A Decision Support Tool using Machine Learning Techniques for Strategic Investment Planning Based on Customer Evaluation Criteria

by Wendy E. Robello

B.S. in Mechanical Engineering, September 2004, University of California, Irvine
M.S. in Systems Architecting and Engineering, May 2009, University of Southern California

A Praxis submitted to

The Faculty of
The School of Engineering and Applied Science
of The George Washington University
in partial fulfillment of the requirements
for the degree of Doctor of Engineering

May 21, 2023

Praxis directed by

Thomas Holzer
Professorial Lecturer of Engineering Management and Systems Engineering

The School of Engineering and Applied Science of The George Washington University certifies that Wendy Robello has passed the Final Examination for the degree of Doctor of Engineering as of February 24, 2023. This is the final and approved form of the Praxis.

A Decision Support Tool using Machine Learning Techniques for Strategic Investment Planning Based on Customer Evaluation Criteria

Wendy E. Robello

Praxis Research Committee:

Thomas Holzer, Professorial Lecturer of Engineering Management and Systems Engineering, Praxis Director

Shahram Sarkani, Professor of Engineering Management and Systems Engineering, Committee Member

Thomas Mazzuchi, Professor of Engineering Management and Systems Engineering & of Decision Sciences, Committee Member

Amir Etemadi, Associate Professor of Engineering and Applied Science, Committee Chair

© Copyright 2023 by Wendy Robello
All rights reserved

Dedication

This research and praxis are dedicated to my family and friends, who have believed in me and supported me throughout this educational journey and beyond. To my moms, who have instilled the passion of exploration and continuous learning. To my dad, who has always taught me to have fun and seek happiness in whatever it is I do in life. To my brother, who has been my best friend and protector from day one. To my second family, the Madrigals, who have opened their home and heart to me, especially the last few extremely challenging years with a global pandemic and the need to be closer to family and cheering for me throughout this academic pursuit. To my colleagues and mentors, Jerry, Angela, Charley, Susan, Christine, and Mary P., for sponsoring me in this program and always believing in me, even when I may have doubted my own capabilities. To my friends, Stephanie, Mara, Justin, Christine, and Tina, for being some of my biggest cheerleaders and always checking on my emotional health along the way. To my amazing little ones, MMM and JJB, who never let me forget that it is ok to take time out of the day to rest, play, and just love relentlessly. And finally, to my dearest husband, Ryan, for always pushing me to never give up, giving me a place of comfort when I need rest, and loving me unconditionally.

Acknowledgements

This endeavor was made possible by a collection of people along the way. First, I would like to thank my husband for encouraging me to pursue a doctorate and acknowledge the sacrifices he made along the way to keep our home functioning while much of my time outside of my professional career was expended by studying for classes, presenting my research at conferences, and the many nights and weekends to develop this research and praxis.

I also am extremely thankful for the support of my close friends, Justin, Stephanie, and Mara, who helped with their strategic analysis and academic research expertise. Their willingness and support to be my personal advisory board throughout this research and praxis development is appreciated beyond words.

Lastly, I would like to thank the faculty and staff of The George Washington University, especially my doctorate advisor, Dr. Thomas Holzer, along with Dr. Mazzuchi and Dr. Sarkani, for all of their guidance and invaluable lessons with regard to academic research.

Abstract

A Decision Support Tool using Machine Learning Techniques for Strategic Investment Planning Based on Customer Evaluation Criteria

Early identification of anticipated customer needs is critical to winning new business in competitive environments. This allows companies to make strategic investment decisions that will improve their core competencies to meet those customer needs. These strategic decisions target enhancements that allow contractors to offer their best proposed solution to meet the anticipated customer's needs, inferred by the customer's prioritized selection criteria. However, contractors do not know the customer's final prioritized selection criteria until release of the final Request for Proposal. A contractor's ability to strengthen their proposed solution decreases by 60% by the release of the customer's selection criteria, reducing their probability of winning.

Current methods leverage limited samples of comparable historical Request for Proposals in conjunction with subject matter experts that have customer insights to predict a customer's prioritized needs and selection criteria. These approaches are subjective, costly, and time-consuming. This research applies supervised machine learning techniques, which are an underutilized approach within strategic management research.

This research yields a decision support tool using supervised machine learning techniques to predict customer needs deduced from historic customer evaluation criteria. This research contributes a robust process for strategic investment predictive modeling, a

newly established database of customer evaluation criteria, and a predictive model with an over 80% accuracy and test quality performance measure.

The impact of these contributions spans various industries in competitive environments and will offer strategic decision-makers the confidence of a robust and rigorous approach that current methods are unable to provide.

Table of Contents

Dedication ..... **iv**

Acknowledgements ..... **v**

Abstract ..... **vi**

List of Figures ..... **xii**

List of Tables ..... **xiii**

List of Acronyms ..... **xv**

Chapter — 1 Introduction 1

1.1 Background 1

1.2 Research Motivation 2

1.3 Problem Statement 3

1.4 Thesis Statement 4

1.5 Research Objectives 4

1.6 Research Questions and Hypotheses 5

1.7 Research Scope and Limitations 7

1.8 Praxis Structure 8

Chapter — 2 Literature Review 10

2.1 Introduction 10

2.2 Strategic Investments within the Defense Industrial Base 11

2.3 Current Methods of Strategic Decision-Making to increase Competitive Positioning....13
2.4 Machine learning (ML) in strategic decision-making....19
2.5 Chapter Summary....22
Chapter — 3 Methodology....23
3.1 Introduction....23
3.2 Data Collection....24
3.2.1 Independent Variable Dataset....26
3.2.2 Dependent Variable Dataset....28
3.2.3 Alteryx Designer x64....30
3.3 Gini Impurity....32
3.4 Classification ML Algorithms....34
3.4.1 Decision Trees (DT) overview....34
3.4.2 Random Forest (RF) overview....35
3.4.3 Extreme Gradient Boost (XGBoost) overview....36
3.5 Validation....36
3.6 Performance Measures....37
3.7 Stage 2: Application of Analytic Hierarchy Process....42
3.8 Summary....47
Chapter — 4 Results....49

4.1 Introduction....49
4.2 Gini Impurity Results....49
4.3 Performance Measure Results....52
4.4 Performance Evaluation....57
4.5 Difference in Performance....59
4.6 Prioritized Evaluation Criteria Results....61
4.6.1 Gini Impurity Results....61
4.6.2 Performance Measures....63
4.6.3 Performance Evaluation....66
4.6.4 Difference in Performance....69
4.7 Summary....71
Chapter — 5 Discussion and Conclusions....73
5.1 Discussion....73
5.2 Contribution to Body of Knowledge....74
5.3 Conclusions....75
5.4 Future Research Recommendations....75
References....76
Appendix A: RFP Database Sample – Factors and Relationships....82
Appendix B: RFP Database Sample – AHP Application....83
Appendix C: RFP Database Sample – Geometric Mean Calculations....84

Appendix D: Alteryx Random Forest Workflow Example 85

Appendix E: ML Algorithm Configuration Parameters 86

Appendix F: Confusion Matrix Results for Prioritized Evaluation Criteria Model 87

Appendix G: Stage 2 AUC Results 89

List of Figures

Figure 1. Research problem statement flow-down..... 7

Figure 2. Company's ability to adjust investments versus their cumulative investment spend aligned to the customer acquisition timeline. 10

Figure 3. Example customer evaluation criteria from RFP – Naval Air Systems Command. (2017). Solicitation No. N0001917R087..... 18

Figure 4. ML model development process.. 24

Figure 5. RFPs reviewed by customer count 25

Figure 6. RFPs reviewed by market area and customer count..... 26

Figure 7. Sample subset of database created for stage 1. 30

Figure 8. Sample workflow to create samples in Alteryx Designer x64..... 31

Figure 9. Sample workflow to create training model in Alteryx Designer x64..... 31

Figure 10. Sample workflow to validate model using test data sample in Alteryx Designer x64 32

Figure 11. Decision Tree Example..... 35

Figure 12. ROC examples and AUC quality of test scale recreated from Trifonova, Lokhov, and Archakov (2013)..... 41

Figure 13. Sample subset of database created for Stage 2..... 46

List of Tables

Table 1. APP Framework for categorizing competitivesnss 14

Table 2. Contract Features collected from RFPs reviewed 27

Table 3. Unique evaluation criterion of major evaluation criteria mapping 29

Table 4. Confusion Matrix 38

Table 5. AHP Scale of Imporance for comparison pair 44

Table 6. Assigned Prioritization Levels 45

Table 7. Summary of Methodologies and Hypothesis Tests 48

Table 8. GI result per Contract Feature per major evaluation criterion 50

Table 9. Statistical results of GI for each Contract Feature 51

Table 10. Performance measures for DT predictions 53

Table 11. Performance measures for RF predictions 54

Table 12. Performance measures for XGBoost predictions 55

Table 13. Statistics of AUC for each ML algorithm 56

Table 14. ANOVA test of ML Algorithms' AUC measure 58

Table 15. Statistics of the difference of AUC_{RF} and AUC_{DT} 60

Table 16. GI result per Contract Feature per major evaluation criterion (with Prioritization) 61

Table 17. Statistical results of GI for each Contract Feature (with Prioritization) 62

Table 18. Performance measures for DT predictions (with Prioritization) 64

Table 19. Performance measures for RF predictions (with Prioritization) 64

Table 20. Performance measures for XGBoost predictions (with Prioritization) 65

Table 21. Statistics of AUC for each ML algorithm and t-test results for H5.2 66

Table 22. ANOVA test of ML Algorithms' AUC measure results for H5.3 67

Table 23. Statistics of the difference of AUC results for H5.4 70

Table 24. Summary of Hypothesis Test Results 72

List of Acronyms

AHP	Analytical Hierarchy Process
AUC	Area Under the Curve
APP	Asset-Process-Performance
BA	Balanced Accuracy
BBP	Better Buying Power
DT	Decision Trees
DACIS	Defense & Aerospace Competitive Intelligence Service
DoD	Department of Defense
DTC	Design to Cost
DRFP	Draft Request for Proposal
XGBoost	Extreme Gradient Boost
FN	False Negative
FP	False Positive
FPR	False Positive Rate
GMM	Geometric Mean Method
GI	Gini Impurity
ML	Machine Learning
NLP	Natural Language Processing
NAICS	North American Industry Classification System
PoP	Period of Performance
PTW	Position to Win

PSC	Product Service Code
RF	Random Forest
ROC	Receiver Operating Characteristics
RFP	Request for Proposal
RBV	Resource Based View
S&P	Standard and Poor
SME	Subject Matter Experts
TN	True Negative
TP	True Positive
TPR	True Positive Rate
US	United States

Chapter — 1 Introduction

1.1 Background

In recent years, companies have been subjected to great pressures to achieve and sustain profitable growth caused by evolving system complexity, rapid technological changes, and customer acquisition trends in both commercial and government sectors (Etemadi & Kamp, 2021). Therefore, companies must make strategic decisions about resource planning and investments in these increasingly competitive times (Blank, 2019; Chatterjee & Wernerfelt, 1991; McGrath & Nerkar, 2004). These strategic decisions allow companies to improve their proposed solution to meet anticipated customers' needs, inferred by the customer's evaluation criteria. However, strategic decision-making is severely hindered due to the lack of insights into how customers will evaluate a proposed solution, as customer evaluation criteria is not known by companies until the release of final solicitation documentation in the form of a request for proposal (RFP). By the time the RFP is released, most companies have already finalized their strategic investments and planning (O'Guin & Kelly, 2012). Insights into customer evaluation criteria allow companies to systematically consider risk, profitability, and investments that impact their probability of winning a contract (Li, Zhang, Chen, & Martek, 2020; Ward & Chapman, 1988). Understanding customer evaluation criteria in relation to the proposed solution is critical for staying competitive (Hari, Kasser, & Weiss, 2007). The earlier customer information is obtained, the bigger the opportunity for companies to influence and adjust investments to secure a more competitive position.

1.2 Research Motivation

Companies lose in competitive environments due to inadequate core competencies within their proposed solution that align to the customer's prioritized needs. Although customer needs are discussed throughout the acquisition phase of a new business opportunity, the customer's ultimate decision is based on a down selection process, rating the company's proposed solutions through review of their proposal against the final customer evaluation criteria. Earlier insights of prioritized customer evaluation criteria enable a company to align strategic investments that strengthen core competencies for better proposed solutions.

Current methods to forecast customer's prioritized evaluation criteria are time consuming, costly, and subjective. Risk averse decision-makers are reluctant of current methods due to lack of their quantitative and objective methods, as much of the insight is based on expensive subject matter experts (SMEs) with intelligence gathering of the customer. This often leads to decision-makers investing in several parallel, and redundant, analyses to provide additional insight before large investments are made to increase their competitive position for the new business opportunity. The question of providing quantitative and objective methods has been avoided by the strategic research literature due to the lack of available structured data (Choi, Menon, & Tabakovic, 2021).

This research hypothesized that customer evaluation criteria can be predicted more effectively using supervised machine learning (ML) techniques. This research provides a systematic, economical, and objective approach to predict customer evaluation criteria that enables strategic investment decisions which will strengthen proposed solutions. Leveraging ML methods and applications from other fields enables strategy managers,

strategic planners, and financial decision-makers to explore which investments will directly influence a customer's down selection to award a contract. This research's implications are applicable to various industries in competitive environments, as it yields a structured process and ML model that can be tailored and refined for future strategy research.

1.3 Problem Statement

A bidder's ability to adjust strategic investments aligned to customer needs decreases by 60% by RFP release, reducing their probability of winning new business.

The competitive landscape across industries continue to increase forcing companies to have more robust investment strategies to stay competitive. Within the defense industrial base, multi-million-dollar new business opportunities take months-to-years of early investments to strengthen core competencies. The risks associated with investing into the wrong technology and resources is realized once the final customer evaluation criteria are released through the RFP document.

Upon release of the final RFP, a company not only has absorbed majority of their cumulative investments for that opportunity, but also their ability to adjust strategic investments decreases. This leaves the company only tactical investment decisions to complete their proposed solution and submission, reducing the prospect of positioning the company to increase their chance to beat competitors.

1.4 Thesis Statement

A decision support tool based on machine learning will better enable a bidder's strategic investment planning by forecasting customers' prioritized evaluation criteria.

Although companies use customer intelligence and analysis methods to forecast the criteria years prior to release of the RFP, it lacks a quantifiable uncertainty value associated with the prediction. ML techniques not only allow a quantifiable measurement of accuracy, but its numerous performance measures provide decision makers visibility into their risk associated with the prediction.

An ML model will also allow for continuous adaptability for any new data collection and changes in customer evaluation criteria over time, which often occurs in numerous draft RFPs before the release of a final RFP. This directly provides an economical and repeatable approach that current methods do not provide.

1.5 Research Objectives

Based on the lack of a repeatable, objective, and means of quantifiable accuracy in customer evaluation criteria forecasting, the purpose of this research is to apply proven ML methods to better forecast evaluation criteria for new business opportunities based on contract features.

Through these ML methods, the research will identify which contract features can and should be used as independent variables to predict the major criteria areas that the customer will use in their final RFP. Based on measuring and comparing different ML algorithm performance measures, the research will identify which ML algorithm should be used for modeling within the strategic investment decision support tool.

The research objectives are as follows:

1. Develop a decision support tool to enable strategic investments decisions.

2. Evaluate and validate the ML performance and accuracy of the support tool.

1.6 Research Questions and Hypotheses

To meet the research objectives, research questions (RQ) and hypotheses (H) were formulated to ensure robustness of the decision support tool development and validation. The research was conducted in two stages: stage 1 – identify if ML techniques produced sufficient performance of forecasting criteria presence for a new business opportunity (RQ1 – RQ 3); stage 2 – replicate tool development with the addition of forecasting the prioritization of customer evaluation criteria for a new business opportunity (RQ4).

Research Questions

RQ1: What contract features are good predictors of customer evaluation criteria?

RQ2: What are the evaluation criteria for a new business opportunity?

RQ3: What ML technique should be used to best predict a customer's evaluation criteria?

RQ4: What is the prioritization of evaluation criteria for a new business opportunity?

The initial three research questions were tested by four distinct hypotheses, whereas the second stage introduces research focused on the decision support tool to forecast customer evaluation criteria with levels of prioritization.

Hypotheses:

H1. Market area, customer type, contract type, period of performance, and product service code are good predictors for customer evaluation criteria.

H2. When applying supervised machine learning techniques for a new business opportunity, evaluation criteria is predicted with an AUC higher than 70%.

H3. When predicting customer evaluation criteria, ML algorithms will not produce a model of equal performance.

H4. When predicting customer evaluation criteria, Random Forest algorithm will perform 10% better than Decision Tree algorithm.

H5. When applying supervised machine learning techniques for a new business opportunity, prioritized evaluation criteria is predicted at same evaluation as H1-H4.

H5.1. Market area, customer type, contract type, period of performance, and product service code are good predictors for prioritized customer evaluation criteria.

H5.2. When applying supervised machine learning techniques for a new business opportunity, prioritized evaluation criteria is predicted with an AUC higher than 70%.

H5.3. When predicting prioritized customer evaluation criteria, ML algorithms will not produce a model of equal performance.

H5.4. When predicting prioritized customer evaluation criteria, Random Forest algorithm will perform 10% better than Decision Tree algorithm

The below diagram shows the flow-down of the problem statement to the research hypotheses, to include the relationships and interdependencies of this research.

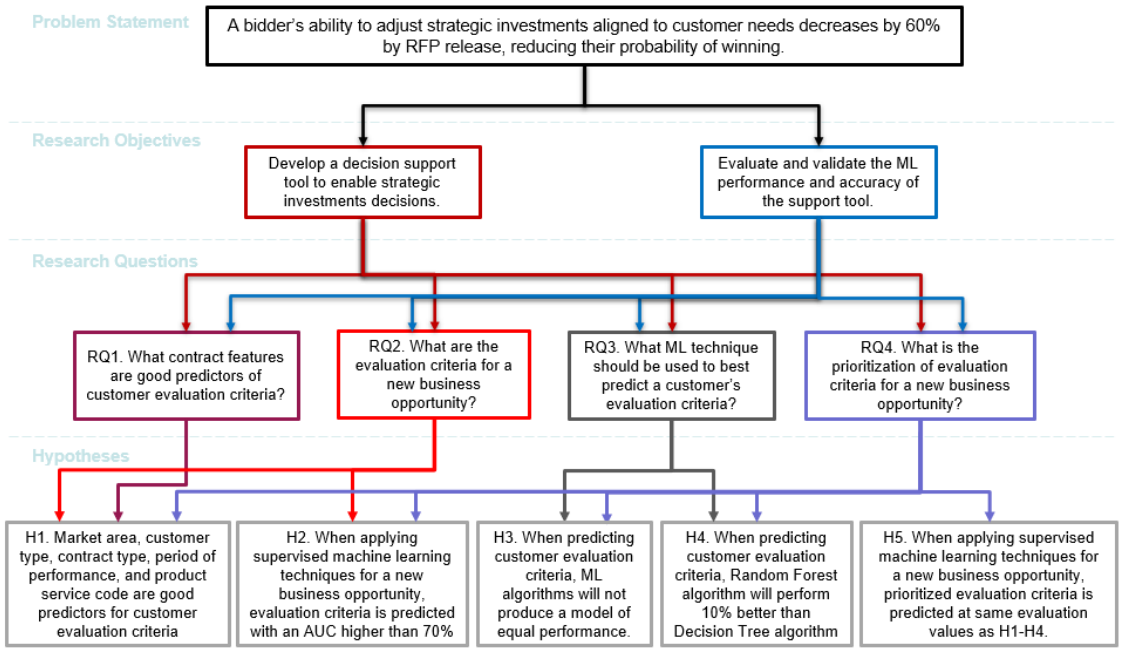


Figure 1. Research problem statement flow-down

1.7 Research Scope and Limitations

The results of this research reflect the creation of a novel collected dataset of 242 publicly released final RFPs from the United States (U.S.) Department of Defense (DoD). These final RFPs aligned from contract award values ranging from $20M-$100M awarded over the years 2016-2020. The DoD market areas analyzed was also limited to the following areas: military aircraft, ground vehicles, space, unmanned aerial vehicles, missiles, radars, weaponry, and satellites. The DoD customers were limited to the following: U.S. Navy, U.S. Air Force, U.S. Army, and Other Defense Agencies. Since

this research is bound by 242 RFPs, any growth in the dataset may impact performance measures of the ML algorithms used in the decision support tool. The benefit of ML algorithms' ability to continuously learn and adapt to expanding datasets allows this tool to increase robustness overtime compared to static models. Further tuning of ML configuration parameters could lead to an increase in prediction accuracy if this occurs, as minimal tuning was used in this research, and thus a limitation in the current results.

