The Utility of Data Science Applied to Military Assessment and Selection for Holistic Systems Improvement

---

A

Thesis

Presented to

the faculty of the School of Engineering and Applied Science

University of Virginia

---

in partial fulfillment

of the requirements for the degree

Master of Science

by

Hayden Deverill

May 2022

APPROVAL SHEET

This

Thesis

is submitted in partial fulfillment of the requirements
for the degree of

Master of Science

Author: Hayden Deverill

This Thesis has been read and approved by the examing committee:

Advisor: William Scherer

Advisor:

Committee Member: Michael Porter

Committee Member: Allan Stam

Committee Member:

Committee Member:

Committee Member:

Committee Member:

Accepted for the School of Engineering and Applied Science:

Jennifer L. West, School of Engineering and Applied Science

May 2022

UNIVERSITY OF VIRGINIA

MASTER'S THESIS

The Utility of Data Science Applied to Military Assessment and Selection for Holistic Systems Improvement

Author:
Hayden Deverill

Thesis Committee:
Michael Porter, Chair
William Scherer, Advisor
Allan Stam

A thesis submitted in fulfillment of the requirements for the degree of Master of Science

in the

Department of Engineering Systems and Environment

May 2022

UNIVERSITY OF VIRGINIA

Abstract

Department of Engineering Systems and Environment

Master of Science

The Utility of Data Science Applied to Military Assessment and Selection for Holistic Systems Improvement

by Hayden Deverill

Elite military units use an in-depth assessment and selection (A&S) process to acquire the most qualified candidates. A unique challenge is to objectively evaluate the human dimension of attributes like leadership, resilience, and initiative in candidates. The A&S process requires significant time and resources to execute. The specific A&S studied for this research is eight weeks long and has a high logistical demand between supplies, personnel, and facilities. Effective screening of candidates prior to the A&S saves resources and selecting the best candidates enables the unit to better conduct highly specialized missions. Improving the system will reap dividends for the military.

Most studies about military A&S have used small data sets, used descriptive statistics for analysis, and focused on identifying predictors of candidate success. This research was broader in scope. We used 11,885 candidate records taken over a five-year period with 89 total features that included administrative, performance, and psychological data on each candidate. We applied a robust data science approach involving feature engineering, feature selection, optimized predictive models, and data subsets analysis to extract meaningful information from the data. Our objective for this research was to evaluate the utility of applying data science techniques to a specific military A&S data set with the goal of improving the holistic A&S system.

We applied ten classification models to a variety of feature, candidate, and feature engineering data subset combinations created using data science techniques. Using all candidates, the best model performance yielded a kappa score of 52 and 77% accuracy. Candidate non-selection prediction accuracy (86% Negative Predictive Value) was higher than candidate selection (68% Positive Predictive Value). The strongest predictors of candidate success were performance features, followed by administrative, and lastly psychological features. Although prediction accuracy was modest (<90%), we discovered utility in applying data science techniques to the A&S data. We extracted valuable insights from the data, found features highly predictive of candidate non-selection, and learned methods to modify the existing data to improve predictive capability. In conclusion, this research 1) validates the importance of an A&S to observe the human dimension of candidates and 2) proposes recommendations to add value to the holistic A&S system.

Acknowledgements

To my Lord and Savior, Jesus Christ. Apart from You, I can do nothing. You have given me new life ever since I placed my faith in You in 2010. I give all glory to You for any good that comes out of this research.

To my family. Faith, another chapter in our eight years of marriage where you've come alongside me, sacrificing your goals and helping me accomplish mine. Without your loving support, I could not have completed this research. Starting grad school with two kids and finishing with three was a feat only you could pull off. **Isaac**, I'll miss telling you about how I "learned about numbers" at school every day. **Titus**, your frequent, unexpected visits to my office during school were huge morale boosts. **Eloise**, each day I looked forward to seeing you before and after your naps to break up the work. I love you all.

To my thesis committee. I am grateful for your investment in both my professional and personal development. I am a better man and scholar as a result. **Professor Scherer**, I remember the first phone conversation we had in the summer of 2019 when I was a prospective UVA student. During the phone call, I was still doing "Army stuff" at a shooting range with bullets cracking in the background. You took the time to answer all my questions and give helpful, transparent feedback as I considered grad school options. Since then you haven't looked back, volunteering to be my academic advisor as a new graduate student and willing to advise me for this thesis research. Thank you for helping an Army guy figure out grad school along with your deliberate mentorship and friendship. **Professor Stam**, thanks for taking the time to get lunch with a random Systems Engineering graduate student in February 2021. That lunch turned into over a year of investment in me and this research. Your perspective has added much value to the research with your unique life experiences and domain knowledge. **Professor Porter**, thank you for giving me the data science foundation I needed to complete this thesis. Your data mining class is top notch and I referenced your notes often.

To the A&S unit. This research literally would not have been possible without you (and the data you collect). Your team does a phenomenal job at collecting and maintaining the data you do - keep up the good work. Thank you for your willingness to support this research along with all the time, feedback, and work to go with it. Everyone I worked with from your unit was eager to support this endeavor. The three day visit to the A&S was one of the most rewarding parts of the research. My goal for this research was to be value added to your organization; I hope that goal was achieved.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Problem Definition

Elite military units use an in-depth assessment and selection (A&S) process to acquire the most qualified candidates. One unique challenge is to objectively evaluate the human dimension of attributes like leadership, resilience, and grit in candidates. Often, data is collected on candidates who attend the A&S that is analyzed to aide in selecting the best candidates.

The specific A&S we researched has detailed data on all candidates who have attended the A&S since 2016 (11,885 candidates). The unit collects administrative, performance, and psychological data on each candidate. Currently, the data is primarily used for robust record keeping and descriptive statistic reports. Likewise, the performance and psychological data is also used to screen candidates prior to entry into the A&S. However, "data science" methods have not been applied to the data with the goal of process improvement.

In recent years, data science has emerged as a field of it's own, distinct from other disciplines such as statistics or machine learning [1]. It is broader and more holistic in scope than any one discipline. Rather, "data science is a new interdisciplinary field that synthesizes and builds on statistics, informatics, computing, communication, management, and sociology to study data and its environments (including domains and other contextual aspects, such as organizational and social aspects) in order to transform data to insights and decisions by following a data-to-knowledge-to-wisdom thinking and methodology" [2]. When we use the term "data science", this is what we mean as it relates to this research.

Data science has led to many breakthroughs and insights across numerous domains [3]. We believe we can gain valuable insights about the holistic A&S system by applying data science to the existing data collected. Our objective for this research was to evaluate the utility of applying data science techniques to a specific military A&S data set with the goal of improving the holistic A&S system.

1.1.1 Research Questions

1. Is application of data science techniques to military A&S data useful to gain insight on ways to holistically improve the system?

2. Are predictive models useful and accurate to model selection in Military A&S?

3. Are there any features that are indicative of a candidate being selected or not selected?

1.2 Motivation

The broad motivation for this research is to add value to the military unit that conducts this A&S. As an Army officer, I am personally invested in trying to improve the organization that I serve. More specifically, there are two motivations for this research:

1. Evaluate the utility of applying data science techniques to this military A&S data set for holistic system improvement. This type of analysis has not been done before and we believe the potential value to be gained is large.

2. Make recommendations based on the results and conclusions that add value to the A&S system. Ultimately, we want our efforts to lead to actionable recommendations that may improve the existing A&S system.

We think this research is significant because any value that is added to A&S system directly translates to selecting more qualified candidates. Selecting better candidates results in a better fighting force to conduct the military's most challenging missions; in turn, returning value to the military on a macro-level scale.

1.3 Results Summary

We applied ten classification models to a variety of feature, candidate, and feature engineering data subset combinations created using data science techniques. Using all candidates, the best model performance yielded a kappa score of 52 and 77% accuracy. Candidate non-selection prediction accuracy (86% Negative Predictive Value [NPV]) was higher than candidate selection (68% Positive Predictive Value [PPV]). The strongest predictors of candidate success were performance features, followed by administrative, and lastly psychological features. The penalized logistic regression and ensemble models performed the best across all data subset combinations, with KNN and CART performing the worst. The best performing feature subset used all the features and resulted in the highest kappa score of 52. All of the 10 feature subsets resulted in higher NPV scores than PPV scores, with sensitivity and specificity varying between subsets. Feature engineering techniques resulted in better results and changed feature importance values compared to using the data as is. Using all candidates yielded higher kappa scores in prediction compared to using only candidates who scored above the minimum fitness scores.

Although prediction accuracy was modest (<90%), we discovered utility in applying data science techniques to the A&S data. We extracted valuable insights from the data, found features highly predictive of candidate non-selection, and learned methods to modify the existing data to improve predictive capability. In conclusion, this research 1) validates the importance of an A&S to observe the human dimension of candidates and 2) proposes recommendations to add value to the holistic A&S system.

Chapter 2

Background

2.1 Understanding the Problem

The following are definitions specific to the A&S process:

1. **Candidate:** A Soldier who is actively participating in the A&S (screening or A&S phase).

2. **Screening:** An objective criteria candidates must meet in the screening phase prior to starting the A&S phase.

3. **Selection:** Evaluation of candidates in the A&S phase for selection into the military unit. Candidates who complete the entire A&S are either selected to join the unit or dropped (removed from the selection and sent home).

Elite units in the military use an A&S process to choose the most qualified candidates for the unit. The A&S process we studied includes a screening phase to ensure candidates meet baseline requirements followed by the A&S phase to assess candidates abilities. Examples of performance tasks candidates may complete in military A&S are shown in Figure 2.1. A systems diagram of the specific A&S that we studied for this research is shown in Figure 2.2.

FIGURE 2.1: Examples of Performance Tasks at Military A&S

Left: Candidates performing push-ups in the water.
Right: Candidates carrying a log as a team.
Note: Photos taken from dvidshub and Military Times.

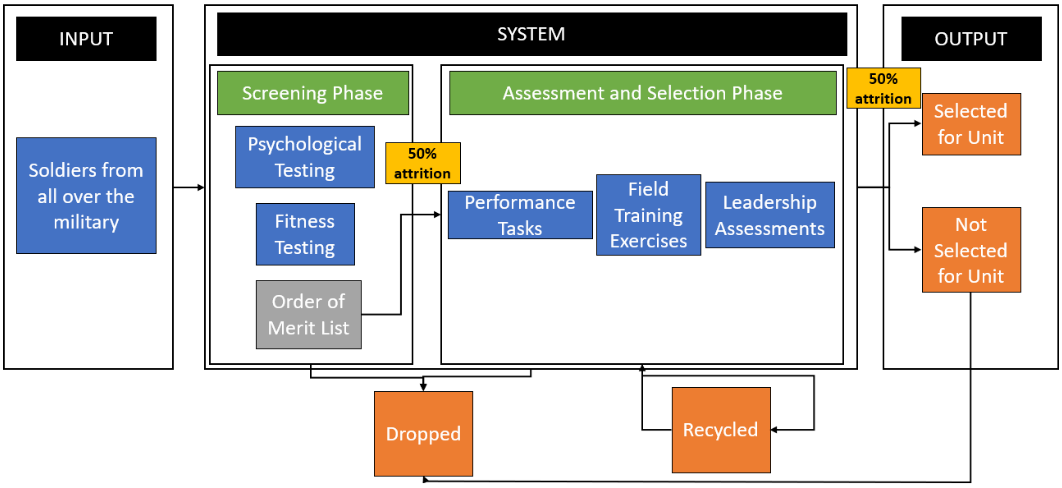


FIGURE 2.2: Military A&S Process

Soldiers from all over the military attend the A&S. In the screening phase, all candidates are required to take psychological tests, comprised of two personality tests and one intelligence quotient (IQ) test. All three tests are written, multiple choice Scantron tests. Candidates also take a fitness test in the screening phase that includes: two minutes of sit-ups and push-ups, a two-mile run, and pull-ups. There are objective minimum standards that candidates must score on the both the psychological tests and fitness tests to advance to the A&S phase. If candidates do not meet the minimum standards, they are dropped. For those candidates who meet the minimum standards, an order of merit list is created that ranks candidates from 1 to "n" based on their performance. The top "n" candidates advance to the A&S phase based on the space available in the specific class (usually about 50% of candidates advance).

The A&S phase is eight weeks long and includes many performance and leadership tasks used to assess candidates. During this phase candidates can be dropped or recycled (removed from the selection to restart at a later time). Candidates who complete the A&S phase are either selected for the unit or not selected for the unit (usually 50% of candidates are selected in the A&S phase; 25% overall of all candidates that arrive during the screening phase). If they are selected, they proceed onto further training. If candidates are not selected, they return to the military unit from which they came.

2.2 Existing Solutions

The existing solutions for screening and selecting candidates in the current A&S process are:

- **Screening:** In the screening phase, candidates are required to meet objective standards on all psychological and fitness tests. If a candidate does not meet the objective standard, that candidate is dropped in the screening phase prior to the A&S phase. Currently, the only two screening criteria used are the psychological and fitness tests.
- **Selection:**
  - **Objective Criteria.** In the A&S phase, candidates are required to meet objective standards on "critical events" (e.g., fitness tests, ruck marches,

and land navigation). If candidates fail a critical event, they are given one opportunity to retest. If candidates fail the retest, they are dropped from the course. Additionally, integrity violations (e.g., a candidate steals, lies, or cheats) are automatic drops.

– **Subjective Criteria.** Candidates complete peer evaluations on each other two times during the A&S phase. The lowest scoring candidates on the peer evaluations could be referred to a Commander's Board if instructors judge the candidate to be unfit for the unit. A Commander's Board requires the instructors submit their recommendations and evidence to have a candidate dropped to the A&S unit commander. At the Board, the commander, instructors, and candidate meet in person to discuss the evidence presented. At the conclusion of the Board, the commander decides if that candidate will be dropped or retained. Candidates can be dropped through a Commander's Board even if they meet all the objective standards.

  The A&S instructors also provide a more subjective evaluation of the candidates' abilities in the human dimension (e.g. leadership, resilience, integrity, initiative, etc). If an instructor thinks a candidate is not fit to be selected for the unit, he can recommend a Soldier be dropped to the commander through a Commander's Board.

2.3 Barriers to Solutions

The barriers to solutions for screening and selecting candidates in the current A&S process are:

- **Screening:**

  1. **Ethical.** Some screening criteria are objective and ethical, such as a minimum fitness standard. However, other screening criteria raise ethical questions. For example, females have historically not performed well at the A&S; however, there are ethical issues that arise if the A&S were to ban females. Even if banning females from the selection resulted in lower attrition rates, that criteria discriminates against females. Ethical considerations such as this example are necessary when deciding on the appropriate screening criteria using empirical data.
  2. **Data.** The objective candidate data is easy to retrieve, record and save in a database (e.g., fitness and psychological test scores). However, the human dimension of candidates is very challenging to measure, record and obtain on candidates (e.g., leadership, resilience, integrity, and initiative). The human dimension data is especially challenging to obtain prior to the screening phase of the A&S, as Soldiers are not required to do any assessments prior to attending the A&S. In the screening phase, there are limited time and resources to assess a candidate's human dimension attributes. This challenge makes robust screening of candidates difficult.

- **Selection:**

  1. **Limited Time and Resources.** The A&S has a required number of classes that it must complete each year. Likewise, for each class, the A&S fills as many candidates as possible given the resources available. The set class

length is eight weeks, the class sizes range between 100-165 candidates, and the number of instructors is between 12-15. Additionally, there are only specific facilities and training areas that can be used for the assessment. These constraints generally cannot be changed and make modifying methods to evaluate candidates difficult.

2. **Instructor to Student Ratio.** An essential part of the A&S is the instructors observing the candidates to evaluate the human dimension. With over 100 candidates and only 12-15 instructors, it is impossible for instructors to observe all the candidates effectively on all tasks.

2.4 Literature Review

2.4.1 Military Special Operations Forces

Military Special Operations Forces (SOF) exist to perform "special operations" on a global scale in order "to protect and advance U.S. policies and objectives" [4]. The missions are considered special operations because they require more specialized training and equipment, are higher risk, are more complex, and contain more sensitive information compared to conventional missions. Given the nature of these special operations, it is imperative that the right Soldiers are selected for the SOF units.

Each military branch in the Department of Defense (DoD) holds high standards and targets specific attributes in Soldiers who wish to join a SOF unit. The Navy SOF (Navy Seals) value traits of maturity, self-assurance, and self-confidence [5]. The Army SOF (Green Berets) uses eight attributes as their "benchmark" for selection: integrity, courage, perseverance, personal responsibility, professionalism, adaptability, team player, and capability [6]. The Army 75th Ranger Regiment (another Army SOF unit) emphasizes four competencies for candidates seeking to join: integrity/honesty, mental and physical fortitude, initiative, and resilience [7]. The Marine SOF seeks Soldiers who are mature, intelligent, mentally agile, determined, ethical, and physically fit [8]. The Air Force SOF espouses 13 "Critical Attributes": integrity, self-motivation, intelligence, self-discipline, perseverance, adaptability, maturity, judgment, selflessness, leadership, skilled, physical fitness, and family strength [9].

All military branches target similar attributes, most of which are in the human dimension such as integrity, maturity, and perseverance. However, physical fitness, an objective attribute, is also required given the grueling physical demands of the missions. To assess these attributes and select the best candidates, each branch has robust and lengthy A&S processes. To aid in the A&S, many data points are collected on candidates that can be used to more objectively assess the attributes. Our research focus deals with how to best use the data collected to holistically improve A&S process.

2.4.2 Previous Military A&S Data Analysis Research

Previous research using Military A&S data have focused on identifying predictors of candidate success. The studies have used physical [10, 11, 12, 13, 14, 15, 16, 17, 18, 19], psychological [10, 14, 16, 18, 19, 20, 21, 22], and demographic [10, 19, 23] predictors to model candidate success. The scope and amount of data varied widely for each study. The smallest scope was for a specific A&S class that was four-weeks long and 104 candidates [11]. The largest scope was for multiple A&S classes over a year time period with 821 candidates [10]. Most studies used a year of data that included multiple A&S classes with over 300 candidates. The majority of the studies only used descriptive statistics and statistical testing between groups for analysis. We found four previous studies that used logistic regression or classification tree modeling techniques [10, 16, 17, 23]. Of the studies using these models, accuracy ranged between 60-78%. We did not find any studies that used data science techniques such as feature engineering, feature selection, or multiple optimally tuned classification models.

The one study that used physical, psychological, and demographic predictors together found that the most predictive features were (in order) physical, demographic, and psychological [10]. Studies that used physical predictors all used fitness test scores as their primary measure. These fitness tests involved push-ups, sit-ups, a 2-mile run, pull-ups, and/or a ruck march. All the study results showed that as physical performance increases, candidate selection success increases [10, 11, 12, 13, 14, 15, 16, 17, 18, 19]. The studies that used psychological predictors collected the predictors through written tests. The studies used the Minnesota Multiphasic Personality Inventory (MMPI) personality test, Jackson Personality Inventory - Revised (JPI-R) personality test, IQ test, and custom "mental toughness" tests. Studies showed specific features from the MMPI, JPI-R, and mental toughness measures to be indicative of candidate success [10, 14, 16, 18, 19, 20, 21, 22]. The studies that used demographic predictors found a candidate's rank and military experience to be the most significant predictors of candidate success [10, 19, 23].

Our research uses similar predictors to the previous studies. We use a combination of administrative (including demographic), performance (primarily physical predictors), and psychological predictors (MMPI, JPI-R, and IQ tests) in our analysis. We will use the knowledge of previous studies as a baseline to compare our results with. We then hope to build on these findings by applying novel data science methods with the goal of extracting meaningful information and conclusions to improve the A&S.

Chapter 3

Methodology

3.1 Holistic Approach

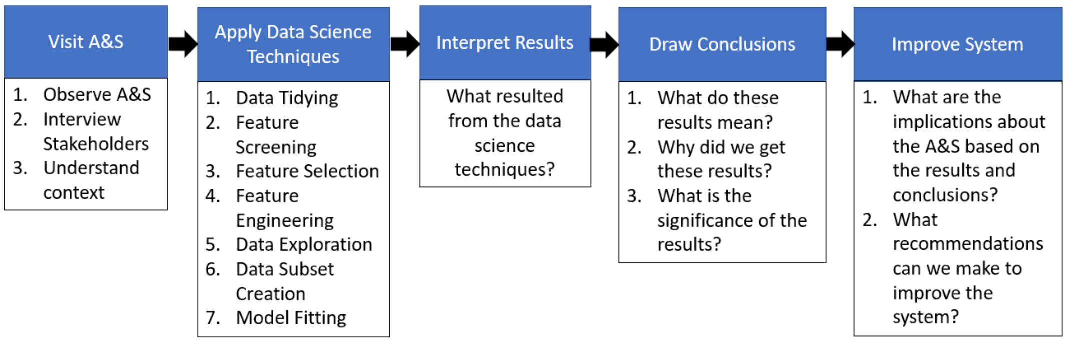


FIGURE 3.1: Research Methodology

Figure 3.1 shows the methodology we used to approach our research. We first visited the A&S prior to doing any data analysis. The goal of the visit was to learn more about the A&S system. During the visit we observed multiple A&S training events and interviewed key stakeholders including the unit leadership, psychologist, instructors, and data managers. The visit equipped us with key insights about the unique A&S nuances that we would not have otherwise understood from only looking at the data. This was arguably the most important step to provide analysis that was useful and actionable to the unit. With the contextual understanding from the visit, we were able to perform a much more informed analysis.

Next we applied data science techniques to the A&S data. The seven steps were used in sequence to holistically analyze the Military A&S data and draw insightful conclusions about the system. These steps were adopted from the general approaches described in [24, 25, 26]. Our approach incorporates multiple data science techniques to extract the most meaningful information out of the data and modeling techniques. Each of the seven steps are described in detail in chapter 4. The data science analysis is the focus of the presentation of work.

Lastly, we interpreted the results, drew conclusions, and made recommendations with the goal of improving the system. Not only did we focus on the results themselves, but also what the results and conclusions imply about the A&S holistically. Our goal was to add value to the A&S from the analysis conducted.

3.2 Data

The data used for this research was provided by the unit that conducts the A&S. The data set includes 11,885 candidates (observations/rows) with 89 features (variables/columns) that was collected over the time period from February 2016 - June 2021 (Note: Not all candidates have all data complete; this data set includes missing data).

There are three primary sources of data that are recorded on each candidate shown in Table 3.1. The response variable is binary: selected or not selected. The full list of features used for analysis are shown in Table 3.2 Complete Feature List.

TABLE 3.1: Primary Data Types

Feature Source	Total Features	# Continuous	# Categorical
Administrative	10	1	9
Performance	5	5	0
Psychological	74	74	0
Total	**89**	**80**	**9**

The administrative data is retrieved from a military human resource database. The performance data is collected during the A&S and manually entered into a digital database by the unit. The psychological data is collected using written, multiple choice Scantron tests and entered directly into a digital database by a Scantron scanner machine.

TABLE 3.2: Complete Feature List

Feature Name	Description	Type	Format/Example	Feature Source
pass	Indicates if a candidate was selected	categorical	Selected	Response Variable
mos	Candidate military occupational specialty (MOS) (i.e. job they do in military)	categorical	11B	Admin
rank	Candidate military rank	categorical	PFC	Admin
race	Candidate race	categorical	WHITE	Admin
arrival_month	Month candidate arrived for A&S	categorical	APRIL	Admin
tis_at_arrival	Time in service candidate at arrival to A&S	continuous	500	Admin
parents_together	States if a candidates parents are married	categorical	TRUE/FALSE	Admin
has_airborne	States if a candidate has attended airborne school	categorical	TRUE/FALSE	Admin
glasses	States if a candidate wears glasses	categorical	TRUE/FALSE	Admin
civilian_education_certification	Civilian education certificate candidate has	categorical	[HIGH SCHOOL DIPLOMA]	Admin
age_at_arrival	Candidate age at arrival	categorical	21	Admin
gt_score	Candidate score on military GT test	continuous	141	Performance
apft_1_pu	Number of push-ups candidate completed on fitness test	continuous	50	Performance
apft_1_su	Number of sit-ups candidate completed on fitness test	continuous	50	Performance
apft_1_run	2-mile run time of candidate on fitness test	period	870	Performance
apft_1_score	Total score on fitness test	continuous	240	Performance
s1	Complexity of thought (abstract vs. concrete)	continuous	10	Psychological
s2	Breadth of Interest (intellectual curiosity)	continuous	10	Psychological
s3	Innovation	continuous	10	Psychological
s4	Tolerance (openness to new beliefs)	continuous	10	Psychological
s5	Empathy	continuous	10	Psychological
s6	Anxiety	continuous	10	Psychological
s7	Cooperativeness	continuous	10	Psychological
s8	Sociability (introverted vs. extroverted)	continuous	10	Psychological
s9	Social Confidence	continuous	10	Psychological
s10	Energy Level	continuous	10	Psychological
s11	Social Astuteness (ability to read and persuade others)	continuous	10	Psychological
s12	Risk Taking (mostly monetary questions)	continuous	10	Psychological
s13	Organization	continuous	10	Psychological
s14	Traditional Values	continuous	10	Psychological
s15	Responsibility	continuous	10	Psychological

Table 3.2 continued from previous page

Feature Name	Description	Type	Format/Example	Feature Source
rci	response consistency index (validity scale)	continuous	0.5	Psychological
inf	infrequent responding scale (validity scale)	continuous	20	Psychological
mis	items left blank	continuous	20	Psychological
verbal_iq	Verbal IQ Standard Score	continuous	70	Psychological
performance_iq	Performance IQ Standard Score	continuous	70	Psychological
full_scale_iq	Full Scale IQ Standard Score	continuous	70	Psychological
vri_nr	Response consistency (higher = less consistent)	continuous	70	Psychological
tri_nr	Portion of answers marked True or False. Too high of one is another indicator of inconsistent responding	continuous	70	Psychological
fr	All of these are validity measures. More endorsed items will elevate these.	continuous	70	Psychological
fpr	All of these are validity measures. More endorsed items will elevate these.	continuous	70	Psychological
fs	All of these are validity measures. More endorsed items will elevate these.	continuous	70	Psychological
fb_sr	All of these are validity measures. More endorsed items will elevate these.	continuous	70	Psychological
rbs	All of these are validity measures. More endorsed items will elevate these.	continuous	70	Psychological
lr	Fake good scales	continuous	70	Psychological
kr	Fake good scales	continuous	70	Psychological
eid	Emotional Dysregulation: a combination of scales that tap into emotional dysregulation	continuous	70	Psychological
thd	Thought Dysfunction: a combination of scales that look at odd thinking styles	continuous	70	Psychological
bxd	Behavioral Dysfunction: a combination of scales that measure acting out behaviors	continuous	70	Psychological
r_cd	demoralization	continuous	70	Psychological
rc1	Somatic complaints	continuous	70	Psychological
rc2	Low positive emotions	continuous	70	Psychological
rc3	Cynicism	continuous	70	Psychological
rc4	Antisocial Behavior (antisocial in the sense of violating others' rights)	continuous	70	Psychological
rc6	Paranoia	continuous	70	Psychological
rc7	Dysfunctional negative emotions	continuous	70	Psychological
rc8	Abberant Thinking	continuous	70	Psychological
rc9	Hypomanic Activation (abnormally high energy)	continuous	70	Psychological
mls	malaise	continuous	70	Psychological
hpc	head pain complaints	continuous	70	Psychological
nuc	neurological complaints	continuous	70	Psychological
gic	gastrointestinal complaints	continuous	70	Psychological
sui	suicidal ideation	continuous	70	Psychological
hlp	Helplessness	continuous	70	Psychological
sfd	self doubt	continuous	70	Psychological
nfc	inefficacy	continuous	70	Psychological
cog	cognitive complaints	continuous	70	Psychological
stw	stress and worry	continuous	70	Psychological
axy	anxiety	continuous	70	Psychological
anp	anger proneness	continuous	70	Psychological
brf	behavioral restricting fears	continuous	70	Psychological
msf	multiple specific fears	continuous	70	Psychological
jcp	juvenile conduct problems	continuous	70	Psychological
sub	substance use problems	continuous	70	Psychological
agg	aggression	continuous	70	Psychological
act	activation: another high energy scale	continuous	70	Psychological
fml	family problems	continuous	70	Psychological
ipp	interpersonal passivity	continuous	70	Psychological
sav	social avoidance	continuous	70	Psychological
shy	shyness	continuous	70	Psychological
dsf	disaffiliativeness	continuous	70	Psychological
aes	artistic interestes	continuous	70	Psychological
mec	mechanical interests	continuous	70	Psychological
agg_rr	aggressive personality (goal-oriented aggression)	continuous	70	Psychological
psy_cr	odd personality (weird thinking and behaviors)	continuous	70	Psychological
dis_cr	discontraint (problems with impulse control)	continuous	70	Psychological
neg_er	personality with negative emotionality	continuous	70	Psychological
int_rr	introverted personality	continuous	70	Psychological
cannot_say	number of items left blank	continuous	20	Psychological

Table 3.2 continued from previous page

Feature Name	Description	Type	Format/Example	Feature Source
pct_true	% marked true	continuous	30	Psychological

3.3 Models

We used 10 binary classification models to compare performance along with pros and cons of each model. For the scope of this research, we will provide a brief summary of how each model works, but not an in-depth mathematical explanation of the models.

