the current study. Second, because human traits are somewhat subjective, it would be helpful to employ unobtrusive measures of human traits. A good example of possible complementary, unobtrusive measure of employees' traits is social media data. Cambridge University, for example, hosts a website called "Apply Magic Sauce" that collects and computes social media data from participants to produce behavioral insights. Third, although performance ratings were provided by direct supervisors, employees' personality traits were measured on a self-reported questionnaire. Results from the current study should be appraised considering that when evaluating observed behavior instead of self-reported behavior, prediction accuracy and variable importance measures may be different. Finally, as machine learning algorithms are constantly evolving, the differences in predictive accuracy and variable importance measures between the models that are tested on this current work would need to be re-evaluated for other types and versions of algorithms.

More research regarding the superior predictive accuracy of machine learning over stochastic data modeling is needed. Future research should pinpoint how more accurate predictions of several work-related outcomes would affect management theories across different frameworks. I suspect that if properly embraced, the combination of effective machine learning algorithms with reliable and valid behavioral constructs will advance the management field to a higher knowledge discovery level at a pace that has not yet been experienced.

REFERENCES

Aguinis, H., Gottfredson, R.K., & Wright, T.A. (2011). Best-practice recommendations for estimating interaction effects using meta-analysis. *Journal of Organizational Behavior, 32*(8), 1033-1043.

Allport, G.W. (1961). *Pattern and growth in personality*. Oxford England: Holt, Reinhart, & Winston.

Anderson, N.H. (1961). Scales and statistics: Parametric and nonparametric. *Psychological Bulletin, 58*(4), 305.

Archer, K.J., & Kimes, R.V. (2008). Empirical characterization of random forest variable importance measures. *Computational Statistics & Data Analysis, 52*(4), 2249-2260.

Ashton, M.C., & Lee, K. (2007). Empirical, theoretical, and practical advantages of the HEXACO model of personality structure. *Personality and Social Psychology Review, 11*(2), 150-166.

Barrick, M.R., & Mount, M.K. (1991). The big five personality dimensions and job performance: A meta-analysis. *Personnel Psychology, 44*(1), 1-26.

Bayat, S., Cuggia, M., Rossille, D., Kessler, M., & Frimat, L. (2009). Comparison of Bayesian network and decision tree methods for predicting access to the renal transplant waiting list. In *Medical Information Europe* (pp. 600-604).

Bickel, P.J., Ritov, Y.A., & Stoker, T.M. (2006). Tailor-made tests for goodness of fit to semiparametric hypotheses. *The Annals of Statistics*, *34*(2), 721-741.

Bleidorn, W., & Hopwood, C.J. (2018). Using machine learning to advance personality assessment and theory. *Personality and Social Psychology Review.*

Booker, L.B. (1988). Classifier systems that learn internal world models. *Machine Learning, 3*(2-3), 161-192.

Breiman, L. (2001a). Random forests. *Machine Learning, 45*(1), 5-32.

Breiman, L. (2001b). Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical Science, 16*(3), 199-231.

Brown, T.J., Mowen, J.C., Donavan, D.T., & Licata, J.W. (2002). The customer orientation of service workers: Personality trait effects on self-and supervisor performance ratings. *Journal of Marketing Research, 39*(1), 110-119.

Carnahan, B., Meyer, G., & Kuntz, L.A. (2003). Comparing statistical and machine learning classifiers: alternatives for predictive modeling in human factors research. *Human Factors, 45*(3), 408-423.