1.8 Praxis Structure

This praxis is divided into five chapters. This chapter, Chapter 1, introduces the background of the research and its objectives further discussed herein, to include the research questions and hypotheses that will be analyzed. Chapter 2 is the literature review that further supports the need for strategic investment enablement for companies in competitive environments, specifically the defense industrial base. This chapter will also review current research within the field of strategic decision-making in competitive environments, and the lack of a robust approach to predict customer evaluation criteria for companies pursuing new business opportunities. Finally, this chapter will address the need for the strategic research field to adopt proven methods from other research fields to increase rigor, like that of ML.

In Chapter 3, a detailed discussion is provided on the methodology used in this research to include data collection, various classification ML techniques applied, and a detailed explanation of evaluating and validating results using these techniques. The ability to evaluate the performance of newer techniques to the strategic research field is critical, given the important business decisions and impacts that are made from experts in this field of study. The application and execution of this methodology is found in the

proceeding chapter, Chapter 4. The results of this research are provided along with an explanation of the hypothesis test results. Finally, Chapter 5 provides a discussion of the findings to highlight the impact of this novel research to industry, future recommended research, and final conclusions.

Chapter — 2 Literature Review

2.1 Introduction

Companies must look beyond a strong technical solution and instead focus on adjusting other considerations that drive business results, such as cost, schedule, and the manufacturing of their product, to stay competitive (Brindley, Wood, & Mullen, 2017). These considerations align with customer priorities and are examples of evaluation criteria provided in an RFP. Predicting evaluation criteria far before release of the RFP directly supports investment decisions that strengthen a bidder's competitive position. However, a company begins their competitive positioning far before the RFP is released, as illustrated in Figure 2. Competitive positioning starts once a company has investigated the business opportunity and decides to pursue the opportunity. A bidder's ability to adjust strategic investments and increase competitiveness is directly inverse to their cumulative investment spent.

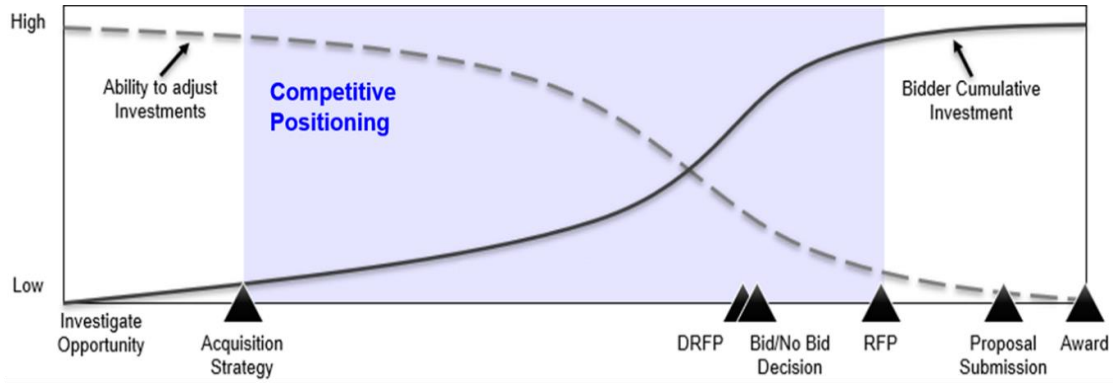


Figure 2. Company's ability to adjust investments versus their cumulative investment spend aligned to the customer acquisition timeline. This graphic is based on O'Guin and Kelly (2012) and is recreated with modifications.

This chapter will discuss the need for focused strategic investments for company competitiveness, current methods used, and the opportunity for application of advanced analytical methods that have not fully matured in the strategy research field.

2.2 Strategic Investments within the Defense Industrial Base

While the findings of this research can be applied to any industrial-base, this research focused on the defense industry as an exemplar for increased competitive pressure from near-peer competitors, non-traditional defense companies, and customer driven initiatives. Over the last decade, the U.S. DoD has published internal guidance to promote effective competition across the defense industrial base and exert pressure on companies to reduce costs. In 2010, the DoD published Better Buying Power (BBP), an internal guidance document, to promote effective competition and reduce the amount of sole source system providers (Carter, 2010). This guidance evolved to version BBP 3.0, with strategic intent to control costs by applying additional pressure to existing industry incumbents by incorporating evaluation criteria beyond the technical offering, such as past performance (Kendall, 2015). Past performance considers a company's existing contract performance health to include staying within the contractual agreement of the customer's budget and product schedule delivery. Companies often make decisions to invest in improving their current work performed to mitigate any negative effects on their probability of winning new business opportunities.

The cultural shift toward a more competitive environment within DoD's acquisition of new systems approach has had inverse implications. In 2020, Duddy, Landucci, and Knechtel conducted a study showing that the DoD increased total contract awards from $285 billion to $381 billion between the years 2014 and 2019, however, the amount of

competitive award money decreased, leaving only half of the $381 billion awarded through effective competition (Duddy, Landucci, and Knechtel, 2020). Duddy, Landucci, and Knechtel found that the customer made numerous exceptions to contracts originally slated for open competition. The documented exceptions argued that companies made extensive strategic investments to strengthen their core competencies, resulting in high start-up costs for non-incumbents, thus outweighing benefits for the DoD to pursue competitive options.

In 2021, Etemadi and Kamp studied statistical association with DoD acquisition strategies and market conditions. They also identified that government decisions in their acquisition approach, such as contract type, development model, and product delivery speed are major influences on companies' investment decisions and financial health to include research intensity, earnings, and revenue within companies (Etemadi and Kamp, 2021).

The companies in the defense industrial-base are making these extensive investments to protect themselves from the ongoing threat of growing competition and losing their market share. They are not only using their strategic investments to solidify core competencies but also increasing their internal research and development investments to mature technologies in anticipation of customer needs that are identified in evaluation criteria (Fanning, 2019). These technology investments also enable company core competencies and infrastructure.

However, a recent study conducted of defense industry leaders showed that 90% participants identified a need to shift strategies to selective investments that increase efficiencies and strengthen existing core competencies, rather than broadening their

product offering (Blank, 2019). Blank also states that defense industry leaders did not believe that an increase in investment was necessary, but that focus on increased selectivity and core competencies was needed to increase their competitiveness for future contract awards.

Due to the defense industry's focus on specific investments, this research analyzed the evaluation criteria from various DoD customers and market areas to provide early insights into anticipated customer needs. The results directly enable strategic investment specificity to increase a company's competitive advantage.

2.3 Current Methods of Strategic Decision-Making to increase Competitive Positioning

Current research fails to address methods to predict evaluation criteria to increase competitive positioning. Instead, related research of strategic decision making with respect to the acquisition of new business focuses primarily on general proposal process considerations and pricing strategies that occur during the late Bid/No Bid Decision milestone, thus not improving a company's competitiveness (Chua & Li, 2000; Kreye, Newnes & Goh, 2014). Early access to information is key to positioning a company to compete for new business and understanding customer evaluation criteria with relation to the proposed solution is critical for staying competitive (Hari, Kasser, and Weiss, 2007).

A company's approach to competitiveness can be categorized by the Asset-Processes-Performance (APP) framework (Table 1), where most research and studies fall into the Processes and Performance categories (Maune, 2014). According to Maune, current research has provided quantitative and qualitative findings but have not yet integrated across the AAP categories. The earlier a company can predict evaluation

criteria will allow the company to adjust investment considerations across the AAP framework.

Table 1. APP Framework for categorizing competitivesnss recreated from Maune (2014)

Main Category	Subcategory
Assets	• Human resources • Firm structure and culture • Technology • Resource Based View (RBV)
Processes	Strategic Management Processes • Competency • Competitive strategy • Flexibility and adaptability HR Processes • Design and deploy talents Technological Processes • Innovation • Systems • IT Operational Processes • Manufacturing • Design • Quality Marketing Processes • Marketing • Managing Relationships • Persuading Power
Performance (Financial and non-financial)	• Productivity • Finance • Market share, differentiation, customer satisfaction • Profitability, Cost, Price • Variety, product range, new product development • Efficiency, Value creation

Several other frameworks have been developed to support strategic processes, such as agent-based modeling simulation to analyze the interaction of decision-making behaviors for effective decision making (Calafut, Sarkani, and Mazzuchi, 2021). Calafut, Sarkani, and Mazzuchi identified an optimal strategy region for effective decision making that results in periodic decision-making as a competitive advantage. This framework provides foundational support to the need for early competitive insights, however, does not address the specific approach to obtain that information.

Other process frameworks, provide systematic approaches to consider performance risk, profitability, and investment impacts to consider once customer evaluation criteria are already known and the impact to probability of winning that new business (Li, Zhang, Chen, and Martek, 2020; Ward and Chapman, 1988). Ward and Chapman (1988) developed a framework that allows strategic decision makers to think through bid preparation from a systematic perspective. They further discuss the efficiencies and effectiveness of adopting a systematic approach and evaluation criteria to consider during this strategic phase. However, they further acknowledge the challenges of quantifying the uncertainty in such an early phase of the acquisition timeline.

Li, Zhang and Martek's (2020) developed a framework based on conducted surveys to identify five factors for Bid/No-Bid consideration: country risk vulnerability; ability to perform the contract; project risk vulnerability; profitability; probability of winning. This was further validated using structured equation modeling. However, their research focuses on a comprehensive framework to assess Bid/No-Bid decisions that occur after a draft version of the evaluation criteria is released to the company, thus making reactive tactical decisions, not upfront strategic decisions. The Bid/No-Bid decision milestone, as

illustrated in Figure 2, has an extremely late-to-need opportunity space for optimizing a company's competitive advantage.

Other studies that support the bid process using empirical methods include Chua and Li's bid reasoning model using the Analytical Hierarchy Process (AHP) technique in examining Bid/No-Bid considerations within a company (Chua and Li, 2000). This research was one of the few identified that used quantitative methods for decision-making during the acquisition timeline, however, also falls short in the ability to address early strategic investment decisions. Other empirical studies focused purely on the price estimating process for a bid, which is an expected and known evaluation criterion for all solicitations according to O'Guin and Kelly (2012). Without insight to the customer's evaluation criteria, uncertainty of investments decisions becomes a critical consideration in the bid process, as limited information is available at that time.

In 2014, a framework to account for unknown evaluation criteria was developed by Kreye, Newnes, and Goh. To account for the unknown evaluation criteria, uncertainties were accounted to include customer budget, available technology, and the competition (Kreye, Newnes, and Goh, 2014). They further state that companies must acknowledge the difference in the decision-making process within a competitive environment. Their research identified three key uncertainty considerations that influence the decision-making process and the lack of information available at that time. The three key uncertainty considerations are: company related - design, estimate, expanded contract portfolio; customer related: budget limitations and evaluation criteria for bids; and competition related: competitor's identity, cost, and available technology. The research

that will be further discussed in this paper will address the evaluation criteria in the customer related key uncertainty area.

The literature provides process considerations throughout the acquisition timeline but are not focused on predicting or providing early insights to customer needs, thus providing reactive solutions. The business case to pursue a new business opportunity decreases if resources are strained or if there is a lack of preparation of technical solution, understanding the customer needs and assessing the competition.

In order to stay competitive, companies must adjust other considerations such as cost, schedule, and the manufacturing of their product that drive business results, often identified as evaluation criteria (Brindley, Wood, and Mullen, 2021). Brindley, Wood, and Mullen go on to discuss that companies must protect revenue for competitions and often restructure their offers to secure profitability and win probability. The value proposition to compete for new business decreases if resources are strained or if there is a lack of preparation in the combination of a technical solution, customer assessment, and competitive assessment. They further discuss the need for rigorous Position to Win (PTW) and Design to Cost (DTC) analyses to avoid late, costly No-Bid decisions, both of which are based on insight how the customer will down select the proposed solution, via the evaluation criteria. Brindley, Wood, and Mullen continue by further elaborating on the critical need for companies to have insight to what levers to adjust beyond a technical solution for today's competitive climate.

Strategy analysts that perform PTW and DTC analyses rely heavily on competitive intelligence practices to understand and analyze customer evaluation criteria (Maune, 2014). However, these practices are limited to small historical samples of comparable

RFPs and tacit knowledge of customers from subject matter experts. Analogous RFPs are collected, reviewed, and analyzed to identify previous customer evaluation criteria and buying behaviors.

In addition to the uncertainty of evaluation criteria, the relationship between criterion also ranges per RFP. The relationship between criterion is stated qualitatively, as shown in Figure 3, introducing an additional complexity to the prioritization of criteria and thus their strategic investments to increase their competitive position.

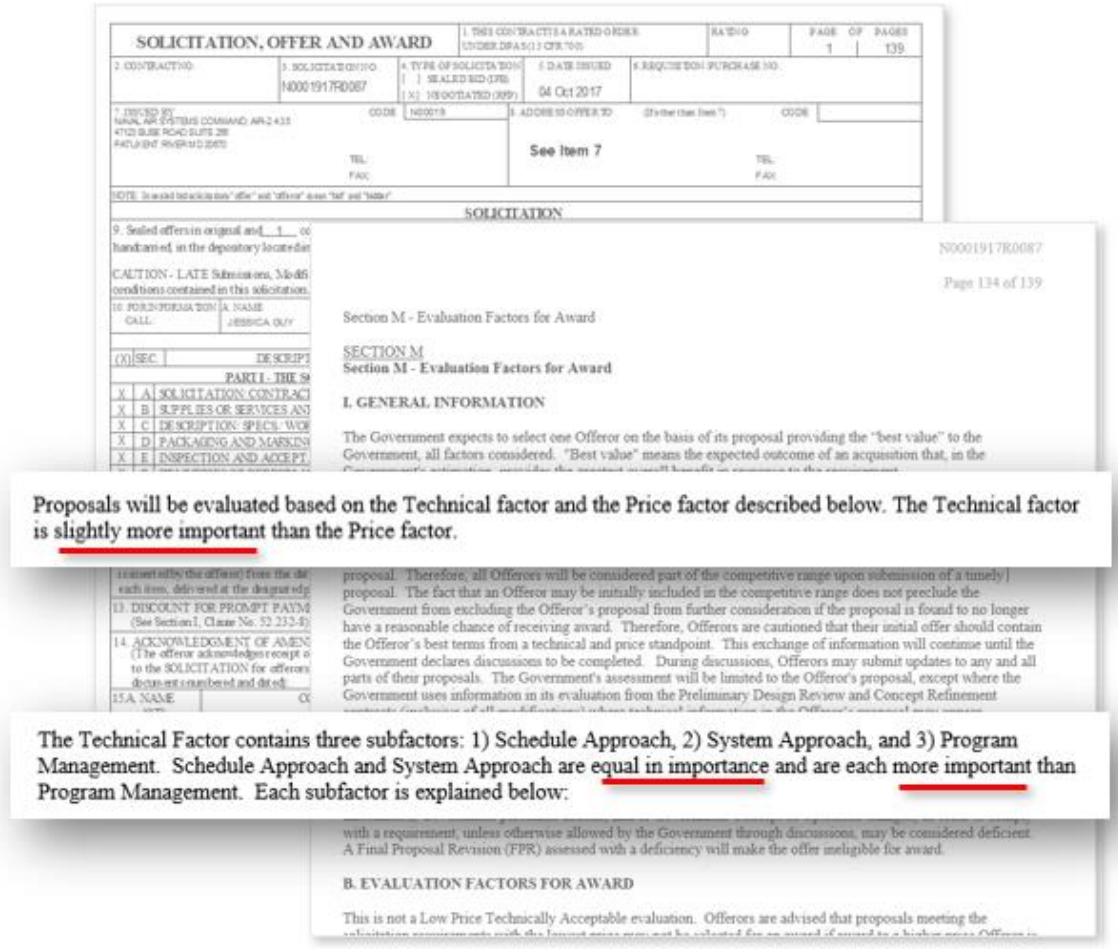


Figure 3. Example customer evaluation criteria from RFP – Naval Air Systems Command. (2017). Solicitation No. N0001917R087

Throughout this process, key decision-makers must ensure that the team also assesses the accuracy and applicability of the information before determining if the findings are adequate to make important strategic investment decisions (Rundall et al., 2007). These methods are time consuming, expensive, and induce bias and subjectivity into the process (Miller, 2020). This lengthy process hinders timely and effective decision-making, a major factor for maintaining a competitive advantage (Calafut, Sarkani, & Mazzuchi, 2021). This research meets the need for early insights and targeted investment considerations, providing decision-makers a novel and cost-effective approach to produce predicted evaluation criteria. This process includes built-in validation of information applicability and quantitative performance values of the prediction, at a fraction of the time in comparison to current methods.

2.4 Machine learning (ML) in strategic decision-making

Combining or applying existing elements from other fields leads to innovation, increased accuracy, and opportunities for exploration. In 2016 the Strategic Management Journal focused on the need for innovative approaches, leveraging other fields, to be used for strategic analysis (Arora et al., 2016). Arora, Gittelman, Kaplan, Lynch, Mitchell, and Siggelkow state that much of strategic research is complex and often times need to have an iterative approach in solving old strategy questions as novel approaches are developed. Innovation consists of new combinations of existing elements or application of existing elements. They further discuss that newer strategy research is demonstrating the utility of external-field methods and comparing the results of the new methods to existing methods, including the increase in accuracy and areas that may not have been explored

due to limiting methods. The crossing of fields allows the research to be driven by questions and not available data (Arora et al., 2016).

Unlike static and linear methods, ML allows iterative and progressive outcomes based on adaptive and dynamic cycles; therefore, ML can be applied to real-world tasks (Provost & Kohavi, 1998). Growing applications of ML capabilities has positioned the technique as an efficient and important addition to applied research. Fields, such as software requirement development, have demonstrated that benefits of ML techniques include reduced costs, shorter time-to-market, and increased product competitiveness (Perini, Susi, & Avesani, 2013). Recent studies in economic fields have shown an increase in ML application to predict financial impacts in the stock market (Fang, Chen, & Xue, 2019). Similarly, Zhang and Hamori forecast gas crises by developing an early warning system using classification ML algorithm. (Zhang & Hamori, 2020).

The literature continues to show that forecasting future market directions and the need for company strategic investments has increased research focus on efficacy of different forecasting procedures. In 2020, Mascio, Fabozzi, and Zumwalt conducted a comparative study to forecast market timing of Standard and Poor's (S&P) 500 Index returns using traditional approaches such as a market sentiment indicator and logistic regression models against ML algorithms. Their research demonstrated that the timing of current computing speeds has allowed for ML models to run more efficiently and affordability to traditional approaches in that field, thus advocating for further exploration for market analysis. They further argue that current studies may result differently from past studies due to the historical limitations of manual processing power (Mascio, Fabozzi, & Zumwalt, 2020).

Strategic research is complex and needs an iterative approach to address strategy-related questions as innovative approaches are developed. In 2021, Choi, Menon, and Tabakovic identified the lack of adequate data and methods hindering strategy research that can be generalized across industries. Their research focused on the application of natural language processing (NLP) to develop a novel diversification-performance relation through analyzing corporate annual reports. Although ML techniques have been used in other fields, much of the data within strategy research is textual. This has led to traditional strategy research avoiding many large-scale empirical analyses due to strategy documentation's lack of structure (Choi, Menon, & Tabakovic, 2021).

Although ML is a growing and trending field, it is currently underutilized by recent strategy and management research (Choudhury, Allen, & Endres, 2021). Choudhury, Allen, and Endres focus on the introduction, or bridging of ML techniques to the field of strategy and management research by providing an in-depth discussion and examples of ML algorithms and the interpretation of performance measures.

Benefits of classification ML algorithms, such as pattern discovery, which may otherwise go unnoticed when dealing with complex and unstructured textual data was highlighted in a recent strategy research study by Miric, Jia, and Huan (2022). Their study demonstrated the ability to analyze unstructured patent text data using ML techniques. Their use of supervised ML algorithms is an early to field demonstration of how the strategy field can leverage historically qualitative information into quantitative variables and classifications (Miric, Jia, & Huan, 2022).

Using ML techniques in this research can reveal customer evaluation criterion patterns based on new business opportunity characteristics, such as market area, specific

customer type, and contract value. This applied method provides decision-makers a novel approach to quantifiable measures to assess core competency health, investment considerations, and risk tolerance.

2.5 Chapter Summary

From the findings of the literature review, competitive pressures across the defense industrial base will lead decision-makers to look for innovative, quantifiable methods to support their strategic investment decisions. The application of ML techniques to strategic investments within the strategy field yield benefits that could not materialize from traditional methods. Furthermore, the variability of customer evaluation criteria in RFPs has inhibited decision-makers' ability to invest with quantifiable risk. This research will enable investment decisions that strengthen proposed solutions for strategy managers, strategic planners, and financial decision-makers in competitive environments.