The models we used were:

1. Penalized Logistic Regression (using Elastic Net)
2. Linear Discriminant Analysis (LDA)
3. Quadratic Discriminant Analysis (QDA)
4. Support Vector Machines (SVM)
5. Classification Tree (from Classification and Regression Trees - CART)
6. K-Nearest Neighbor (KNN)
7. Random Forest (Ensemble Model)
8. xgboost (Ensemble Model)
9. Stack - Logistic Regression Aggregation (Ensemble Model)
10. Stack - Random Forest Aggregation (Ensemble Model)

3.3.1 Penalized Logistic Regression (using elastic net)

Penalized Logistic regression is used to predict the posterior probability, $\boldsymbol{p}$, of a categorical feature being a certain class using the logistic function. An example of what the function looks like for number of hours studying for a test (input feature) and probability of passing the test (response feature) is shown in Figure 3.2. For our model, the predicted probability, $\hat{\boldsymbol{p}}$, represents the **probability of a candidate being selected** or $\hat{\boldsymbol{p}} = Pr(\text{Candidate Selected} \mid \text{Data}) \iff Pr(Y = 1 \mid X = x)$.

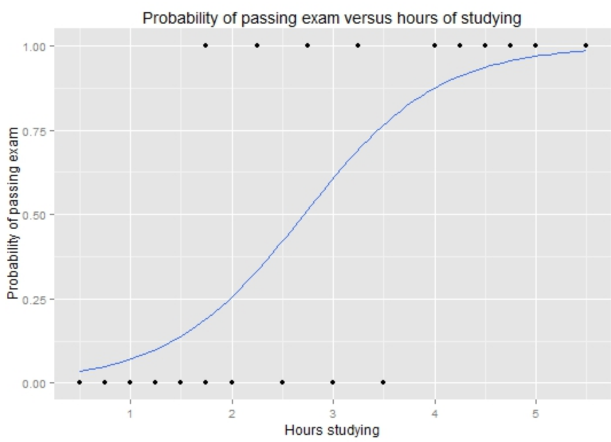


FIGURE 3.2: Logistic Regression Example

The added benefit to using Penalized Logistic Regression compared to baseline Logistic Regression is that the model incorporates embedded feature selection by penalizing the coefficients (β). The specific variant of penalized logistic regression we used was the the elastic net model shown in Equation 3.1. The model seeks to optimize the loss ($\ell(\beta)$) plus the penalty (λ) applied to coefficients (β) with the goal to decrease coefficients of features that are less important. The elastic net allows for a distinct advantage in choosing the optimal coefficient penalty by weighting the lasso and ridge penalties using the alpha (α) tuning parameter. We used the "glmnet" R package to implement this model [27]. Specific details about the model tuning parameters are below.

The creators of the elastic net penalized logistic regression model note a key strength in using the model:

> It is known that the ridge penalty shrinks the coefficients of correlated predictors towards each other while the lasso tends to pick one of them and discard the others. The elastic net penalty mixes these two: if predictors are correlated in groups, an $\alpha = 0.5$ tends to either select or leave out the entire group of features. This is a higher level parameter, and users might pick a value upfront or experiment with a few different values. One use of α is for numerical stability; for example, the elastic net with $\alpha = 1 - \epsilon$ for some small $\epsilon > 0$ performs much like the lasso, but removes any degeneracies and wild behavior caused by extreme correlations. [28]

$$\underset{\beta}{\text{argmax}}\ \ell(\beta) + \lambda \sum_{j=1}^{p} \left[(1-\alpha)\frac{|\beta_j|^2}{2} + \alpha|\beta_j| \right] \tag{3.1}$$

The model tuning parameters are:

- α = The weight assigned to each penalty (value between 0 and 1).
  - α = 0: Ridge penalty
  - α = 1: Lasso penalty
- λ = The penalty applied to the model coefficients.
  - λ = 0: No penalty applied to model coefficients. Same outcome as maximizing the negative log loss function ($\ell(\beta)$).
  - λ > 0: The larger the penalty, the smaller the model coefficients become (i.e. coefficients shrink closer to zero). If all model coefficients = 0, model becomes intercept only (extreme case of large penalty).

3.3.2 Linear Discriminant Analysis (LDA)

LDA is a classification model that uses linear combinations of features to separate (or discriminate) between two or more classes (categories). The LDA model results in a **linear decision boundary** to choose the predicted class, as shown on the **left** in Figure 3.3. There are no tuning parameters for this model. We used the "MASS" R package to implement this model [29].

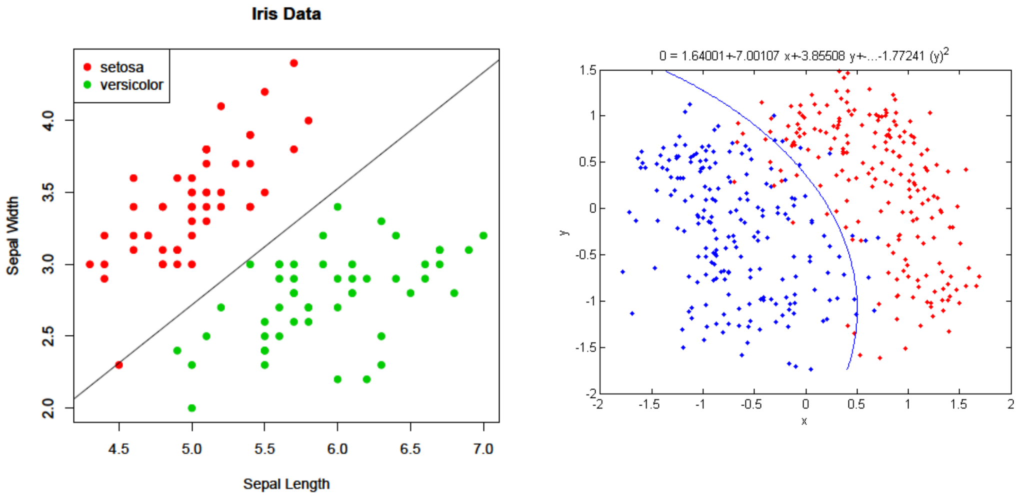


FIGURE 3.3: LDA and QDA Decision Boundaries

Left: LDA.
Right: QDA.

The LDA method makes the following assumptions:

1. **Data Normality**. The model assumes that all the data is normally distributed (or Gaussian; i.e. bell curve distribution). This assumption is *generally* true for the data (see subsection 4.5.1).

   Note: LDA models can only use continuous features based on this assumption. Thus, all data subsets used for this model only used continuous features.

2. **Data Independence**. The model assumes that the data is independent from one another. We assume this assumption is generally true, as each candidate is a different person with unique circumstances.

3. **Equal Class Covariance**. The model assumes that feature covariances are the same for both classes: selected and not selected. We compared the class covariances and discovered this assumption is true for most features, but not all. We expect it may have some impact on model performance, but that performance will still be reliable.

3.3.3 Quadratic Discriminant Analysis (QDA)

This method is very similar to LDA. QDA seeks to discriminate using a decision boundary like the LDA model does and has the same assumptions with the exception of the **Equal Class Covariance**. This assumption is relaxed for QDA and the classes can have different covariances. This allows for more flexible, non-linear decision boundaries as shown on the **right** in Figure 3.3. There are no tuning parameters for this model. We used the "MASS" R package to implement this model [29].

3.3.4 Support Vector Machines (SVM)

SVM seeks to find the optimal decision boundary between classes by maximizing the "margin" or distance between observations from each class. SVM can use a linear (as shown in Figure 3.4) or a non-linear decision boundary. There is also a "kernel trick" that can be applied to the SVM model resulting in a transformed, non-linear feature space. In the transformed feature space, the SVM classifier is still a linear hyperplane, but in the original feature space it may be non-linear. We chose to use the SVM model with a linear kernel. We used the "kernlab" R package to implement the model [30]. For our analysis, we only used the SVM model with a linear kernel.

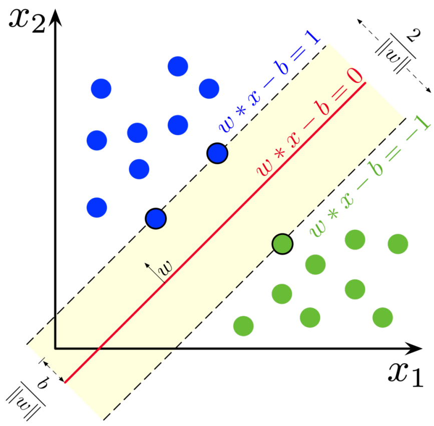


FIGURE 3.4: SVM visualization

Note: Figure taken from Wikipedia

The model tuning parameters for the SVM linear kernel model are (taken from the "kernlab" R package):

- **cost** = cost of constraints violation (default: 1) this is the 'C'-constant of the regularization term in the Lagrange formulation.

3.3.5 Classification Tree (from Classification and Regression Trees - CART)

The classification tree model conducts binary splits of input features (branches) that maximize a specified metric to make predictions about the response feature (leaves). An example of this using our data set is: splitting the number of push-ups candidates perform such that candidates with less than the split value are more likely to be not selected and candidates with more than the split value are more likely to be selected. Splits like this are performed with multiple features to form a final tree with multiple, subsequent splits. An example of a regression tree (continuous response variable) that estimates the probability of kyphosis after spinal surgery, given patient age and vertebra started on is shown in Figure 3.5.

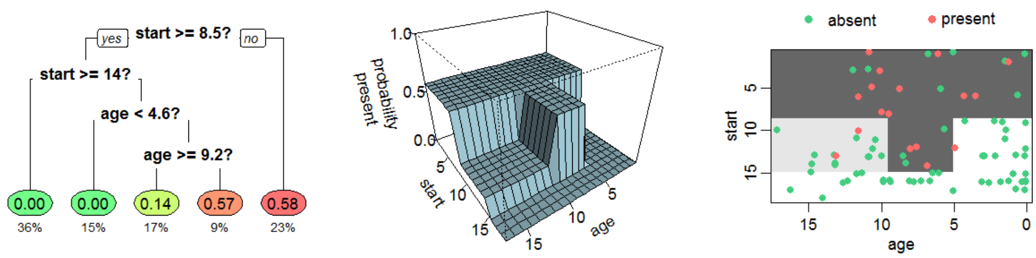


FIGURE 3.5: Regression Tree example

Left: The colored leaves show the probability of kyphosis after spinal surgery (inside leaf), and percentage of patients in each category under the leaf.
Middle: The tree as a 3d perspective plot, showing probability of kyphosis given surgery start and patient age.
Right: Top view of the middle plot. The probability of kyphosis after surgery is higher in the darker areas.
Note: Figure and explanation taken from Wikipedia.

Tree-based models are very flexible with no *formal* assumptions. For example, normality of data does not matter. Likewise, it can handle both continuous and categorical variables without any standardizing or encoding, respectively. These models are easy to interpret as one can follow the tree "rules" to make decisions. However, the models can be sensitive to changes in data and over fitting, especially with limited data. For example, if the test and train data is modified slightly, the classification tree could produce very different results. A classification tree (what we used) is the same as the regression tree example in Figure 3.5, except it has a categorical response feature instead of continuous. We used the "rpart" R package to implement this model [31].

The model tuning parameters are (taken from the "rpart" R package):

- **minsplit** = The minimum number of observations that must exist in a node in order for a split to be attempted.
- **minbucket** = The minimum number of observations in any terminal leaf node. If only one of minbucket or minsplit is specified, the code either sets minsplit to minbucket*3 or minbucket to minsplit/3, as appropriate.
- **complexity parameter (cp)** = Any split that does not decrease the overall lack of fit by a factor of cp is not attempted. For instance, with anova splitting, this means that the overall R-squared must increase by cp at each step. The main role of this parameter is to save computing time by pruning off splits that are obviously not worthwhile. Essentially, the user informs the program that any split which does not improve the fit by cp will likely be pruned off by cross-validation, and that hence the program need not pursue it.
- **maxcompete** = The number of competitor splits retained in the output. It is useful to know not just which split was chosen, but which variable came in second, third, etc.
- **maxsurrogate** = the number of surrogate splits retained in the output. If this is set to zero the compute time will be reduced, since approximately half of the computational time (other than setup) is used in the search for surrogate splits.
- **usesurrogate** = How to use surrogates in the splitting process. 0 means display only; an observation with a missing value for the primary split rule is not sent further down the tree. 1 means use surrogates, in order, to split subjects missing the primary variable; if all surrogates are missing the observation is not split. For value 2, if all surrogates are missing, then send the observation in the majority direction. A value of 0 corresponds to the action of tree, and 2 to the recommendations of Breiman et.al (1984).
- **surrogatestyle** = Controls the selection of a best surrogate. If set to 0 (default) the program uses the total number of correct classification for a potential surrogate variable, if set to 1 it uses the percent correct, calculated over the non-missing values of the surrogate. The first option more severely penalizes covariates with a large number of missing values.
- **maxdepth** = Set the maximum depth of any node of the final tree, with the root node counted as depth 0.
- **xval** = Number of cross-validations.

3.3.6 K-Nearest Neighbor (KNN)

The K-Nearest Neighbor model is a non-parametric model that predicts classifications based on how many "k-neighbors", or training observations, are closest to the test observation. The model estimates the conditional probability (soft classification) of the test observation being a specific class as $\frac{\text{points in class j}}{\text{k neighbors}}$. To predict a class (hard classification), the model selects that class with the most neighbors or highest conditional probability [32]. We used the "FNN" R package to implement this model [33]. Figure 3.6 shows an example of how a specific test observation would be classified using KNN as k changes.

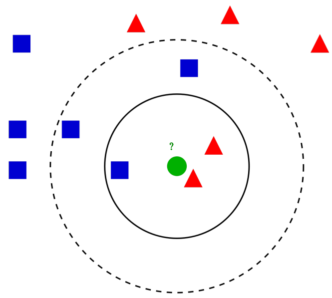


FIGURE 3.6: KNN visualization

The test observation (green dot) should be classified either to blue squares or to red triangles. If k = 3 (solid line circle) it is assigned to the red triangles because there are 2 triangles and only 1 square inside the inner circle. If k = 5 (dashed line circle) it is assigned to the blue squares (3 squares vs. 2 triangles inside the outer circle). Note: Figure and explanation taken from Wikipedia

The model tuning parameter is: **k** = number of neighbors.

3.3.7 Ensemble Models

Ensemble models are models that use multiple "base" models and combine their strengths to form an aggregate or "ensemble" final model. There are two tasks in ensemble learning: 1) develop a variety of base models trained on the training data and 2) combine the models to form a holistic model to evaluate test data with one outcome prediction [34].

The three types of ensemble methods are:

1. Bootstrap Aggregating (Bagging)
2. Boosting
3. Stacking

1. Bootstrap Aggregating (Bagging)

The bootstrap is a powerful resampling technique used for a variety of purposes. When applied to bagging, the bootstrap can help reduce the variance of machine learning models. More specifically, when bagging is applied to classification trees, it is often very effective at reducing the variance as classification trees can have a lot of variance depending on the variability of test/train data splits [32]. For the bagging ensemble technique, we used the **Random Forest** model with the "randomForest" R package implementation [35].

The random forest model is a method of bootstrap aggregating (creating multiple classification trees using bootstrap samples of same data and aggregating the results) with the goal of building de-correlated trees (created in parallel) by only using a random sub-sample of features at each tree split. Based on the results of the random forest, the classification prediction is made by the aggregated output. This can be in the form of majority vote (hard classification, i.e. selected or not selected) or aggregated probabilities (soft classification, i.e. probability between [0,1]) for each class. A visualization of how the random forest model works is shown in Figure 3.7.

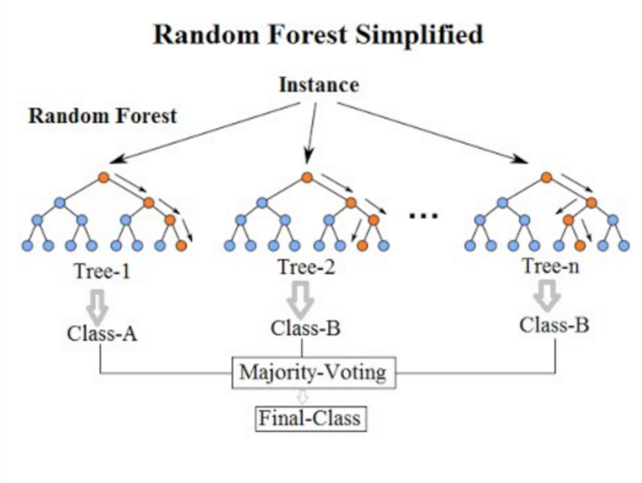


FIGURE 3.7: Random forest visualization

Note: Figure taken from Wikipedia

The model tuning parameters are (taken from the "randomForest" R package):

- **ntree.** Number of trees to grow. This should not be set to too small a number, to ensure that every input row gets predicted at least a few times.
- **mtry.** Number of variables randomly sampled as candidates at each split. Note that the default values are different for classification (sqrt(p) where p is number of variables in x) and regression (p/3).
- **nodesize.** Minimum size of terminal nodes. Setting this number larger causes smaller trees to be grown (and thus take less time). Note that the default values are different for classification (1) and regression (5).
- **maxnodes.** Maximum number of terminal nodes trees in the forest can have. If not given, trees are grown to the maximum possible (subject to limits by nodesize). If set larger than maximum possible, a warning is issued.

2. Boosting

The main idea behind boosting ensemble models is that many "weak learners" (or base models) can be combined to form a "strong" final model while also reducing bias. Boosting is a sequential learning technique where each successive model learns from the previous model performance. As a result, using different weights or loss functions allow model performance to be continually improved with each iteration.

We used **Extreme Gradient Boosting (xgboost)** for this ensemble model. xgboost uses parallel shallow depth trees as weak learners to fit models at each step. The model optimizes performance by minimizing a loss function from each tree using gradient descent. The loss function includes a penalty term that seeks to minimize over fitting. When the algorithm is complete, an aggregate ensemble is created from all trees to predict the optimal classifications. Figure 3.8 shows a visualization of xgboost. We used the "xgboost" R package implementation [36].

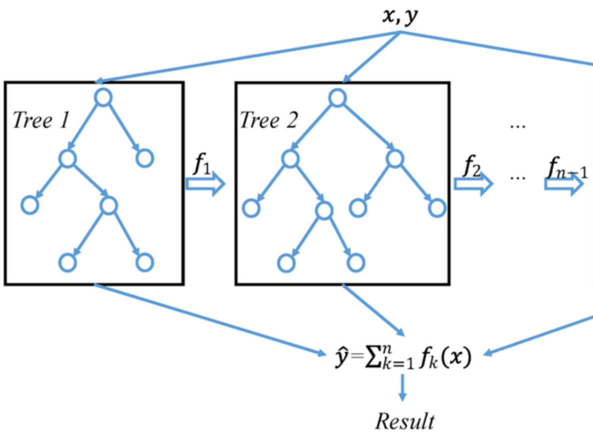


FIGURE 3.8: xgboost visualization

Note: Figure taken from ResearchGate

The model tuning parameters are (taken from "xgboost" package documentation):

- **nrounds.** max number of boosting iterations.
- **max_depth.** maximum depth of a tree. Default: 6
- **eta.** control the learning rate: scale the contribution of each tree by a factor of 0 < eta < 1 when it is added to the current approximation. Used to prevent overfitting by making the boosting process more conservative. Lower value for eta implies larger value for nrounds: low eta value means model more robust to overfitting but slower to compute. Default: 0.3
- **gamma.** minimum loss reduction required to make a further partition on a leaf node of the tree. The larger, the more conservative the algorithm will be.
- **colsample_bytree.** subsample ratio of columns when constructing each tree. Default: 1
- **min_child_weight.** minimum sum of instance weight (hessian) needed in a child. If the tree partition step results in a leaf node with the sum of instance weight less than min_child_weight, then the building process will give up further partitioning. In linear regression mode, this simply corresponds to minimum number of instances needed to be in each node. The larger, the more conservative the algorithm will be. Default: 1
- **subsample.** subsample ratio of the training instance. Setting it to 0.5 means that xgboost randomly collected half of the data instances to grow trees and this will prevent overfitting. It makes computation shorter (because less data to analyze). It is advised to use this parameter with eta and increase nrounds. Default: 1

3. Stacking
Stacking is an ensemble method that uses multiple base learners that all individually are fit to training data. After each model is fit, the models are combined using some kind of "generalizer" model that uses all the base learner predictions to make an

aggregate final prediction. Figure 3.9 shows a visualization of stacking. We used the "caretEnsemble" R package implementation of the model [37]. There are no tuning parameters for this model. Our stack ensemble model used LDA, SVM, penalized logistic regression, CART, KNN, Random Forest, and xgboost as base models with both logistic regression and random forest generalizer functions.

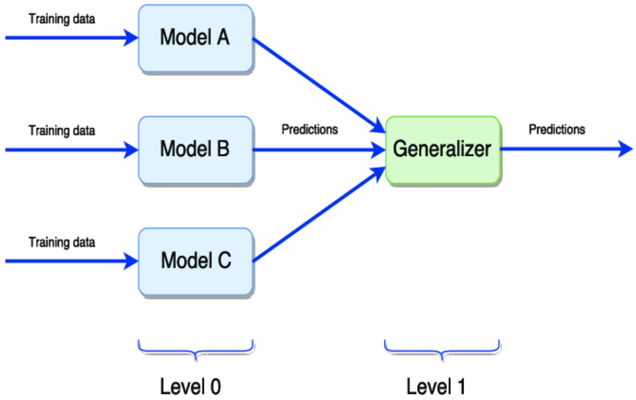


FIGURE 3.9: Stacking visualization

Note: Figure taken from ResearchGate

3.4 Evaluation Methods

Model performance will be evaluated using metrics from a confusion matrix. An example confusion matrix is shown in Table 3.3.

TABLE 3.3: Example Confusion Matrix

		Observed (True Outcome)		
		Selected	**Not Selected**	**Total**
Predicted (Model Outcome)	**Selected**	True Positive (TP)	False Positive (FP)	TP + FP (Predicted Selected)
	Not Selected	False Negative (FN)	True Negative (TN)	FN + TN (Predicted Not Selected)
	Total	TP + FN (Observed Selected)	FP + TN (Observed Not Selected)	N (Total Observations)

The diagonals in the confusion matrix ("True Positive" (TP) and "True Negative" (TN)) represent those candidates correctly classified by the model for both Selected (TP) and Not Selected (TN). The off-diagonals represent those candidates that were not classified incorrectly by the model for both Selected ("False Positive" (FP)) and Not Selected ("False Negative" (FN)). More specifically:

- **TP**: Candidate predicted selected and observed selected.

- **FP**: Candidate predicted selected and observed not selected.
- **TN**: Candidate predicted not selected and observed not selected.
- **FN**: Candidate predicted not selected and observed selected.

Table 3.4 shows the metrics that we used to evaluate models from a confusion matrix and their interpretation.