Chapman, D.V. (1996). Water quality assessments: A guide to the use of biota, sediments and water in environmental monitoring. World Health Organization, Unesco & United Nations Environment Programme.http://www.who.int/iris/handle/10665/41850
Chen, G., & Klimoski, R.J. (2003). The impact of expectations on newcomer performance in teams as mediated by work characteristics, social exchanges, and empowerment. *Academy of Management Journal, 46*, 591–607.
Christensen, R. (2011). *Plane answers to complex questions: The theory of linear models*. Springer Science & Business Media.
Christiansen, N., Sliter, M., & Frost, C. (2014). What employees dislike about their jobs: Relationship between personality-based fit and work satisfaction. *Personality and Individual Differences, 71*, 25-29.
Cogliser, C.C., Gardner, W.L., Gavin, M.B., & Broberg, J.C. (2012). Big five personality factors and leader emergence in virtual teams. *Group & Organization Management, 37*(6), 752-784.
Collins, J. (2007). Level 5 leadership. *The Jossey-Bass reader on educational leadership*, *2*, 27-50.
Cohen, J. (1983). The cost of dichotomization. *Applied Psychological Measurement, 7*(3), 249-253.
Cooper, L.H.N. (1937). Some conditions governing the solubility of iron. *Proceedings of Royal Society of London. B, 124*(836), 299-307.
Cortez, P., & Silva, A.M.G. (2008). Using data mining to predict secondary school student performance. *Proceedings of the 5th Annual Future Business Technology Conference*.
De Baar, H.J.W. (1994). von Liebig's law of the minimum and plankton ecology (1899–1991). *Progress in Oceanography, 33*(4), 347-386.
De Campos, M.F., Lopes, L.R., Magina, P., Tavares, F.L., Kunioshi, C.T., & Goldenstein, H. (2005). Texture and microtexture studies in different types of cast irons. *Materials Science and Engineering, 398*(1-2), 164-170.
DeCoster, J., Iselin, A.M.R., & Gallucci, M. (2009). A conceptual and empirical examination of justifications for dichotomization. *Psychological Methods, 14*(4), 349.
Delen, D. (2009). Analysis of cancer data: A data mining approach. *Expert Systems, 26*(1), 100-112.
Delen, D. (2011). Predicting student attrition with data mining methods. Journal of College Student Retention: *Research, Theory & Practice, 13*(1), 17-35.
Delen, D., & Al-Hawamdeh, S. (2009). A holistic framework for knowledge discovery and management. *Communications of the ACM, 52*(6), 141-145.
Delen, D., Cogdell, D., & Kasap, N. (2012). A comparative analysis of data mining methods in predicting NCAA bowl outcomes. *International Journal of Forecasting, 28*(2), 543-552.
Delen, D., Kuzey, C., & Uyar, A. (2013). Measuring firm performance using financial ratios: A decision tree approach. *Expert Systems with Applications, 40*(10), 3970-3983.
Delen, D., Oztekin, A., & Kong, Z.J. (2010). A machine learning-based approach to prognostic analysis of thoracic transplantations. *Artificial Intelligence in Medicine, 49*(1), 33-42.
Delen, D., Sharda, R., & Bessonov, M. (2006). Identifying significant predictors of injury severity in traffic accidents using a series of artificial neural networks. *Accident Analysis & Prevention, 38*(3), 434-444.