Chapter — 3 Methodology

3.1 Introduction

In order for a company to make strategic investments that will position their competitiveness in alignment with the customer evaluation criteria, the company must have early insight to what that criteria will be. The criteria, however, depends on multiple parameters to include what market the new business opportunity is within, which customer is acquiring the new system, how large of a contract would be awarded – both monetarily and temporally. Identifying the combination of these parameters and their influence on the criteria has not been explored due to the complexity of unorganized, textual-based data, making the initial steps of this completed research novel. Furthermore, once the parameters are identified, the predictive performance measures provided through ML techniques allow this research to continue to be refined over time and directly support strategy analysts, enterprise architects, systems engineering leads, and other decision-makers insight to where focused investments and trades must be made.

This chapter will detail the methodology used in this research to identify the parameters, or contract features (independent variables) needed to develop a decision support tool, the different ML algorithms used and how to understand their accuracy and other performance measures, and validation process. As noted in Chapter 1, the decision support tool was developed in two stages. Stage 1, predicting the presence of evaluation criteria, will be discussed in section 3.2 through 3.6, whereas the differentiation of Stage 2, predicting prioritized evaluation criteria, will be discussed in section 3.7. The

methodology overview is illustrated in Figure 4 with details further discussed in the sections identified.

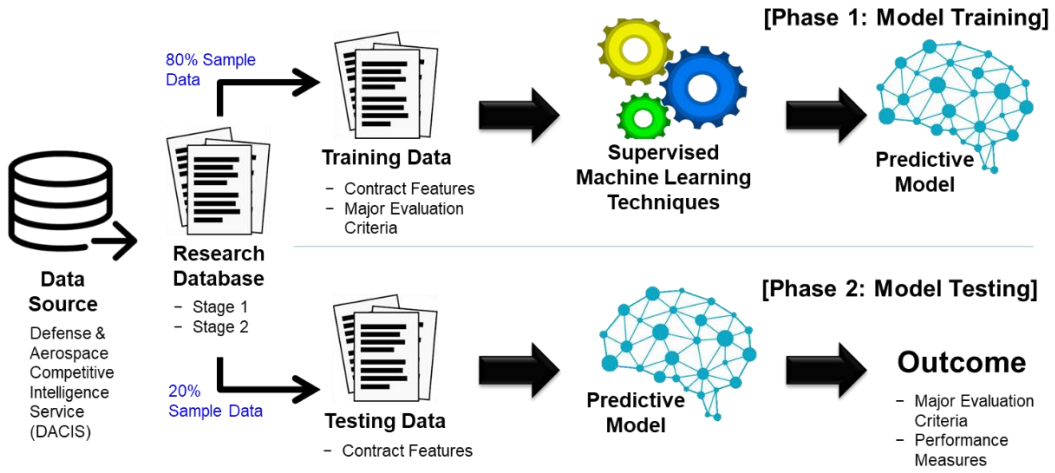


Figure 4. ML model development process. This graphic is based on Sarker (2021) and has been recreated with modifications.

3.2 Data Collection

The data collected to support this research consisted of 242 DoD RFPs with contract award values between $20M-$100M awarded over 2016-2020. The historic RFPs were collected using the Defense & Aerospace Competitive Intelligence Service (DACIS) system, which provides open and unclassified solicitation documentation released by the U.S. government. The four DoD customers used in this research included: U.S. Navy, U.S. Air Force, U.S. Army, and Other Defense Agencies. The RFP documents from these four DoD customers spanned eight market areas. These markets were military aircraft, ground vehicles, space, unmanned aerial vehicles, missiles, radars, weaponry, and satellites. The collected RFP count by customer and market area are illustrated in the figures below (Figure 5 and Figure 6, respectively). This is the first

known data collection of DoD RFPs, reviewed and sorted by descriptive information and evaluation criteria, as several strategy experts and leaders in the field were consulted prior to data collection.

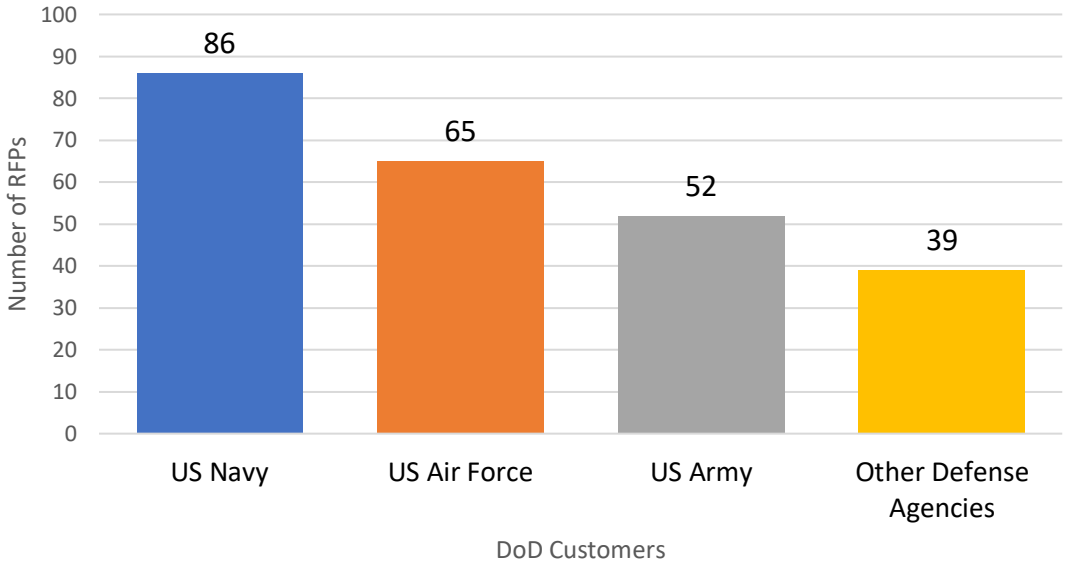


Figure 5. RFPs reviewed by customer count

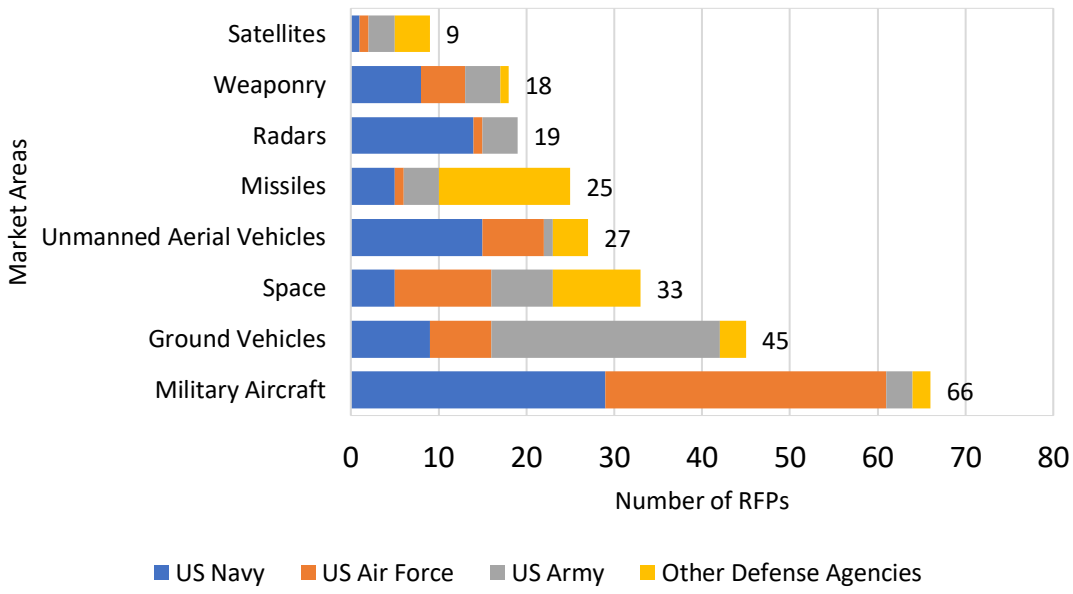


Figure 6. RFPs reviewed by market area and customer count

3.2.1 Independent Variable Dataset

For each RFP, descriptive award information was collected to analyze patterns in the customer evaluation criteria. The descriptive award information, referred to as Contract Features in the remainder of this paper, was used as independent variables in the model to predict the presence of a customer evaluation criterion (Table 2).

Table 2. Contract Features collected from RFPs reviewed

Contract Features	Data Type	Description
ID #	ID	ID # is a unique identifier value assigned to each RFP reviewed and used in this research.
Market Area	Categorical	Market area specifies categories that are supported by the purchased product or service and include, but are not limited to, military aircraft, ground vehicles, space, unmanned aerial vehicles, and satellites.
Customer	Categorical	Customer specifies the major United States DoD branch and/or agency that issued the solicitation and contract award.
Contract Type	Categorical	Contract Type specifies the type of contract awarded by the government customer, often known as part of the customer acquisition strategy. (Federal Acquisition Regulation, 2022).
Contract Value ($M)	Numeric	Contract value specifies the total contract amount awarded in dollars-million.
Period of Performance (PoP)	Numeric	PoP specifies the duration of the effort awarded by the contract in unit months.
Product Service Code (PSC)	Categorical	PSC are codes to describe products, services, and research and development efforts purchased by the United States federal government and reported in the Federal Procurement Data System. (Product and Service Code Manual, 2022)
North American Industry Classification System (NAICS) Code	Categorical	NAICS codes are the standard to classify business establishments used by Federal statistics agencies and developed under the United States Office of Management and Budget. Examples of NAICS codes include, but are not limited to, Engineering Services (541330), Explosives Manufacturing (325920), and Flight Training (611512) (NAICS Manual, 2022)

3.2.2 Dependent Variable Dataset

Each RFP was reviewed to identify and document evaluation criteria. For all 242 RFPs reviewed, 18 unique evaluation criteria were identified. These "unique criterion" were consolidated into nine "major criterion" groups due to the commonality of some of the unique criterion meanings. This consolidation of unique criterion mapped to major criterion is shown with the designation of an "X" in Table 3. For example, the unique criterion of "Technical" and "Technical Merit" was consolidated to the major criterion of "Technical," whereas the major criterion of "Past Performance" only has one unique criterion – "Past Performance." The Price criterion has been omitted from this research, as it is a required customer evaluation criterion and is thus present in all RFPs.

Table 3. Unique evaluation criterion of major evaluation criteria mapping

Major Criterion / Unique Criterion	Price	Technical	Program Management	Past Performance	Small Business	Importance to Customer	Schedule	Mission Capability	Quality	Security Plan
Technical		X								
Program Management			X							
Past Performance				X						
Small Business					X					
Price	X									
Technical Merit		X								
Importance to Customer						X				
Transition Plan & Capability									X	
Cost & Schedule Realism	X						X			
Mission Capability								X		
Funds Availability	X									
Quality									X	
Org. Conflict of Interest Plan			X							
Security Plan										X
Product Sample Acceptability									X	
Schedule							X			
Business Viability			X							
Participants					X					

For each line item (RFP) in the database, the contract features and criterion were documented in preparation for the model development. The figure below is an example subset of the database created for the first stage of the research, that is, to identify the presence of a criterion based on the contract features of a new business opportunity (Figure 7). Further examples of the raw database can be found in Appendix A – C.

Market Area	Customer	PoP (months)	Contract Type	Contract Value ($M)	PSC	NAICS	Technical	Program Management	Past Performance	Small Business	Price	Importance to Customer	Schedule	Mission Capability	Quality	Security Plan
Aircraft (Military)	US Navy	60	Fixed Price	49.9	13	325998	1	0	1	1	1	0	0	0	0	0
Aircraft (Military)	US Air Force	62	Fixed Price	43.1	U009	611512	1	1	0	0	1	0	0	0	0	0
Aircraft (Military)	US Air Force	60	Fixed Price	99	1680	336413	1	0	1	0	1	0	0	0	0	0
Aircraft (Military)	US Air Force	63	ID/IQ	44.5	N016	336411	1	0	0	0	1	0	0	0	0	0
Aircraft (Military)	US Air Force	64	Fixed Price	25.3	1650	336413	1	0	0	0	1	0	0	0	0	0
Aircraft (Military)	US Air Force	60	Fixed Price	35	28	336412	1	0	0	0	1	0	0	0	0	0
Aircraft (Military)	US Air Force	73	Fixed Price	40	15	336413	1	0	0	0	1	0	0	0	0	0
Aircraft (Military)	US Air Force	60	Fixed Price	34.8	R425	561320	1	0	0	0	1	0	0	0	0	0
Aircraft (Military)	US Air Force	27	Fixed Price	36.7	15	336411	1	0	0	1	1	0	0	0	0	0
Aircraft (Military)	US Navy	60	Cost Plus Fee	68.7	K	541330	1	0	1	0	1	0	0	0	0	0
Aircraft (Military)	US Air Force	66	Fixed Price	46.3	U	611512	1	1	0	0	1	0	0	0	0	0
Aircraft (Military)	US Air Force	60	Fixed Price	20.5	15	336413	1	0	0	0	1	0	0	0	0	0
Aircraft (Military)	US Air Force	132	ID/IQ	65.7	15	336413	1	0	1	1	1	0	0	0	0	0
Aircraft (Military)	US Air Force	120	Fixed Price	68.6	L	541330	1	0	0	0	1	0	0	0	0	0
Aircraft (Military)	US Air Force	60	Fixed Price	28.5	17	336412	1	0	1	0	1	0	0	0	0	0
Aircraft (Military)	US Air Force	53	Cost Plus Fee	39.4	16	336411	1	0	1	0	1	0	0	0	0	0
Aircraft (Military)	US Navy	73	Fixed Price	94	13	325998	1	0	1	0	1	0	0	0	0	0
Aircraft (Military)	US Air Force	36	Fixed Price	68.5	23	333924	1	0	0	0	1	0	0	0	0	0
Aircraft (Military)	US Air Force	51	Fixed Price	38.8	99	488190	1	0	1	0	1	0	0	0	0	0
Aircraft (Military)	US Navy	61	Fixed Price	39.9	13	325920	1	0	1	0	1	0	0	0	0	0
Aircraft (Military)	US Air Force	56	Fixed Price	36	15	334511	1	0	0	0	1	0	0	0	0	0
Aircraft (Military)	US Air Force	72	Fixed Price	48.8	J	336411	1	0	1	0	1	0	0	0	0	0
Aircraft (Military)	US Air Force	48	Fixed Price	45	U	611512	1	0	1	1	1	0	0	0	0	0
Aircraft (Military)	US Air Force	90	Fixed Price	68	15	334511	1	0	1	1	1	0	0	0	0	0
Aircraft (Military)	US Air Force	60	ID/IQ	55.7	R	561320	1	0	1	0	1	0	0	0	0	0

Figure 7. Sample subset of database created for stage 1.

3.2.3 Alteryx Designer x64

Several commercial ML tools exist to support the development of ML models. This research used Alteryx Designer x64 and its Assisted ML modeling feature, which provided a guided approach to ML model building. Once the collected RFP database was entered into Alteryx, the data was partitioned to construct one sample for Phase One (model training) and another for Phase Two (model testing) using an 80%-train and 20%-test partition, respectively (Figure 8).

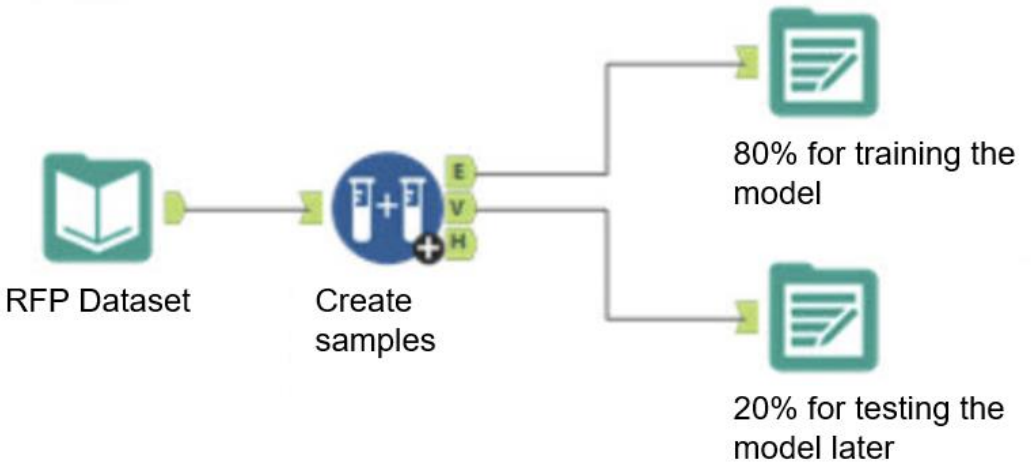


Figure 8. Sample workflow to create samples in Alteryx Designer x64 recreated from Alteryx overview site *Assisted modeling tool*. Modeling Tool. (2023)

After partitioning the dataset, the Alteryx Assisted ML model stepped through building the workflow pipeline to create a predictive model with different ML algorithms (Figure 9). These steps included setting the data type of each parameter within the database, accounting for any missing values in the database, feature selection using Gini Impurity calculations, coding the selected supervised ML algorithm selected, then training the model. The database used in this research is complete and had no missing values.

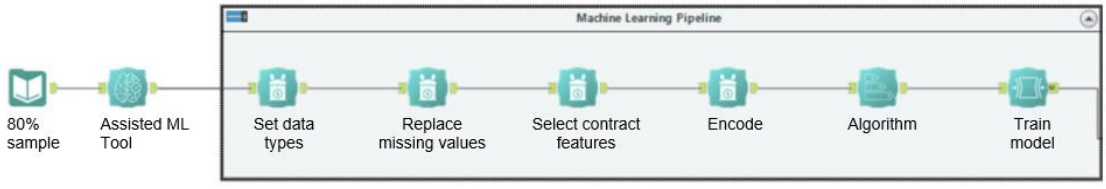


Figure 9. Sample workflow to create training model in Alteryx Designer x64 recreated from Alteryx overview site *Assisted modeling tool*. Modeling Tool. (2023)

The last step in the Assisted ML modeling feature within Alteryx Designer x64 was to validate the model using cross-validation with the sample test data setup in the initial data partition step (Figure 10). Cross-validation method was applied to reduce systematic error caused by incorrect assumptions, known as ML algorithm bias. A predictive ML model was created for each major criterion, as the dependent variable, to predict its presence based on the Contract Features. An example of the full model workflow can be found in Appendix D with a listing of configuration parameters in Appendix E.

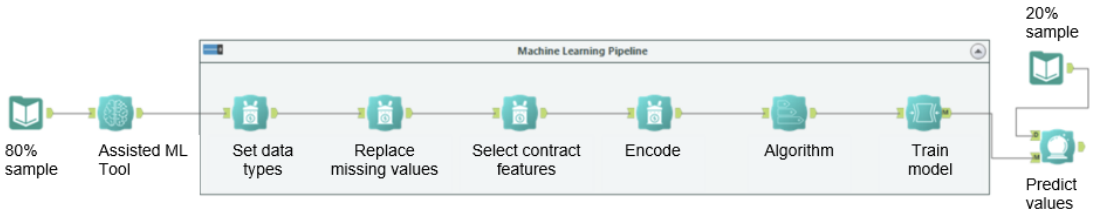


Figure 10. Sample workflow to validate model using test data sample in Alteryx Designer x64 recreated from Alteryx overview site *Assisted modeling tool*. Modeling Tool. (2023)

The following subsection describes the details of each analysis and ML technique applied during the Assisted ML modeling feature of the model development and how to evaluate the performance of each ML algorithm.

3.3 Gini Impurity

The independent variables, Contract Features (Table 1), were first analyzed for applicability of predicting evaluation criteria. This analysis is a common ML practice known as Feature Selection. Feature Selection supports lower computational cost, increased prediction performance, and easier model interpretability (Guyon & Elisseef, 2003; Miao & Niu, 2016). Within classification ML algorithms, information is

continuously split by features into categories until the data is completely compartmentalized. Selecting which features to use is a key step to safeguard prediction accuracy. Suppose we were predicting if an observed bird were a duck, we could split the dataset into birds that can fly and birds that cannot fly. If the observed bird could fly, it is likely that predicting the bird is a duck is wrong because not all ducks can fly. Misclassifications can dramatically reduce prediction accuracy. The quantitative means used to assess the likelihood of misclassified data in this research is known as the Gini Impurity (GI). The GI is calculated by the following equation:

$$GI = \sum_{i=1}^{n} (p_i)^2$$

where n is the number of class labels and p_i is the proportion of i^{th} class label

The GI can range between a value of 0 and 1, where an independent variable is deemed too weakly (GI < .02) or too highly (GI > .5) associated with the dependent variable and deemed as a "good predictor" when the GI is between .02 and .5. The lower the GI result, the lower the likelihood of misclassifying the targeted predicted output. The results of this analysis were used to test H1.