TABLE 3.4: Confusion Matrix Metrics

Metric	Interpretation	Equation
Accuracy	For overall classification (both Selection and Non-Selection), how accurate was the model? Total of TP and TN divided by Total Observations.	$\frac{TP+TN}{N}$
Precision Positive Predicted Value (PPV)	Given the model predicts Selected, how accurately does the model predict Predicted Selected? The number of TP predictions divided by total Predicted Selected.	$\frac{TP}{TP+FP}$
Sensitivity Recall True Positive Rate (TPR)	Given the model predicts Selected, how accurately does the model predict Observed Selected? The number of TP predictions divided by total Observed Selected.	$\frac{TP}{TP+FN}$
Specificity True Negative Rate (TNR)	Given the model predicts Not Selected, how accurately does the model predict Observed Not Selected? The number of TN predictions divided by total Observed Not Selected.	$\frac{TN}{TN+FP}$
Negative Predicted Value (NPV)	Given the model predicts Not Selected, how accurately does the model predict Predicted Not Selected? The number of TN predictions divided by total Predicted Not Selected.	$\frac{TN}{TN+FN}$
Receiver Operator Characteristic (ROC) Score	A plot that shows true positive rate (TPR) and false positive rate (FPR) as a function of the classification threshold. The closer the curve is to TPR=1, the better the performance. Figure 3.10 shows an example ROC curve.	Maximum score on ROC Curve
ROC Area Under Curve (ROC AUC)	Calculates the total area under the ROC curve, creating an average performance metric for the whole ROC curve.	Area under ROC curve
F1 Score	Combines precision and recall into one metric by taking the mean of the two.	$\frac{2(TP)}{2(TP)+FP+FN}$
Kappa	Uses TP, TN, FN, and FP with the goal to compare "expected" to "observed" accuracy. Goal is to account for the accuracy that happens "by chance" vs. what the model actually predicted.	$\frac{2[(TPxTN)+(FNxFP)]}{[(TP+FP)x(FP+TN)]+[(TP+FN)x(FN+TN)]}$

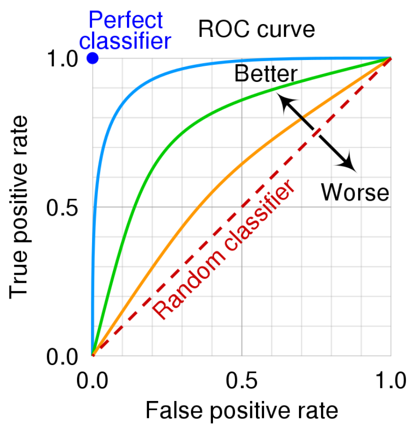


FIGURE 3.10: Receiver Operating Characteristic (ROC) Curve Example

Note: Figure taken from Wikipedia

The data set we are using has a large class imbalance, meaning one class has far more observations than the other. Specifically, selected candidates only represent 25% of the data, while not selected candidates represent 75%. The class imbalance varies given the specific data subset, but in general, the selected candidate class is much less than the not selected class. As a result, it is necessary to select an evaluation metrics that account for this class imbalance.

Likewise, choosing evaluation metrics to meet the goals and objectives of the stakeholder is important. For example, if the military unit only prioritized correctly classifying selected candidates, then we would tune models by maximizing precision and sensitivity. Likewise, if the military only prioritized correctly classifying candidates not selected, then we would tune models by maximizing specificity and NPV. Often, the goal is to optimize models using an evaluation metric that can balance a variety of metrics and best classify both classes. A metric that optimizes classifying both selection and non-selection is our goal for this research.

The accuracy metric can be misleading with a class imbalance. For example, our data set contains 75% of candidates who were not selected and 25% of candidates who were selected. If all of the not selected candidates were correctly classified, and none of the selected candidates were correctly classified, the accuracy would still be 75%. However, the overall accuracy is misleading, because while the model is perfect at classifying non-selection, it is useless at classifying selection.

Considering the class imbalance and stakeholder input, we sought to choose a "well-balanced" metric to use for optimizing our models. Due to the class imbalance, accuracy is not a good choice as it will be biased. ROC and AUC are other common metrics; however, these too are sensitive to class imbalances and can be misleading. F1 Score is a better metric for our data, but there are still some drawbacks. It is effective at providing a metric specifically for the TP classification, but does not account for the TN classification. For our objectives, considering the TN classification is important because there is a cost associated with not screening those candidates out.

As a result, we chose to use the **kappa** evaluation metric. The kappa metric is holistic in that it incorporates TP, TN, FP, and FN scores. Likewise, it inherently accounts for class imbalances based on its calculations using all four metrics to normalize the score [38]. The kappa metric can range from -1 to 1 and can be understood similarly to a correlation statistic. Kappa can be interpreted as the accuracy that happens "by chance" vs. what the model actually predicted. A negative score indicates that the model prediction is worse than random chance, a score of 0 indicates that the model has no agreement with the data, and a positive score indicates some level of agreement between the predicted outcome and the true outcome. The larger the kappa score, the greater the agreement (or disagreement).

Chapter 4

Presentation of Work

4.1 Data Tidying

Tidying the data involved the the following steps:

1. Merge Separate Data
2. Address Missing Data
3. Fix Incorrect Data

4.1.1 Merge Separate Data

We received three data sets (administrative, performance, and psychological) that had different features on each candidate. Each data set contained a common "key" to combine a candidate's data from all three data sets into one holistic data set. This was done using functions in R that join or merge data using a common key. After merging all three data sets, the data was composed of **369 features** and **11,885 observations** (this included missing data).

4.1.2 Address Missing Data

Missing data proved to be a challenge as 55% of the observations from the data set were missing. A helpful visualization of this is shown in Figure 4.1, where the observation number (candidate) is on the x-axis and the features on the y-axis.

While each individual feature on the y-axis cannot be easily read, the overall trends are evident. The top 3/4 of the plot contains features from the three psychological tests. That is why there are sets of features that have the same shape of missing data: a candidate either took the entire personality test or not at all. The bottom 1/4 of the chart represents administrative and performance features, which, in general had more complete fields. It is also easy to visualize how the missing data relates across all features (e.g. all candidates in a certain range missing the psychological features, but may have the performance and psychological features). Candidates having all the features complete with no missing data are represented by a black vertical line.

The reason most data is missing is due to data collection methods and requirements changing over the five-year period. For example, one of the psychological tests was implemented in 2018; therefore, the candidates that attended between 2016-2018 are all missing those features.

To deal with the missing data, we created two data sets.

1. **Non-imputed Data Set.**

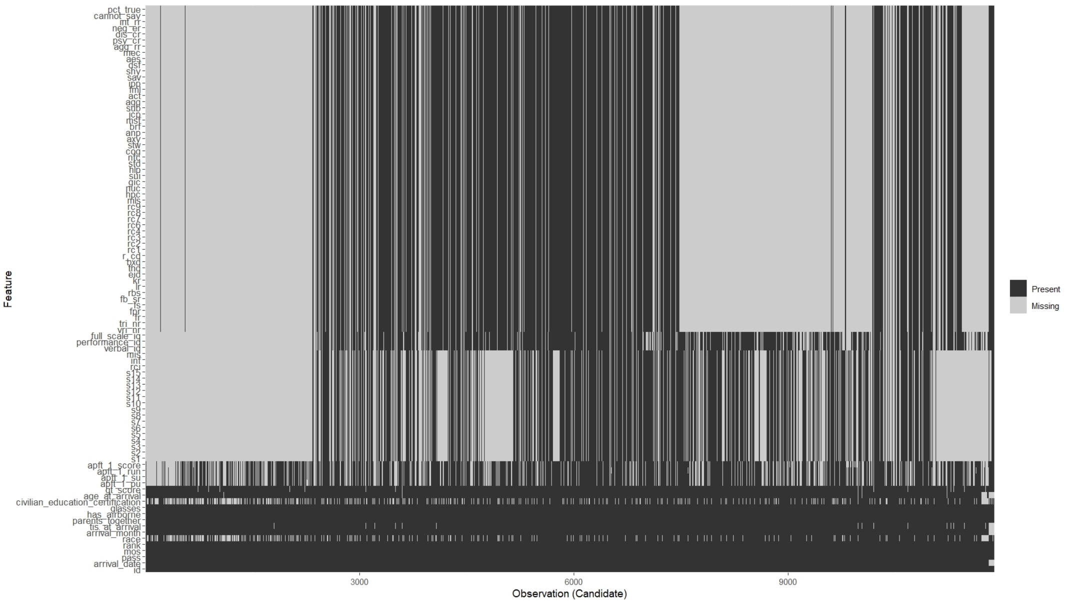


FIGURE 4.1: Missing data visualization

Note 1: The chart is in chronological order from 2016-2021.
Note 2: This data is the final data set after feature screening.

- **Categorical Features.** We created an "UNKNOWN" category for all missing values of categorical variables. This allowed us to still retain some information about the categories without removing the missing data.
- **Continuous Features.** We removed all missing observations.

  Note: After doing this, there were **3,095 / 11,885** observations that were complete with all features.

2. **Imputed Data Set.**

   - We imputed all missing data (both continuous and categorical) using a bagged tree implementation. We used the "step_impute_bag" method in the caret R package [25]. Using this method, features are imputed individually using all other features (except the response feature: selection) as the inputs using bootstrap aggregated trees. We chose this method because it can use input predictors that are missing data and can impute both categorical and continuous features. Likewise, the method consistently outperforms methods like mean imputation and better maintains the feature distributions and relationships [26].

For the final results, we only used the **non-imputed** data to ensure a non-biased analysis. However, we did compare performance of the non-imputed data to the imputed data to understand what impact data imputation had on our results.

4.1.3 Fix Incorrect Data

After examining the data, we discovered errors (less than 100 observations) that were likely due to an inaccurate manual entry (e.g. fitness score outside of 0-100 range). To ensure only accurate data is used, we went through each feature and removed the data that was inaccurate. Likewise, there were some categorical features that had "typos" (e.g., rank of "PVTT" instead of "PVT") or redundant categories (e.g. education level of "Bachelor's" and "Bachelor's Degree" both meaning the same thing). To address this, we standardized the categorical features by fixing all the "typos" and redundancies. After doing this, we assume the data is accurate.

4.2 Feature Screening

Feature screening is defined as removing features that are not informative prior to doing any analysis. This is valuable in the modeling process because screening removes features that add noise to the model. The goal in removing them is to yield more representative, accurate and interpretative results. We used multiple feature screening techniques described below.

4.2.1 All Features

We applied the following methods to both categorical and continuous features. The number next to each method is the number of features we screened using that specific method.

1. **Incomplete Features (77):** These were features that had a large amount of missing data (greater than 75% of the data). Most of the data that was this incomplete were performance scores collected during the A&S phase. Due to the high attrition rates, a small amount of candidates have all the scores collected in each event. Likewise, these features were outside the scope as our analysis uses features common to all candidates in the screening phase.

2. **Character Features (20):** These features were free text entry (e.g. reason why a candidate quit). While these features could be informative by transforming the information, it was outside the scope for this research.

3. **Response Features (18):** Response features that are only known upon a candidate completing the assessment (e.g. reason a candidate dropped or the candidate's class number) were removed as this is data leakage.

4. **Redundant Features (12):** Features that captured the same information (e.g. civilian education certification, civilian education degree, and number of years of civilian education all contained redundant information) were consolidated to more concisely represent the information. For cases like these, we reduced the number of features to what was necessary capture the redundant information.

4.2.2 Continuous Features

Specifically for continuous features, we screened for **highly correlated features (131).** Features that had a correlation of more than 90% were removed due to collinearity, as these features contain the same information and can lead to unstable model results. In our data set, some examples included raw fitness scores and point fitness scores. For example, doing 80 push ups is equal to 100 points; the values are a function of each other and represent the same information. Likewise, most of the psychological features were highly correlated. One of the psychological tests included 15 main attributes and 128 other features that were percentiles scores or combinations of those 15 main attributes, all of which were highly correlated. Figure 4.2 shows an example of one of the personality tests where the 15 main attributes (s1-s15) are perfectly correlated (correlation=1) to the percentile scores (mts1-mts15).

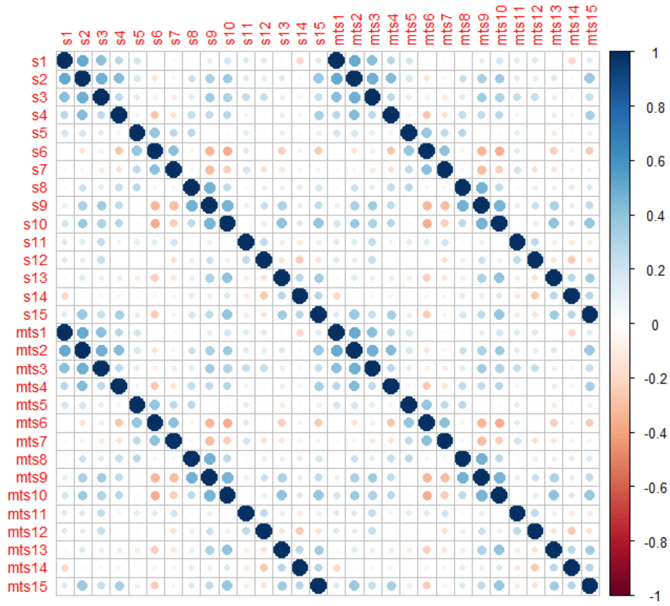


FIGURE 4.2: Example of highly correlated features

4.2.3 Categorical Features

Specifically for categorical features, we used **contingency table analysis (22).** For each categorical feature, we created a contingency table to see the relationship between the input feature and the response feature (selected or not selected). Table 4.1 shows an example of screening the "pov", or "privately owned vehicle" column which indicates whether a candidate drove themselves to the assessment or not. We screened categorical features for the following reasons:

1. **Many categories.** Features that 10 or more categories that couldn't easily be "lumped" (see subsection 4.3.2) required screening, because large dimensionality makes modeling results unreasonable. For example, "city_of_birth" had over 3,000 categories with most having 20-100 observations. While there are feature engineering techniques to handle categorical features with many categories, it was outside the scope of this research.

2. **Similar Ratio to Response.** Features where all the categories that had similar ratio to the selection rate required screening. This is comparable to having almost no correlation to the response. There is little information in these features that is valuable to prediction.

3. **Most Observations in a Single Category.** Features that had almost all observations in a single category required screening. This is comparable to having a low variance. For example, less than 1% of candidates were female. Keeping this feature would add little information to the modeling process and therefore was dropped.

TABLE 4.1: Example of categorical feature screening using contingency tables

pov	Selected	Not Selected	Total	Percent Selected	Percent of Data
FALSE	2,874	8,727	11,601	24.8	97.6
TRUE	65	219	284	22.3	2.4
Total	2,939	8,946	11,885	24.7	100

Remove "pov" feature because:
1. "FALSE" and "TRUE" categories have similar ratio to response (within 2%) [Blue Highlight]
2. "FALSE" category has almost all candidates in category (97%) [Gray Highlight]

After screening all the features, we removed 280 features total. These techniques **reduced the total features from 369 to 89** that were used for analysis.

4.3 Feature Engineering

The goal of feature engineering is to modify existing features or create new features to better represent the information contained in the data. Ideally, the changes made should reveal additional insight about the data and/or improve model performance. We applied the following feature engineering methods.

4.3.1 Continuous Features

1. **Log transformation.** We applied a log transformation of all the continuous features. With skewed data, this can help make the data more normal (i.e., bell curve shaped distribution) and improve model performance.

2. **Quantile (Binned) Transformation.** For each continuous feature, we divided the values into four equal quantiles or bins. Each bin contained approximately 25% of the data in each feature. Although with this transformation information is lost by aggregating values, it can be effective at returning more interpretable results.

3. **Principle Component Analysis (PCA) Transformation.** The goal of the PCA transformation is to reduce the data from a high-dimensional space into a low-dimensional space, such that the low-dimensional representation retains some meaningful properties of the original data. The method involves orthogonal projections of the data onto a specified number of principle components (PCs). We applied PCA to all the continuous features.

   Figure 4.3 shows an example of a biplot used to show the relationship between number of PCs and amount of variance in the data accounted for. As evident in the biplot, the vast majority of the variance in the data is contained in the first three PCs (where the "elbow" is in the line).

   PCA also allows us to visualize many features projected on a two-dimensional space. For example, Figure 4.4 shows two scatter plots of a personality test with 53 features projected onto the first two PCs. Sub figure (a) is the data projected without any other transformations. It appears that there is not a large difference between selected and non selected candidates (one large grouping of blue and red dots mixed together). Sub figure (b) shows the same psychological test with the binned transformation prior to projecting it onto two

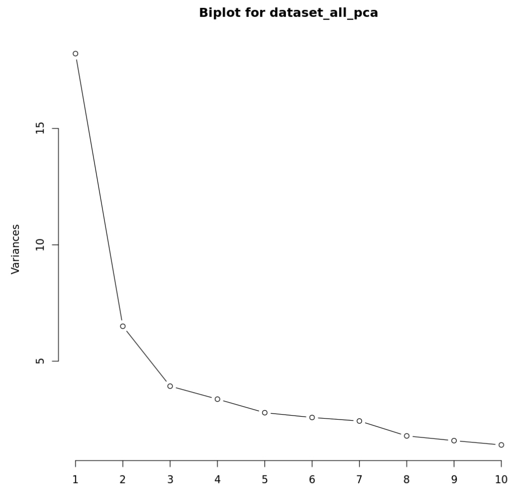
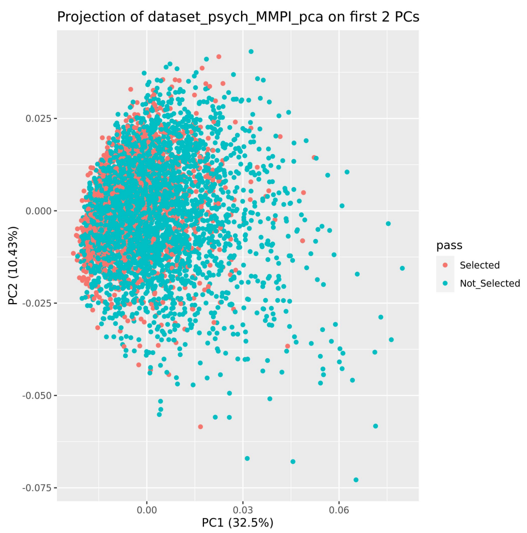
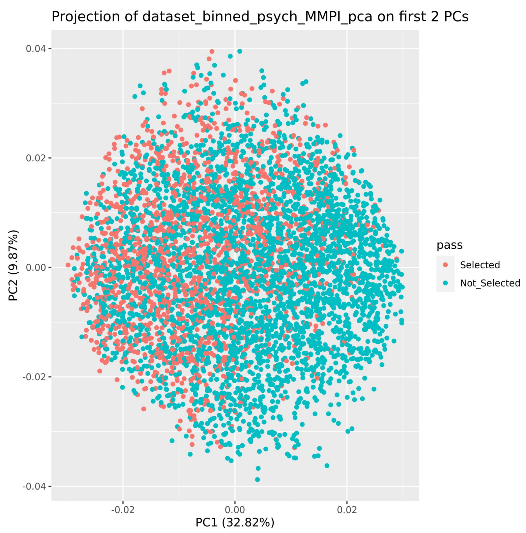


FIGURE 4.3: Biplot showing number of PCs and cumulative variance

PCs. Interestingly, applying PCA on that same personality test with the binned transformation shows are more clear groupings between selected and not selected candidates. Most of the candidates who were selected are in the top left of the plot (red dots), while those who were not selected are in the bottom right (blue dots).



(a)



(b)

FIGURE 4.4: Scatter plots of personality test with 52 features projected on first two PCs

(a) Projection on first two PCs with no feature engineering
(b) Projection on first two PCs with binned transformation

To choose the optimal number of PCs for each data set, we fit a LDA model with 1 to up to 40 PCs on each data subset (using only continuous features). After getting all the results, we then found the PC number with the highest kappa value. After plotting the results and visually inspecting them, it was evident that for most data subsets the performance stayed about the same after the first 5 PCs. Given this, we chose the lowest number of PCs with a kappa score within 2% of the highest kappa score to avoid over fitting. Figure 4.5 shows a plot of the results for the data subset using all features. For this specific data set, the performance is optimized at 7 PCs and then does not improve as PCs increase. The complete PCA analysis results are in section C.1.

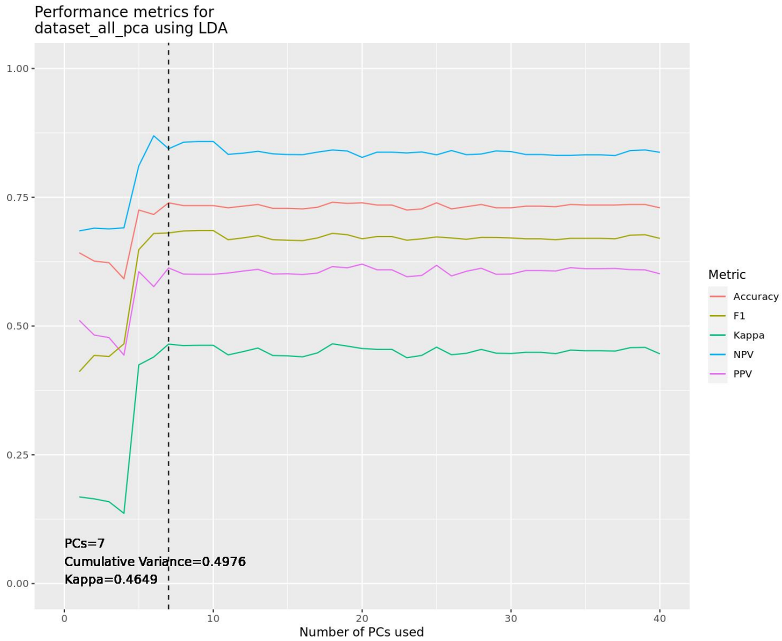


FIGURE 4.5: Results of predicting on PCs 1-40 for dataset_all using LDA

4.3.2 Categorical Features

We transformed categorical features using **lumping.** Lumping is reducing the number of categories in a categorical feature by combining (or "lumping") them together. To do this, we created a contingency table with each categorical feature and the response feature. We then compared the percent pass rate of each category and grouped categories within a 2% pass rate together (with some exceptions if the group size was less than 50 candidates). For example, see how the "race" feature was reduced from 11 categories to three in Table 4.2. After lumping all categorical features, we reduced the total categories for all nine categorical features from 167 to 30. Table 4.3 shows the feature and number of categories before and after lumping.

TABLE 4.2: Lumping example using the "race" feature

race	Selected	Not Selected	Total	Percent Selected	Percent of Data
UNREPORTED	2	4	6	33.3	0.05
WHITE	2,347	6,266	8,613	27.2	72.5
ASIAN/PACIFIC ISLANDER	75	273	348	21.6	2.9
AMER INDIAN OR ALASKA NATIVE	15	55	70	21.4	0.5
ASIAN	1	4	5	20	0.04
BLACK OR AFRICAN AMERICAN	1	4	5	20	0.04
BLACK	158	663	821	19.2	6.9
UNKNOWN	333	1,634	1,967	16.9	16.6
OTHER	7	39	46	15.2	0.4
AMERICAN INDIAN OR ALASKA NATIVE	0	1	1	0	0.008

Table 4.2 continued from previous page

race	Selected	Not Selected	Total	Percent Selected	Percent of Data
Z	0	3	3	0	0.03
Total	2,939	8,946	11,885	24.7	100

New categories: WHITE OTHER UNKNOWN

TABLE 4.3: Categorical Feature Lumping Summary

Feature	Categories Before	Categories After
mos	87	4
rank	6	4
race	12	3
arrival_month	12	5
parents_together	2	2
has_airborne	2	2
glasses	2	2
civilian_education_certification	19	5
age_at_arrival	25	3
Total	**167**	**30**

4.3.3 Creating New Features

We created two features using existing features in the data.

1. time_in_service = arrival date to A&S - date joined military
   • goal is to capture military experience (i.e., how long has the candidate been in the military?)

2. arrival_month = arrival date month
   • goal is to capture seasonality impact (e.g., difference in summer vs. winter months)

4.4 Feature Selection

The goal of feature selection to choose the subset of features that best represent the data. Feature selection methods are not a definitive answer on what the most important features are, but rather give an intuition of what *may* be the most important features given a specific selection method. Feature importance can only fully be understood after applying predictive models and analyzing the results. The three primary techniques for feature selection we used were:

1. **Filter.** The main idea behind filter based feature selection is to use some type of statistical method or criterion to filter "important" features prior to any type of model being applied to the data. Often this is done by comparing individual input features to the response feature and computing a metric [38].

2. **Wrapper.** The wrapper based approach works by first searching for a subset of features from the data's feature space to be used to fit the model. After the model is fit, model performance is measured with an evaluation metric. These steps are iterated for all the different feature combinations available and a final "best feature subset" is returned based on the number of features chosen for the subset size.

3. **Embedded.** Embedded feature selection techniques are "embedded" into machine learning models. For example, penalized models (e.g., lasso or ridge) and tree-based models have an embedded way to assign feature importance in the model itself. As a result, machine learning models like these do feature selection by means of fitting the model. We applied the following models that use embedded feature selection: penalized logistic regression, classification trees, random forest, and xgboost.

4.4.1 Filter Methods

We used the following filter feature selection methods. All methods were implemented using the "FSelector" R packages [39].

1. **Chi Square.** This method compares each individual input feature with the response feature (pass) using Pearson's chi-square test. The result given is the Cramer's V coefficient. Figure 4.6 shows the output of each features importance. The fitness test scores were the top four most important features and almost double as important as the fifth most important feature. There were 20 features that had a Cramer's V score of 0, indicating those features have little relation to the response feature.

2. **Information Gain.** This method is an entropy based method that compares each individual input feature with the response feature (pass) using information gain. The result is transformed using log.

3. **Gain Ratio.** This method is an entropy based method that compares each individual input feature with the response feature (pass) using gain ratio. The result is transformed using log.

4. **Symmetrical Uncertainty.** This method is an entropy based method that compares each individual input feature with the response feature (pass) using symmetrical uncertainty. The result is transformed using log.

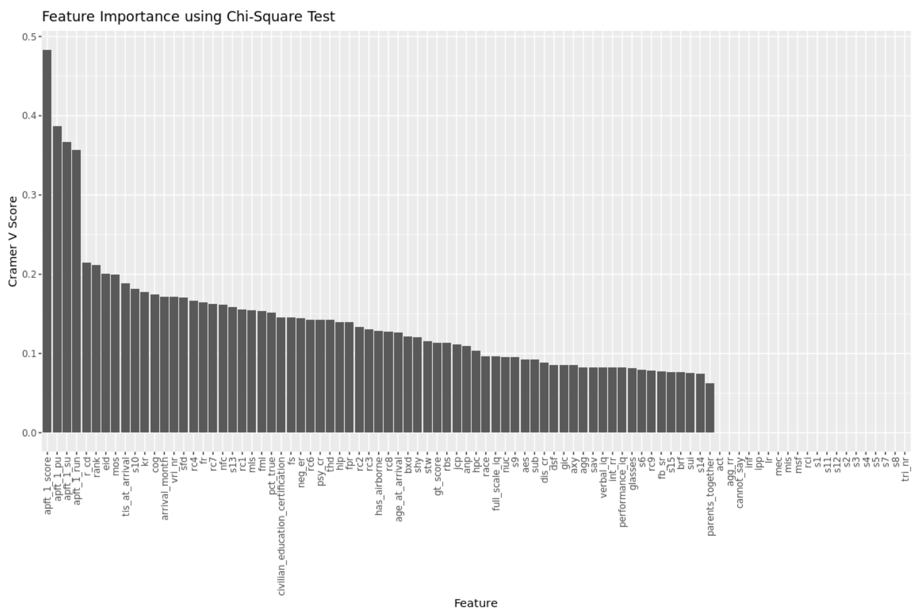


FIGURE 4.6: Feature selection results using chi-square

After doing all these methods, we compared the results and found that most of the methods returned similar feature importance rankings.