Delen, D., Tomak, L., Topuz, K., & Eryarsoy, E. (2017). Investigating injury severity risk factors in automobile crashes with predictive analytics and sensitivity analysis methods. *Journal of Transport & Health,* 4, 118-131.
Delen, D., Walker, G., & Kadam, A. (2005). Predicting breast cancer survivability: A comparison of three data mining methods. *Artificial Intelligence in Medicine, 34*(2), 113-127.
Do, M.H., & Minbashian, A. (2014). A meta-analytic examination of the effects of the agentic and affiliative aspects of extraversion on leadership outcomes. *The Leadership Quarterly, 25*(5), 1040-1053.
Dreiseitl, S., & Ohno-Machado, L. (2002). Logistic regression and artificial neural network classification models: a methodology review. *Journal of Biomedical Informatics, 35*(5-6), 352-359.
Eichstaedt, J.C. (2017). Predicting and characterizing the health of individuals and communities through language analysis of social media. *Dissertations available from ProQuest*. AAI10273705.
Eichstaedt, J.C., Schwartz, H.A., Kern, M.L., Park, G., Labarthe, D.R., Merchant, R.M., & Weeg, C. (2015). Psychological language on Twitter predicts county-level heart disease mortality. *Psychological Science, 26*(2), 159-169.
Farrington, D.P., & Loeber, R. (2000). Some benefits of dichotomization in psychiatric and criminological research. *Criminal Behaviour and Mental Health, 10*(2), 100-122.
Fulkerson, B. (1995). Machine learning, neural and statistical classification. *Technometrics, 37*(4), 459.
George, L., (2016). Another missing link to burnout and job performance: The mediator role of work self compassion Dissertation, University of Denver, Daniels College of Business.
Gorban, A.N., Pokidysheva, L.I., Smirnova, E.V., & Tyukina, T.A. (2011). Law of the minimum paradoxes. *Bulletin of Mathematical Biology, 73*(9), 2013-2044.
Gorban, A.N., Smirnova, E.V., & Tyukina, T.A. (2010). Correlations, risk and crisis: From physiology to finance. *Physica A: Statistical Mechanics and its Applications, 389*(16), 3193-3217.
Gray, K. (2017). How to map theory: Reliable methods are fruitless without rigorous theory. *Perspectives on Psychological Science, 12*(5), 731-741.
Greenwald, A.G. (2012). There is nothing so theoretical as a good method*. Perspectives on Psychological Science, 7*(2), 99-108.
Grefenstette, J.J. (1988). Credit assignment in rule discovery systems based on genetic algorithms. *Machine Learning, 3*(2-3), 225-245.
Grijalva, E., & Harms, P.D. (2014). Narcissism: An integrative synthesis and dominance complementarity model. *Academy of Management Perspectives, 28*(2), 108–127.
Grimm, S.S., Paris, Q., & Williams, W.A. (1987). A von Liebig model for water and nitrogen crop response. *Western Journal of Agricultural Economics*, *12*(2):182-192.
Hambrick, D.C., & Chatterjee, A. (2007). It's all about me: Narcissistic chief executive officers and their effects on company strategy and performance. *Administrative Science Quarterly*, *52*(3) 351–386.
Harari, M.B., Rudolph, C.W., & Laginess, A.J. (2015). Does rater personality matter? A meta-analysis of rater Big Five–performance rating relationships. *Journal of Occupational and Organizational Psychology, 88*(2), 387-414.
Harvey, H.W., Cooper, L.H.N., Lebour, M.V., & Russell, F.S. (1935). Plankton production and its control. *Journal of the Marine Biological Association of the United Kingdom, 20*(2), 407-441.

Hunt, E., Marin, J., & Stone, P., (1966). *Experiments in induction*, New York: Academic Press.

James, L.R., & Mazerolle, M.D. (2001). *Personality in work organizations*: Thousand Oaks, CA: Sage.

Johannesen, J.K., Bi, J., Jiang, R., Kenney, J.G., & Chen, C.M.A. (2016). Machine learning identification of EEG features predicting working memory performance in schizophrenia and healthy adults. *Neuropsychiatric Electrophysiology, 2*(1), 3.

Keen, P.G., & Morton, S. (1978). *Decision support systems: An organizational perspective* (No. 4; QA76. 6, K44). Reading MA: Addison Wesley.

Kononenko, I. (2001). Machine learning for medical diagnosis: history, state of the art and perspective. *Artificial Intelligence in Medicine, 23*(1), 89-109.

Kotsiantis, S., Pierrakeas, C., & Pintelas, P. (2004). Predicting student's performance in distant learning using machine learning techniques. *Applied Artificial Intelligence*, *18*(5), 411-426.

Langley, P. (1989). Toward a unified science of machine learning. *Machine Learning, 3*(4), 253-259.

Lee, K., & Ashton, M.C. (2004). Psychometric properties of the HEXACO personality inventory. *Multivariate Behavioral Research, 39*(2), 329-358.