H1: Market area, customer type, contract type, period of performance, and product service code are good predictors for customer evaluation criteria.

3.4 Classification ML Algorithms

Supervised ML algorithms can be broken into two categories: regression and classification. Regression algorithms are used to predict continuous values, whereas classification algorithms are used to predict discrete values. This research used classification algorithms to predict the presence of customer evaluation criteria, thus providing critical insights to companies to facilitate early strategic investment decisions. Using ML algorithms enables the model to produce dynamic outputs based on continuously evolving datasets, allowing this research to grow and adapt to future defense industry and any impact to future evaluation criteria.

ML algorithms selected for this research were Decision Trees (DT), Random Forest (RF), and Extreme Gradient Boost (XGBoost). The DT algorithm is one of the easiest classification algorithms to interpret. RF and XGBoost algorithms are also DT algorithms but are more complex and take longer to perform than the simpler DT algorithm. The following subsections provide a brief overview of each algorithm used in this research.

3.4.1 Decision Trees (DT) overview

The DT algorithm is a supervised ML technique used when independent and dependent variables have classification labels. The algorithm trains the model by producing splits at a decision node forming branches and new nodes until the data can no longer split and become what is known as leaf nodes – the final classification of the data. In the animal feature selection example, we assessed if a bird was a duck based on an attribute feature. The feature specifying the bird's flight capability would be considered a decision node. If the bird dataset were revisited to start a root node with the feature of swim capability, this feature would then be split into continuous branches to predict the

bird type, as illustrated in Figure 11. If the test resulted “true,” a new decision node would assess if the bird had webbed feet. If this test resulted “true,” the final leaf node would predict the animal was a duck. A major benefit of the DT algorithm is its intuitive flowchart-like structure allowing for high interpretability. The DT algorithm best performs when modeling nonlinear associations between features. Although this method is computationally efficient, it is subject to aligning predictions too closely to the data used in Phase One (model training), known as overfitting.

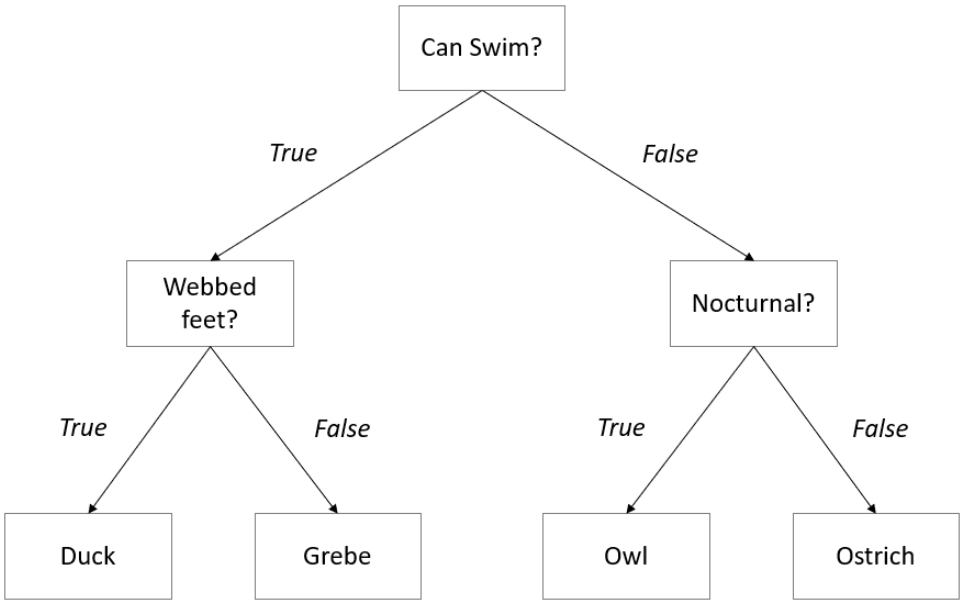


Figure 11. Decision Tree Example

3.4.2 Random Forest (RF) overview

A RF classification algorithm is known as an ensemble method, in that it builds numerous, randomized DTs to obtain results. As RF is an ensemble DT algorithm, it carries many characteristics of DT and performs best when modeling nonlinear associations between features. RF is known to manage imbalanced data better than a

single DT and for capturing subtle feature effects, while being less prone to under and overfitting. However, due to its complexity, it takes longer estimation time and is computationally more expensive.

3.4.3 Extreme Gradient Boost (XGBoost) overview

XGBoost is an algorithm that builds many DTs to model the association between independent and dependent variables, including weak associations, known as gradient boosting (Friedman, 2001). Due to this algorithm's boosting capability, it is often referred to as an improvement algorithm, a method where DTs improve each other by combining weak models into stronger ones, thus reducing bias-error in ML (Chatzis, Siakoulis, Petropoulos, Stravroulakis, & Vlachogiannakis, 2018).

3.5 Validation

As mentioned in previous sections, the model uses cross-validation to ensure validity of the predictive model created. This attribute of ML modeling is one of the additional benefits of the application of this methodology in comparison to current methods. The cross-validation method is setup on the basis of the partitioned dataset for training the model (80% sample data) and the testing dataset (20% sample data).

Cross-validation develops the model by training the ML algorithm based on the training data, leaving out a portion of the full database. The remaining data is then used to test the algorithms, which produces many of the performance measures values. This helps the decision-maker understand the uncertainty innate to the model. The hypotheses for this research are based on the performance measures produced from the validated model, of which is further discussed in the proceeding sections.

3.6 Performance Measures

Different performance measures are used to evaluate ML models based on the specific question the ML model is solving. It is important to understand the several ways of assessing the performance of ML models to mitigate bias in ML results (Choudhury, Starr, & Agarwal, 2020). These are the quantifiable measures that support decision-makers in understanding the usefulness and uncertainty of the prediction model. This research is the first and only known predictive modeling of customer evaluation criteria with model performance measures, providing decision-makers a quantifiable assessment of early insight into customer evaluation criteria.

The Confusion Matrix is a standard ML performance measure table that summarizes the classification results of correct and incorrect predictions. It captures where the model's predicted target (i.e., major evaluation criterion) was truly present or absent, known as True Positive (TP) or True Negative (TN), respectively. The matrix also displays if the model was confused and predicted that the target would be present when it was truly absent, or vice versa, known as False Positive (FP) and False Negative (FN), respectively. These performance measures are the foundational performance measures that will enable more advanced performance measures discussed later in this section. Table 4 below provides the layout of a confusion matrix where the actual target values are compared to the predicted values.

Table 4. Confusion Matrix

		Predicted Values: Positive	Predicted Values: Negative
Actual Values	Positive	True Positive (TP) *Correct Prediction*	False Negative (FN) *Incorrect Prediction*
	Negative	False Positive (FP) *Incorrect Prediction*	True Negative (TN) *Correct Prediction*

- Accuracy – accuracy is defined as the total value of the data for which the model had a correct prediction, yielding the equation:

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

The next set of performance values leverage the results from the confusion matrix and address a more useful performance diagnostic (Metz, 1978). These values were developed to account for imbalanced data, which is a common challenge in classification problems (Bekkar, Djemaa, & Alitouche, 2013).

- Sensitivity – provides the percent of correctly predicted positive values among all actual positive values:

$$Sensitivity\ or\ True\ Positive\ Rate\ (TPR) = \frac{TP}{(TP + FN)}$$

- Specificity – provides the percent of correctly predicted negative values among all actual negative values:

$$Specificity\ or\ True\ Negative\ Rate = \frac{TN}{TN + FP}$$

- Balanced Accuracy (BA) – unlike Accuracy, BA leverages the usefulness of sensitivity and specificity by calculating the average of the two performance measures:

$$BA = \frac{1}{2}\ (Sensitivity + Specificity)$$

- False Positive Rate (FPR) – provides the probability of incorrectly predicted positive values among all actual negative values:

$$False\ Positive\ Rate\ (FPR) = \frac{FP}{TN + FP}$$

The previously discussed performance measures reflect the direct observations of the algorithms' predictability. However, the final selection of an algorithm, based on these performance measures, is subjective and dependent on decision thresholds defined by the decision-maker and objective of the study.

Unlike other performance measures where decision thresholds are chosen arbitrarily, the Receiver Operating Characteristics (ROC) curve deliberately varies the decision

threshold, accounting for changes in TPR and FPR (Metz, 1978). The ROC curve is the most widely used performance measure when comparing algorithm performance and provides a graphic representation of the comparison of cost-benefit and impact between misclassified predictions and correctly classified predictions.

The ROC curve, as depicted in Figure 12, is plotted on a vertical (TPR) and horizontal axis (FPR) that range between zero and one. A 45-degree reference line is often included to illustrate a curve that produces random true or false positive results. Results above this reference line provide better-than-random predictions, whereas points below the reference line represent worse-than-random results (Choudhury et al., 2021). However, this approach is often deemed impractical for human interpretation, thus an Area Under the Curve (AUC) is calculated as a summary metric for ROC (Metz, 1978).

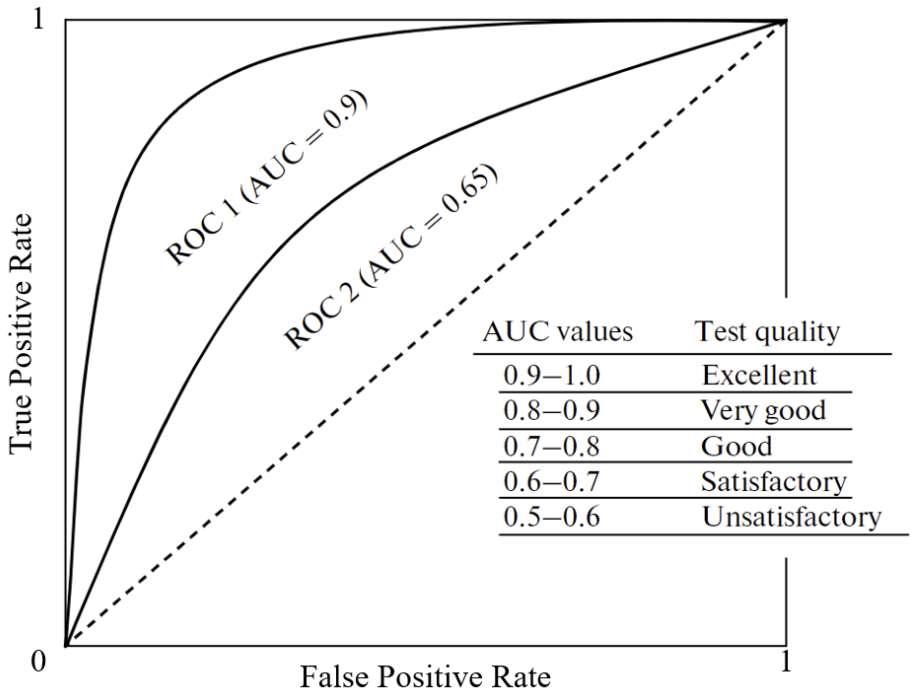


Figure 12. ROC examples and AUC quality of test scale recreated from Trifonova, Lokhov, and Archakov (2013)

Comparing AUC values supports algorithm selection, where a higher AUC value corresponds to a more accurate prediction. Threshold definitions to assess the quality of these algorithms are based on heuristics and are subjective to the objective of the study; however, the goal of this research is to develop a predictive model that obtains results that are better-than-random, or an AUC above 0.5 (Trifonova et al., 2013).

The AUC values produced per ML algorithm, from this methodology were used to test H2 – H4.

H2. When applying supervised machine learning techniques for a new business opportunity, evaluation criteria is predicted with an AUC higher than 70%.

H3. When predicting customer evaluation criteria, ML algorithms will not produce a model of equal performance.

H4. When predicting customer evaluation criteria, Random Forest algorithm will perform 10% better than Decision Tree algorithm

3.7 Stage 2: Application of Analytic Hierarchy Process

As mentioned in previous sections, the above methodology was used to develop a model that identified the presence of evaluation criteria within binary constraints. Although early prediction of which evaluation criteria will be used to down-select a winning bid is helpful to companies for investment considerations, it lacks the level of prioritization to provide awareness of how much investment should be directed to the various criteria. For example, if Schedule is an evaluation criterion, but a low priority amongst the other criteria, a company may focus their efforts on the other criteria to increase their competitiveness. In that same example, another company may have already invested in areas within their enterprise, solution architecture, or resources that would enable their Schedule to be their strongest attribute in their proposed solution. This insight would allow the company to have an opportunity to shift their investment strategy to a different criterion that has not reached maximum optimized capacity. Another option is for that same company to work with the customer prior to final release of the RFP to influence the customer to increase the prioritization of the Schedule criterion, knowing their ability to meet or surpass Schedule is a differentiator amongst the competition.

The need for prioritized evaluation criteria prediction initiates the second stage of this research. The same database used for Stage 1 was revisited and the relationships between each major criterion, per RFP, were identified. As discussed in Chapter 2, DoD RFPs use qualitative descriptors when stating the prioritization of evaluation criteria.

That is, criterion 1 is “more important” than criterion 2, whereas criterion 2 is “significantly more important” than criterion 3, of which is equal to criterion 4. Current methods used for identifying the weighting of each criterion in an RFP are objective and left to the company’s strategic analyst to estimate. The process heavily relies on “best guess” when assigning weighting factors to the criteria.

The AHP is a recognized technique in decision making and analyzing complex decisions based on identifying the importance of criteria through pairwise comparison, then calculating a weighting score per criterion using geometric means of those relationships (Saaty, 1987). Within the database, the relationships between each major criterion were documented into intensity of importance using the AHP Scale (Table 5). A detailed example of the database with assigned intensity of importance values can be found in Appendix A and B.

Table 5. AHP Scale of Imporance for comparison pair

Intensity of importance	Explanation	Definition	Research Mapping of DoD RFPs
1	Equal Importance	Two factors contribute equally to the objective	Equal
3	Somewhat more important	Experience and judgment slightly favor one over the other.	Slightly more important
5	Much more important	Experience and judgement strongly favor one over the other	More important
7	Very much more important	Experience and judgement very strongly favor one over the other. Its importance is demonstrated in practice	Significantly more important
9	Absolutely more important	The evidence favoring one over the other is of the highest possible validity.	
2, 4, 6, 8		When compromise is needed	

Once all intensity of importance levels were documented, geometric means were calculated to identify the weighting per criterion. The geometric mean method (GMM) is an average of a set of numbers, indicating the central tendency of the value distribution. This is calculated using the product of the values, in comparison to the sum of values often used for arithmetic mean. The geometric mean is calculated by the following equation:

$$\left(\prod_{i=1}^{n} x_i\right)^{\frac{1}{n}} = \sqrt[n]{x_1 \cdot x_2 \cdots x_n}$$

for n numbers and a set of numbers x_1, x_2, …, x_n.

A detailed example of the database with calculated GMM values can be found in Appendix C.

Once the geometric means were calculated into weighted values per criterion, values were grouped into the following prioritization levels for this research (Table 6).

Table 6. Assigned Prioritization Levels

Weighted Values from GMM	Prioritization Levels
0	None
Less than 0.2	Low
Less than 0.5	Medium
Less than 1	High

This process was repeated for each of the 242 RFP database entries prior to the development of the ML model for Stage 2. An example of the final database used for Stage 2 model development is illustrated below in Figure 13.

Market Area	Customer	PoP	Contract Type	Contract Value ($M)	PSC	NAICS	Technical	Program Management	Past Performance	Small Business	Price	Importance to Customer	Schedule	Mission Capability	Quality	Security Plan
Aircraft (Military)	US Navy	60	Fixed Price	50	13	325998	Medium	None	Medium	Low	Low	None	None	None	None	None
Aircraft (Military)	US Air Force	62	Fixed Price	50	U009	611512	High	Low	None	None	Low	None	None	None	None	None
Aircraft (Military)	US Air Force	60	Fixed Price	100	1680	336413	High	None	Medium	None	Low	None	None	None	None	None
Aircraft (Military)	US Air Force	62	ID/IQ	50	N016	336411	High	None	None	None	Low	None	None	None	None	None
Aircraft (Military)	US Air Force	63	Fixed Price	30	1650	336413	Medium	None	None	None	Medium	None	None	None	None	None
Aircraft (Military)	US Air Force	60	Fixed Price	40	28	336412	Medium	None	None	None	Medium	None	None	None	None	None
Aircraft (Military)	US Air Force	73	Fixed Price	50	15	336413	Medium	None	None	None	Medium	None	None	None	None	None
Aircraft (Military)	US Air Force	60	Fixed Price	40	R425	561320	Medium	None	None	None	Medium	None	None	None	None	None
Aircraft (Military)	US Air Force	26	Fixed Price	40	15	336411	High	None	None	Low	Low	None	None	None	None	None
Aircraft (Military)	US Navy	60	Cost Plus Fee	70	K	541330	High	None	Medium	None	Low	None	None	None	None	None
Aircraft (Military)	US Air Force	66	Fixed Price	50	U	611512	Medium	Medium	None	None	Low	None	None	None	None	None
Aircraft (Military)	US Air Force	60	Fixed Price	30	15	336413	Medium	None	None	None	Medium	None	None	None	None	None
Aircraft (Military)	US Air Force	132	ID/IQ	70	15	336413	Medium	None	Medium	Low	Low	None	None	None	None	None
Aircraft (Military)	US Air Force	120	Fixed Price	70	L	541330	Medium	None	None	None	Medium	None	None	None	None	None
Aircraft (Military)	US Air Force	60	Fixed Price	30	17	336412	High	None	Medium	None	Low	None	None	None	None	None
Aircraft (Military)	US Air Force	53	Cost Plus Fee	40	16	336411	High	None	Low	None	Low	None	None	None	None	None
Aircraft (Military)	US Navy	73	Fixed Price	100	13	325998	High	None	Low	None	Low	None	None	None	None	None
Aircraft (Military)	US Air Force	36	Fixed Price	70	23	333924	Medium	None	None	None	Medium	None	None	None	None	None
Aircraft (Military)	US Air Force	50	Fixed Price	40	99	488190	Medium	None	Medium	None	Medium	None	None	None	None	None
Aircraft (Military)	US Navy	60	Fixed Price	40	13	325920	Low	None	High	None	Medium	None	None	None	None	None
Aircraft (Military)	US Air Force	56	Fixed Price	40	15	334511	Medium	None	None	None	Medium	None	None	None	None	None
Aircraft (Military)	US Air Force	72	Fixed Price	50	J	336411	High	None	Low	None	Low	None	None	None	None	None
Aircraft (Military)	US Air Force	48	Fixed Price	50	U	611512	High	None	Medium	Low	Low	None	None	None	None	None
Aircraft (Military)	US Air Force	90	Fixed Price	70	15	334511	High	None								
Aircraft (Military)	US Air Force	60	ID/IQ	60												

Figure 13. Sample subset of database created for Stage 2

Once the database for Stage 2 was completed, the methodology applied for Stage 1 was repeated using the detailed approach outlined in Sections 3.2 through 3.6. A separate set of results were produced, and all hypotheses were tested for the new prioritized evaluation criteria prediction model. The hypotheses tested for Stage 2 are designated as sub-hypotheses to H5.

H5. When applying supervised machine learning techniques for a new business opportunity, prioritized evaluation criteria is predicted at same evaluation as H1-H4.

H5.1. Market area, customer type, contract type, period of performance, and product service code are good predictors for prioritized customer evaluation criteria.

H5.2. When applying supervised machine learning techniques for a new business opportunity, prioritized evaluation criteria is predicted with an AUC higher than 70%.

H5.3. When predicting prioritized customer evaluation criteria, ML algorithms will not produce a model of equal performance.

H5.4. When predicting prioritized customer evaluation criteria, Random Forest algorithm will perform 10% better than Decision Tree algorithm.

3.8 Summary

This chapter provided a detailed discussion of the multiple analyses and methodologies applied to this research and the expected quantitative results that will be used to support and test its hypotheses. The GI values of the contract features (independent variables) will support H1 testing while the performance measures calculated through the different classification ML algorithms will be used to result in a final AUC per evaluation criterion for each ML algorithm. These AUC values will support testing of H2 – H4. Finally, the AHP is applied to the database to develop a secondary database used for Stage 2 of this research. In Stage 1 of this research, a model is developed to predict the presence of evaluation criteria for a new business opportunity based on contract features, while Stage 2 will predict the presence and prioritization level of a criterion for a new business opportunity based on contract features. This second stage will result in a new set of data results to include GI and AUC values that will be tested by the same methodologies used in Stage 1. Stage 2 testing is reflected in H5 and its sub-hypotheses. The summary of methodology and hypothesis testing is listed in Table 7.