4.4.2 Wrapper Method

We used two wrapper methods for feature selection.

1. **Backward Feature Selection.** We used the **recursive feature elimination** method that implements backward feature selection for feature selection. Using backward feature selection, all of the features are used to build the initial model and the performance is evaluated. Then, a single feature is removed and the remaining features are used to build the model. All different combinations of removing one feature are used to build the model. The N-1 subset of features that yields the best model performance is then selected. This process of removing a single feature at each step is repeated until the specified number features are chosen. We used the "caret" R package implementation of the recursive feature elimination method [25].

   We used the recursive feature elimination method to find the best feature subsets using 1-10, 15, 20, 25, 30, and 89 features. The results are summarized graphically in Figure 4.7. Optimal performance was when n=25 features, while 20 and 30 features were near the best. The caret function outputs a ranked 1-n output of feature importance. Like filter-based methods, the fitness scores were the top four most important features.

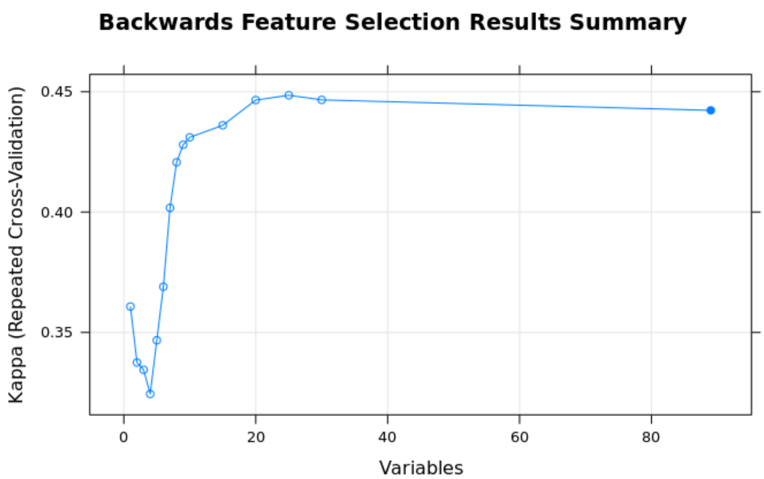


FIGURE 4.7: Backwards Feature Selection Results Using Caret

2. **Forward Feature Selection.** Using this method, a null model is initiated. A model is built for each individual feature. The feature that yields the best model performance is selected. Then, all the different combinations between the first selected feature and the rest of the features are used to build the model. The two features that yield the best model performance are selected. This process of adding a single feature at each step is repeated until the specified number features are chosen. We implemented this method using the "varrank" R package [40]. The "varrank" method uses mutual information as the criteria to evaluate the importance of features at each step.

While it is difficult to see all the specific features and forward steps used for the forward feature selection method, Figure 4.8 provides some useful insights. The features on the y axis show the most important features ranked 1-n (most important on top). The features on the x axis show the features added at each step and associated redundancy/relevancy to the response feature. The key takeaway is most features appeared to be redundant using this approach. In other words, as more and more features were added at each step, there was little increase in information gain. This feature selection method also returns a ranked list of most important features from 1-n. The most important feature was apft_1_score like the rest of the methods; however, the rest of the top five differed. The next most important features were mos, glasses, s10, and rank.

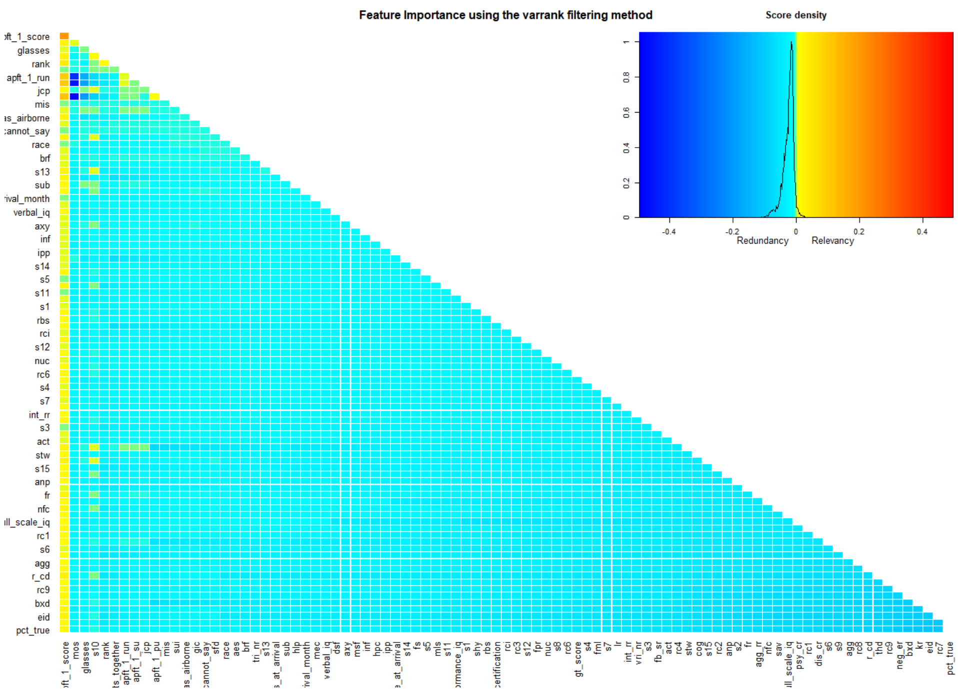


FIGURE 4.8: Feature selection results using "varrank" method

4.4.3 Best Feature Subsets

After completing all these feature selection methods, we created two final best feature subsets by aggregating all the results. One subset has the 10 best features and the other subset has the 20 best features. To choose the features for the subsets, we aggregated all the results by adding all the 1-n ranked results from each method and selecting the top n results with the lowest scores. Table 4.4 shows the features in each subset. Figure 4.9 shows the plot of the aggregated feature importance analysis. Doing this analysis revealed 1) the methods seemed to rank feature importance similarly (length of stacked bars for each feature about the same) and 2) the fitness test features (apft_1_score, apft_1_su, apft_1_pu, and apft_1_run) were the top 4 highest ranked features most methods. We used these feature subsets as part of the analysis.

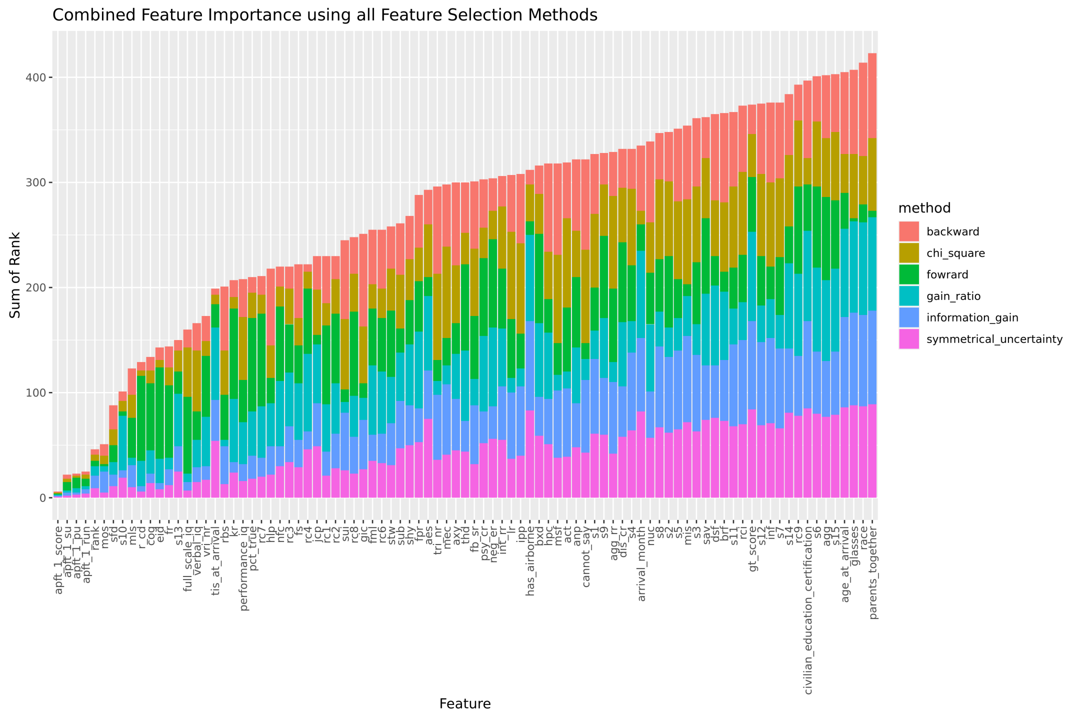


FIGURE 4.9: Feature selection results using combined method ranks

TABLE 4.4: Best feature subsets

Subset	Features
best_10	apft_1_score, apft_1_su, apft_1_pu, apft_1_run, rank, mos, sfd, s10, mls, r_cd
best_20	apft_1_score, apft_1_su, apft_1_pu, apft_1_run, rank, mos, sfd, s10, mls, r_cd, cog, eid, fr, s13, full_scale_iq, verbal_iq, vir_nr, tis_at_arrival, rbs, kr

4.5 Data Exploration

To explore the data, we looked at a variety of descriptive statistics and created visualizations to better understand the data. We explored the following:

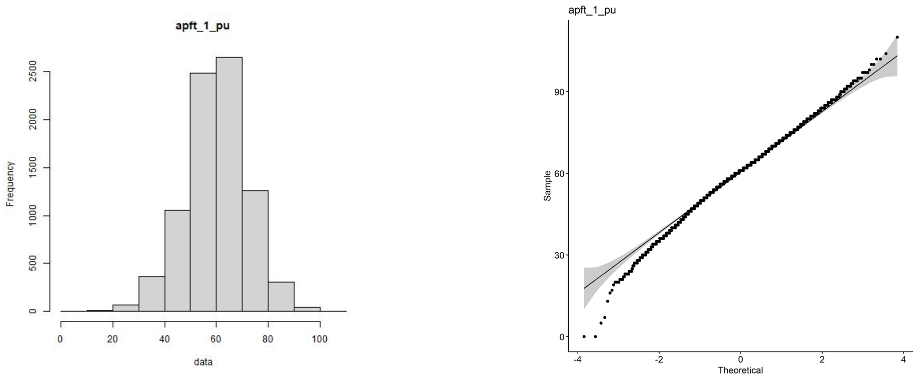
1. Descriptive Statistics and Plots
2. Comparisons Between Groups
3. Feature Comparison to Response
4. Indicator Variables Analysis

4.5.1 Descriptive Statistics and Plots

Data Normality

The data normality assumption is not always required, but may impact model performance when using some parametric modeling approaches such as the Linear Discriminant Analysis (LDA) model used in this research. Other models such as random forest do not require this assumption, however. Regardless, to better understand the data, we chose to check normality of all the features. To visually check, we created a histogram and quantile-quantile (QQ) plot for each feature.

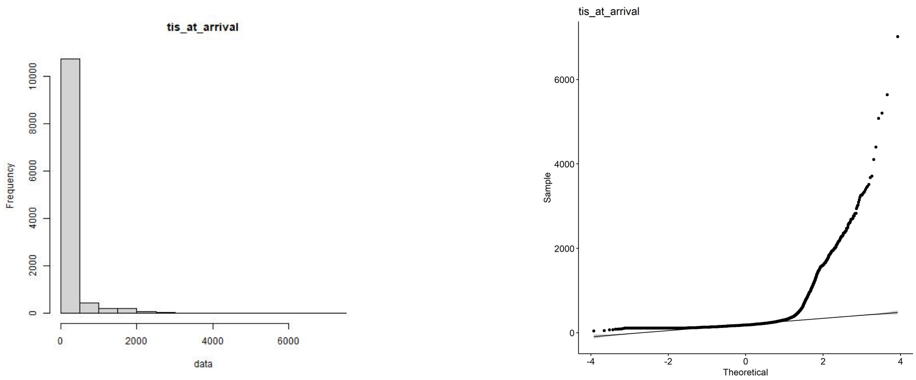
Most of the data appeared to be normal, such as the number of push ups done on the fitness test shown in Figure 4.10. It is clear that the number of push ups can be assumed to be normal based on the bell shaped histogram and QQ plot with the data empirical quantiles largely in line with the theoretical quantiles.



(a) (b)

FIGURE 4.10: Histogram (a) and QQ Plot (b) of Number of Push Ups Completed on the Fitness Test

Some of the data did not look normal, however. For example, the "tis_at_arrival" feature, or "Time in service at arrival" that is the amount days a Soldier has served in the military. The data appeared to be highly right skewed (far more lower values than higher) as shown in Figure 4.11.

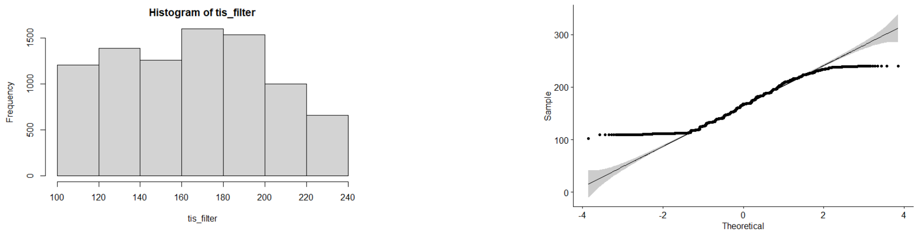


(a) (b)

FIGURE 4.11: Histogram (a) and QQ Plot (b) of Time in Service at Arrival

Data with a skewed distribution requires further investigation to understand. Based on knowledge about the A&S, this distribution makes sense as over 70% of the Soldiers who attend the A&S have limited experience in the military. After further examination of the data, we discovered that 74% of the candidates (8,624/11,689 records) had between 100 - 240 days time in service, with the rest being largely outliers. A histogram and QQ plot showing the time in service values between 100 -

240 days is shown in Figure 4.12 where the data appears to be more normally distributed.



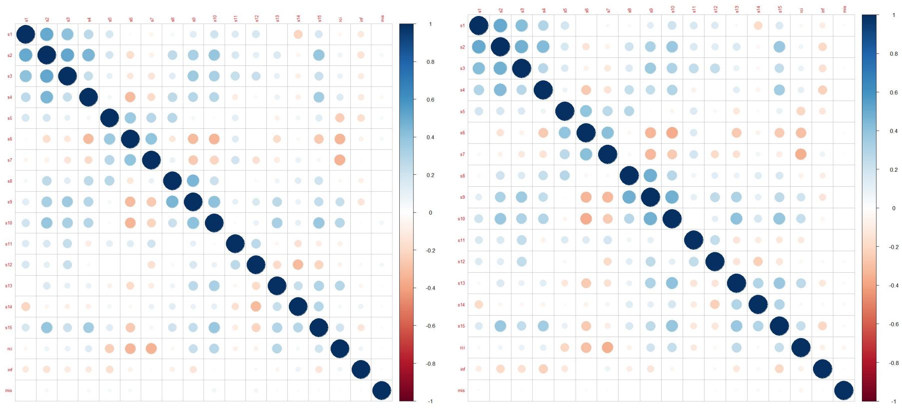
(a) (b)

FIGURE 4.12: Histogram (a) and QQ Plot (b) of Time in Service at Arrival for values between 100-240 days

After visually inspecting all of the continuous features and investigating features that had non-normal looking distributions to understand our data, we moved forward in analysis. We understand that all the data may not be *perfectly* normal, and this may impact the results in the models that assume this.

Correlation

To improve our understanding of how the features related to each other, we looked at correlation plots for different subsets of features and candidates. For example, Figure 4.13 shows side-by-side correlation plots for the JPI_R psychological test for candidates who were selected (a) and not selected (b). Both groups appear to have similar correlations, indicating that both groups answered questions similarly on the psychological test.

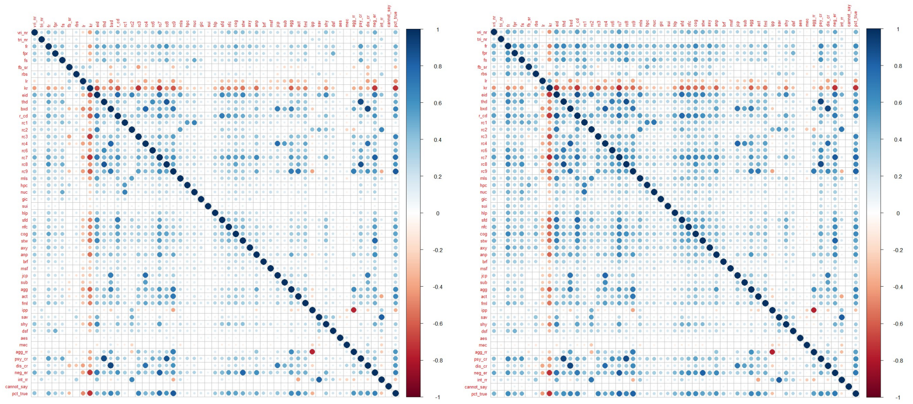


(a) (b)

FIGURE 4.13: JPI_R Psychological Test Correlation Plots

(a) Candidates who were selected
(b) Candidates who were not selected

However, looking at the MMPI psychological test, there was an observable difference in correlation between candidates selected (a) and not selected (b) as shown in Figure 4.14. In general, selected candidates responses were less correlated (i.e. lighter in color) compared to candidates who were not selected. This indicates that there is a difference between the two groups in how they answer questions on the psychological test.



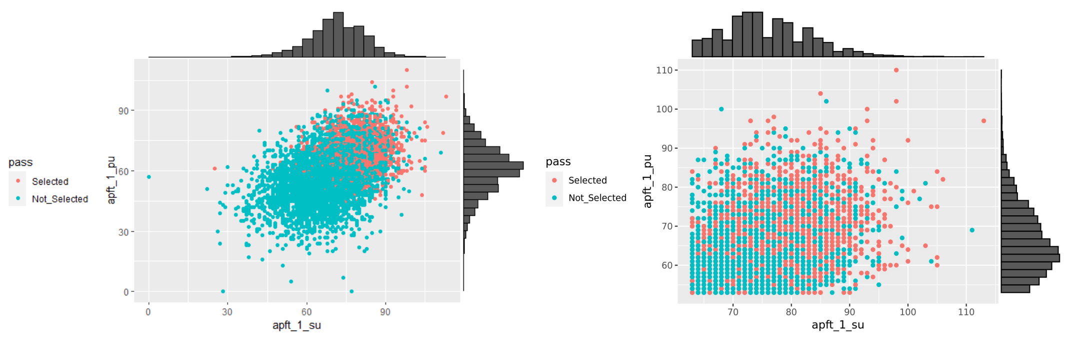
(a) (b)

FIGURE 4.14: MMPI Psychological Test Correlation Plots

(a) Candidates who were selected
(b) Candidates who were not selected

Plots

We created plots with many combinations of features to learn more about feature relationships. For continuous features, we primarily used scatter plots. The fitness test was an area of interest as candidates are screened using this (see section 2.2). Figure 4.15 shows two scatter plots of fitness test events. Sub figure (a) shows a scatter plot for all candidates number of sit ups (x axis) and push ups (y axis) with associated histograms on the top and right of the plots. Sub figure (b) uses the same features, but only for those candidates who met the minimum fitness test screening criteria.



(a) (b)

FIGURE 4.15: Scatter plot of push ups and sit ups

(a) All candidates
(b) Candidates who passed the minimum fitness test screening criteria

From sub figure (a), it is evident that only candidates who perform well on sit ups and push ups (top right area on plot) are selected. This makes sense, because 1) candidates are dropped who do not meet the minimum screening criteria score and 2) generally the more fit a candidate is, the more disciplined and motivated he or she is and more likely to be selected. From sub figure (b), it is evident that, of the candidates who met the minimum fitness score, candidates who performed more sit ups and push ups were selected at a higher rate compared to those with lower scores. However, there are still approximately 50% candidates who met the minimum fitness standards who were not selected. This includes some of the candidates who performed in the top quantile on both push ups and sit ups. This is interesting because it reveals that even if a candidate performs exceptionally on the fitness test, it does not imply he or she will be selected.

For categorical features, we primarily used box plots for analysis. Figure 4.16 shows a box plot of a candidates rank (x axis) and number of push ups performed (y axis). Candidates who were selected are represented by the purple box and candidates not selected are represented by the yellow box for each rank. Candidate ranks from lowest to highest are: PVT_PV2, PFC, SPC, CPL_SGT. The small black dots show the number of candidates who were in each category (i.e. the more black dots, the more candidates in that category).

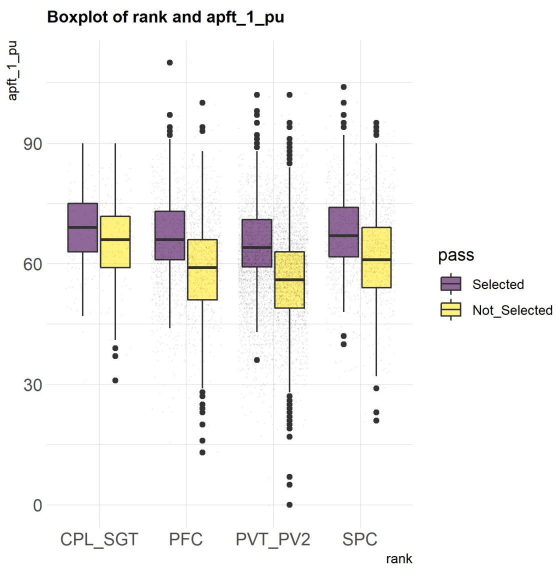


FIGURE 4.16: Box plot of candidate rank and number of push ups

A box plot showing a candidates rank and the number of push ups performed. Selected candidates are the purple boxes. Not selected candidates are the yellow boxes.

Some key observations from this box plot are:

1. On average, the higher a candidate's rank, the larger percentage of candidates were selected (larger proportion of black dots in selected box for higher rank).

2. On average, the higher a candidate's rank, the more push ups he or she performed (purple and yellow boxes higher on y axis as rank gets higher).

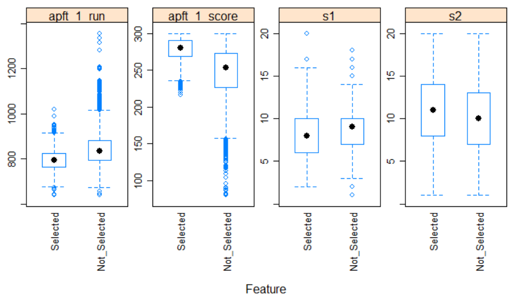
3. On average, for each rank, candidates who were selected performed more push ups compared to those candidates not selected (purple box higher than yellow box for each rank).

4. The vast majority of candidates are lower rank (PVT_PV2 and PFC). SPC and CPL_SGT are the minority of candidates. (Almost all of the total black dots are in PVT_PV2 and PFC. Very few are in CPL_SGT and SPC.)

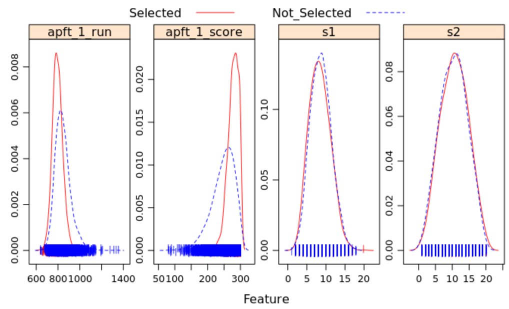
4.5.2 Comparisons between Groups

To compare different groups we used visualizations and statistical tests. For visualizations, we plotted each feature with candidates split into a Selected and Not Selected group using both a box plot and a density plot. These two plots show similar information, but offer unique insights. A box plot clearly shows the quantiles, means, and outliers; however, understanding the distribution of the data is less clear. Whereas a density plot clearly shows the distribution of the data, but the quantiles, mean, and outliers are less apparent. Comparing both of these plots gives a more full picture of the data distributions.

Figure 4.17 shows an example of both of a box plot (sub figure (a)) and density plot (sub figure (b)) for the features apft_1_run, apft_1_score, s1, and s2 using all data. apft_1_run and apft_1_score show a visually clear difference between the group scores and distributions, while s1 and s2 had almost no difference.



(a)



(b)

FIGURE 4.17: Comparison between Selected and Not Selected groups using all data

(a) Box plot
(b) Density plot

After visually inspecting each feature, we conducted a Welch's two sample t-test on each feature to compare the means of candidates selected and not selected to more objectively understand the differences. Figure 4.18 shows a plot of the T-test statistic scores between candidate selected and not selected using all candidates (blue) and using only those who met the minimum fitness test standards (red).

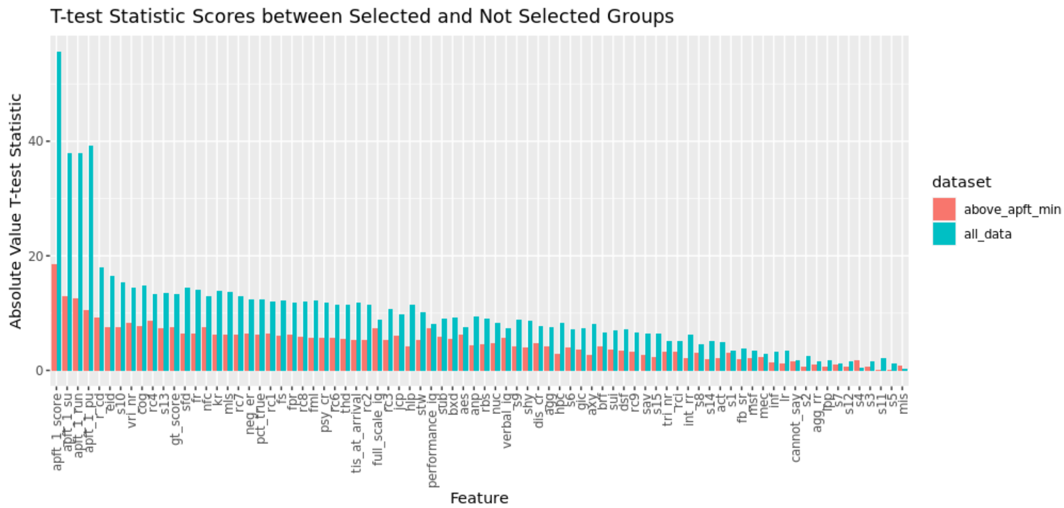


FIGURE 4.18: T-test Comparison Between Selected and Not Selected Groups

For both candidate subsets, there were statistically significant differences between group means for a majority of the features at the $\alpha = 0.05$ level (test statistic value >= 1.96). Specifically, for all candidates, 71/89 features and for candidates who met the minimum fitness test screening criteria, 66/89 features. Some key observations from the t-test comparison are:

1. All of the features that were significant were more significant using all data compared to only the subset of candidates who met the minimum fitness testing screening criteria (blue bars larger than red bars).