Lievens, F., Chasteen, C.S., Day, E.A., & Christiansen, N.D. (2006). Large-scale investigation of the role of trait activation theory for understanding assessment center convergent and discriminant validity. *Journal of Applied Psychology, 91*(2), 247-258.

Loftus, G.R. (1996). Psychology will be a much better science when we change the way we analyze data. *Current Directions in Psychological Science, 5*(6), 161-171.

MacCallum, R. ., Zhang, S., Preacher, K.J., & Rucker, D.D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods, 7*(1), 19.

Minaei-Bidgoli, B., Kashy, D.A., Kortemeyer, G., & Punch, W.F. (2003). Predicting student performance: An application of data mining methods with an educational web-based system. *Frontiers in Education*, *1*, T2A-13.

Mischel, W. (2008). The toothbrush problem. *Association for Psychological Science Observer, 21*(11).

Morgeson, F.P., Campion, M.A., Dipboye, R.L., Hollenbeck, J.R., Murphy, K., & Schmitt, N. (2007). Are we getting fooled again? Coming to terms with limitations in the use of personality tests for personnel selection. *Personnel Psychology, 60*(4), 1029-1049.

Mussel, P., & Spengler, M. (2015). Investigating intellect from a trait activation perspective: Identification of situational moderators for the correlation with work-related criteria. *Journal of Research in Personality, 55*, 51-60.

Nilsson, N.J., Sejnowski, T.J., & White, H. (1965). *Learning machines*. New York: McGraw-Hill.

Novais, R.F. (2007). Fertilidade do solo. Sociedade Brasileira de Ciência do Solo, 760-850.

Olson, D., & Delen, D. (2008). *Advanced data mining techniques*. New York: Springer.

Owens, B.P., Wallace, A.S., & Waldman, D.A. (2015). Leader narcissism and follower outcomes: The counterbalancing effect of leader humility. *Journal of Applied Psychology, 100*(4), 1203–1213.

Paris, Q. (1992). The von Liebig hypothesis. *American Journal of Agricultural Economics, 74*(4), 1019-1028.

Parker, P.L., McDaniel, H.S., & Crumpton-Young, L.L. (2002). Do research participants give interval or ordinal answers in response to Likert scales? In *IIE Annual Conference. Proceedings* (p. 1). Institute of Industrial and Systems Engineers (IISE).