Table 7. Summary of Methodologies and Hypothesis Tests

Hyp.	Methodology Summary
H1	**Analysis Technique:** Gini Impurity calculation of business opportunity feature per criterion **Hypothesis Test:** – T-test of Gini Impurity (GI) per feature – H1.1: Gini Impurity of features are not weakly associated (per criteria prediction) ▪ H_o: $\mu_{FEATURE_GI} \leq .02$ ▪ H_a: $\mu_{FEATURE_GI} > .02$ – H1.2: Gini Impurity of features are not highly associated (per algorithm) ▪ H_o: $\mu_{FEATURE_GI} \geq .5$ ▪ H_a: $\mu_{FEATURE_GI} < .5$
H2	**Analysis Technique:** Classification ML Algorithm's AUC: – Decision Tree – Random Forest – XGBoost **Hypothesis Test:** – T-test of ML algorithm AUC per criterion – H_o: $\mu_{AUC_Algorithm} \leq 0.7$ – H_a: $\mu_{AUC_Algorithm} > 0.7$
H3	**Analysis Technique:** AUC measure from H2 results **Hypothesis Test:** – ANOVA test of AUC of ML algorithms ▪ H_o: $\mu_{AUC_DT} = \mu_{AUC_RF} = \mu_{AUC_XGBoost}$ ▪ H_a: $\mu_{AUC_DT} \neq \mu_{AUC_RF} \neq \mu_{AUC_XGBoost}$
H4	**Analysis Technique:** AUCs from H2 results **Hypothesis Test: T-test** of Random Forest to Decision Tree comparison – RF has an AUC performance metric 10% better than DT ▪ H_o: $\mu_{AUC_RF} - \mu_{AUC_DT} \leq 10\%$ ▪ H_a: $\mu_{AUC_RF} - \mu_{AUC_DT} > 10\%$
H5	**Analysis Technique:** AHP and GMM combined with analysis techniques from H1-H4 **Hypothesis Test:** • Same as H1 through H4

Chapter — 4 Results

4.1 Introduction

This chapter discusses the results of the GI analysis and performance measures of the ML techniques, along with major findings of the research through various hypothesis testing methods, based on the methodologies discussed in Chapter 3. For Stage 1 of the research, predicting the presence of evaluation criteria for new business opportunities, the model developed resulted in selecting the RF algorithm based on performance measures, whereas Stage 2 of the research, predicting the presence and prioritization of evaluation criteria, the model shows similar performance for both RF and XGBoost algorithms based on the AUC values.

The decision support tool will allow decision-makers a quantifiable, repeatable, and economic approach to understand how a customer will evaluate their proposed solution. This will enable strategic analysts and decision makers earlier insight into strategic investment decisions with known uncertainty based on the performance of the model, and ability for the model to adopt and adapt to more information over time.

4.2 Gini Impurity Results

In Stage 1 of this research, the GI was calculated for each Contract Feature per major evaluation criterion to assess if the selected Contract Feature would result in an adequate node (Table 8). This step mitigates the risk of generalization error in ML algorithms.

Table 8. GI result per Contract Feature per major evaluation criterion

		Contract Features GI Result						
		Market Area	Customer	Contract Type	Contract Value ($M)	PoP	PSC	NAICS
Major Evaluation Criterion	Technical	.06	.06	.03	.05	.05	.29	.28
	Program Management	.12	.10	.07	.05	.04	.20	.19
	Past Performance	.09	.09	.05	.06	.07	.25	.22
	Small Business	.09	.08	.05	.04	.04	.25	.26
	Importance to Customer	.10	.11	.08	.03	.05	.19	.17
	Schedule	.11	.15	.04	.03	.06	.23	.15
	Mission Capability	.12	.13	.04	.03	.02	.20	.19
	Quality	.11	.07	.04	.06	.07	.28	.16
	Security Plan	.12	.11	.05	.04	.04	.21	.20

Once the GI values were calculated, these values were used to test H1. This research hypothesized that the Contract Features in

Table 2 are applicable independent variables to predict major evaluation criteria based on their calculated GI. To evaluate each Contract Feature's applicability, an unpaired one sample t-test was performed on the GI results across all major evaluation criteria to test H1 (Table 9).

Hypothesis (H1). Contract Features from historical RFPs are good predictors of customer evaluation criteria.

- H1.1. Contract Features are not too weakly associated with evaluation criteria.

$$H_o: \mu_{CONTRACT_FEATURE_GI}: \leq 0.02;\ H_a\ \mu_{CONTRACT_FEATURE_GI} > 0.02$$

- H1.2. Contract Features are not too highly associated with evaluation criteria.

$$H_o: \mu_{CONTRACT_FEATURE_GI}: \geq 0.5;\ H_a\ \mu_{CONTRACT_FEATURE_GI} < 0.5$$

Table 9. Statistical results of GI for each Contract Feature and unpaired one-sample t-test results

Hypothesis (H1) Results	Contract Features GI Statistical Results						
	Market Area	Customer	Contract Type	Contract Value ($M)	PoP	PSC	NAICS
Mean GI (μ)	0.103	0.101	0.049	0.044	0.048	0.234	0.202
Variance	0.0003	0.0008	0.0002	0.0001	0.0002	0.0013	0.0019
Standard Deviation	0.018	0.029	0.015	0.012	0.015	0.036	0.044
df	8	8	8	8	8	8	8
(H1.1) H_o: $\mu_{CONTRACT_FEATURE_GI} \leq .02$; H_a: $\mu_{CONTRACT_FEATURE_GI} > .02$							
t Stat	13.67	8.45	5.85	6.00	5.77	17.76	12.46
t Critical	1.86	1.86	1.86	1.86	1.86	1.86	1.86
p-value	3.95E-07	1.46E-05	1.92E-04	1.61E-04	2.10E-04	5.16E-08	8.03E-07

(H1.2) H_o: $\mu_{CONTRACT_FEATURE_GI} \geq 0.5$; H_a: $\mu_{CONTRACT_FEATURE_GI} < 0.5$

t Stat	-65.18	-41.95	-90.39	-115.27	-91.74	-22.13	-20.38
t Critical	-1.86	-1.86	-1.86	-1.86	-1.86	-1.86	-1.86
p-value	1.71E-12	5.74E-11	1.25E-13	1.79E-14	1.11E-13	9.19E-09	1.76E-08

The t-test results of each Contract Feature show the extremely low probability (p-value) that these independent variables would result in misclassification. Therefore, each Contract Feature has been assessed as a good predictor for all customer evaluation criteria, and the null hypothesis is rejected. This critical step is the foundation to predicting customer's evaluation criteria and facilitating early strategic investment decisions.

4.3 Performance Measure Results

As discussed in Chapter 3, ML performance measures provide insight to the goodness of the model and its validity. Performance measures were calculated for each ML algorithm used to predict major evaluation criterion (Tables 10-12). Once performance measures were calculated, an AUC was calculated to measure the quality of the prediction model for each ML algorithm.

Table 10. Performance measures for DT predictions per customer evaluation criterion

	DT Performance Results	Total Values		Confusion Matrix Results				Diagnostic Results				
		Correct	Incorrect	TP	TN	FP	FN	Sensitivity	Specificity	Accuracy	BA	AUC
Customer Evaluation Criterion	Technical	174	20	.97	.13	.88	.03	.97	.13	.90	.55	.61
	Program Management	144	50	.16	.88	.12	.84	.16	.88	.74	.52	.58
	Past Performance	125	69	.77	.46	.54	.23	.77	.46	.64	.61	.69
	Small Business	154	40	.14	.9	.1	.86	.14	.90	.79	.52	.59
	Importance to Customer	181	13	.64	.97	.03	.36	.64	.97	.93	.80	.82
	Schedule	179	15	.53	.96	.04	.47	.53	.96	.92	.75	.83
	Mission Capability	178	16	.38	.98	.02	.62	.38	.98	.92	.68	.69
	Quality	179	15	.15	.98	.02	.85	.15	.98	.92	.57	.61
	Security Plan	187	7	.57	.98	.02	.43	.57	.98	.96	.78	.77

Table 11. Performance measures for RF predictions per customer evaluation criterion

	RF Performance Results	Total Values		Confusion Matrix Results				Diagnostic Results				
		Correct	Incorrect	TP	TN	FP	FN	Sensitivity	Specificity	Accuracy	BA	AUC
Customer Evaluation Criterion	Technical	178	16	1	0	1	0	1	0	.92	.50	.63
	Program Management	164	30	.27	.98	.02	.73	.27	.98	.85	.63	.80
	Past Performance	136	58	.86	.46	.54	.14	.86	.46	.70	.66	.72
	Small Business	165	29	.07	.98	.02	.93	.07	.98	.85	.53	.66
	Importance to Customer	182	12	.55	.99	.01	.45	.55	.99	.94	.77	.97
	Schedule	179	15	.18	.99	.01	.82	.18	.99	.92	.59	.92
	Mission Capability	182	12	.52	.99	.01	.48	.52	.99	.94	.76	.90
	Quality	180	14	0	.99	.01	1	0	.99	.93	.50	.75
	Security Plan	189	5	.29	1	0	.71	.29	1	.97	.64	.90

Table 12. Performance measures for XGBoost predictions per customer evaluation criterion

	XGBoost Performance Results	Total Values		Confusion Matrix Results				Diagnostic Results				
		Correct	Incorrect	TP	TN	FP	FN	Sensitivity	Specificity	Accuracy	BA	AUC
Customer Evaluation Criterion	Technical	174	20	.98	0	1	.02	.98	0	.90	.49	.49
	Program Management	165	29	.46	.94	.06	.54	.46	.94	.85	.70	.82
	Past Performance	136	58	.79	.56	.44	.21	.79	.56	.70	.68	.72
	Small Business	160	34	.18	.93	.07	.82	.18	.93	.83	.56	.66
	Importance to Customer	186	8	.77	.98	.02	.23	.77	.98	.96	.88	.94
	Schedule	185	9	.65	.98	.02	.35	.65	.98	.95	.82	.81
	Mission Capability	177	17	.48	.97	.03	.52	.48	.97	.91	.72	.85
	Quality	182	12	.23	.99	.01	.77	.23	.99	.94	.61	.71
	Security Plan	185	9	.14	.98	.02	.86	.14	.98	.95	.56	.89

To assess the quality of each ML algorithm for predicting major evaluation criteria, an unpaired one-sample t-test was performed (Table 13). This research hypothesized that the mean-value AUC per ML algorithm would result in a "Good" test quality (Figure 12) to evaluate for usefulness and contribution of the research.

Hypothesis (H2). When applying supervised machine learning techniques, ML algorithms will predict evaluation criteria with an AUC higher than 70%.

$$Ho: \mu_{AUC_Algorithm} \leq .7; Ha: \mu_{AUC_Algorithm} > .7$$

Table 13. Statistics of AUC for each ML algorithm and unpaired one-sample t-test results

Hypothesis (H2) Results	**Algorithm**		
	DT	RF	XGBoost
Mean AUC (μ_{AUC})	.688	.806	.766
Variance	0.0097	0.0151	0.0188
Standard Deviation	0.099	0.123	0.137
Df	8	8	8
H_o: $\mu_{AUC_Algorithm} \leq 0.7$; H_a: $\mu_{AUC_Algorithm} > 0.7$			
t Stat	0.372	2.581	1.435
t Critical	1.860	1.860	1.860
p-value	.360	.016	.095

The AUC results show that RF and XGBoost perform above the "Good" quality test level (70%), with a mean-value of .806 and .766, respectively. However, the RF algorithm has a higher probability of performing above the "Good" quality test level with a p-value of .016 compared to the XGBoost's p-value of .095. DT algorithm has a high

probability (p-value = .360) of performing below the 70% level; however, it performs within the "satisfactory" quality test level (> 60%) with a mean-value of .688.

4.4 Performance Evaluation

As previously mentioned in Chapters 2 and 3, the quality test thresholds are subjective and based on the problem statement. To ensure rigor and further evaluate the performance of predicting customer evaluation criteria, an ANOVA test was performed for statistical difference between algorithms to identify if selection of a particular algorithm increases prediction performance. This research hypothesized that there is a statistical difference in prediction performance when comparing ML algorithms' mean-value AUC.

Hypothesis (H3). When predicting customer evaluation criteria, ML algorithms will not produce a model of equal performance.

- H3.1. DT algorithm will not perform equal to RF algorithm.

$$H_o: \mu_{AUC_DT} = \mu_{AUC_RF}; H_a: \mu_{AUC_DT} \neq \mu_{AUC_RF}$$

- H3.2. DT algorithm will not perform equal to XGBoost algorithm.

$$H_o: \mu_{AUC_DT} = \mu_{AUC_XGBoost}; H_a: \mu_{AUC_DT} \neq \mu_{AUC_XGBoost}$$

- H3.3. RF algorithm will not perform equal to XGBoost algorithm.

$$H_o: \mu_{AUC_RF} = \mu_{AUC_XGBoost}; H_a: \mu_{AUC_RF} \neq \mu_{AUC_XGBoost}$$

Table 14. ANOVA test of ML Algorithms' AUC measure

Hypothesis (H3)

Results	**ANOVA SUMMARY**				
Groups	Count	Sum	Average	Variance	Standard Deviation
DT	9.000	6.190	.688	0.0097	0.0986
RF	9.000	7.250	.806	0.0151	0.1227
XGBoost	9.000	6.890	.766	0.0188	0.1370

(H3.1) H_o: $\mu_{AUC_DT} = \mu_{AUC_RF}$; H_a: $\mu_{AUC_DT} \neq \mu_{AUC_RF}$

Source of Variance	SS	df	MS	F	F Critical	p-value
Between Groups	0.062	1.000	0.062	5.040	4.494	.039
Within Groups	0.198	16.000	0.012			

(H3.2) H_o: $\mu_{AUC_DT} = \mu_{AUC_XGBoost}$; H_a: $\mu_{AUC_DT} \neq \mu_{AUC_XGBoost}$

Source of Variance	SS	df	MS	F	F Critical	p-value
Between Groups	0.027	1.000	0.027	1.911	4.494	.186
Within Groups	0.228	16.000	0.014			

(H3.3) H_o: $\mu_{AUC_RF} = \mu_{AUC_XGBoost}$; H_a: $\mu_{AUC_RF} \neq \mu_{AUC_XGBoost}$

Source of Variance	SS	df	MS	F	F Critical	p-value
Between Groups	0.007	1.000	0.007	0.426	4.494	.523
Within Groups	0.271	16.000	0.017			

The ANOVA test resulted in a statistical difference only between RF and DT algorithms' performance with a low probability of equal performance (p-value = .039),

whereas the comparison between DT:XGBoost and RF:XGBoost did not result in a statistically different performance, as the p-value = 0.185 and 0.523, respectively. These results lead to the conclusion of the RF algorithm outperforming the other two algorithms, and thus selected for the Stage 1 of the predictive model development.

4.5 Difference in Performance

The last consideration in this research was to identify the performance difference between the algorithms that resulted in a statistical difference from the ANOVA results in the previous section. This will enable future research to improve upon the current findings and provide decision-makers a more robust predictive model. This research hypothesized that amongst the algorithms with a statistical difference, a 10% performance difference existed and was evaluated by an unpaired one-sample t-test. To provide robustness to the model, the study further performed a series of unpaired one-sample t-tests to analyze the hypothesized mean difference denoted by the variable z, reducing z by an increment of 1% to calculate the probabilities of performance difference between the two ML algorithms.

Hypothesis (H4). When predicting customer evaluation criteria, Random Forest algorithm will perform 10% better than Decision Tree algorithm.

- H4.1. RF algorithm performs 10% better than DT algorithm.

$$H_o: \mu_x \leq .1;\ H_a: \mu_x > .1;\ \text{where } x = AUC_{RF} - AUC_{DT}$$

- H4.2. RF algorithm performs z-value better than DT algorithm.

$$H_o: \mu_x \leq z;\ H_a: \mu_x > z;\ \text{where } z = .09; .08; .07$$

Table 15. Statistics of the difference of AUC_{RF} and AUC_{DT} and unpaired one-sample t-test results

Hypothesis (H4) Results

$x = AUC_{RF} - AUC_{DT}$	
Mean (μ_x)	0.118
Variance	0.0051
Standard Deviation	0.072
Df	8

(H4.1) H_o: $\mu_x \leq .1$; H_a: $\mu_x > .1$

t Stat	0.745
t Critical	1.860
p-value	.239

(H4.2) H_o: $\mu_x \leq z$; H_a: $\mu_x > z$

z values	.09	.08	.07
t Stat	1.165	1.584	2.003
p-value	.139	.076	.040

The t-test (Table 15) showed that RF algorithm has a low probability of performing 10% better than DT algorithm with a p-value of .239, thus fails to reject the null hypothesis. However, when reducing the null hypothesis value by z, the RF algorithm results in a high probability of outperforming DT algorithm by more than 7% (p-value= .04; μ_x =0.118).

4.6 Prioritized Evaluation Criteria Results

As mentioned in Chapter 3, a second stage of the research was performed to induce a secondary set of results that reflected the prediction of prioritized evaluation criteria. Once every criterion, per RFP had a calculated weighting and assigned categorized by prioritization level, a second ML model was developed. The results of Stage 2 are discussed in the sections below.

4.6.1 Gini Impurity Results

As mentioned in Section 4.2, GI values were calculated for each Contract Feature per major evaluation criterion to mitigate the risk of generalization error in ML algorithms.

Table 16. GI result per Contract Feature per major evaluation criterion (with Prioritization)

		Contract Features GI Results (with Prioritization)						
		Market Area	Customer	Contract Type	Contract Value ($M)	PoP	PSC	NAICS
Major Evaluation Criterion	Technical	0.08	0.04	0.03	0.02	0.02	0.10	0.11
	Program Management	0.08	0.06	0.04	0.03	0.02	0.16	0.14
	Past Performance	0.07	0.06	0.03	0.04	0.03	0.15	0.17
	Small Business	0.06	0.05	0.03	0.02	0.03	0.18	0.20
	Importance to Customer	0.07	0.06	0.05	0.01	0.03	0.13	0.14
	Schedule	0.08	0.08	0.02	0.02	0.03	0.15	0.11
	Mission Capability	0.07	0.08	0.03	0.02	0.03	0.17	0.17
	Quality	0.09	0.05	0.03	0.02	0.03	0.20	0.10
	Security Plan	0.07	0.08	0.02	0.02	0.02	0.13	0.11

Once the GI values were calculated, these values were then used to test H.5.1.

H5.1. Market area, customer type, contract type, period of performance, and product service code are good predictors for prioritized customer evaluation criteria.

Table 17. Statistical results of GI for each Contract Feature and unpaired one-sample t-test results (with Prioritization) for H5.1

Hypothesis (H5.1) Results	**Contract Features GI Statistical Results (with Prioritization)**						
	Market Area	Customer	Contract Type	Contract Value ($M)	PoP	PSC	NAICS
Mean (μ)	0.073	0.062	0.031	0.023	0.026	0.153	0.139
Variance	0.0001	0.0002	0.0001	0.0001	0.0000	0.0009	0.0011
Standard Deviation	0.009	0.016	0.009	0.009	0.005	0.030	0.034
df	8	8	8	8	8	8	8
H_o: $\mu_{FEATURE_GI} \leq 0.02$; H_a: $\mu_{FEATURE_GI} > 0.02$							
t Stat	18.37	8.11	3.76	1.02	3.94	13.13	10.54
t Critical	1.86	1.86	1.86	1.86	1.86	1.86	1.86
p-value	3.97E-08	1.98E-05	2.78E-03	1.68E-01	2.14E-03	5.37E-07	2.86E-06
H_o: $\mu_{FEATURE_GI} \geq 0.5$; H_a: $\mu_{FEATURE_GI} < 0.5$							
t Stat	-147.25	-83.97	-154.58	-161.15	-304.74	-34.41	-32.15
t Critical	-1.86	-1.86	-1.86	-1.86	-1.86	-1.86	-1.86
p-value	2.53E-15	2.26E-13	1.72E-15	1.23E-15	7.53E-18	2.78E-10	4.77E-10

Based on an unpaired one sample t-test across all major evaluation criteria, all Contract Features were found to be applicable independent variables to predict prioritized evaluation criteria, with the exception of Contract Value having a p-value of 0.0168. This

value indicates that there is a 0.0168 probability that this contract feature will have too weak of a correlation to the prioritization criteria for use in the prediction model. However, all other features have an extremely low p-value and thus the null hypothesis can be rejected.

4.6.2 Performance Measures

Similarly, to section 4.3, performance measures were analyzed for each ML algorithm in Stage 2. However, with several prioritization levels that could be predicted in comparison to the binary approach in Stage 1, performance values were analyzed for each criterion, per prioritization level, per algorithm. The full confusion matrix results for Stage 2 can be found in Appendix F.