2. For both candidate subsets, fitness test scores were the four most significant features between selected and not selected candidates.

4.5.3 Feature Comparison to Response

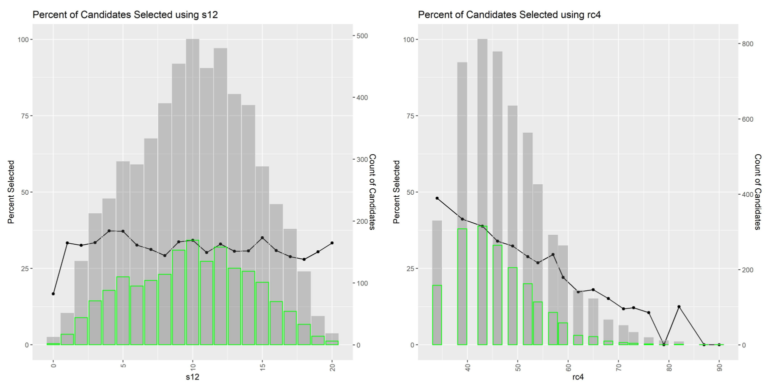
To understand input feature relationships to the response feature, we created contingency tables with each category (categorical feature) or score (continuous feature) and how it related to the response feature (pass). An example table is shown in Table 4.5. Using each table, we created plots that included: 1) a histogram with all candidates (grey bars), 2) a histogram with only selected candidates (green bars), and 3) a line showing the percent of candidates selected at each value (black line). These plots make clear the input feature relationships to the response feature and underlying distributions. section C.2 shows the plots for all features.

TABLE 4.5: Contingency Table for age_at_arrival

age_at_arrival	Selected	Not_Selected	total	percent_pass	percent_data
UNKNOWN	87	107	194	44.85	1.63
>20	1,479	3,441	4,920	30.06	41.4
<=20	1,373	5,398	6,771	20.28	56.97
Total	2,939	8,946	11,885	24.73	100

Figure 4.19 shows an example of two of these plots. (a) is the Risk Taking ("s12") feature from one of the psychological tests and (b) is the Anti Social Behavior ("rc4")

feature from another psychological test. s12 did not seem to have strong correlation to the response feature (black line is horizontal across all s12 scores), while rc4 did seem to have a correlation (the lower the score, the greater percent of candidates who were selected). It is also clear the distributions are different. The s12 feature had a normal distribution, while the rc4 had a skewed distribution.

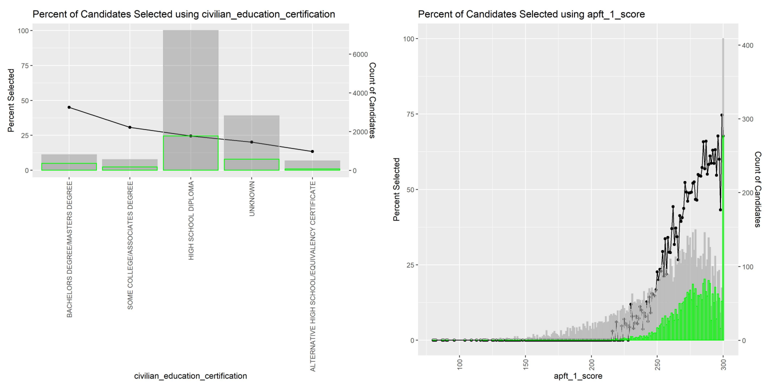


(a) (b)

FIGURE 4.19: Psychological Features Compared to Selected Candidates

(a) Risk Taking ("s12")
(b) Anti Social Behavior ("rc4")

Figure 4.20 shows two other plots of education level (a) and fitness test score (b). In both cases, there is a clear linear trend between category/score and percent of candidates who were selected. More specially, the higher a candidate's education, the greater percentage of candidates who were selected. Likewise, the higher a candidate's fitness test score, the greater percent of candidates who were selected.



(a) (b)

FIGURE 4.20: Admin and Performance Features Compared to Selected Candidates

(a) Civilian Education Certification
(b) Total Fitness Score

We also did this same analysis using the quantile (binned) data (see section 4.3). Using the binned features, we combined multiple features together to better understand the interaction effects. Figure 4.21 shows a plot of a candidate's sit up, push up, and 2 mile run score combined bins compared to candidate selection rate. There is a clear linear trend that shows the higher a candidate's score in all three events, the greater percent of candidates were selected.

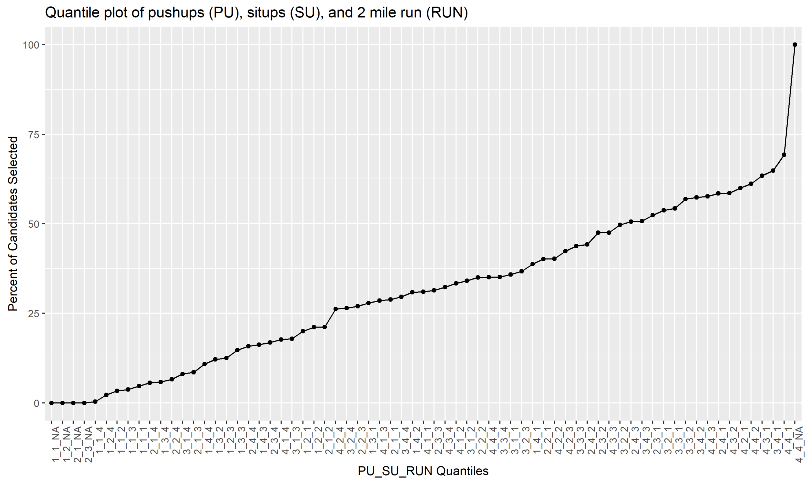


FIGURE 4.21: Push up, sit up, and run bin feature comparison to "selected" candidates

While comparing the input feature to response feature uses descriptive statistics and is not predictive in any way, it is useful in providing key relationships between the features that can help interpret model results.

4.5.4 Indicator Variables Analysis

Due to the amount of missing data, we performed analysis on the value of the missing data to see if it was missing at random or not. To do this, we first created binary indicator features for each feature. If the observation was complete, the indicator feature value was 1 and if missing, 0. Then for each indicator feature, we calculated the percent of candidates who were selected for values of 1 and 0, respectively. This allowed us to examine if the data for each feature was missing at random or not. If missing at random, we would expect the selection rates to be close to even for both missing and complete data for each feature. However, if the data was not missing at random, we would expect a large difference in selection rates (indicating that there is a systematic reason why the data is missing).

Figure 4.22 shows a plot of each indicator feature (x axis), the percent of candidates who were selected with the feature data missing ("Percent Pass 0"; red line), the percent of candidates who were selected with the feature data complete ("Percent Pass 1"; green line), and the percent of overall observations complete from each feature ("Percent Data Complete"; black line). Notably, all of the features appear to be missing systematically except for civilian education certification. Our hypotheses as to why each feature is missing systematically are:

- **age_at_arrival, arrival_month.** These two features have nearly all the data complete (98%). For both features, the percent of selected candidates is higher for candidates missing the data compared to candidates with the data. Interestingly for the arrival_month feature, all 86 candidates missing the data were all selected. We think that in both cases there was systematic missed data entry for groups of candidates who were selected.

- **apft_1_pu, apft_1_su, apft_1_run, apft_1_score.** For all candidates missing fitness test scores, there was a 3% selection rate. For candidates with fitness

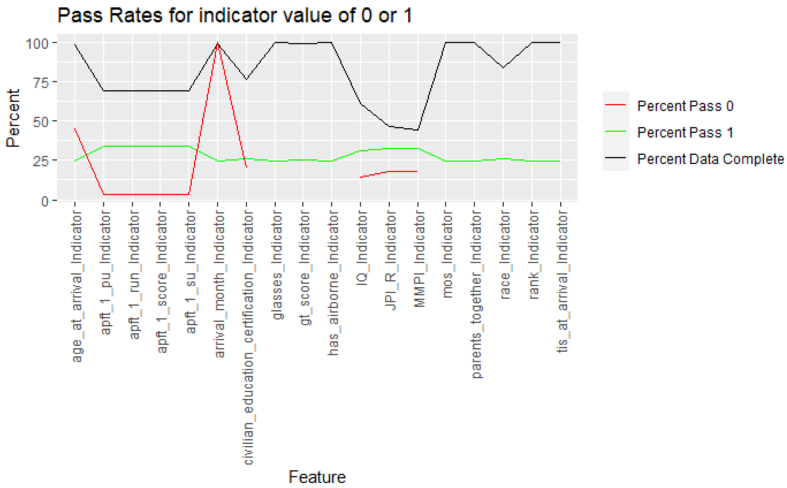


FIGURE 4.22: Analysis of indicator features

test scores, there was a 34% selection rate. The data is 68% complete. We think that the vast majority of candidates missing this data were likely screened out early in the screening phase and their scores not recorded as only 3% were selected.

- **IQ, JPI_R, and MMPI.** These features are the three psychological tests. Each test is missing between 45-60% of the observation on candidates. Likewise, candidates with the psychological test data were selected 15% more compared to candidates missing the data. We know that some of the psychological tests were not implemented until 2018 (two years after the data starts: 2016). This explains some of the missing data. However, after 2018 we believe that, similar to the fitness data, the candidates missing the data were dropped early in the screening phase prior to taking the psychological tests.

After conducting this indicator feature analysis, we now understand that the missing data is almost all systematically missing.

4.6 Create Data Subsets

We created data subsets using combinations of feature subsets, candidate subsets, and feature engineering methods to better understand specific dynamics of each of these aspects in analysis. We did not conduct analysis on every possible combination of these subsets, but instead created specific subsets for specific analysis purposes. Figure 4.23 shows a diagram of how the final data subsets were created. We labeled the data in the following way: "candidate subset"_"feature subset"_"feature engineering technique". For example, a data subset name could be "dataset_below_apft_minimum_psych_log". In total, we used 66 different combinations of subsets shown in Table A.2.

4.6.1 Feature Subsets

We created 10 feature subsets shown in Table 4.6. The subsets represent the different types of data collected on each candidate. Specifically for the psychological data, the "Psych" data set is all three psychological tests combined ("Psych_JPI_R",

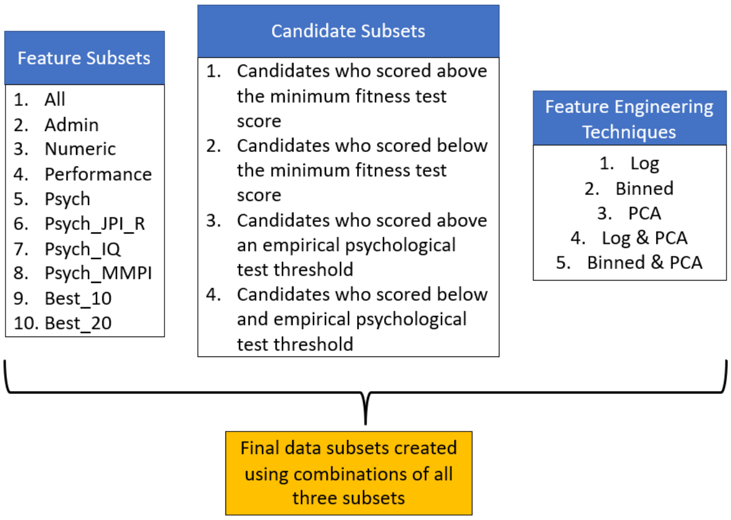


FIGURE 4.23: Visualization of creating final data subset combinations

"Psych_IQ", and "Psych_MMPI"). For a full list of all features in each subset, see Table A.1. To create each subset, we selected the features and then removed all missing data from the non-imputed data set to maximize observations in each data subset. This resulted in different amounts of observations and percent of candidates selected in each data subset. All of the data subsets still had 3,095 or more observations and a class imbalance with selected candidates being only around 30%.

TABLE 4.6: Data Subsets Summary

Number	Data Subset	Total Features	Continuous	Categorical	Observations	Percent of Candidates Selected
1	All	89	80	9	3,095	36%
2	Admin	10	1	9	11,689	24%
3	Numeric	80	80	0	3,095	36%
4	Performance	5	5	0	8,084	34%
5	Psych	74	74	0	3,544	33%
6	Psych_JPI_R	18	18	0	5,506	32%
7	Psych_IQ	3	3	0	7,286	31%
8	Psych_MMPI	53	53	0	5,298	33%
9	Best_10	10	8	2	3,295	37%
10	Best_20	20	18	2	3,096	36%

4.6.2 Candidate Subsets

1. **Above minimum fitness screening score. (n=4,969)** In the screening phase, there is a minimum fitness score candidates must achieve (see section 2.2 Existing Solutions) or they will be dropped. To better understand how to model fully qualified candidates, we wanted to analyze this candidate subset.

2. **Below minimum fitness screening score. (n=3,272)** We created this data subset to explore if there are any additional features that are predictive of a candidate performing poorly on the fitness test.

3. **Above psychological screening scores. (n varied based on chosen threshold)** We chose specific psychological features to set empirical thresholds. Our

goal with analyzing this candidate subset is to see the relationship that scoring above the threshold has with other features.

4. **Below psychological screening scores. (n varied based on chosen threshold)** We chose specific psychological features to set empirical thresholds. We then fit models to the data subset with candidates who scored below that threshold to understand how effective different thresholds are as screening criteria.

4.7 Model Fitting

We fit each model (see section 3.3) to each data subset combination (see section 4.6) to gain specific insights into model performance with different combinations of features, observations, and feature engineering. We used the following steps to fit each model to each data set.

1. **Data Preprocessing**
2. **Model Training and Parameter Tuning**
3. **Model Testing**

4.7.1 Data Preprocessing.

Each model requires the data be preprocessed in a way that is compatible for the model to use. All of the models that use distance-based loss functions required standardizing continuous features and one-hot encoding the categorical features. These models require standardizing (centering [mean = 0] and scaling [standard deviation = 1)]) to prevent the scale of some features from dominating others when estimating model coefficients. Likewise, the models require one-hot encoding (making each level of the categorical features a binary vector of 1 and 0) to represent the categorical relationships in numerical form. After the categorical features were one-hot encoded, the features increased from 9 to 30. Table 4.7 shows the required preprocessing steps required for each specific model.

TABLE 4.7: Data pre-processing required for each model

Model	Standardized	One Hot Encoding	Type of Features
LDA	No	NA	Continuous Only
QDA	No	NA	Continuous Only
Penalized Logistic Regression	Yes	Yes	Continuous and Categorical
SVM	Yes	Yes	Continuous and Categorical
CART	No	No	Continuous and Categorical
KNN	Yes	Yes	Continuous and Categorical
Random Forest Ensemble	No	No	Continuous and Categorical
xgboost Ensemble	No	No	Continuous and Categorical
Stacked Ensemble	Yes	NA	Continuous Only

4.7.2 Model Training and Parameter Tuning

We did the following steps to train and tune the model for each data subset. We used the "caret" R package to tune and train all of our models [25]. It has a "train" function that allows customization of specific training and tuning parameters. Figure 4.24 shows an overview of the model training and testing procedure for each model applied to each data subset combination.

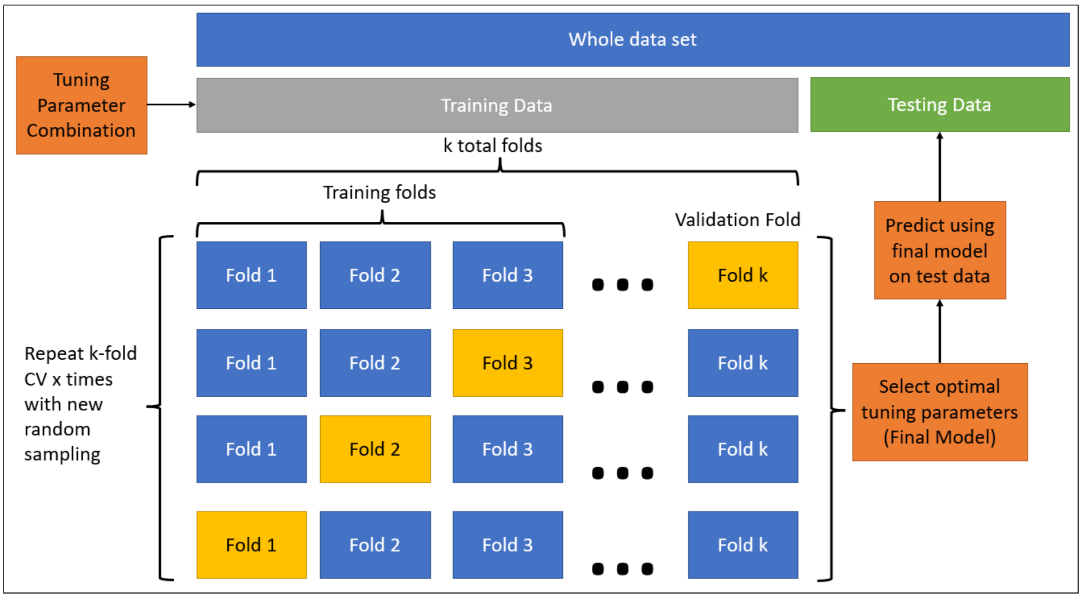


FIGURE 4.24: Model training and testing procedure

1. **Split the data into a training and test set.** We performed random, stratified sampling to split the data into 70% training data and 30% test data. We used stratified sampling because this keeps the class distribution the same for both training and test data sets based on the class distribution in using all the data. This was important for our data due to the class imbalances, as some data subsets had less than 25% of selected candidates in them. We chose to train on 70% of the training data because most of the data subsets had more than 3,000 observations. Training on 70% of 3,000 or more observations provides adequate data to train a robust model, but avoids over fitting by using too much of the data. Likewise, using the model to test on 30% of the data provides enough data to have a representative sample of the whole data set.

2. **Select the best model tuning parameters.** For each model, with each combination of tuning parameters, we trained the model on the training data set using a resampling method to find the optimal tuning parameters. For most models, we 10 fold Cross Validation (CV) with 5 repeats. We chose to use 10-fold CV because there is enough observations to use 10 folds and still have it be representative of the data in each fold. We chose to use 5 repeats of 10 fold CV to account for any bias that could exist in only tuning the parameters on one 10 fold CV iteration. To find the optimal tuning parameters, we calculated the average performance of each tuning parameter combination on each validation fold for all five 10-fold CV repeats and chose the combination with the highest kappa value. We did not use 10 fold CV with 5 repeats on the ensemble models due to 1) embedded methods of resampling and 2) computation time. Figure 4.25 shows the procedure as an algorithm that is used by the caret package for tuning parameter selection. Table 4.8 shows the training resampling method used for each model.

   When training the model using k fold CV, we also used "up sampling" for each training fold to account for the class imbalance. Up sampling is randomly sampling (with replacement) the minority class (selected) to be the same size as the majority class (not selected). This is the best option for our data set because we do not lose information from the majority class (not selected), but also have

```
Define sets of model parameter values to evaluate
for each parameter set do
    for each resampling iteration do
        Hold–out specific samples
        [Optional] Pre–process the data
        Fit the model on the remainder
        Predict the hold–out samples
    end
    Calculate the average performance across hold–out predictions
end
Determine the optimal parameter set
Fit the final model to all the training data using the optimal parameter set
```

FIGURE 4.25: Caret Model Tuning Parameter Procedure

enough observations from the minority class (selected) with which to train the model. We only used up sampling for training folds, not for validation fold, as up sampling on the validation fold would be data leakage and not represent the true distribution. The "caret" R package contains a function (upSample) to do this.

Each model has different tuning parameters used (see section 3.3). As a result, we used different tuning parameter combinations for each model. The "caret" package has a built in parameter "tuneLength" where you can set the number of values to try for each tuning parameter. We used this for some of the models, where other models we set our own tuning parameter grid to search. Table 4.8 shows the tuning parameter combinations used for each model.

TABLE 4.8: Model training and tuning method for each model

Model	Training Resampling Method	Tuning Parameter Search
LDA	5 x 10 Fold CV	N/A (No tuning parameters)
QDA	5 x 10 Fold CV	N/A (No tuning parameters)
Penalized Logistic Regression	5 x 10 Fold CV	tuneLength=10
SVM	5 x 10 Fold CV	tuneLength=10
CART	5 x 10 Fold CV	tuneLength=10
KNN	5 x 10 Fold CV	k=seq(3,63,by=4)
Random Forest	10 Fold CV	mtry=seq(3,13,by=2) ntree=seq(500,1500,by=500)
xgboost	10 Fold CV	tuneLength=3
Stacked	10 Fold CV	base model tune

3. **Determine the Optimal Classification Threshold Setting.** In order to make best classification from the model predicted posterior probability (soft classification) for the validation folds, it is necessary to understand the impact of changing the classification threshold from 0.5 (default) to a range of values 0-1. The goal of the model is what drives the classification threshold, as the classification threshold is linked to the idea of classification costs [32]. Classification costs are set based on the cost of making FP and FN classifications. After specifying the costs given to the FP and FN based on domain knowledge, we can then set the best classification threshold. To understand the relationship, see Table 4.9 that shows how they relate.

    Our goal is to have the best holistic model for both selection and non-selection prediction. There is a high cost for both FN (predicted not selected, actually selected) and FP (predicted selected, actually not selected). A FN error

TABLE 4.9: Example Classification Cost Matrix

		Observed (True Outcome)	
		Selected	**Not Selected**
Predicted (Model Outcome)	**Selected**	TP Cost=0	FP Cost=1
	Not Selected	FN Cost=5	TN Cost=0

To minimize expected cost, set threshold to:
$$\Pr(\text{Selected}) > \frac{FN}{FP+FN}$$
In this example, the optimal threshold would be: $\frac{5}{1+5} = \frac{5}{6} = 0.833$

results in a candidate that is qualified for the unit that was missed. A FP error results in a candidate being an additional cost the system. As a result, we allowed the data to empirically choose the threshold that maximizes the kappa value. We then use the same probability threshold chosen based on the 10 fold CV validation folds in the training data on our test data.

Figure 4.26 shows an example plot of the classification threshold impact on different metrics. In this example, the optimal classification threshold was 0.58 based on the kappa metric. As the threshold increases, the NPV and Sensitivity decrease while PPV and Specificity increases. This makes sense, as when the threshold is near 0, nearly all classifications will be Selected and only the candidates with a very low posterior probability will be classified as not selected. As a result, most of the candidates classified as not selected will be accurate, causing high NPV and Sensitivity scores. As the threshold increases, a balance happens (usually near 0.5) and then the opposite effect happens as the threshold approaches 1.

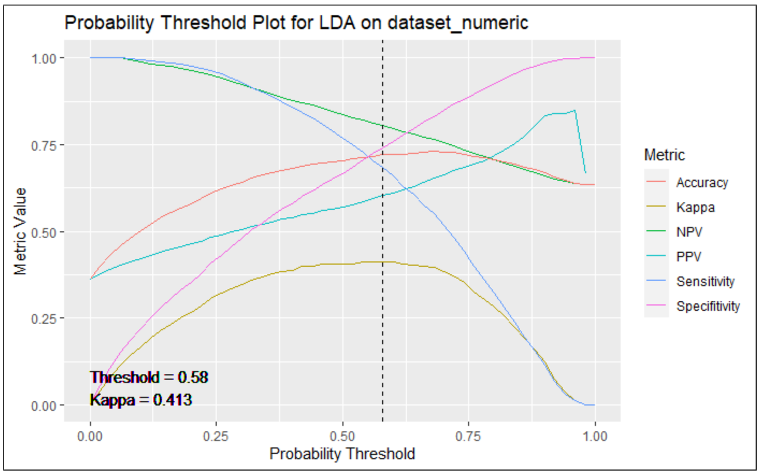


FIGURE 4.26: Optimal Threshold Analysis Example

4.7.3 Model Testing and Evaluation

Model Testing. To test each model, we used the optimal final model trained on the training data. We tested (or predicted) the models on the testing data (30%) using the optimal classification threshold from the training data.

Model Evaluation. Based on the model prediction results from the test data, we created a confusion matrix to evaluate using the metrics discussed in section 3.4. Additionally, we saved the following results from each model and data subset combination:

1. Optimal Tuning Parameters Search Results
2. Model Coefficients (if applicable to model)
3. Feature Importance
4. Classification Cost and Threshold Analysis
5. Test Data Predictions (both soft and hard)
6. Confusion Matrices
7. Visualizations (e.g. Classification Tree)
8. Final Results Summary (using optimal tuning parameters and classification threshold)

As part of the model evaluation, we reviewed all the relevant output data from the model to understand how the model performed, why it performed that way, and if the results made sense. The process was iterative, and if something needed to be adjusted, we made the adjustment and fit the model again.

Chapter 5

Results

We will use the abbreviations in Table 5.1 for brevity.

TABLE 5.1: Abbreviations

Name	**Model**
pen_log_reg	Penalized Logistic Regression
LDA	LDA
QDA	QDA
SVM_lin	SVM (using linear kernel)
CART	CART
KNN	KNN
rf	Random Forest
stack_glm	Stack Ensemble (using logistic regression generalizer)
stack_rf	Stack Ensemble (using random forest generalizer)

Name	**Feature Engineering Type**
none	None
log	Log
binned	Binned (4 quanitles)
pca	Principle Component Analysis
log_pca	Log first, then PCA
binned_pca	Binned first, then PCA

Name	**Candidate Subset**
all	All Candidates
above_apft	Candidates who scored above the minimum fitness test screening criteria

5.1 Results Summary

The results for all models and data subset combinations are in section C.3. **For all sections of the results, we only used non-imputed data subsets to avoid creating bias in our results.** In section 5.7 we specifically discuss the impact of data imputation on results.