Quinlan, J.R. (1986). Induction of decision trees. *Machine Learning, 1*(1), 81-106.
Quinlan, J.R. (1996). Improved use of continuous attributes in C4. *Journal of Artificial Intelligence Research, 4*, 77-90.
Ramesh, V., Parkavi, P., & Ramar, K. (2013). Predicting student performance: A statistical and data mining approach. *International Journal of Computer Applications, 63*(8), 35 – 43.
Rosenblatt, F. (1962). *Principals of neurodynamics*. Washington, DC: Spartan Books.
Saarela, M., Yener, B., Zaki, M.J., & Kärkkäinen, T. (2016). Predicting math performance from raw large-scale educational assessments data: A machine learning approach. In *JMLR Workshop and Conference Proceedings,* 48. JMLR.
Seni, G., & Elder, J.F. (2010). Ensemble methods in data mining: improving accuracy through combining predictions. *Synthesis Lectures on Data Mining and Knowledge Discovery, 2*(1), 1-126.
Shaffer, J.A., & Postlethwaite, B.E. (2013). The validity of conscientiousness for predicting job performance: A meta-analytic test of two hypotheses. *International Journal of Selection and Assessment, 21*(2), 183-199.
Sharda, R., & Delen, D. (2006). Predicting box-office success of motion pictures with neural networks. *Expert Systems with Applications, 30*(2), 243-254.
Sharda, R., Delen, D., & Turban, E. (2016). Business intelligence, analytics, and data science: A managerial perspective (Fourth edition). London: Pearson.
Shearer, C. (2000). The CRISP-DM model: the new blueprint for data mining. *Journal of Data Warehousing, 5*, 13–22.
Sitser, T., Van der Linden, D., & Born, M.P. (2013). Predicting sales performance criteria with personality measures: The use of the general factor of personality, the Big Five and narrow traits. *Human Performance, 26*(2), 126-149.
Stevens, S.S. (1951). Mathematics, measurement, and psychophysics. In S.S. Stevens (Ed.), Handbook of experimental psychology (pp. 1–49). New York: Wiley.
Van Hoye, G., & Turban, D.B. (2015). Applicant–employee fit in personality: Testing predictions from similarity-attraction theory and trait activation theory. *International Journal of Selection and Assessment, 23*(3), 210-223.
Warsi, O.M., & Dykhuizen, D.E. (2017). Evolutionary implications of Liebig's law of the minimum: Selection under low concentrations of two nonsubstitutable nutrients. *Ecology and Evolution*, *7*(14), 5296-5309.
Wallace, J.C., Edwards, B.D., Arnold, T., Frazier, M.L., & Finch, D.M. (2009). Work stressors, role-based performance, and the moderating influence of organizational support. *Journal of Applied Psychology, 94*(1), 254-262.
Wang, Y., & Kosinski, M. (2017). Deep neural networks are more accurate than humans at detecting sexual orientation from facial images, *Journal of Personality and Social* Psychology, *114*(2), 246-257.
Welbourne, T., Johnson, D., & Erez, A. (1998). The role-based performance scale: Validity analysis of a theory-based measure. *Academy of Management Journal, 41*(5), 540-555.
Wille, B., De Fruyt, F., & Feys, M. (2013). Big Five traits and intrinsic success in the new career era: A 15-year longitudinal study on employability and work-family conflict. *Applied Psychology, 62*(1), 124-156.
Wu, D., Jennings, C., Terpenny, J., Gao, R.X., & Kumara, S. (2017). A comparative study on machine learning algorithms for smart manufacturing: Tool wear prediction using random forests. *Journal of Manufacturing Science and Engineering, 139*(7), 071018.

Xu, J., Moon, K.H., & Van Der Schaar, M. (2017). A machine learning approach for tracking and predicting student performance in degree programs. *IEEE Journal of Selected Topics in Signal Processing, 11*(5), 742-753.

Youyou, W., Kosinski, M., & Stillwell, D. (2015). Computer-based personality judgments are more accurate than those made by humans. *Proceedings of the National Academy of Sciences, 112*(4), 1036-1040.

VITA

Marcio Lima

Candidate for the Degree of

Doctor of Philosophy

Dissertation: MACHINE LEARNING AND PERSONALITY TRAITS: A DISTURBING CONTRIBUTION FROM THE ALGORITHMIC CULTURE TO BEHAVIORAL SCIENCE

Major Field: Business Administration

Biographical:

Education:

Completed the requirements for the Doctor of Philosophy in Business Administration at Oklahoma State University, Stillwater, Oklahoma in December, 2018.

Completed the requirements for the Master of Science in Agronomy Engineering at Universidade Federal de Minas Gerais/Belo Horizonte, Minas Gerais/Brazil in 2004.

Completed the requirements for the Bachelor of Science in Metallurgical Engineering at Universidade Federal de Viçosa/Viçosa, Minas Gerais/Brazil in 2002.

Experience:

In charge of several, research, planning, and implementation projects in the fields of agriculture, forest engineering (*Eucaliptus ssp.)*, metallurgical industry, cast-iron foundry, and automotive parts industry at the nine-digit figure. Master theses won the Voest Alpine International Award in 2004 related to industrial waste management studies. Invited to the Phi Kappa Phi academic-based honor society in 2018. Board member of the Daniels College of Business, Denver University in 2018.

Professional Memberships:

Metalsider S.A. Corp. Shareholder
Member of the Phi Kappa Phi Honor Society.
Board Member of Daniels of College of Business – Denver University