The ROC results were documented per major criterion and an average was calculated to assess the overall AUC performance value identified in Table 18-20. Detailed ROC for all Contract Features and algorithms curves can be found in Appendix G.

Table 18. Performance measures for DT predictions per prioritized customer evaluation criterion

DT Algorithm		Total Values		Diagnostic Results		
		Correct	Incorrect	Accuracy	BA	AUC
Major Criterion	Technical	103	91	0.53	0.37	0.61
	Program Management	149	45	0.77	0.38	0.59
	Past Performance	92	102	0.47	0.33	0.57
	Small Business	152	42	0.78	0.46	0.67
	Importance to Customer	181	13	0.93	0.90	0.92
	Schedule	175	19	0.90	0.51	0.66
	Mission Capability	159	38	0.82	0.35	0.61
	Quality	178	16	0.92	0.45	0.63
	Security Plan	182	12	0.94	0.37	0.62

Table 19. Performance measures for RF predictions per prioritized customer evaluation criterion

RF Algorithm		Total Values		Diagnostic Results		
		Correct	Incorrect	Accuracy	BA	AUC
Major Criterion	Technical	122	72	0.63	0.33	0.69
	Program Management	161	33	0.83	0.31	0.71
	Past Performance	120	74	0.62	0.42	0.78
	Small Business	168	26	0.87	0.36	0.85
	Importance to Customer	183	11	0.94	0.77	0.98
	Schedule	185	9	0.95	0.56	0.81
	Mission Capability	181	13	0.93	0.42	0.91
	Quality	180	14	0.93	0.33	0.75
	Security Plan	187	7	0.96	0.25	0.97

Table 20. Performance measures for XGBoost predictions per prioritized customer evaluation criterion

XGBoost Algorithm		Total Values		Diagnostic Results		
		Correct	Incorrect	Accuracy	BA	AUC
Major Criterion	Technical	114	80	0.59	0.39	0.69
	Program Management	155	39	0.80	0.31	0.76
	Past Performance	128	66	0.66	0.47	0.79
	Small Business	171	23	0.88	0.46	0.90
	Importance to Customer	182	12	0.94	0.79	0.96
	Schedule	181	13	0.93	0.57	0.77
	Mission Capability	178	16	0.92	0.39	0.84
	Quality	179	15	0.92	0.39	0.89
	Security Plan	188	6	0.97	0.37	0.98

This research hypothesized that the mean-value AUC per ML algorithm would result in a "Good" test quality to evaluate for usefulness and contribution of the research.

Hypothesis (H5.2). When applying supervised machine learning techniques, ML algorithms will predict prioritized customer evaluation criteria with an AUC higher than 70%.

$$Ho: \mu_{AUC_Algorithm} \leq .7; \; Ha: \mu_{AUC_Algorithm} > .7$$

Table 21. Statistics of AUC for each ML algorithm and unpaired one-sample t-test results for H5.2

Hypothesis (H5.2) Results	**Algorithm**		
	DT	RF	XGBoost
Mean (μ)	0.654	0.827	0.842
Variance	0.0108	0.0113	0.0089
Standard Deviation	0.104	0.107	0.095
df	8	8	8
H_o: $\mu_{AUC_Algorithm} \leq 0.7$; H_a: $\mu_{AUC_Algorithm} > 0.7$			
t Stat	-1.343	3.580	4.506
t Critical	1.860	1.860	1.860
p-value	0.108	0.004	0.001

The AUC results show that RF and XGBoost perform above the "Good" quality test level (70%), with a mean-value of .827 and .842, respectively, thus above the "Very Good" quality test level, as discussed in Section 3.6. The DT algorithm, similar to Stage 1, has a higher probability (p-value = .108) of performing below the 70% level; however, it performs within the "satisfactory" quality test level (> 60%) with a mean-value of .654.

4.6.3 Performance Evaluation

An ANOVA test was performed on the AUC performance results from the previous section to test for statistical difference between algorithm performance and to support identifying which algorithm increases prediction performance for the Stage 2 model. This research hypothesized that there is a statistical difference in prediction performance when comparing ML algorithms' mean-value AUC.

Hypothesis (H5.3). When predicting customer evaluation criteria, ML algorithms will not produce a model of equal performance.

- H5.3.1. DT algorithm will not perform equal to RF algorithm.

$$H_o: \mu_{AUC_DT} = \mu_{AUC_RF}; H_a: \mu_{AUC_DT} \neq \mu_{AUC_RF}$$

- H5.3.2. DT algorithm will not perform equal to XGBoost algorithm.

$$H_o: \mu_{AUC_DT} = \mu_{AUC_XGBoost}; H_a: \mu_{AUC_DT} \neq \mu_{AUC_XGBoost}$$

- H5.3.3. RF algorithm will not perform equal to XGBoost algorithm.

$$H_o: \mu_{AUC_RF} = \mu_{AUC_XGBoost}; H_a: \mu_{AUC_RF} \neq \mu_{AUC_XGBoost}$$

Table 22. ANOVA test of ML Algorithms' AUC measure results for H5.3

Hypothesis (H5.3) Results	**ANOVA SUMMARY**				
Groups	Count	Sum	Average	Variance	Standard Deviation
DT	9.0000	5.8818	0.6535	0.0108	0.1038
RF	9.000	7.444	0.827	0.011	0.1065
XGBoost	9.000	7.577	0.842	0.009	0.0945

(H5.3.1) H_o: $\mu_{AUC_DT} = \mu_{AUC_RF}$; H_a: $\mu_{AUC_DT} \neq \mu_{AUC_RF}$

Source of Variance	SS	df	MS	F	F Critical	p-value
Between Groups	0.136	1.000	0.136	12.256	4.494	0.003
Within Groups	0.177	16.000	0.011			

(H5.3.2) H_o: $\mu_{AUC_DT} = \mu_{AUC_XGBoostF}$; H_a: $\mu_{AUC_DT} \neq \mu_{AUC_XGBoost}$

Source of Variance	SS	df	MS	F	F Critical	p-value
Between Groups	0.160	1.000	0.160	16.207	4.494	0.001
Within Groups	0.158	16.000	0.010			

(H5.3.3) H_o: $\mu_{AUC_RF} = \mu_{AUC_XGBoost}$; H_a: $\mu_{AUC_RF} \neq \mu_{AUC_XGBoost}$

Source of Variance	SS	df	MS	F	F Critical	p-value
Between Groups	0.001	1.000	0.001	0.097	4.494	0.759
Within Groups	0.162	16.000	0.010			

Unlike Stage 1 results discussed in Section 4.4, the ANOVA test results in Stage 2 showed a statistical difference for DT between RF and XGBoost algorithms' performance with a low probability of equal performance (p-value = .003 and .001, respectively). However, the results of the ANOVA results for the RF and XGBoost comparison show no statistical difference in the algorithms' AUC, and thus are of equal performance for use in the prediction model.

4.6.4 Difference in Performance

Based on the results from the previous section and ANOVA results, the RF and XGBoost algorithms have a significant difference in their performance results in comparison to the DT algorithm. This research hypothesized that RF would have a 10% performance measure above that of the DT algorithm.

To provide robustness to the model, the study further performed a one-sample t-test to identify if XGBoost also performed 10% better than the DT algorithm, allowing a comparison between RF and XGBoost algorithms to complete the development of the Stage 2 model.

Hypothesis (H5.4). When predicting prioritized customer evaluation criteria, Random Forest algorithm will perform 10% better than Decision Tree algorithm (expansion of XGBoost identified in sub-hypothesis 5.4.2).

- H5.4.1 RF algorithm performs 10% better than DT algorithm.

$$H_{o:}\, \mu_x \leq .1;\ H_a{:}\ \mu_x > .1;\ \textit{where}\ x = AUC_{RF} - AUC_{DT}$$

- H5.4.2 XGBoost algorithm performs 10% better than DT algorithm.

$$H_{o:}\, \mu_x \leq .1;\ H_a{:}\ \mu_x > .1;\ \textit{where}\ x = AUC_{XGBoost} - AUC_{DT}$$

Table 23. Statistics of the difference of RF and XGBoost AUC values and DT AUC with unpaired one-sample t-test results for H5.4

Hypothesis Results (H5.4)	$\mathbf{x = AUC_{Algorithm} - AUC_{DT}}$	
	RF	XGBoost
Mean (μ_x)	0.174	0.188
Variance	0.0096	0.0092
Standard Deviation		
df	8	8
H_o: $\mu_x \leq 0.1$; H_a: $\mu_x > 0.1$		
t Stat	2.26	2.76
t Critical	1.86	1.86
p-value	0.027	0.012

The t-test (Table 23) showed that RF algorithm has a solid probability of performing 10% better than DT algorithm with a p-value of .027 and a μ=0.174. This results in rejecting the null hypothesis and shows a quantifiable increase in predictive performance using the RF algorithm. In addition, the XGBoost algorithm also shows to outperform the DT algorithm with a μ=0.188 and a p-value of 0.012. This shows that XGBoost is also a high consideration for use of the Stage 2 model for final algorithm selection. Contrasting the two results and reflecting on the ANOVA results from the previous section, there is no statistical and significant difference in performance of the RF and XGBoost algorithm, as their average outperformance of the DT algorithm is within 1.4% of each other. However, the XGBoost has a lower p-value of outperforming DT (p-value = 0.012) and thus the selected ML algorithm used to complete the development of the prioritized evaluation criteria prediction model.

4.7 Summary

The resulting prioritized evaluation criteria prediction model has identified the independent variables needed to predict prioritized evaluation criteria with quantifiable performance measures that current methods do not offer. A summary of the hypothesis test results in this research, and described in detail throughout this chapter, can be found in Table 24. This model can be used to enable strategic decisions around resource planning and investments to strengthen a company's competitive position to win new business opportunities.

Table 24. Summary of Hypothesis Test Results

Hyp.	Hypothesis Test Results Summary	Reject Null Hypothesis
H1	**Hypothesis Test:** T-test of Gini Impurity (GI) per feature – H1.1: Gini Impurity of features are not weakly associated (per criteria prediction) ▪ H_o: $\mu_{FEATURE_GI} \leq .02$ ▪ H_a: $\mu_{FEATURE_GI} > .02$ – H1.2: Gini Impurity of features are not highly associated (per algorithm) ▪ H_o: $\mu_{FEATURE_GI} \geq .5$ ▪ H_a: $\mu_{FEATURE_GI} < .5$	Yes, Reject
H2	**Hypothesis Test:** T-test of ML algorithm AUC per criterion – H_o: $\mu_{AUC_Algorithm} \leq 0.7$ – H_a: $\mu_{AUC_Algorithm} > 0.7$	H_o: $\mu_{AUC_DT} \leq 0.7$; No, Fail to Reject H_o: $\mu_{AUC_RF} \leq 0.7$; Yes, Reject H_o: $\mu_{AUC_XGBoost} \leq 0.7$; Yes, Reject
H3	**Hypothesis Test:** ANOVA test of AUC of ML algorithms – H_o: $\mu_{AUC_DT} = \mu_{AUC_RF} = \mu_{AUC_XGBoost}$ – H_a: $\mu_{AUC_DT} \neq \mu_{AUC_RF} \neq \mu_{AUC_XGBoost}$	H_o: $\mu_{AUC_DT} = \mu_{AUC_XGBoost}$; No, Fail to Reject H_o: $\mu_{AUC_RF} = \mu_{AUC_XGBoost}$; No, Fail to Reject H_o: $\mu_{AUC_DT} = \mu_{AUC_RF}$; Yes, Reject
H4	**Hypothesis Test:** T-test of Random Forest to Decision Tree comparison – RF has an AUC performance metric 10% better than DT ▪ H_o: $\mu_{AUC_RF} - \mu_{AUC_DT} \leq 10\%$ ▪ H_a: $\mu_{AUC_RF} - \mu_{AUC_DT} > 10\%$	H_o: $\mu_{AUC_RF} - \mu_{AUC_DT} \leq 10\%$; No, Fail to Reject
H5	Stage 2: Prioritized Evaluation Criteria	
H5.1	**Hypothesis Test:** T-test of Gini Impurity (GI) per feature – H1.1: Gini Impurity of features are not weakly associated (per criteria prediction) ▪ H_o: $\mu_{FEATURE_GI} \leq .02$ ▪ H_a: $\mu_{FEATURE_GI} > .02$ – H1.2: Gini Impurity of features are not highly associated (per algorithm) ▪ H_o: $\mu_{FEATURE_GI} \geq .5$ ▪ H_a: $\mu_{FEATURE_GI} < .5$	Yes, Reject
H5.2	**Hypothesis Test:** T-test of ML algorithm AUC per criterion – H_o: $\mu_{AUC_Algorithm} \leq 0.7$ – H_a: $\mu_{AUC_Algorithm} > 0.7$	H_o: $\mu_{AUC_DT} \leq 0.7$; No, Fail to Reject H_o: $\mu_{AUC_RF} \leq 0.7$; Yes, Reject H_o: $\mu_{AUC_XGBoost} \leq 0.7$; Yes, Reject
H5.3	**Hypothesis Test:** ANOVA test of AUC of ML algorithms – H_o: $\mu_{AUC_DT} = \mu_{AUC_RF} = \mu_{AUC_XGBoost}$ – H_a: $\mu_{AUC_DT} \neq \mu_{AUC_RF} \neq \mu_{AUC_XGBoost}$	H_o: $\mu_{AUC_DT} = \mu_{AUC_XGBoost}$; Yes, Reject H_o: $\mu_{AUC_DT} = \mu_{AUC_RF}$; Yes, Reject H_o: $\mu_{AUC_RF} = \mu_{AUC_XGBoost}$; No, Fail to Reject
H5.4	**Hypothesis Test:** T-test of Random Forest to Decision Tree comparison – RF has an AUC performance metric 10% better than DT ▪ H_o: $\mu_{AUC_RF} - \mu_{AUC_DT} \leq 10\%$ ▪ H_a: $\mu_{AUC_RF} - \mu_{AUC_DT} > 10\%$	H_o: $\mu_{AUC_RF} - \mu_{AUC_DT} \leq 10\%$; Yes, Reject

Chapter — 5 Discussion and Conclusions

5.1 Discussion

With an increasingly competitive environment in defense and commercial industries, pressure continues to grow on decision-makers to make sound strategic investments for their firms that strengthen core competencies and provide a competitive advantage. Strategic managers and planners need robust tools and rigorous methods to mitigate the systematic risk that impacts their ability to register profitable growth on a sustainable basis. This research provides decision-makers early insights into customer needs inferred by customer evaluation criteria using ML techniques.

The research found that the Contract Features collected from the 242 RFPs were applicable independent variables to use in predicting evaluation criteria by applying GI calculations, specific to ML classification algorithms. In sections 4.2 and 4.6.1, the results of the research showed that all major evaluation criteria had an acceptable level of correlation to be used as predictive independent variables in the model. Once the independent variables were set, the research results demonstrated that DT, RF, and XGBoost algorithms can predict evaluation criteria with quantifiable measures and a test quality level that exceeds better-than-random predictions for both the identification and prioritization of evaluation criteria for new business opportunities. Once it was observed that the algorithms would perform at an acceptable quality, further results show that both RF and XGBoost algorithms produced AUC measures within the “good” test quality scale in predicting the presence of evaluation criteria. Consequently, when further classifications were introduced to measure the prioritization of criteria with application of

the AHP methodology discussed in section 3.7, RF and XGBoost increased their test quality to above an 80% AUC performance measure, resulting in “very good” quality. Finally, applying robust analysis of comparative analysis among the three supervised ML algorithms allowed the identification of the algorithm to use for the best quality model that results in significant cost and time savings for predicting the prioritized customer evaluation criterion with a mean-accuracy of 85% and an above 80% AUC. The overall approach and methodology of this research can be used recursively as this research continues to evolve.

Current methods using competitive intelligence to predict customer evaluation criteria introduces subjectivity and lacks the ability to provide decision-makers a quantifiable measurement of uncertainty (Maune, 2014; Miller, 2020). These findings will enable decision-makers to focus on their firms’ specific core competencies to improve their competitiveness with a measured level of confidence.

5.2 Contribution to Body of Knowledge

This study makes a number of valuable contributions through the development of a newly established database of customer evaluation criteria, a robust process for strategic investment predictive modeling, and an evaluated and validated quantitative decision support tool with high accuracy and test quality.

Further, this research addresses one of the most fundamental problems in the field of strategy management – How to anticipate customers’ needs and make appropriate investment decisions? Although this research focuses on the U.S. defense sector, its implications are applicable to several other sectors and countries worldwide.

5.3 Conclusions

Recent years have witnessed an advancement and growth in the use of ML in applied research. Its methods of solving complex problems can benefit the strategy field. This novel study sets the foundation for future research that will continue to integrate methods from other fields yielding innovative solutions that are cheaper, quicker, and more reliable than current methods for strategic investment decisions.

5.4 Future Research Recommendations

The results of this research are limited to the DoD dataset used in this study to include the customer set, contract value, market areas, and award dates as outlined in section 1.7. Future research should consider how to have an efficient data collection process for evaluation criteria from RFPs, as the data were collected manually in this research. Automation of this step will help save time through error reduction and rework; it will also allow to grow the database quickly.

As stated in section 1.7, this research consisted of minimal tuning of ML configuration parameters and could also be further researched to advance the model performance, which could lead to an increase in performance measures and model quality. This is a major consideration as the dataset increases and may have an impact on performance measures.

Lastly, as previously discussed, the dataset spanned a 2016-2020 time-window. Expanding the dataset time-window and identifying fluctuation in evaluation criteria could also be a future research consideration that would be beneficial to strategic investment planning.

References

Arora, A., Gittelman, M., Kaplan, S., Lynch, J., Mitchell, W., & Siggelkow, N. (2016). Question-based innovations in strategy research methods. *Strategic Management Journal 37*(1), 3–9. https://doi.org/10.1002/smj.2465

Assisted modeling tool. Modeling Tool. (n.d.). Retrieved January 8, 2023, from https://downloads.alteryx.com/betawh_xnext/AssistedModeling.htm

Bekkar, M., Djemaa, H. K., & Alitouche, T. A. (2013). Evaluation measures for models assessment over imbalanced data sets. *Journal of Information Engineering and Applications 3*(10).

Blank, J. (2019). The impacts of DoD acquisition initiatives on defense industry business strategy. *Defense Acquisition Research Journal 26*(3), 230–248. https://doi.org/10.22594/dau.18-813.26.03

Brindley, A., Wood, B., & Mullen, T. (2017). Facing new realities in the defense market. *National Defense 101*(762), 26–29.

Calafut, M., Sarkani, S., & Mazzuchi, T. A. (2021). Effective decision-making behaviors for defense R&D: Accounting for dynamic competition. *Defense Acquisition Research Journal 28*(2), 190–233. https://doi.org/10.22594/dau.20-859.28.02

Carter, A. B. (2010). Better buying power: Guidance for obtaining greater efficiency and productivity in defense spending [Memorandum]. Office of the Under Secretary of Defense for AT&L.

https://www.acq.osd.mil/fo/docs/usd_atl_guidance_memo_september_14_2010_final.pdf

Chatterjee, S. & Wernerfelt, B. (1991). The link between resources and type of diversification: Theory and evidence. *Strategic Management Journal 12*(1), 33–48.

Chatzis, S. P., Siakoulis, V., Petropoulos, A., Stavroulakis, E., & Vlachogiannakis, N. (2018). Forecasting stock market crisis events using deep and statistical machine learning techniques. *Expert Systems with Applications 112*, 353-371.

Choi, J., Menon, A., & Tabakovic, H. (2021). Using machine learning to revisit the diversification–performance relationship. *Strategic Management Journal 42*(9), 1632–1661. https://doi.org/10.1002/smj.3317

Choudhury, P., Allen, R. T., & Endres, M. G. (2021). Machine learning for pattern discovery in management research. *Strategic Management Journal 42*(1), 30–57. https://doi.org/10.1002/smj.3215

Choudhury, P., Starr, E., & Agarwal, R. (2020). Machine learning and human capital complementarities: Experimental evidence on bias mitigation. *Strategic Management Journal 41*(8), 1381–1411. https://doi.org/10.1002/smj.3152

Chua, D. K. H. & Li, D. (2000). Key factors in bid reasoning model. *Journal of Construction Engineering and Management 126*(5), 349–357.

Duddy, B., Landucci, T., & Knechtel, J. (2020). Delivering capability through competition in defense contracting: Does policy drive results? *Defense Acquisition Research Journal 27*(94), 436–474. https://doi.org/10.22594/dau.20-855.27.04

Etemadi, A. & Kamp, J. (2021). Market and contractor factors affecting rapid acquisition strategies. *Systems Engineering*. https://doi.org/10.1002/sys.21577

Fang, Y., Chen, J., & Xue, Z. (2019). Research on quantitative investment strategies based on deep learning. *Algorithms 12*(2). https://doi.org/10.3390/a12020035

Fanning, E. (2019). *2019 Facts & Figures: U.S. Aerospace & Defense*. https://www.aia-aerospace.org/publications/2019-facts-figures/

Federal Acquisition Regulation [FAR] 16.1, 2022

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *The Annals of Statistics*, 1189-1232.

Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research 3*, 1157–1182.

Hari, A., Kasser, J. E., & Weiss, M. P. (2007). How lessons learned from using QFD led to the evolution of a process for creating quality requirements for complex systems. *Systems Engineering 10*(1), 45–63. https://doi.org/10.1002/sys.20065

Kendall, F. (2015). *Better Buying Power 3.0 Implementation Guidance* [Memorandum]. Office of the Under Secretary of Defense for AT&L. https://www.acq.osd.mil/fo/docs/betterbuyingpower3.0(9Apr15).pdf

Kreye, M. E., Newnes, L. B., & Goh, Y. M. (2014). Uncertainty in competitive bidding – a framework for product–service systems. *Production Planning & Control 25*(6), 462–477. https://doi.org/10.1080/09537287.2012.705354

Li, G., Zhang, G., Chen, C., & Martek, I. (2020). Empirical bid or no bid decision process in international construction projects: Structural equation modeling framework. *Journal of Construction Engineering and Management 146*(6), 04020050. https://doi.org/10.1061/(asce)co.1943-7862.0001830

Mascio, D. A., Fabozzi, F. J., & Zumwalt, J. K. (2020). Market timing using combined forecasts and machine learning. *Journal of Forecasting*. https://doi.org/10.1002/for.2690

Maune, A. (2014). Competitive intelligence as an enabler for firm competitiveness: An overview. *Journal of Governance and Regulation 3*(2).

McGrath, R. G. & Nerkar, A. (2004). Real options reasoning and a new look at the R&D investment strategies of pharmaceutical firms. *Strategic Management Journal 25*(1), 1–21. https://doi.org/10.1002/smj.358

Metz, C. E. (1978). Basic principles of ROC analysis. *Seminars in Nuclear Medicine 8*(4), 283–298.

Miao, J. & Niu, L. (2016). A survey on feature selection. *Procedia Computer Science 91*(2016), 919–926. https://doi.org/10.1016/j.procs.2016.07.111

Miller, C. W. (2020). An investigation into subject matter expert elicitation in cost risk analysis. *Theses and Dissertations 3247*

Miric, M., Jia, N., & Huang, K. G. (2022). Using supervised machine learning for large-scale classification in management research: The case for identifying artificial

intelligence patents. *Strategic Management Journal*. https://doi.org/10.1002/smj.3441

Naval Air Systems Command. (2017). Solicitation No. N0001917R087

North American industry classification system - census.gov. NAICS Manual. (n.d.). Retrieved December 23, 2022, from https://www.census.gov/naics/reference_files_tools/2022_NAICS_Manual.pdf

O'Guin, M. & Kelly, K. (2012). *Winning the big ones: how teams capture large contracts*. Lulu Com.

Perini, A., Susi, A., & Avesani, P. (2013). A machine learning approach to software requirements prioritization. *IEEE Transactions on Software Engineering 39*(4), 445–461. https://doi.org/10.1109/TSE.2012.52

Provost, F., & Kohavi, R. (1998). Guest editors' introduction: on applied research in machine learning. *Machine Learning 30,* 127–132.

Product and service code manual. Product and Service Code Manual | Acquisition.GOV. (n.d.). Retrieved December 22, 2022, from https://www.acquisition.gov/psc-manual

Rundall, T. G., Martelli, P. F., Arroyo, L., Mccurdy, R., Graetz, I., Neuwirth, E. B., Curtis, P., Schmittdiel, J., Gibson, M., & Hsu, J. (2007). The informed decisions toolbox: Tools for knowledge transfer and performance improvement. *Journal of Healthcare Management 52*(5), 325.

Saaty, R. W. (1987). The analytic hierarchy process—what it is and how it is used. *Mathematical modelling, 9*(3-5), 161-176.

Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and research directions. *SN Computer Science 2*(3). https://doi.org/10.1007/s42979-021-00592-x

Trifonova, O. P., Lokhov, P. G., & Archakov, A. I. (2013). Metabolic profiling of human blood. *Biochemistry (Moscow) Supplement Series B: Biomedical Chemistry 7*(3), 179–186. https://doi.org/10.1134/S1990750813030128

Ward, S. C. & Chapman, C. B. (1988). Developing competitive bids: A framework for information processing. *Journal of the Operational Research Society 39*(2), 123–134. https://doi.org/10.1057/jors.1988.24

Zhang, W. & Hamori, S. (2020). Do machine learning techniques and dynamic methods help forecast US natural gas crises? *Energies 13*(9). https://doi.org/10.3390/en13092371

Appendix A: RFP Database Sample – Factors and Relationships

Market	Defens	Contract #	Solicitation #	PoP (mi	Contrac	Contrac	PSC	NAICS	Factor_1	Factor_2	Factor_3	Factor_4	Factor_5	Factor_6	Factor_7	Factor_1_2	Factor_2_3	Factor_3_4	Factor_4_5	Factor_5_6	Factor_5_7	Factor_6_7	Checkp
Vehicles	US Army	W56HZV-20-C-0216	W56HZV-19-R-0043	35.96712	Cost Plus f	48.6	J025 - MAII	811198 - A	Technical	Price	Small Business					More	Sig_More						Check
Vehicles	US Army	W56HZV-20-D-0039	W56HZV-19-R-0079	30.47671	Firm Fixed	56.9	23 - MOTO	336120 - H	Technical	Price						Equal							Check
Vehicles	US Army	W50RAJ-20-9-C014	W56JSR-18-S-0001_BHEV	23.96712		32.3		5417 - Scie	Technical Me	Importance t	Schedule	Price				More	More	More					Check
Vehicles	US Army	SPRDL1-20-D-0067	SPRDL1-19-R-0120	60	Firm Fixed	61.8	5998 - ELEC	334418 - Pr	Technical	Price						Equal							Check
Vehicles	US Army	SPRDL1-20-C-0109	SPRDL1-20-R-0033	27.61644	Firm Fixed	58	2815 - DIES	333618 - O	Technical	Price						Equal							Check
Vehicles	US Army	SPRDL1-20-C-0062	SPRDL1-19-R-0060	20.87671	Firm Fixed	26.1	2530 - VEH	336992 - M	Technical	Price						Equal							Check
Vehicles	US Army	W56HZV-20-D-0023	W56HZV-18-R-0173	84	Firm Fixed	51.5	23 - MOTO	336212 - Tr	Past Performa	Price						Sig_More							Check
Vehicles	US Army	SPRDL1-20-D-0038	SPRDL1-19-R-0038	74.46575	Firm Fixed	50.5	2815 - DIES	336310 - M	Technical	Price						Equal							Check
Vehicles	US Army	W56HZV-20-D-0016	W56HZV-19-R-0089	36.03288	Indefinite	24.5	23 - MOTO	336120 - H	Technical	Price						Equal							Check
Vehicles	US Army	W90BRJ-19-C-0001	W90BRJ-19-R-0005	60.03288	Firm Fixed	21.3	S - UTILITII	561210 - Fa	Technical	Past Performa	Price					Equal	Equal						Check
Vehicles	US Army	W52P1J-17-G-0100	W52P1J-18-R-0070	60.65753	Firm Fixed	81.5	R - SUPPOF	561210 - Fa	Technical	Past Performa	Price					Equal	Sig_More						Check
Vehicles	US Army	W56HZV-19-D-0041	W56HZV-17-R-0106	58.06027	Firm Fixed	60.7	25 - VEHICI	336992 - M	Past Performa	Price						Sig_More							Check
Vehicles	US Army	W56HZV-19-D-0013	W56HZV-16-R-0210	84.03288	Firm Fixed	66.7	23 - MOTO	336212 - Tr	Technical	Price	Past Performance					Sig_More	Slightly_More						Check
Vehicles	US Army	W56HZV-19-D-0010	W56HZV-17-R-0092	35.47397	Firm Fixed	71.7	23 - MOTO	336111 - Au	Technical	Price						Equal							Check
Vehicles	US Army	SPRDL1-18-C-0353	SPRDL1-17-R-0254	46.91507	Firm Fixed	28.3	12 - FIRE C	333314 - O	Technical	Price						Equal							Check
Vehicles	US Army	W900KK-18-C-0049	W900KK-17-R-0016	42.04932	Cost Plus f	32.8	69 - TRAIN	333318 - O	Technical	Program Man	Past Performa	Price				Sig_More	Sig_More	Sig_More					Check
Vehicles	US Army	W911NF-18-D-0005	W911NF-18-R-0001	59.76986	Cost Plus f	25	R - SUPPOF	541330 - En	Business Via	Technical	Program Man	Past Performa	Price			Equal	Sig_More	Equal	More				Check
Vehicles	US Army	W56HZV-18-D-0032	W56HZV-16-R-0149	26.79452	Indefinite	26	23 - MOTO	336120 - H	Technical	Price						Equal							Check
Vehicles	US Army	SPRDL1-18-C-0142	SPRDL1-17-R-0430	17.75342	Firm Fixed	20.5	25 - VEHICI	336992 - M	Technical	Price						Equal							Check
Vehicles	US Army	W56HZV-17-D-0108	W56HZV-16-R-0096	60	Firm Fixed	88.6	25 - VEHICI	334111 - El	Technical	Price						Equal							Check
Vehicles	US Army	W911KF-17-D-0005	W911KF-16-R-0276	61.4137	Firm Fixed	93	R - SUPPOF	336992 - M	Technical	Past Performa	Price					More	More						Check
Vehicles	US Army	SPRDL1-17-C-0116	SPRDL1-16-R-0360	8.646575	Firm Fixed	20.2	25 - VEHICI	336992 - M	Technical	Price						Equal							Check
Vehicles	US Army	W56HZV-17-D-0032	W56HZV-16-R-0149	36.03288	Firm Fixed	27.9	23 - MOTO	336120 - H	Technical	Price						Equal							Check
Vehicles	US Army	W911SR-17-C-0017	W911SR-16-R-0001	24	Cost Plus f	26.5	A - RESEAR	541712 -	Technical	Transition Pl	Past Performa	Program Man	Small Busine	Price		More	More	More	More	More			Check
Vehicles	US Army	W31P4Q-17-C-0063	W31P4Q-16-R-0124	36.03288	Firm Fixed	23.4	13 - AMMU	336419 - O	Technical	Past Performa	Price					Equal	More						Check
Vehicles	US Army	W911SR-16-C-0039	W911SR-16-R-0002	47.86849	Cost Plus I	20	A - RESEAR	541712 -	Business Via	Technical	Transition Pl	Program Man	Past Performa	Small Busine	Price	More	More	More	More	More		More	Check
Vehicles	US Army	W56HZV-16-C-0174	W56HZV-16-R-0051	11.83562	Firm Fixed	32.9	25 - VEHICI	336992 - M	Technical	Price													Check
Vehicles	US Air For	FA8534-20-D-0002	FA8534-18-R-0015	60.9863	Firm Fixed	22.1	1740 - AIRF	333924 - In	Technical	Past Performa													
Vehicles	US Air For	FA8505-18-D-0003	FA8505-18-R-0007	36.03288	Firm Fixed	68.5																	

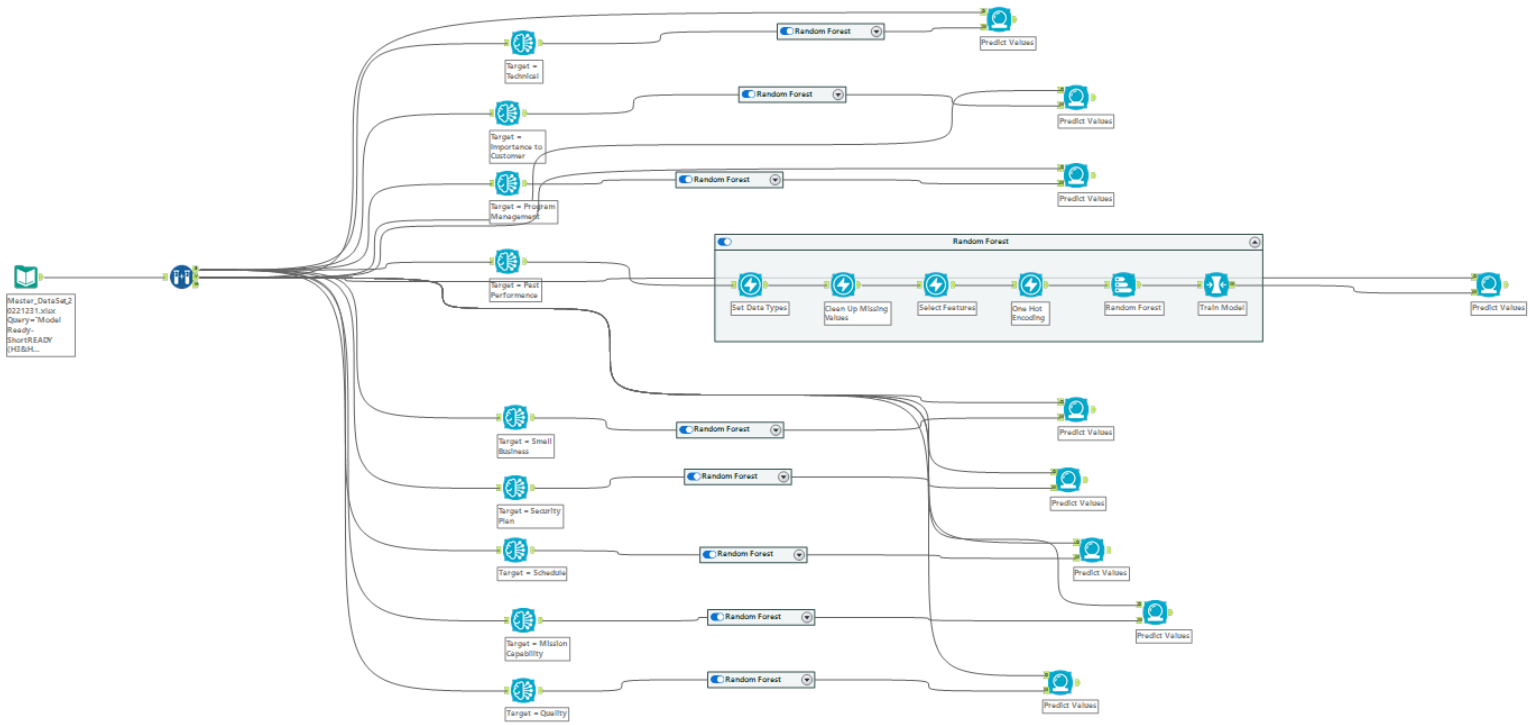
Appendix B: RFP Database Sample – AHP Application

Market	Defens	Contract #	Solicitation #	PoP (m	Contrac	Contrac	PSC	NAICS	Factor_1_2	Factor_	Factor_	Factor_	Factor_	Factor_	Factor_	Factor_	Factor_	Factor_	Factor_	Factor_	Factor_	Factor_	Factor_	Factor_	Factor_	Factor_
Vehicles	US Army	W56HZV-20-C-0216	W56HZV-19-R-0043	35.96712	Cost Plus I	48.6	J025 - MAII	811198 - A	5	5.5					0.2	1	7					0.181818	0.142857	1		
Vehicles	US Army	W56HZV-20-D-0039	W56HZV-19-R-0079	30.47671	Firm Fixed	56.9	23 - MOTO	336120 - H	1						1	1										
Vehicles	US Army	W50RAJ-20-9-C014	W56JSR-18-S-0001_BHEV	23.96712		32.3	5417 - Scie		5	5.5	6				0.2	1	5	5.5				0.181818	0.2	1	5	
Vehicles	US Army	SPRDL1-20-D-0067	SPRDL1-19-R-0120	60	Firm Fixed	61.8	5998 - ELEC	334418 - Pr	1						1	1										
Vehicles	US Army	SPRDL1-20-C-0109	SPRDL1-20-R-0033	27.61644	Firm Fixed	58	2815 - DIES	333618 - O	1						1	1										
Vehicles	US Army	SPRDL1-20-C-0062	SPRDL1-19-R-0060	20.87671	Firm Fixed	26.1	2530 - VEH	336992 - M	1						1	1										
Vehicles	US Army	W56HZV-20-D-0023	W56HZV-18-R-0173	84	Firm Fixed	51.5	23 - MOTO	336212 - Tr	7						0.142857	1										
Vehicles	US Army	SPRDL1-20-D-0038	SPRDL1-19-R-0038	74.46575	Firm Fixed	50.5	2815 - DIES	336310 - M	1						1	1										
Vehicles	US Army	W56HZV-20-D-0016	W56HZV-19-R-0089	36.03288	Indefinite	24.5	23 - MOTO	336120 - H	1						1	1										
Vehicles	US Army	W90BRJ-19-C-0001	W90BRJ-19-R-0005	60.03288	Firm Fixed	21.3	S - UTILITII	561210 - Fa	1	1					1	1	1					1	1	1		
Vehicles	US Army	W52P1J-17-G-0100	W52P1J-18-R-0070	60.65753	Firm Fixed	81.5	R - SUPPOI	561210 - Fa	1	7					1	1	7					0.142857	0.142857	1		
Vehicles	US Army	W56HZV-19-D-0041	W56HZV-17-R-0106	58.06027	Firm Fixed	60.7	25 - VEHICI	336992 - M	7						0.142857	1										
Vehicles	US Army	W56HZV-19-D-0013	W56HZV-16-R-0210	84.03288	Firm Fixed	66.7	23 - MOTO	336212 - Tr	7	7.5					0.142857	1	3					0.133333	0.333333	1		
Vehicles	US Army	W56HZV-19-D-0010	W56HZV-17-R-0092	35.47397	Firm Fixed	71.7	23 - MOTO	336111 - Au	1						1	1										
Vehicles	US Army	SPRDL1-18-C-0353	SPRDL1-17-R-0254	46.91507	Firm Fixed	28.3	12 - FIRE C	333314 - Op	1						1	1										
Vehicles	US Army	W900KK-18-C-0049	W900KK-17-R-0016	42.04932	Cost Plus I	32.8	69 - TRAIN	333318 - O	7	7.5	8				0.142857	1	7	7.5				0.133333	0.142857	1	7	
Vehicles	US Army	W911NF-18-D-0005	W911NF-18-R-0001	59.76986	Cost Plus I	25	R - SUPPOI	541330 - En	1	7	7	7.5			1	1	7	7	7.5			0.142857	0.142857	1	1	5
Vehicles	US Army	W56HZV-18-D-0032	W56HZV-16-R-0149	26.79452	Indefinite	26	23 - MOTO	336120 - H	1						1	1										
Vehicles	US Army	SPRDL1-18-C-0142	SPRDL1-17-R-0430	17.75342	Firm Fixed	20.5	25 - VEHICI	336992 - M	1						1	1										
Vehicles	US Army	W56HZV-17-D-0108	W56HZV-16-R-0096	60	Firm Fixed	88.6	25 - VEHICI	334111 - El	1						1	1										
Vehicles	US Army	W911KF-17-D-0005	W911KF-16-R-0276	61.4137	Firm Fixed	93	R - SUPPOI	336992 - M	5	5.5					0.2	1	5					0.181818	0.2	1		
Vehicles	US Army	SPRDL1-17-C-0116	SPRDL1-16-R-0360	8.646575	Firm Fixed	20.2	25 - VEHICI	336992 - M	1						1	1										
Vehicles	US Army	W56HZV-17-D-0032	W56HZV-16-R-0149	36.03288	Firm Fixed	27.9	23 - MOTO	336120 - H	1						1	1										
Vehicles	US Army	W911SR-17-C-0017	W911SR-16-R-0001	24	Cost Plus I	26.5	A - RESEAR	541712 -	5	5.5	6	6.5	7		0.2	1	5	5.5	6	6.5		0.181818	0.2	1	5	5.5
Vehicles	US Army	W31P4Q-17-C-0063	W31P4Q-16-R-0124	36.03288	Firm Fixed	23.4	13 - AMMU	336419 - O	1	5					1	1	5					0.2	0.2	1		
Vehicles	US Army	W9115R-16-C-0039	W9115R-16-R-0002	47.86849	Cost Plus I	20	A - RESEAR	541712 -	5	5.5	6	6.5	7	7.5	0.2	1	5	5.5	6	6.5	7	0.181818	0.2	1	5	5.5
Vehicles	US Army	W56HZV-16-C-0174	W56HZV-16-R-0051	11.83562	Firm Fixed	32.9	25 - VEHICI	336992 - M	1						1	1										
Vehicles	US Air Forc	FA8534-20-D-0002	FA8534-18-R-0015	60.9863	Firm Fixed	22.1	1740 - AIRI	333924 - In	1	1																
Vehicles	US Air Forc	FA8505-18-D-0003	FA8505-18-R-0007	36.03288	Firm Fixed																					