Figure 5.1 shows the results for all models on all candidate, feature, and feature engineering data subset combinations using kappa score. Kappa scores ranged from -0.56 (negative score indicates that prediction was worse compared to what is expected by random chance) to 52.16. The black line indicates the range in scores between models for each individual data subset. In general, the more features in the subset, the wider the kappa range.

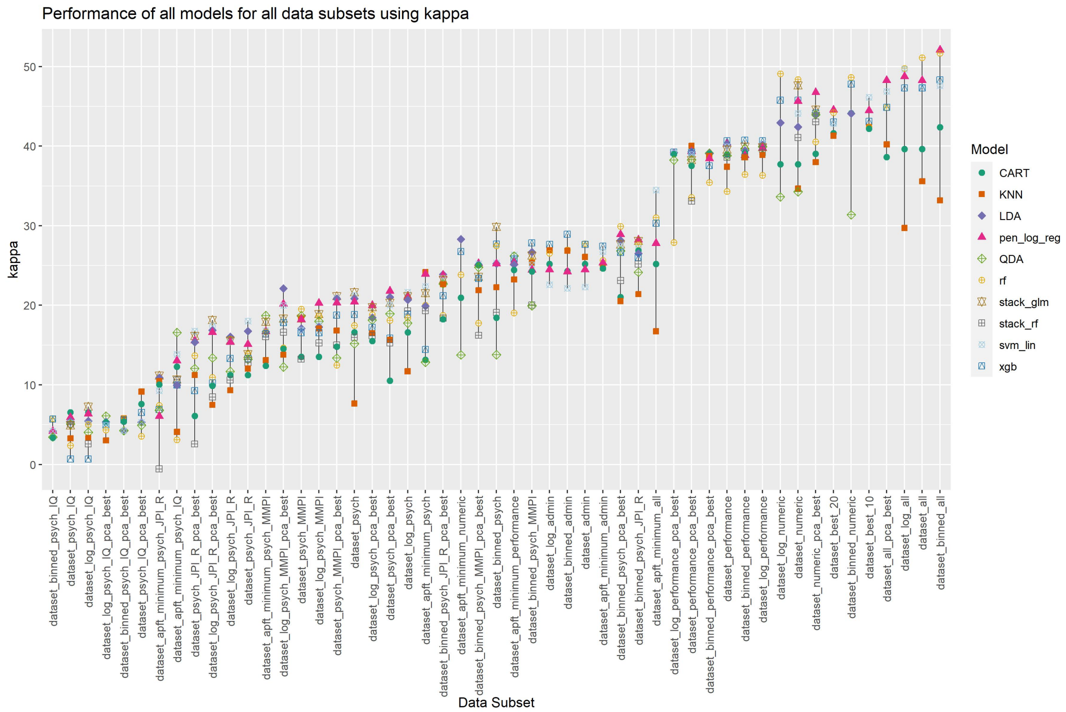


FIGURE 5.1: Results for all models on all candidate, feature, and feature engineering data subset combinations using kappa

Figure 5.2 shows the best results for all models on all candidate, feature, and feature engineering data subset combinations using kappa. The model results were overall generally modest (< 80% accuracy and < 0.5 kappa). Likewise, for almost all subsets, candidate non-selection prediction accuracy (77% NPV average) was much higher than candidate selection (47% PPV average). Despite the modest model performance, however, the results revealed other useful insights that we will discuss.

Notably, 9 of the 10 models scored best on at least one of the data subsets. Likewise, all of the five feature engineering methods used produced optimal results on at least one of the data subsets. The variation in scores between accuracy, NPV, specificity, PPV, sensitivity, and kappa are also evident. Overall, the scores trended the same (when one increased so did the others). However, there are observable differences for the other metric scores with similar kappa scores.

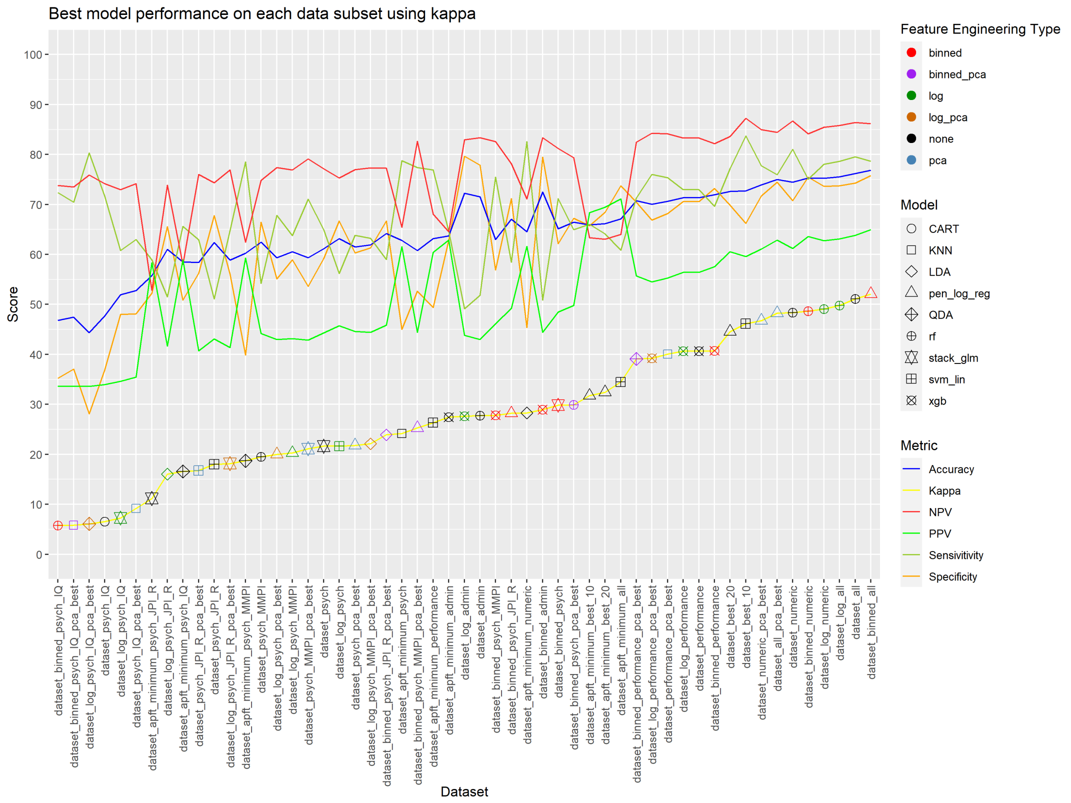


FIGURE 5.2: Best results for all models on all candidate, feature, and feature engineering data subset combinations using kappa

Kappa balanced all metrics and was best when all confusion matrix categories were all collectively better (i.e. TN and TP were high; FP and FN were low). Kappa was not impacted by the peaks and troughs in other metrics, but instead normalized all metrics by aggregation. This validates our choice for a performance metric described in section 3.4 to provide a balanced, holistic metric.

5.2 Models

The model tuning parameters and optimal thresholds for all models are in section B.1. The model coefficients for all applicable models are in section B.3. The CART trees for each data subset are in section B.4.

Figure 5.3 shows the total counts of best and worst performance for each model on all data subsets using kappa. This plot informs an aggregate idea of model performance.

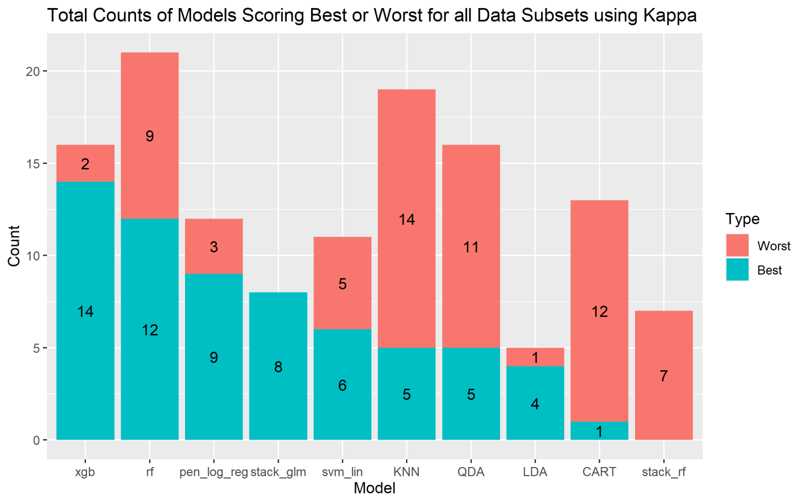


FIGURE 5.3: Count of Best and Worst Model Performance for all Data Subsets

Key Findings:

- Overall, the ensemble models performed best (with the exception of stack_rf). The aggregation and weighting methods of the ensemble models consistently outperformed other models.
- **xgboost, penalized logistic regression, stack_glm, SVM, and LDA** performed most consistently across all subsets (worst model <= 5 times).
- **Random forest, KNN, QDA, CART, and stack_rf** performance varied significantly across subsets (worst model >= 7 times).

Table 5.2 shows the best model kappa score on all feature, subset, and feature engineering combinations with all candidates. All of the models scored the best kappa score using the "all", "numeric", or "best_10" feature subsets.

TABLE 5.2: Best kappa scores on all models using all feature subset and feature engineering combinations with all candidates

Model	Feature Subset	Feature Engineering Type	Kappa	Accuracy	Precision PPV	Sensitivity	NPV	Specificity
pen_log_reg	all	binned	52.08	76.83	64.95	78.64	86.15	75.8
rf	all	binned	51.69	76.62	64.63	78.64	86.1	75.47
svm_lin	all	log	49.67	75.75	63.93	76.26	84.79	75.47
xgb	all	binned	48.33	75.11	63.18	75.37	84.22	74.96
stack_glm	numeric	none	47.62	74.46	61.85	77.45	84.98	72.76
CART	numeric	binned	44.14	72.41	59.14	77.74	84.54	69.37
LDA	numeric	binned	44.09	73.17	61.11	71.81	82.14	73.94
QDA	numeric	pca	43.97	72.2	58.76	78.64	84.91	68.53
stack_rf	numeric	pca	43.05	73.49	63.11	64.99	79.69	78.34
KNN	best_10	none	42.42	70.92	57.89	80.71	85.02	65.11

Figure 5.4 shows Table 5.2 plotted. The best model performance shown in Figure 5.4 is characteristic of model performance on other data subsets where kappa was sub-optimal. By visualizing model performance, we can understand the strengths and weaknesses of each model.

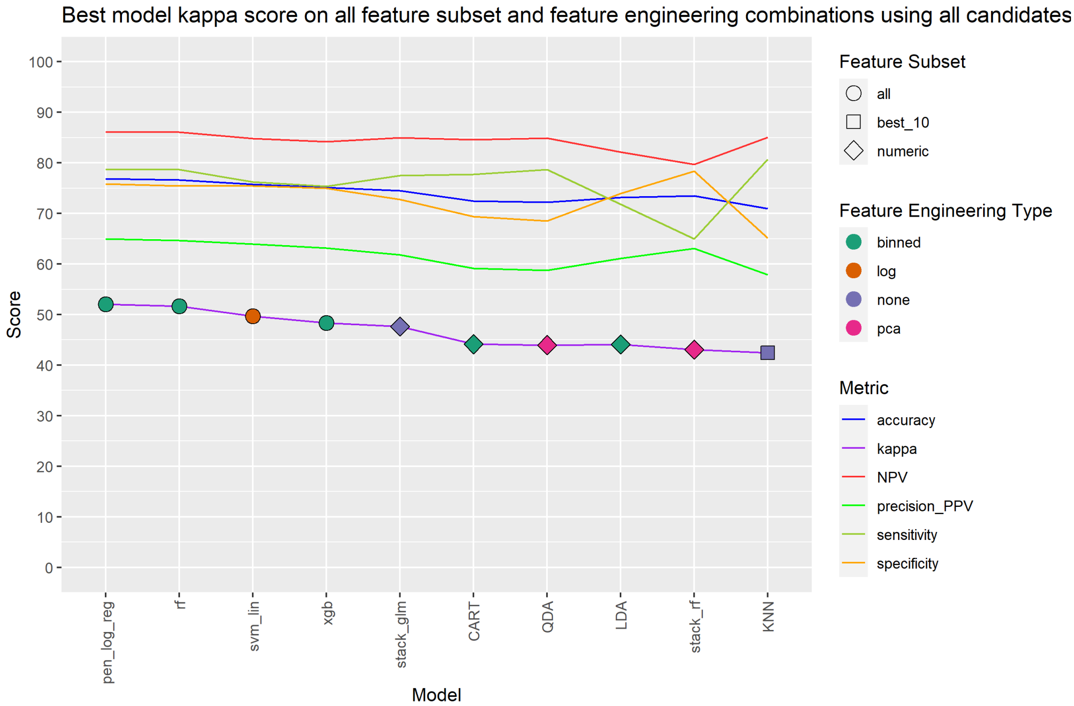


FIGURE 5.4: Best model kappa score on all feature subset and feature engineering combinations using all candidates

Key Findings:

- **pen_log_reg** performed best and **KNN** worst.
- All models scored within 7% NPV, PPV, and accuracy using the best kappa model score. These metrics indicate that models performed similarly in predicting both selection and non-selection.
- **xgb, pen_log_reg, stack_glm, random forest, SVM, and LDA** all scored over 70% and had 5% or less difference between specificity and sensitivity. These models were best at detecting which candidates actually were selected and not selected.
- **KNN, QDA, CART, and stack_rf** had a difference of 8% or more between sensitivity and specificity. **KNN, QDA, CART** had a higher sensitivity than specificity (i.e., the models classified more candidates as not selected than observed not selected), while **stack_rf** had a higher specificity than sensitivity (i.e., the model classified more candidates as selected than observed selected).
- Kappa scores ranged from 42-52. The higher the NPV, PPV, sensitivity, and specificity; the better the kappa score.

Figure 5.5 shows the best model performance on each feature subset using kappa with associated NPV, PPV, sensitivity, and specificity scores. There was not any specific model that dominated performance for all subsets. Instead, model performance varied based on the candidate, feature, and feature engineering subset combination.

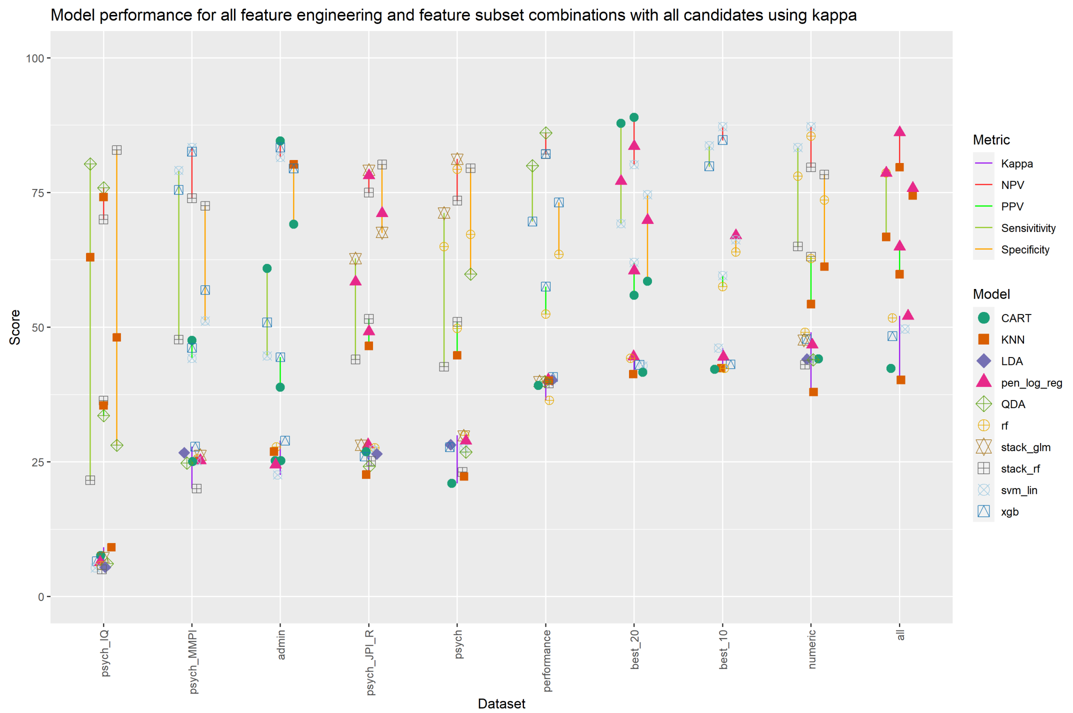


FIGURE 5.5: Best kappa score for all models on all feature subsets

The best model performance for all candidate, feature, and feature engineering type subsets were comparable on most subsets using kappa, NPV, and PPV (within a range of 10%). However, the sensitivity and specificity varied widely between models (12% or more) on all feature subsets with the exception of the "best_10" feature subset. This means that models performed similarly predicting selection and non-selection (NPV and PPV), but differed in the amount observed selected (sensitivity) and not selected candidates (specificity) classified correctly. **pen_log_reg** scored within the top three kappa scores on all feature subsets. **KNN** scored in the bottom three for all subsets except "psych_IQ", "psych_MMPI", and "admin" feature subsets.

5.3 Feature Subsets

Even though we are discussing feature subsets, it is important to note the distinctions between the candidate subsets as results varied significantly. Table 5.3 shows the best model score using kappa on all feature subset and feature engineering combinations for both "all" and "apft_min" candidate subsets.

TABLE 5.3: Best model score using kappa on all feature subset, candidate subset, and feature engineering combinations

Model	Feature Subset	Feature Engineering Type	Candidate Subset	Kappa	Accuracy	Precision PPV	Sensitivity	NPV	Specificity
xgb	admin	binned	all_candidates	28.94	72.5	44.43	50.88	83.38	79.48
xgb	admin	none	above_apft	27.43	63.71	62.9	64.64	64.54	62.8
pen_log_reg	all	binned	all_candidates	52.08	76.83	64.95	78.64	86.15	75.8
svm_lin	all	none	above_apft	34.48	67.13	71.09	60.87	64	73.76
svm_lin	best_10	none	all_candidates	46.14	72.75	59.57	83.7	87.23	66.24
pen_log_reg	best_10	none	above_apft	31.74	65.91	68.35	66.06	63.37	65.75
pen_log_reg	best_20	none	all_candidates	44.55	72.59	60.53	77.12	83.6	69.88

Table 5.3 continued from previous page

Model	Feature Subset	Feature Engineering Type	Candidate Subset	Kappa	Accuracy	Precision PPV	Sensitivity	NPV	Specificity
pen_log_reg	best_20	none	above_apft	32.42	66.16	69.42	64.13	63.07	68.44
rf	numeric	log	all_candidates	49.08	75.22	62.77	78.04	85.46	73.6
LDA	numeric	none	above_apft	28.29	64.54	61.6	82.61	71.11	45.39
xgb	performance	binned	all_candidates	40.74	71.95	57.54	69.61	82.18	73.18
svm_lin	performance	none	above_apft	26.33	63.19	60.45	76.92	68.04	49.39
rf	psych	binned_pca	all_candidates	29.92	66.48	49.78	64.97	79.33	67.23
KNN	psych	none	above_apft	24.16	62.81	61.54	78.73	65.46	45.04
KNN	psych_IQ	pca	all_candidates	9.16	52.75	35.48	63	74.15	48.1
QDA	psych_IQ	none	above_apft	16.58	58.52	58.94	65.62	57.95	50.88
pen_log_reg	psych_JPI_R	binned	all_candidates	28.22	67.05	49.21	58.43	78.17	71.17
stack_glm	psych_JPI_R	none	above_apft	11.16	55.79	58.32	58.92	52.86	52.24
xgb	psych_MMPI	binned	all_candidates	27.84	63	46.14	75.48	82.59	56.89
QDA	psych_MMPI	none	above_apft	18.71	60.23	59.27	78.51	62.45	39.85

Figure 5.6 shows Table 5.3 with all candidates (left) and "above_apft" (right) plotted side by side. The best subsets using all data are in order as shown on the left plot in Figure 5.6.

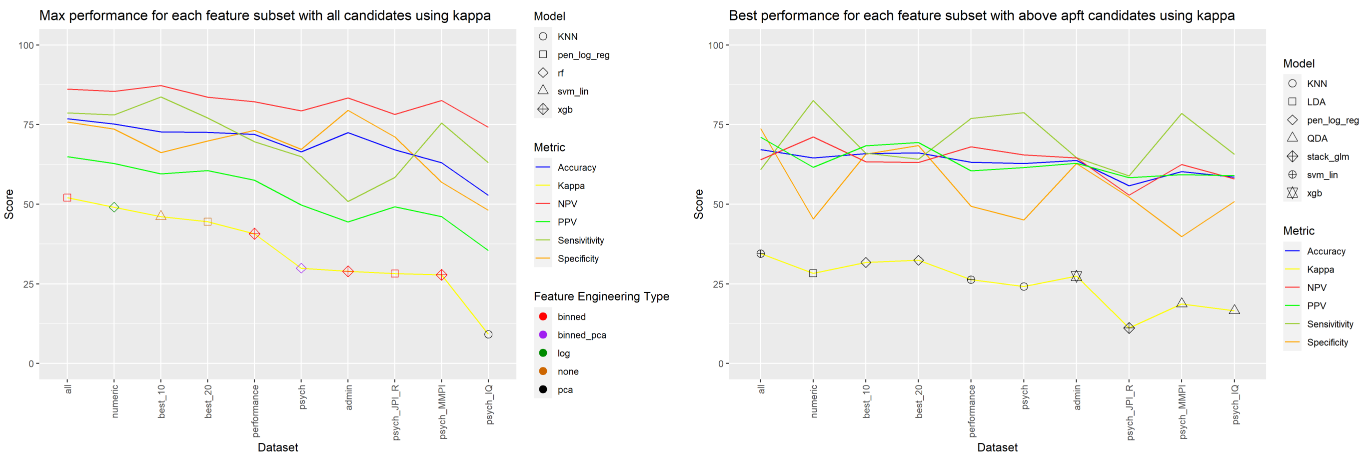


FIGURE 5.6: Best kappa score for each feature subset

Left: all candidates
Right: above_apft candidates

Key findings for candidate subset differences:

- For all feature subsets, the "all" candidate subset performed better at predicting non-selection (higher NPV and specificity) compared to the "above_apft" subset. This is because 1) there are more observations to classify non-selection using all candidates and 2) it is more difficult to discriminate non-selection after removing candidates screened out for fitness test scores.
- For all feature subsets, the "above_apft" candidate subset performed better at predicting selection (higher PPV and similar sensitivity) compared to the "all" subset. The classes were nearly balanced (51.5% selected; 48.5% not selected) in this candidate subset (n = 4,969). Given 1) the fewer observations and class balance and 2) having only fully qualified candidates after removing those screened out for fitness scores, this candidate subset had more variation in the selected class to inform better predictions.
- The "all" feature subset resulted in the best overall performance using both "all" and "apft_min" candidate subsets.

Key findings for feature subsets using all candidates:

- All candidate subsets had high NPV scores (74%-87%), while other metrics varied. This is significant because it shows that all subsets have unique information effective at discriminating not selected candidates.
- The **"all"** and **"numeric"** subsets were holistically the best feature subsets on almost all metrics. We think this is due to finding unique interaction effects using all the data compared only to specific subsets.
- The **"best_10"** and **best_20"** feature subsets performed only slightly worse than "all" and "numeric". This indicates that the feature selection methods (section 4.4) were effective at selecting the most important features in the data. The "best_10" subset had a lower specificity than "best_20", meaning it did not correctly classify as many candidates who were observed not selected compared to "best_20".
- The **"performance"** subset metrics almost mirrored the "best_20", with a small decrease in sensitivity and increase specificity. This is significant because the "performance" subset only has five features (4 of the 5 being fitness test related) and still performed comparable to the best subsets. This indicates physical fitness is highly predictive and is consistent with what we discovered in section 4.5 and section 4.4 about the importance of performance features.
- The **"admin"** subset performed best at non-selection (high NPV and highest specificity), but performed worst on selection classification (lowest sensitivity and second lowest PPV). The "admin" feature subset had the most observations (n=11,689) which is why the accuracy was still so high. Despite PPV being 44%, the model NPV was 83%. With the 75% class imbalance of candidates being not selected, the number of negative classifications caused the accuracy to still be greater than 70%, which is misleading. This is significant because only using administrative features was best at predicting non-selection.
- **Psych feature subsets.** Combining all three psych tests (the **"psych"** feature subset) yielded better results compared to any of the three individually. The **"psych"** resulted in a marginally better kappa score (2% increase) with a closer range between sensitivity and specificity compared to the other three tests. All psych tests performed poorly at candidate selection (< 50% PPV).

  "psych_IQ" resulted in the worst kappa across all subsets and is the least predictive psych test.

  Comparing the MMPI and JPI_R tests. "psych_MMPI" had a slightly better NPV (83%; 4% improvement over **"psych_JPI_R"**), but a lower specificity (56%; 14% lower than **"psych_JPI_R"**) compared to **"psych_JPI_R"**. Given the high NPV and specificity of the **"psych_JPI_R"** test, it yielded the best results for classifying non-selection for all psych tests, including compared to the **"psych"** subset that uses all three tests.

5.4 Feature Importance

The feature importance for all models using the "all" feature subset for both "all" and "above_apft" candidate subsets are in section B.2. Feature importance is calculated different ways for different models. The main idea is to give the most importance to the features that most impacted the model prediction. The feature importance score is standardized to a scale between 0-100, with the most important feature having a score of 100. Features with a score of 0 means that that feature was not used in model prediction at all. The full details for how feature importance is calculated for each model is in the caret package documentation.

Although this section is focused on feature importance, we will compare feature importance between "all" and "above_apft" candidate subsets. Aggregating feature importance between these two groups results in a loss of many important insights. Table 5.4 shows the top three most important features for all feature subsets using both "all" and "above_apft" candidate subsets for the "pen_log_reg" model. This model resulted in the best results for the "all" feature subset (see Figure 5.6) and performed in the top three on all subsets (see Figure 5.5). We argue this model provides a more objective interpretation of feature importance because the embedded feature selection removes features that are not important.

TABLE 5.4: Top 3 most important features for each candidate subset feature subset using Penalized Logistic Regression

	All Candidates			Above APFT Minimum Candidates		
Feature Subset	**1**	**2**	**3**	**1**	**2**	**3**
All	mos_18X	rank_CPL_SGT	mos_OTHER_MOS	rank_PVT_PV2	mos_OTHER_MOS	apft_1_score
Admin	mos_OTHER_MOS	arrival_month_MAY	rank_PVT_PV2	mos_OTHER_MOS	arrival_month_MAY	rank_PVT_PV2
Numeric	apft_1_score	rc4	apft_1_run	apft_1_score	rc4	s10
Performance	apft_1_score	gt_score	apft_1_pu	apft_1_score	gt_score	apft_1_run
Psych	s10	sfd	rc4	s4	rc4	s13
Psych_JPI_R	mis	s10	s13	s4	s10	s1
Psych_IQ	full_scale_iq	performance_iq	verbal_iq	performance_iq	full_scale_iq	verbal_iq
Psych_MMPI	pct_true	rc4	vri_nr	rc4	aes	rbs
Best_10	apft_1_score	mos_OTHER_MOS	rank_PVT_PV2	rank_PVT_PV2	mos_OTHER_MOS	rank_SPC
Best_20	apft_1_score	mos_OTHER_MOS	rank_PVT_PV2	rank_PVT_PV2	mos_OTHER_MOS	rank_SPC

Key Findings:

- **After aggregation (Figure 5.9), the overall the most important feature categories (in order) were 1) performance, 2) administrative, and 3) psychological.** The most important performance feature was "apft_1_score" (higher score better; lower score worse). The most important administrative feature was "mos_OTHER_MOS" (combat specialty better; non-combat specialty worse). The most important psychological feature was "s10" or organization (higher score better; lower score worse).

- Using all candidates, "apft_1_score" was most important for all feature subsets that had that feature except the "all" subset. For the "above_apft" candidates, "apft_1_score" decreased in importance to the third most important. Removing all candidates who score low on the fitness test makes the fitness score have less variance and less predictive power.

- Using "above_apft" candidates, "rank_PVT_PV2" and "mos_MOS_OTHER" were the two most important features. This means that low rank and a non-combat specialty are highly predictive of candidate non-selection.

- The top three most important features changed between "all" candidates and "above_apft" candidates for all feature subsets except the performance subset. This is significant because it means different things matter between those candidate subsets.

Figure 5.7 shows both the top 20 feature importance for the "all" feature subset for both "all" (left) and "above_apft" (right) candidate subsets. This plot makes clear the magnitude of difference between the feature importance. Using "all" candidates (left), the feature importance distribution is more balanced. However, using "above_apft" candidates (right), the top three features were two times as important as the fourth feature ("rank_PVT_PV2" = 100, "mos_OTHER_MOS"=98, "apft_1_score"=76, and arrival_month_MAY=41).

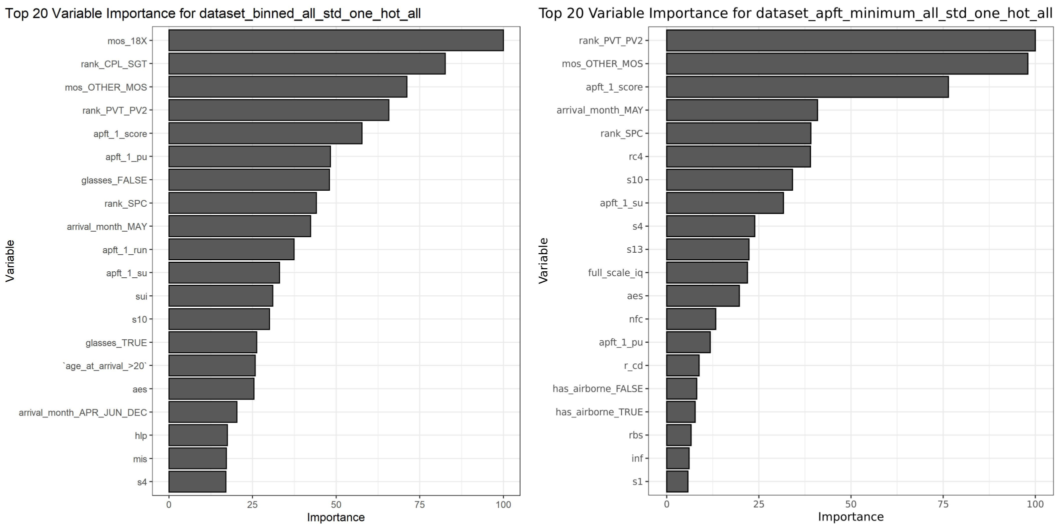


FIGURE 5.7: Top 20 Most Important Features for dataset_all using Penalized Logistic Regression

Left: "all" candidate subset
Right: "above_apft" candidate subset

Figure 5.8 shows both the top 20 feature importance for the "all" feature subset for both "all" (red) and "above_apft" (blue) candidate subsets. This shows similar information to Figure 5.7, but shows the features side by side to better visualize the change in feature importance between candidate subsets. "rank_PVT_PVT2", "mos_OTHER_MOS", and "apft_1_score" are the three most important for the "above_apft" candidate subset, but are less important for the "all" subset.

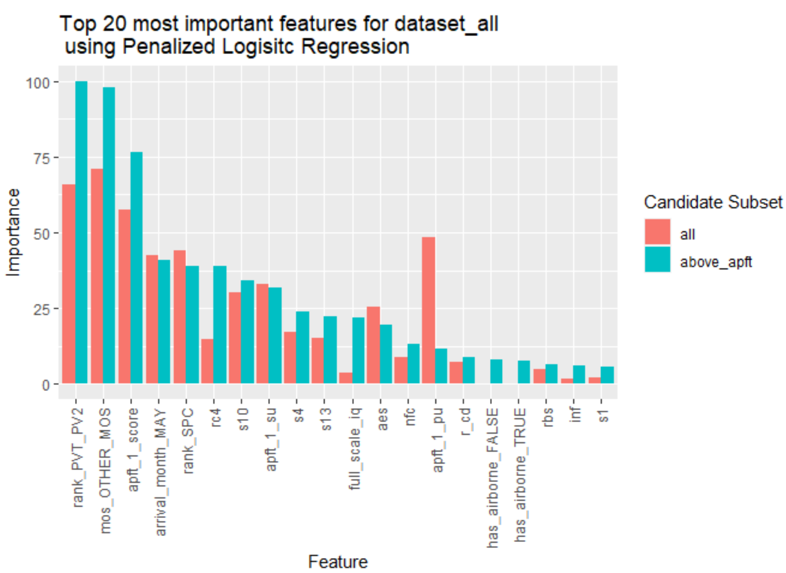


FIGURE 5.8: Top 20 Most Important Features for dataset_all using Penalized Logistic Regression

Figure 5.9 shows the combined feature importance for all models with all candidates for "dataset_all". The feature importance was aggregated by adding up all the values from each of the six models (max score = 600; 6 different models). The plot makes clear how important each feature was in each respective model. In general, feature importance was similar across all models with "apft_1_score" being the best ranked feature by all models except pen_log_reg (which performed the best).

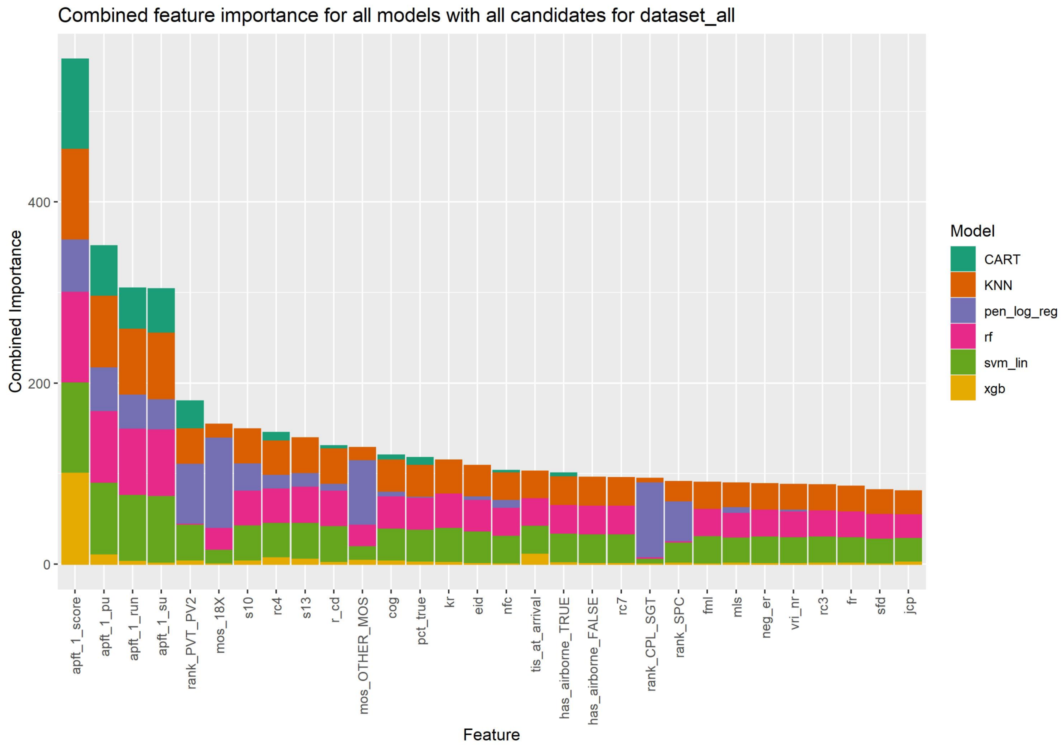


FIGURE 5.9: Combined Feature Importance for all models with all candidates for dataset_all (Top 30 Features)

It is evident how the "pen_log_reg" model only selected features that are most important using the embedded feature selection. For example, all other models ranked "apft_1_score" as the most important (100); the "pen_log_reg" instead assigned it an importance of 58. With the "psych" features, "pen_log_reg" assigned most importance to "rc4", "s10", and "s13"; while other models assigned similar importance to 15 other psych features. By removing features that provide the same predictive power from the model, the "pen_log_reg" model was able to identify other features that were important (e.g., "rank_PVT_PV2", "mos_18X" and "mos_OTHER_MOS") and outperform the other models. For these reasons, we conclude that the "pen_log_reg" model provides the best representation of most important features.

Figure 5.10 shows the combined feature importance for all models for the "psych" feature subset. The feature importance was aggregated by adding up all individual model feature importance scores (max score = 800; 8 different models). With the above comments in mind about the "pen_log_reg" model's ability to best represent feature importance, this plot still informs an overall idea of aggregate feature importance. In general, psych feature importance was similar (both blue and red lines overall trend the same way); however, there were clear changes between candidate subsets. For example, "full_scale_iq" and "s13" were more important for "above_apft" candidates, while "axy" and "mis" were more important for "all" candidates. There are other features like "nfc" that had similar importance for both candidate subsets.

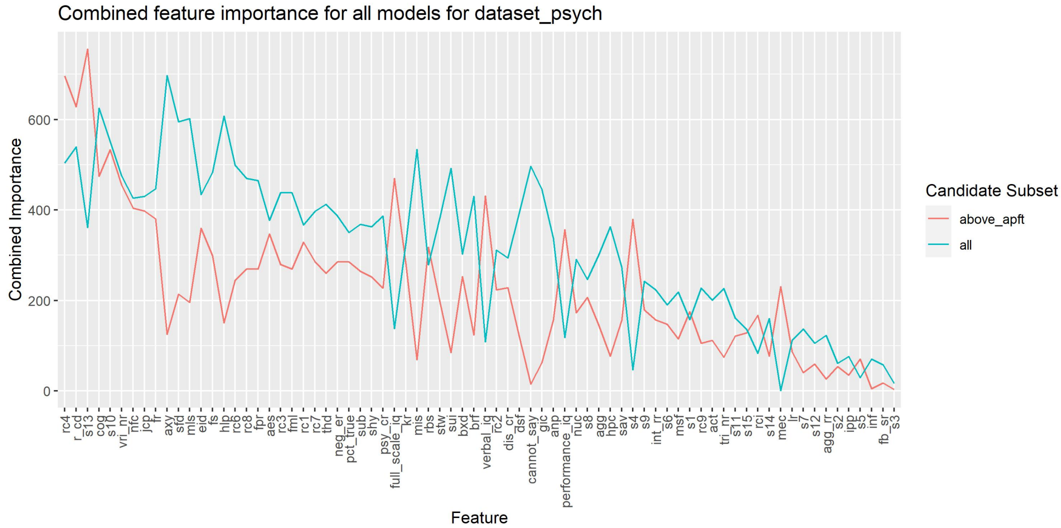


FIGURE 5.10: Combined feature importance for all models for dataset_psych

5.5 Candidate Subsets

We performed analysis on four candidate subsets (see subsection 4.6.2). While we learned valuable insights from all four subsets, we discuss the "above_apft" candidate subset in this section, as this subset informed the most valuable insights.

Table 5.5 shows the best performance of all models on all feature and candidate subsets using kappa. Figure 5.11 shows Table 5.5 visually. Each feature subset is on the x axis and the respective metric score on the y axis. The red line represents candidates who scored above the minimum APFT score and the blue line represents all candidates.

TABLE 5.5: Best model performance using kappa for all feature subsets and candidate subsets

feature_subset	candidate_subset	feature_engineering_type	model	kappa	accuracy	NPV	specificity	precision_PPV	sensitivity
admin	above_apft	none	xgb	27.43	63.71	64.54	62.8	62.9	64.64
admin	all_candidates	binned	xgb	28.94	72.5	83.38	79.48	44.43	50.88
all	above_apft	none	svm_lin	34.48	67.13	64	73.76	71.09	60.87
all	all_candidates	binned	pen_log_reg	52.08	76.83	86.15	75.8	64.95	78.64
best_10	above_apft	none	pen_log_reg	31.74	65.91	63.37	65.75	68.35	66.06
best_10	all_candidates	none	svm_lin	46.14	72.75	87.23	66.24	59.57	83.7
best_20	above_apft	none	pen_log_reg	32.42	66.16	63.07	68.44	69.42	64.13
best_20	all_candidates	none	pen_log_reg	44.55	72.59	83.6	69.88	60.53	77.12
numeric	above_apft	none	LDA	28.29	64.54	71.11	45.39	61.6	82.61
numeric	all_candidates	log	rf	49.08	75.22	85.46	73.6	62.77	78.04
performance	above_apft	none	svm_lin	26.33	63.19	68.04	49.39	60.45	76.92
performance	all_candidates	binned	xgb	40.74	71.95	82.18	73.18	57.54	69.61
psych	above_apft	none	KNN	24.16	62.81	65.46	45.04	61.54	78.73
psych	all_candidates	binned_pca	rf	29.92	66.48	79.33	67.23	49.78	64.97
psych_IQ	above_apft	none	QDA	16.58	58.52	57.95	50.88	58.94	65.62
psych_IQ	all_candidates	pca	KNN	9.16	52.75	74.15	48.1	35.48	63
psych_JPI_R	above_apft	none	stack_glm	11.16	55.79	52.86	52.24	58.32	58.92
psych_JPI_R	all_candidates	binned	pen_log_reg	28.22	67.05	78.17	71.17	49.21	58.43
psych_MMPI	above_apft	none	QDA	18.71	60.23	62.45	39.85	59.27	78.51

Table 5.5 continued from previous page

feature_subset	candidate_subset	feature_engineering_type	model	kappa	accuracy	NPV	specificity	precision_PPV	sensitivity
psych_MMPI	all_candidates	binned	xgb	27.84	63	82.59	56.89	46.14	75.48

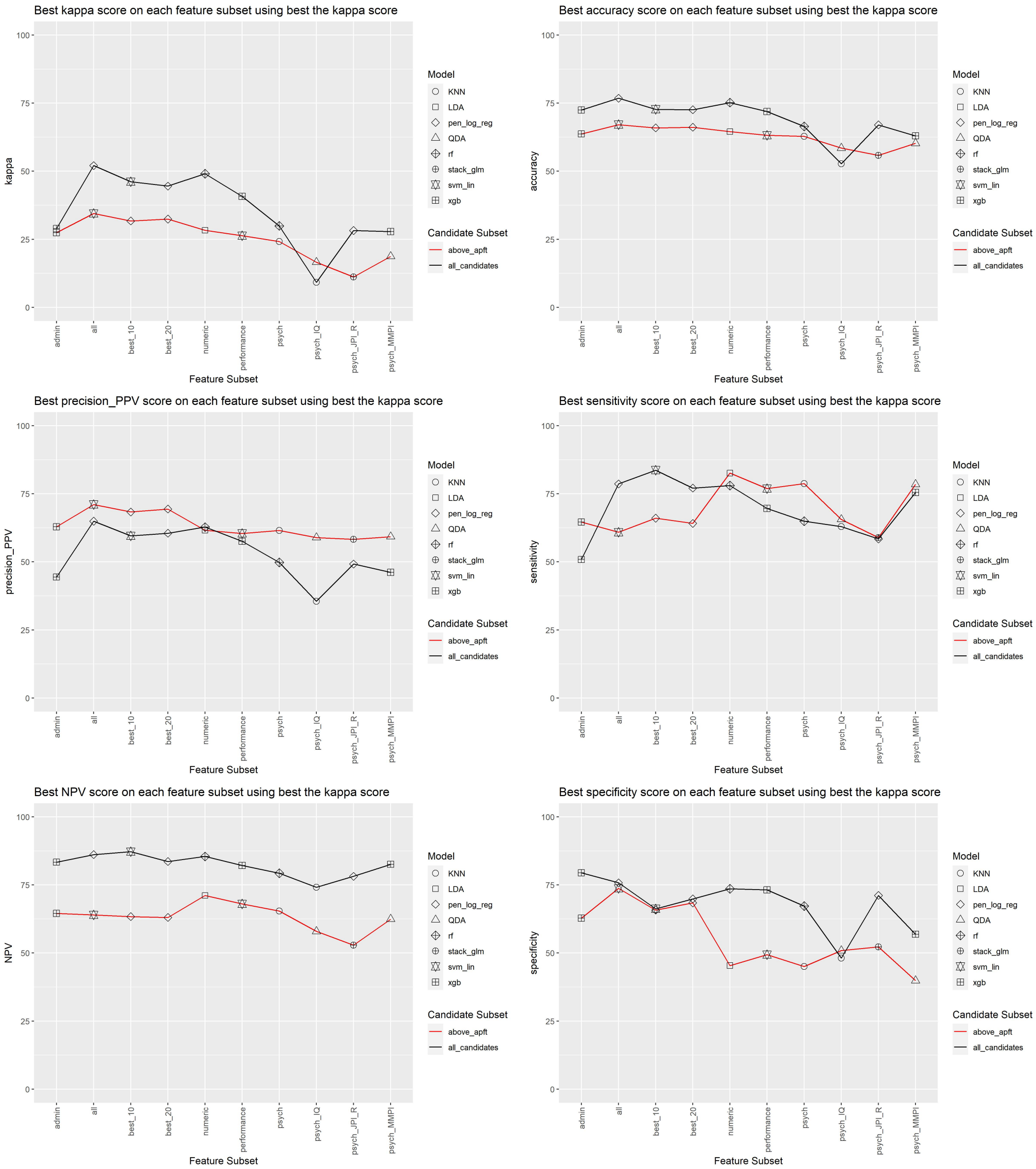


FIGURE 5.11: Max performance of feature subsets comparing candidate subsets using kappa, accuracy, PPV, sensitivity, NPV, and specificity

Key findings:

- Kappa was higher using "all" candidates for all subsets except "psych_IQ". This is because the "psych_IQ" subset increased in PPV by 26% in the "above_apft" candidates, while other metrics were similar. Despite this improvement, the "psych_IQ" still had an overall low kappa score (17).
- PPV was higher for all feature subsets and sensitivity was higher for all subsets except the "all", "best_10", and "best_20" feature subsets for "above_apft" candidates. This indicates that the "above_apft" candidate subset provided more predictive power for selected candidates.

- NPV and specificity was higher for "all" candidates for all feature subsets (except specificity for the "psych_IQ" feature subset). This means models could better discriminate and predict not selected candidates using "all" candidates.

5.6 Feature Engineering

This section only includes the "all" candidate subset as we only performed feature engineering on that candidate subset. Additionally, we did not perform feature engineering on the "best_10" or "best_20" feature subsets. Figure 5.12 shows the feature engineering method impact on performance for the "dataset_psych" feature subset. All methods except "log_pca" resulted in the same or better kappa score with "binned" and "binned_pca" improving results in all metrics.

FIGURE 5.12: Feature engineering method impact on performance for the dataset_psych feature subset

Table 5.6 shows shows the best model performance using kappa for all feature subsets using feature engineering compared to not using feature engineering. Figure 5.13 shows Table 5.6 visually. The red line shows scores using feature engineering and the black line shows scores without feature engineering for each metric.

TABLE 5.6: Best model performance using kappa for all feature subsets and feature engineering types

feature_subset	feature_engineering_type	model	kappa	accuracy	NPV	specificity	precision_PPV	sensitivity
admin	binned	xgb	28.94	72.5	83.38	79.48	44.43	50.88
admin	none	rf	27.75	71.51	83.36	77.86	43.01	51.81
all	binned	pen_log_reg	52.08	76.83	86.15	75.8	64.95	78.64
all	none	rf	51.1	76.19	86.42	74.28	63.81	79.53
numeric	log	rf	49.08	75.22	85.46	73.6	62.77	78.04
numeric	none	rf	48.37	74.46	86.72	70.73	61.21	81.01
performance	binned	xgb	40.74	71.95	82.18	73.18	57.54	69.61
performance	none	xgb	40.66	71.38	83.32	70.57	56.41	72.94
psych	binned_pca	rf	29.92	66.48	79.33	67.23	49.78	64.97
psych	none	stack_glm	21.63	61.11	77.16	59.18	44.32	64.97
psych_IQ	pca	KNN	9.16	52.75	74.15	48.1	35.48	63
psych_IQ	none	CART	6.54	47.71	74.16	36.86	33.96	71.66

Table 5.6 continued from previous page

feature_subset	feature_engineering_type	model	kappa	accuracy	NPV	specificity	precision_PPV	sensitivity
psych_JPI_R	binned	pen_log_reg	28.22	67.05	78.17	71.17	49.21	58.43
psych_JPI_R	none	svm_lin	18.02	62.39	74.36	67.77	43.13	51.12
psych_MMPI	binned	xgb	27.84	63	82.59	56.89	46.14	75.48
psych_MMPI	none	rf	19.53	62.43	74.79	66.45	44.15	54.21

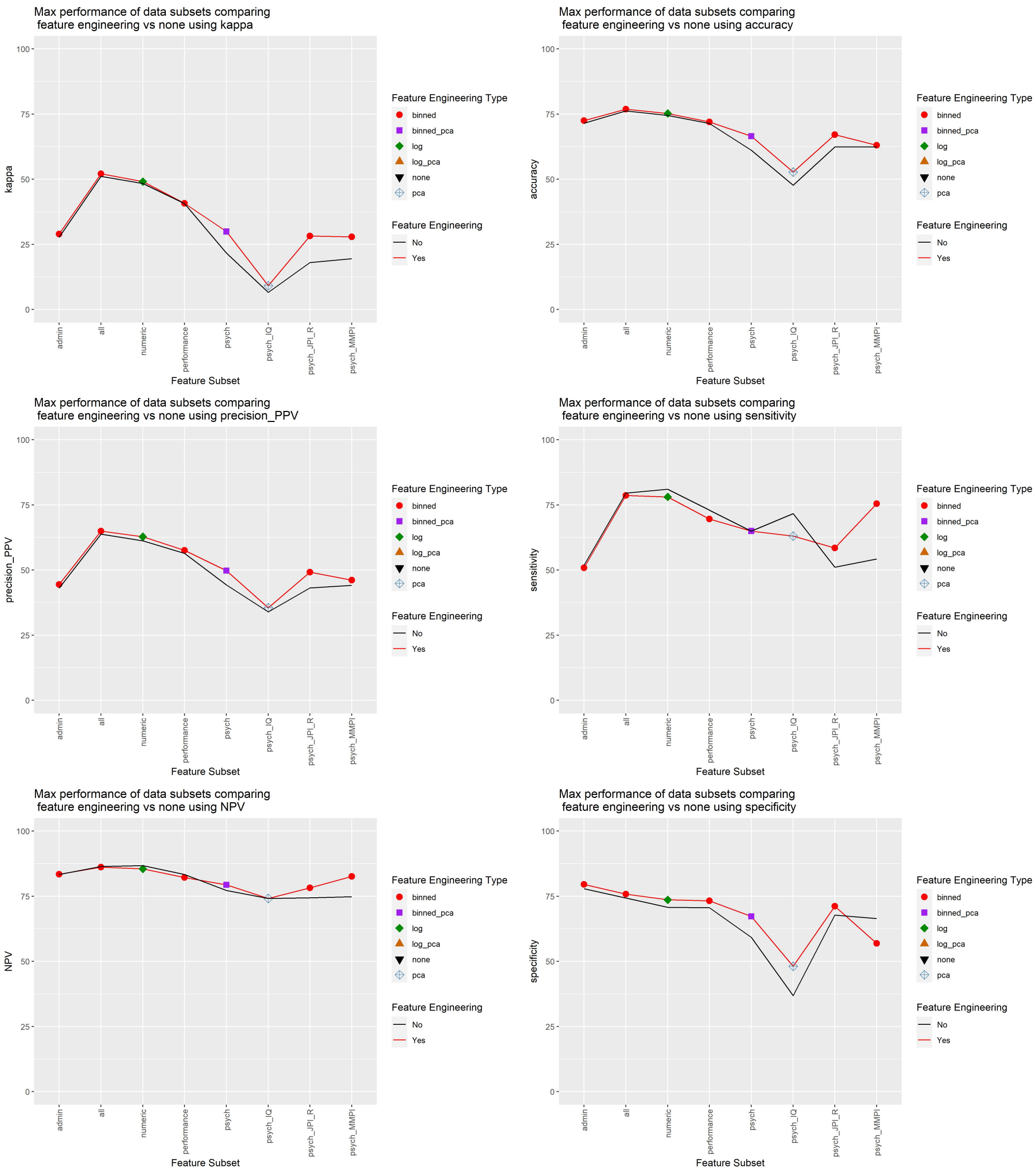


FIGURE 5.13: Max performance of data subsets comparing feature engineering vs. none using kappa, accuracy, PPV, sensitivity, NPV, and specificity for all candidates

Key findings:

- For all feature subsets, feature engineering resulted in better kappa, accuracy, and PPV scores.
- The "psych_JPI_R" and "psych_MMPI" feature subsets had the greatest improvement using feature engineering. The "binned" transformation improved the "psych_JPI_R" in all metrics and the "psych_MMPI" in all but specificity.