Appendix C: RFP Database Sample – Geometric Mean Calculations

Market	Defens	Contract #	Solicitation #	PoP (mi	Contrac	Contrac	PSC	NAICS	GMM1	GMM2	GMM3	GMM4	GMM5	GMM6	GMM7	Sum_G	Wt_1	Wt_2	Wt_3	Wt_4	Wt_5	Wt_6	Wt_7	Wt_Sur
Vehicles	US Army	W56HZV-20-C-0216	W56HZV-19-R-0043	35.96712	Cost Plus I	48.6	J025 - MAII	811198 - Al	3.018405	1.118689	0.296151	0	0	0	0	4.433245	0.680857	0.252341	0.066802	0	0	0	0	1
Vehicles	US Army	W56HZV-20-D-0039	W56HZV-19-R-0079	30.47671	Firm Fixed	56.9	23 - MOTO	336120 - H	1	1	0	0	0	0	0	2	0.5	0.5	0	0	0	0	0	1
Vehicles	US Army	W50RAJ-20-9-C014	W56JSR-18-S-0001_BHEV	23.96712		32.3		5417 - Scie	3.584025	1.531407	0.652994	0.279016	0	0	0	6.047442	0.592651	0.253232	0.107979	0.046138	0	0	0	1
Vehicles	US Army	SPRDL1-20-D-0067	SPRDL1-19-R-0120	60	Firm Fixed	61.8	5998 - ELEC	334418 - Pr	1	1	0	0	0	0	0	2	0.5	0.5	0	0	0	0	0	1
Vehicles	US Army	SPRDL1-20-C-0109	SPRDL1-20-R-0033	27.61644	Firm Fixed	58	2815 - DIES	333618 - O	1	1	0	0	0	0	0	2	0.5	0.5	0	0	0	0	0	1
Vehicles	US Army	SPRDL1-20-C-0062	SPRDL1-19-R-0060	20.87671	Firm Fixed	26.1	2530 - VEH	336992 - M	1	1	0	0	0	0	0	2	0.5	0.5	0	0	0	0	0	1
Vehicles	US Army	W56HZV-20-D-0023	W56HZV-18-R-0173	84	Firm Fixed	51.5	23 - MOTO	336212 - Tr	2.645751	0.377964	0	0	0	0	0	3.023716	0.875	0.125	0	0	0	0	0	1
Vehicles	US Army	SPRDL1-20-D-0038	SPRDL1-19-R-0038	74.46575	Firm Fixed	50.5	2815 - DIES	336310 - M	1	1	0	0	0	0	0	2	0.5	0.5	0	0	0	0	0	1
Vehicles	US Army	W56HZV-20-D-0016	W56HZV-19-R-0089	36.03288	Indefinite	24.5	23 - MOTO	336120 - H	1	1	0	0	0	0	0	2	0.5	0.5	0	0	0	0	0	1
Vehicles	US Army	W90BRJ-19-C-0001	W90BRJ-19-R-0005	60.03288	Firm Fixed	21.3	S - UTILITII	561210 - Fa	1	1	1	0	0	0	0	3	0.333333	0.333333	0.333333	0	0	0	0	1
Vehicles	US Army	W52P1J-17-G-0100	W52P1J-18-R-0070	60.65753	Firm Fixed	81.5	R - SUPPOF	561210 - Fa	1.912931	1.912931	0.273276	0	0	0	0	4.099138	0.466667	0.466667	0.066667	0	0	0	0	1
Vehicles	US Army	W56HZV-19-D-0041	W56HZV-17-R-0106	58.06027	Firm Fixed	60.7	25 - VEHICI	336992 - M	2.645751	0.377964	0	0	0	0	0	3.023716	0.875	0.125	0	0	0	0	0	1
Vehicles	US Army	W56HZV-19-D-0013	W56HZV-16-R-0210	84.03288	Firm Fixed	66.7	23 - MOTO	336212 - Tr	3.744436	0.753947	0.35422	0	0	0	0	4.852603	0.771635	0.15537	0.072996	0	0	0	0	1
Vehicles	US Army	W56HZV-19-D-0010	W56HZV-17-R-0092	35.47397	Firm Fixed	71.7	23 - MOTO	336111 - Au	1	1	0	0	0	0	0	2	0.5	0.5	0	0	0	0	0	1
Vehicles	US Army	SPRDL1-18-C-0353	SPRDL1-17-R-0254	46.91507	Firm Fixed	28.3	12 - FIRE C	333314 - O	1	1	0	0	0	0	0	2	0.5	0.5	0	0	0	0	0	1
Vehicles	US Army	W900KK-18-C-0049	W900KK-17-R-0016	42.04932	Cost Plus I	32.8	69 - TRAIN	333318 - O	4.527019	1.654875	0.604275	0.220896	0	0	0	7.007066	0.646065	0.236172	0.086238	0.031525	0	0	0	1
Vehicles	US Army	W911NF-18-D-0005	W911NF-18-R-0001	59.76986	Cost Plus I	25	R - SUPPOF	541330 - En	3.258753	3.258753	0.633512	0.633512	0.234632	0	0	8.019162	0.406371	0.406371	0.079	0.079	0.029259	0	0	1
Vehicles	US Army	W56HZV-18-D-0032	W56HZV-16-R-0149	26.79452	Indefinite	26	23 - MOTO	336120 - H	1	1	0	0	0	0	0	2	0.5	0.5	0	0	0	0	0	1
Vehicles	US Army	SPRDL1-18-C-0142	SPRDL1-17-R-0430	17.75342	Firm Fixed	20.5	25 - VEHICI	336992 - M	1	1	0	0	0	0	0	2	0.5	0.5	0	0	0	0	0	1
Vehicles	US Army	W56HZV-17-D-0108	W56HZV-16-R-0096	60	Firm Fixed	88.6	25 - VEHICI	334111 - El	1	1	0	0	0	0	0	2	0.5	0.5	0	0	0	0	0	1
Vehicles	US Army	W911KF-17-D-0005	W911KF-16-R-0276	61.4137	Firm Fixed	93	R - SUPPOF	336992 - M	3.018405	1	0.331301	0	0	0	0	4.349706	0.693933	0.229901	0.076166	0	0	0	0	1
Vehicles	US Army	SPRDL1-17-C-0116	SPRDL1-16-R-0360	8.646575	Firm Fixed	20.2	25 - VEHICI	336992 - M	1	1	0	0	0	0	0	2	0.5	0.5	0	0	0	0	0	1
Vehicles	US Army	W56HZV-17-D-0032	W56HZV-16-R-0149	36.03288	Firm Fixed	27.9	23 - MOTO	336120 - H	1	1	0	0	0	0	0	2	0.5	0.5	0	0	0	0	0	1
Vehicles	US Army	W911SR-17-C-0017	W911SR-16-R-0001	24	Cost Plus I	26.5	A - RESEAR	541712 -	4.425027	2.446646	1.348006	0.741836	0.408723	0.225987	0	9.596226	0.461122	0.254959	0.140473	0.077305	0.042592	0.02355	0	1
Vehicles	US Army	W31P4Q-17-C-0063	W31P4Q-16-R-0124	36.03288	Firm Fixed	23.4	13 - AMMU	336419 - O	1.709976	1.709976	0.341995													1
Vehicles	US Army	W911SR-16-C-0039	W911SR-16-R-0002	47.86849	Cost Plus I	20	A - RESEAR	541712 -																
Vehicles	US Army	W56HZV-16-C-0134	W56HZV-15-R-0001																					

Appendix D: Alteryx Random Forest Workflow Example



Appendix E: ML Algorithm Configuration Parameters

Configuration Parameter \ Algorithm	DT	RF	XGBoost
Bootstrap	N/A	On	N/A
Class Weight	None	None	N/A
Criterion	Gini Impurity	Gini Impurity	N/A
Max Depth	Unlimited	Limited: 10	3
Max Features	Auto	Auto	N/A
Max Leaf Nodes	None	N/A	N/A
Min Impurity Decrease	0	0	N/A
Min Samples Split	Integer: 2	Integer: 2	N/A
Min Weight Fraction Leaf	0	0	N/A
Presort Data	Off	N/A	N/A
Number of Estimators	N/A	100	100
Random Seed	Seed: 10	Seed: 10	10
Splitter	Best	N/A	N/A
Subsample	N/A	N/A	1
Sample Columns by Level	N/A	N/A	1
Sample Columns by Node	N/A	N/A	1
Sample Columns by Tree	N/A	N/A	1
Gamma	N/A	N/A	0
Learning Rate	N/A	N/A	0.05
Minimum Child Weight	N/A	N/A	1

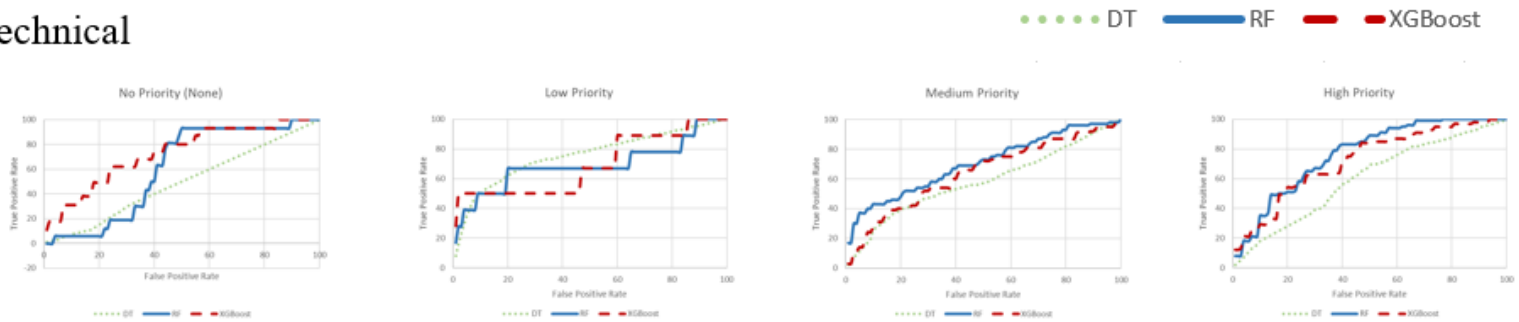
Appendix F: Confusion Matrix Results for Prioritized Evaluation Criteria Model

Confusion Matrix (with Prioritization)			Predicted											
			Decision Tree				Random Forest				XGBoost			
			None	Low	Medium	High	None	Low	Medium	High	None	Low	Medium	High
Actual	Technical	None	0%	6%	25%	69%	0%	0%	38%	63%	6%	6%	56%	31%
		Low	0%	38%	13%	50%	0%	0%	25%	75%	0%	25%	38%	38%
		Medium	6%	4%	42%	48%	0%	1%	40%	58%	0%	3%	46%	51%
		High	3%	6%	21%	70%	0%	0%	8%	92%	1%	0%	21%	78%
	Program Management	None	90%	4%	4%	2%	99%	1%	1%	0%	95%	4%	1%	0%
		Low	54%	29%	8%	8%	67%	25%	4%	4%	67%	21%	8%	4%
		Medium	80%	20%	0%	0%	80%	20%	0%	0%	50%	40%	10%	0%
		High	33%	33%	0%	33%	67%	33%	0%	0%	67%	33%	0%	0%
	Past Performance	None	77%	8%	14%	1%	85%	3%	13%	0%	85%	1%	12%	5%
		Low	48%	18%	35%	0%	33%	15%	53%	0%	13%	30%	58%	0%
		Medium	49%	14%	36%	1%	23%	9%	69%	0%	4%	21%	74%	0%
		High	33%	33%	33%	0%	67%	0%	33%	0%	83%	0%	17%	0%
	Small Business	None	84%	13%	2%		100%	0%	0%		98%	2%	0%	
		Low	43%	52%	4%		87%	9%	4%		52%	39%	9%	
		Medium	60%	40%	0%		80%	20%	0%		60%	40%	0%	
		High												
	Importance to Customer	None	95%	1%	4%		99%	1%	1%		98%	0%	2%	
		Low	0%	100%	0%		0%	83%	17%		0%	83%	17%	
		Medium	19%	6%	75%		50%	0%	50%		38%	6%	56%	
		High												

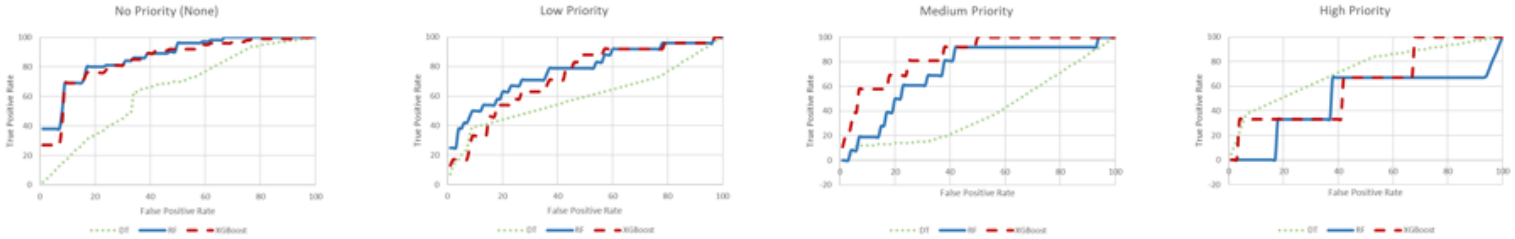
	Schedule	None	95%	3%	2%		100%	0%	0%		97%	3%	0%	
		Low	42%	58%	0%		33%	67%	0%		25%	75%	0%	
		Medium	80%	20%	0%		100%	0%	0%		80%	20%	0%	
		High												
	Mission Capability	None	88%	1%	4%	3%	100%	0%	0%	0%	99%	1%	1%	0%
		Low	42%	50%	0%	8%	33%	67%	0%	0%	42%	58%	0%	0%
		Medium	100%	0%	0%	0%	100%	0%	0%	0%	40%	60%	0%	0%
		High	75%	25%	0%	0%	100%	0%	0%	0%	75%	0%	25%	0%
	Quality	None	96%	4%	0%		99%	1%	0%		98%	2%	0%	
		Low	50%	40%	10%		90%	0%	10%		80%	20%	0%	
		Medium	67%	33%	0%		67%	33%	0%		67%	33%	0%	
		High												
	Security Plan	None	97%	1%	2%	0%	100%	0%	0%	0%	99%	0%	1%	0%
		Low	0%	50%	50%	0%	50%	0%	50%	0%	0%	0%	100%	0%
		Medium	50%	25%	0%	25%	75%	25%	0%	0%	25%	25%	50%	0%
		High	100%	0%	0%	0%	100%	0%	0%	0%	100%	0%	0%	0%

Appendix G: Stage 2 AUC Results

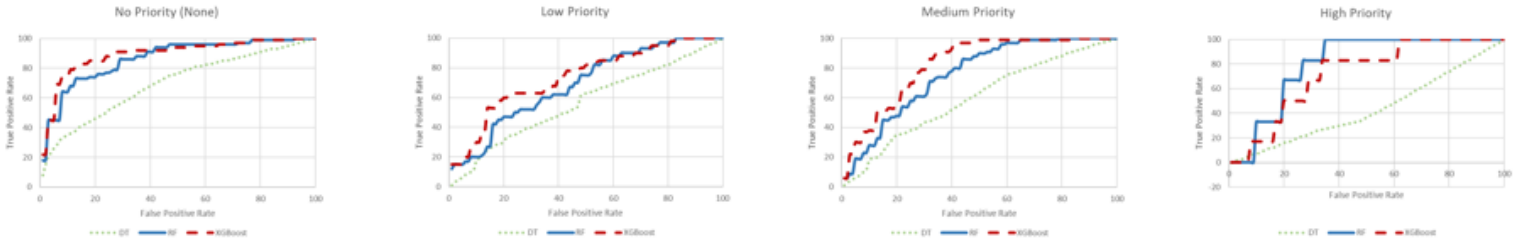
Technical



Program Management

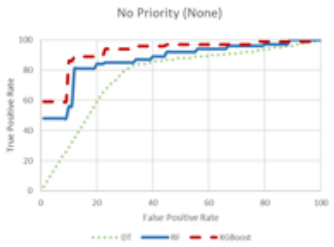


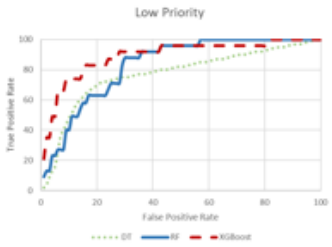
Past Performance



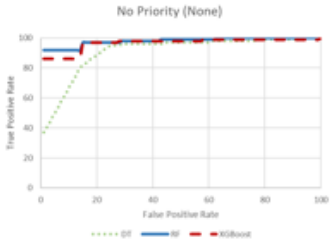
Small Business

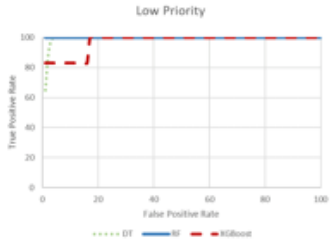
DT RF XGBoost

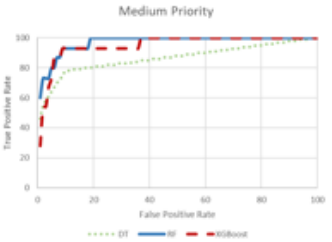




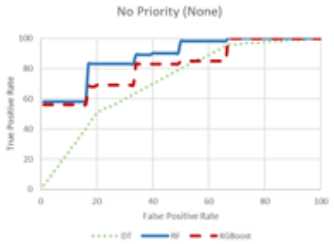
Importance to the Customer

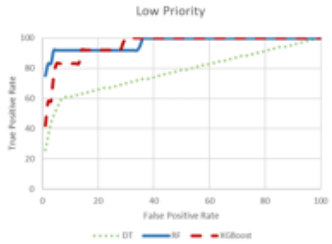


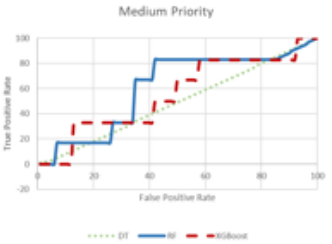




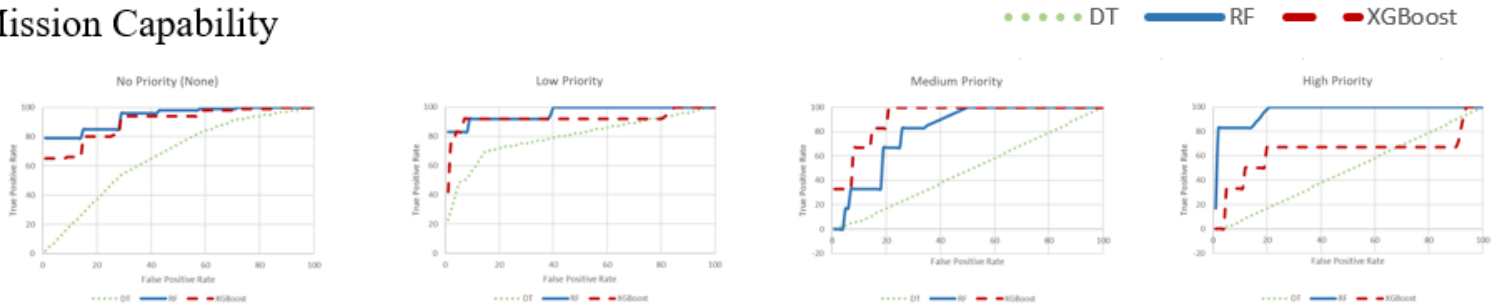
Schedule



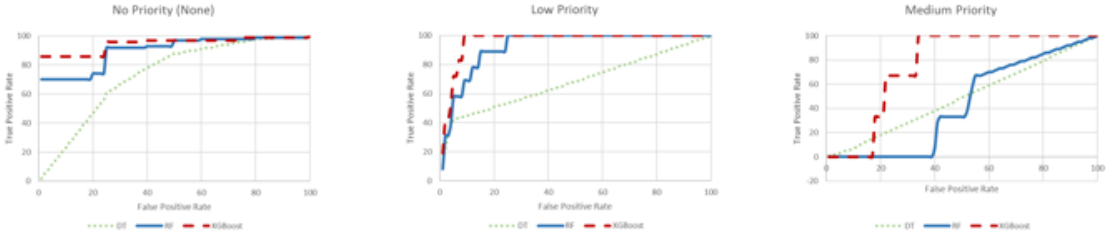




Mission Capability



Quality



Security Plan