This finding is significant, because it means that predictive power is increased. We think this is because all the features were related (i.e., come from the same psych test) so binning and/or PCA was effective at extracting additional information from the data.

- Feature engineering led to least improvement on the "numeric", "all", "admin", and "performance" feature subsets.

Feature engineering also impacted feature importance. Figure 5.14 shows feature importance with no feature engineering (left) compared to using the "binned" technique (right) for the "all" feature subset using the "pen_log_reg" model. For this specific model and feature subset, binning improved kappa, accuracy, NPV, PPV, sensitivity, and specificity by 2% (and resulted in the best overall performance out of all data subset combinations). Binning the continuous features changed their distributions such that their interaction effects made other features more important for an overall better model. In this case, the feature distribution was more evenly distributed with admin features becoming more important than performance.

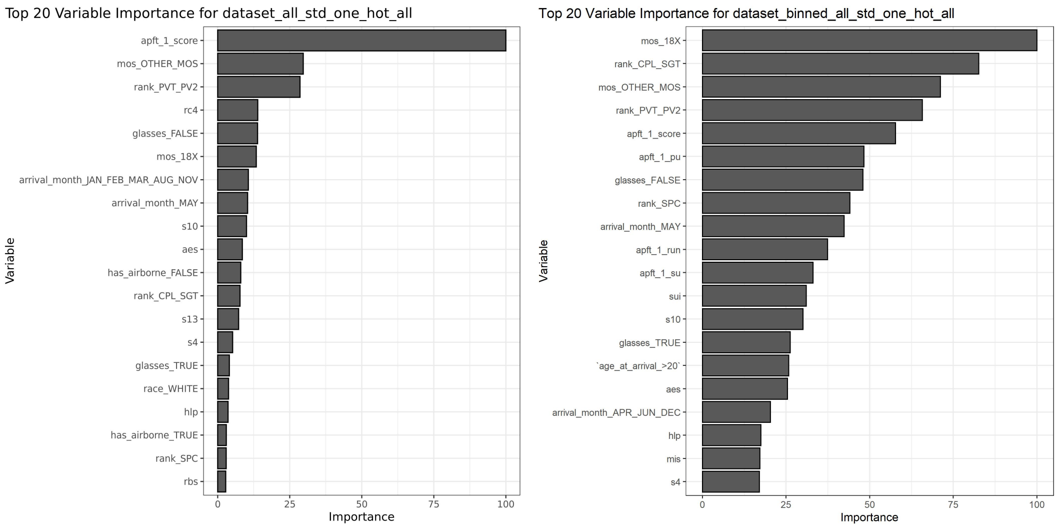


FIGURE 5.14: Feature importance for "all" feature subset with no feature engineering vs. binned using pen_log_reg

Feature engineering methods prove to be highly valuable. They enabled us to discover more meaningful feature relationships and improve model performance.

5.7 Data Imputation

While we did not use imputed data for any of our final analysis results, we wanted to understand the impact it had on model performance. Table 5.7 shows the best model performance on all feature subsets using kappa with imputed versus missing data with associated accuracy, PPV, sensitivity, NPV, and specificity metrics. Figure 5.15 shows Table 5.7 visually for each respective metric. The black line represents missing data and the red line represents imputed data.

TABLE 5.7: Best model performance using kappa for all feature subsets using imputed data vs. missing data

feature_subset	data_imputation	feature_engineering_type	model	kappa	accuracy	NPV	specificity	precision_PPV	sensitivity
admin	imputed	none	KNN	31.02	73.71	83.32	81.36	47.03	50.4
admin	missing	binned	xgb	28.94	72.5	83.38	79.48	44.43	50.88
all	imputed	none	rf	52.16	80.61	90.65	82.78	58.53	74.01
all	missing	binned	pen_log_reg	52.08	76.83	86.15	75.8	64.95	78.64
numeric	imputed	none	xgb	47.62	77.89	90.65	78.76	53.77	75.26
numeric	missing	log	rf	49.08	75.22	85.46	73.6	62.77	78.04
performance	imputed	pca	stack_glm	47.76	76.18	93.83	73.16	51.09	85.36
performance	missing	binned	xgb	40.74	71.95	82.18	73.18	57.54	69.61
psych	imputed	none	rf	23.49	69.14	82.14	75.4	40.05	50.06
psych	missing	binned_pca	rf	29.92	66.48	79.33	67.23	49.78	64.97
psych_IQ	imputed	none	xgb	13.5	59.79	80.79	61.13	32.01	55.73
psych_IQ	missing	pca	KNN	9.16	52.75	74.15	48.1	35.48	63
psych_JPI_R	imputed	none	xgb	20.7	64.59	82.91	66.72	36.44	58.12
psych_JPI_R	missing	binned	pen_log_reg	28.22	67.05	78.17	71.17	49.21	58.43
psych_MMPI	imputed	none	rf	23.32	71.02	81.23	79.99	41.76	43.7
psych_MMPI	missing	binned	xgb	27.84	63	82.59	56.89	46.14	75.48

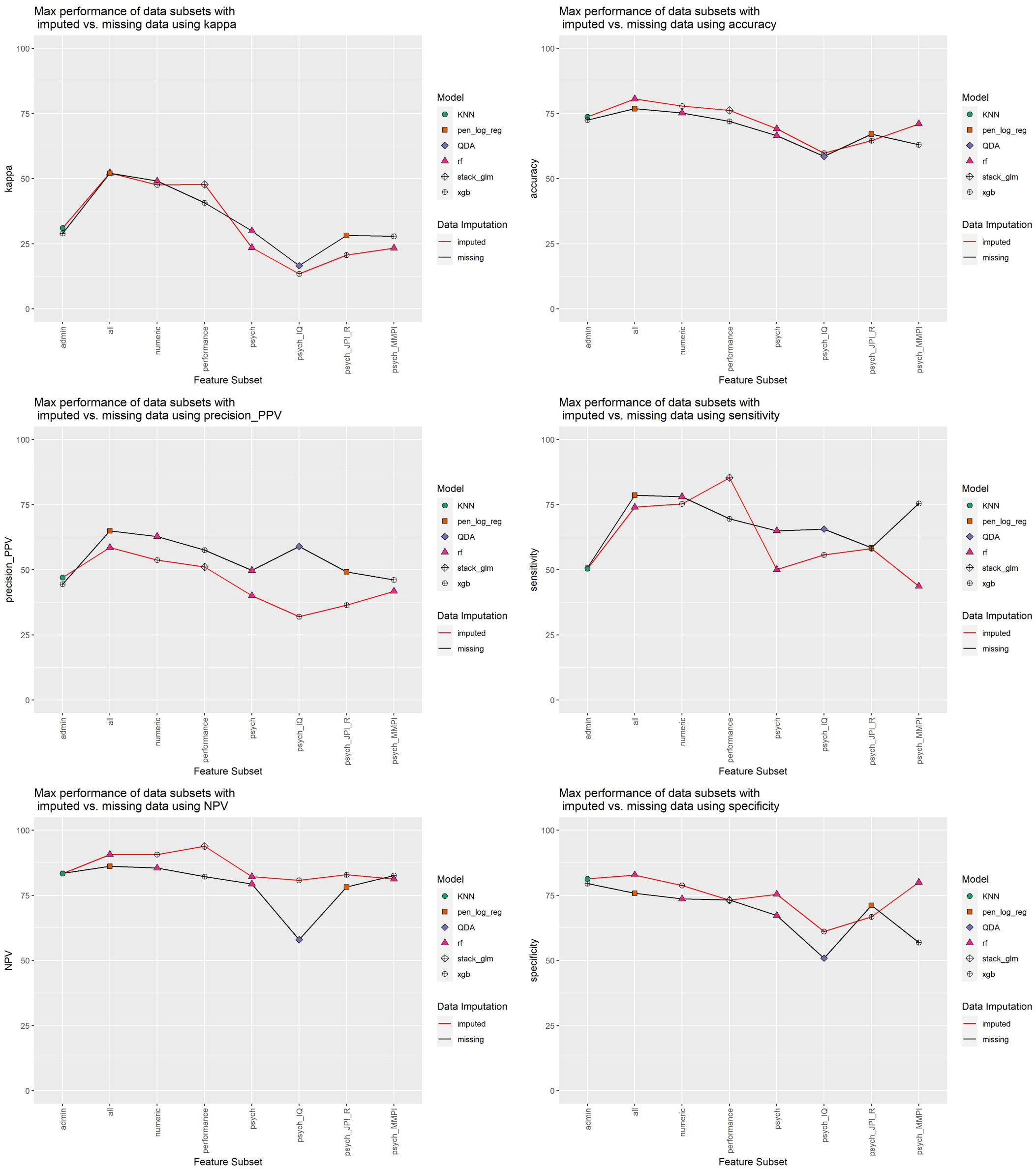


FIGURE 5.15: Best model performance on all feature subsets using kappa with imputed vs. missing data with associated accuracy, PPV, sensitivity, NPV, and specificity metrics

Key findings:

- Imputed data improved NPV, specificity, and accuracy on 8/9 feature subsets. We think that the imputed data created similar observations to existing candidates, enabling models to better predict non-selection. Because of the class imbalance, it improved accuracy.
- Non-imputed data outperformed imputed data for PPV and sensitivity on 8/9 subsets. This likely means the imputed data added more noise to the selected candidates to make them more difficult to discriminate.
- The kappa score was better in all non-imputed data subsets except "performance" and "admin" feature subsets. This is because, for both subsets, even though PPV decreased, sensitivity, specifity, and NPV increased. The amount of observations in each subset caused kappa to slightly increase.

Chapter 6

Conclusions and Recommendations

6.1 Conclusions

Answers to research question:

1. Is application of data science techniques to military A&S data useful to gain insight on ways to holistically improve the system?

   Yes. Although prediction accuracy is modest (<90%), applying data science techniques adds value to the A&S.

   The following are value added to the A&S:

   - **Data Tidying.** By creating a combined data set using administrative, performance, and psychological features (section 4.1), we learned holistic feature relationships (section 4.5) and importance (section 5.4) across all features combined instead of "stove-piped". Analysis on the three types of feature subsets together had not been done before using this data.
   - **Feature Screening.** By screening features (section 4.2), we removed noise from the data, using only features containing the most useful information and returning the best results.
   - **Data Exploration**. By exploring the data (section 4.5), we visualized the data in powerful ways to observe different data distributions, meaningful relationships between the features, and why data was missing. We also performed statistical testing to provide a more objective understanding of the differences between selected and not selected candidates. All of these insights are valuable for the A&S unit to make informed decisions.
   - **Feature Engineering.** By performing feature engineering (section 4.3), we modified the existing features and created new features to better represent the information contained in the data. Through this analysis we were: 1) able to extract and visualize unique underlying relationships of features that we would not have otherwise and 2) improve holistic model performance on all feature subsets (section 5.6).
   - **Feature Selection.** By performing feature selection (section 4.4), we developed an informed intuition behind what features were most important and why. This importance was later confirmed by the model results for feature importance (section 5.4).
   - **Creating Data Subsets**. By creating numerous subset combinations using feature subsets, candidate subsets, and feature engineering techniques

(section 4.6); we learned unique relationships that exist and what combinations achieved the best results (section 5.3, section 5.5, and section 5.6).

- **Model Fitting.** By fitting ten optimally tuned models to the data (section 4.7), we learned more objective information about what models perform best given the unique subset combinations and how to best predict candidate selection and non-selection (section 5.1).

*** **After considering these insights from the holistic data science approach applied to that A&S data, we made actionable recommendations to the A&S unit (section 6.2).**

2. Are predictive models useful and accurate to model selection in Military A&S?

Useful: Yes. We demonstrated how the predictive models revealed meaningful and useful things about the data to add value to the A&S. See answer to research question 1 above.

Accurate: Yes and No.

Yes for predicting non-selection. Using all candidates, the best model had a NPV of 86% and sensitivity of 75% (see Figure 5.6). The results indicate that the features within our data set have a more objective predictive capability to predict non-selection. This idea is also visually evident by comparing features to the response, such as in Figure 4.19. Looking at the "rc4" feature, it is evident that candidates who had low scores were selected at a lower rate, making it easier for models to correctly classify candidates. We submit that the non-selection prediction can be value added for the unit to incorporate into its current screening criteria (see section 6.2).

No for predicting selection. Using all candidates, the best model had a PPV of 65% and specificity of 78% (see Figure 5.6). We conclude that models performed poorly on classifying candidate selection due to intangible features in the human dimension that are highly variable and difficult to model. For example, a candidate's grit may be indicative of if a candidate will quit or not. Likewise, a candidate may fit the mold of an "all-star" candidate, but get injured or unexpectedly quit. Figure 4.15 highlights this idea using fitness features. This idea is also evident looking at Figure 4.20. While the fitness score shows a clear linear trend of the higher a candidate's score, the greater percent of candidates were selected; only 68% of the candidates who scored the highest fitness score were selected. Most other features only had around a 50% maximum of candidates who were selected at best.

The modest selection accuracy is significant because it validates the need for an in-person assessment to observe the human dimension of how candidates respond to challenges.

Candidates who scored above the fitness test screening score. Using "above_apft" candidates, the best model had a PPV of 71% and NPV of 64% (see Figure 5.6). In this candidate subset, the results indicate there is more information to classify candidate selection (6% higher PPV compared to all candidates), but less information in the data to classify non-selection (22% lower NPV compared to all candidates) compared to all candidates. Removing all candidates who scored above the minimum fitness test screening criteria removed information in the data to discriminate non-selection, but added more information to discriminate selection. We think this further points to the importance of the human dimension for this subset, because of the decreased model performance compared to using all candidates (17% lower kappa score).

3. Are there any features that are indicative of a candidate being selected or not selected?

   Yes. See section 5.4.

6.2 Recommendations

1. **Screening criteria.** We recommend using our models for consideration in the A&S unit screening criteria. The best performing models classified more than 85% of not selected candidates correctly (NPV), that included more than 75% of the candidates who were actually selected (sensitivity) (see Figure 5.6). We think using a soft classification (i.e., percent chance of a candidate being selected) for candidates would be value added to consider when screening candidates. For example, if a candidate's soft classification is 0.05, that candidate should be screened. However, if a candidates soft classification is 0.95, that candidate should be retained.

   We acknowledge there is an ethical consideration to using a predictive model as screening criteria (section 2.3). As a result, we are not suggesting the models be used as the primary screening criteria, but rather as another tool to consider. Further, the models are not "black box" and provide associated feature importance in making the classification.

   We have some other specific recommendations for objective screening criteria with different combinations of features (e.g., candidate scores more than X score on Y feature); however, the specific details are outside the scope of this report.

2. **Psychological Tests**. The psychological testing consumes time and resources. If the purpose of the psychological testing is exclusively for screening candidates, we recommend using only the JPI_R as it proved to be holistically the best test at screening candidates section 5.3.

   Additionally, we discovered which specific features were most important from the psychological features (see Table 5.4). We think it is worth exploring other options to measure only the most important psychological features that are predictive. There may be a more efficient solution possible compared to using an entire psychological test.

3. **Additional Data Collection.** It is evident that the predictors using the data set for this research does not contain the information needed for high predictive capability (>90% PPV, NPV, sensitivity, and specificity).

   Does the modest model performance (best model: 86% NPV, 75% specificity, 65% PPV, and 79% specificity) imply anything about the A&S? We think it may imply that there remains additional data that could be collected that may be more predictive of candidate selection and non-selection. We think these predictors would likely be in the human dimension. There has been research on how to best capture these features (e.g., grit or mental toughness) that we recommend the unit investigating. This data could be collected during the screening phase and considered when ranking candidates on the order of merit list.

4. **Evaluation Criteria.** We think the feature importance may provide insight to evaluation criteria (see section 5.4). Using "pen_log_reg", for "all" candidates (the best model results), feature importance was (in order) "mos_18X" (100%), "rank_CPL_SGT" (83%), and "mos_OTHER_MOS" (71%). For "above_apft" candidates, feature importance was (in order) "rank_PVT_PV2" (100%), "mos_OTHER_MOS" (98%), and "apft_1_score" (76%) (Figure 5.7). For both candidate subsets, fitness test score was in the top five most important features, which aligns with physical fitness being an important attribute in candidates for military A&S (see subsection 2.4.1). However, for both candidate subsets, combat specialty and senior ranking candidates were most important for selection, while non-combat specialty and junior ranking candidates were most important in non-selection.

   Does feature importance imply anything about the A&S? Could junior Soldiers be lacking general military experience to be successful? Are non-combat specialties expected to do the same things combat specialties do in the unit? Are there systematic reasons why these two features are among the most important? Does this align with unit expectations and goals for the A&S? Could this be accounted for somehow in the screening criteria, evaluation criteria, and/or modifying the A&S? While we do not have answers to these questions, we think the feature importance warrants an investigation of these questions.

6.3 Future Work

We plan to explore the following items for future work:

1. **Long-Term Candidate Data.** The scope of this research focused only on candidate selection or non-selection during the A&S. While this is important, the question remains: Are the right candidates being selected at that A&S? To more objectively answer to this question, long-term data about selected candidate performance in the elite military unit needs to be analyzed to verify the candidate performance in the unit in comparison to performance at the A&S.

2. **Apply Data Science to other A&S Data.** We think there would be value in doing similar analysis to other military A&S and comparing results. This would add value to other A&S and better inform what matters in A&S. We would also like to apply these data science methods using additional features not in our data to explore other features that may be predictive.

Appendix A

Data Details

A.1 Complete List of Feature Subsets

TABLE A.1: Complete list of features in data subsets

Number	Feature	All	Admin	Numeric	Performance	Psych	Psych_JPI_R	Psych_IQ	Psych_MMPI	Best_10	Best_20
1	mos	x	x							x	x
2	rank	x	x							x	x
3	race	x	x								
4	arrival_month	x	x								
5	tis_at_arrival	x	x	x							x
6	parents_together	x	x								
7	has_airborne	x	x								
8	glasses	x	x								
9	civilian_education_certification	x	x								
10	age_at_arrival	x	x								
11	gt_score	x		x	x						
12	apft_1_pu	x		x	x					x	x
13	apft_1_su	x		x	x					x	x
14	apft_1_run	x		x	x					x	x
15	apft_1_score	x		x	x					x	x
16	s1	x		x		x	x				
17	s2	x		x		x	x				
18	s3	x		x		x	x				
19	s4	x		x		x	x				
20	s5	x		x		x	x				
21	s6	x		x		x	x				
22	s7	x		x		x	x				
23	s8	x		x		x	x				
24	s9	x		x		x	x				
25	s10	x		x		x	x			x	x
26	s11	x		x		x	x				
27	s12	x		x		x	x				
28	s13	x		x		x	x				x
29	s14	x		x		x	x				
30	s15	x		x		x	x				
31	rci	x		x		x	x				
32	inf	x		x		x	x				
33	mis	x		x		x	x				
34	verbal_iq	x		x		x		x			x
35	performance_iq	x		x		x		x			
36	full_scale_iq	x		x		x		x			x
37	vri_nr	x		x		x			x		
38	tri_nr	x		x		x			x		
39	fr	x		x		x			x		x
40	fpr	x		x		x			x		
41	fs	x		x		x			x		
42	fb_sr	x		x		x			x		
43	rbs	x		x		x			x		x
44	lr	x		x		x			x		
45	kr	x		x		x			x		x
46	eid	x		x		x			x		x
47	thd	x		x		x			x		
48	bxd	x		x		x			x		

Table A.1 continued from previous page

Number	Feature	All	Admin	Numeric	Performance	Psych	Psych_JPI_R	Psych_IQ	Psych_MMPI	Best_10	Best_20
49	r_cd	x		x		x			x	x	x
50	rc1	x		x		x			x		
51	rc2	x		x		x			x		
52	rc3	x		x		x			x		
53	rc4	x		x		x			x		
54	rc6	x		x		x			x		
55	rc7	x		x		x			x		
56	rc8	x		x		x			x		
57	rc9	x		x		x			x		
58	mls	x		x		x			x	x	x
59	hpc	x		x		x			x		
60	nuc	x		x		x			x		
61	gic	x		x		x			x		
62	sui	x		x		x			x		
63	hlp	x		x		x			x		
64	sfd	x		x		x			x	x	x
65	nfc	x		x		x			x		
66	cog	x		x		x			x		x
67	stw	x		x		x			x		
68	axy	x		x		x			x		
69	anp	x		x		x			x		
70	brf	x		x		x			x		
71	msf	x		x		x			x		
72	jcp	x		x		x			x		
73	sub	x		x		x			x		
74	agg	x		x		x			x		
75	act	x		x		x			x		
76	fml	x		x		x			x		
77	ipp	x		x		x			x		
78	sav	x		x		x			x		
79	shy	x		x		x			x		
80	dsf	x		x		x			x		
81	aes	x		x		x			x		
82	mec	x		x		x			x		
83	agg_rr	x		x		x			x		
84	psy_cr	x		x		x			x		
85	dis_cr	x		x		x			x		
86	neg_er	x		x		x			x		
87	int_rr	x		x		x			x		
88	cannot_say	x		x		x			x		
89	pct_true	x		x		x			x		

A.2 Complete list of Data Subsets

TABLE A.2: Complete List of Data Subset Combinations

Number	Data Subset
1	dataset_numeric
2	dataset_performance
3	dataset_psych
4	dataset_psych_IQ
5	dataset_psych_JPI_R
6	dataset_psych_MMPI
7	dataset_imputed_numeric
8	dataset_imputed_performance
9	dataset_imputed_psych
10	dataset_imputed_psych_JPI_R
11	dataset_imputed_psych_IQ
12	dataset_imputed_psych_MMPI
13	dataset_log_numeric
14	dataset_log_performance
15	dataset_log_psych
16	dataset_log_psych_JPI_R
17	dataset_log_psych_IQ
18	dataset_log_psych_MMPI

Table A.2 continued from previous page

Number	Data Subset
19	dataset_binned_numeric
20	dataset_binned_performance
21	dataset_binned_psych
22	dataset_binned_psych_JPI_R
23	dataset_binned_psych_IQ
24	dataset_binned_psych_MMPI
25	dataset_numeric_pca_best
26	dataset_performance_pca_best
27	dataset_psych_pca_best
28	dataset_psych_IQ_pca_best
29	dataset_psych_JPI_R_pca_best
30	dataset_psych_MMPI_pca_best
31	dataset_imputed_performance_pca_best
32	dataset_imputed_psych_pca_best
33	dataset_imputed_psych_JPI_R_pca_best
34	dataset_imputed_psych_IQ_pca_best
35	dataset_imputed_psych_MMPI_pca_best
36	dataset_log_performance_pca_best
37	dataset_log_psych_pca_best
38	dataset_log_psych_JPI_R_pca_best
39	dataset_log_psych_IQ_pca_best
40	dataset_log_psych_MMPI_pca_best
41	dataset_binned_performance_pca_best
42	dataset_binned_psych_pca_best
43	dataset_binned_psych_JPI_R_pca_best
44	dataset_binned_psych_IQ_pca_best
45	dataset_binned_psych_MMPI_pca_best
46	dataset_apft_minimum_numeric
47	dataset_apft_minimum_performance
48	dataset_apft_minimum_psych
49	dataset_apft_minimum_psych_JPI_R
50	dataset_apft_minimum_psych_IQ
51	dataset_apft_minimum_psych_MMPI
52	dataset_admin
53	dataset_all
54	dataset_imputed_all
55	dataset_imputed_admin
56	dataset_log_all
57	dataset_log_admin
58	dataset_binned_all
59	dataset_binned_admin
60	dataset_all_pca_best
61	dataset_apft_minimum_all
62	dataset_apft_minimum_admin
63	dataset_best_10
64	dataset_best_20
65	dataset_apft_minimum_best_10
66	dataset_apft_minimum_best_20

Appendix B

Model Details

B.1 Model Optimal Tuning Parameters and Thresholds

B.1.1 Penalized Logistic Regression (using Elastic Net)

TABLE B.1: Penalized Logistic Regression Optimal Tuning Parameters and Thresholds

number	dataset	alpha	lambda	threshold	FP_Cost	FN_Cost
6	dataset_admin_std_one_hot_all	0.7	0.029133	0.56	56	44
12	dataset_all_std_one_hot_all	0.9	0.015311	0.52	52	48
17	dataset_numeric_std	0.5	0.035371	0.5	50	50
19	dataset_performance_std	0.3	0.002812	0.5	50	50
33	dataset_psych_IQ_std	0.5	0.003239	0.5	50	50
35	dataset_psych_JPI_R_std	0.8	0.000536	0.52	52	48
37	dataset_psych_MMPI_std	0.3	0.001314	0.54	54	46
38	dataset_psych_std	0.4	0.007387	0.54	54	46
48	dataset_imputed_all_std_one_hot_all	1	0.000073	0.6	60	40
55	dataset_imputed_admin_std_one_hot_all	0.5	0.030442	0.56	56	44
57	dataset_imputed_performance_std	0.8	0.000391	0.56	56	44
59	dataset_imputed_psych_std	0.8	0.000319	0.5	50	50
61	dataset_imputed_psych_JPI_R_std	0.2	0.001133	0.52	52	48
63	dataset_imputed_psych_IQ_std	1	0.001386	0.52	52	48
65	dataset_imputed_psych_MMPI_std	0.4	0.001702	0.5	50	50
74	dataset_log_all_std_one_hot_all	0.8	0.014836	0.5	50	50
81	dataset_log_admin_std_one_hot_all	0.7	0.029133	0.56	56	44
83	dataset_log_performance_std	0.2	0.014240	0.5	50	50
85	dataset_log_psych_std	0.7	0.017196	0.52	52	48
87	dataset_log_psych_JPI_R_std	0.8	0.001207	0.54	54	46
89	dataset_log_psych_IQ_std	0.6	0.040898	0.5	50	50
91	dataset_log_psych_MMPI_std	0.5	0.003070	0.52	52	48
100	dataset_binned_all_std_one_hot_all	0.9	0.006675	0.52	52	48
107	dataset_binned_admin_std_one_hot_all	0.7	0.029133	0.54	54	46
109	dataset_binned_performance_std	0.3	0.002940	0.5	50	50
111	dataset_binned_psych_std	0.8	0.008134	0.48	48	52
113	dataset_binned_psych_JPI_R_std	1	0.049354	0.52	52	48
115	dataset_binned_psych_IQ_std	0.7	0.006334	0.5	50	50
117	dataset_binned_psych_MMPI_std	0.5	0.000294	0.48	48	52
141	dataset_all_pca_best	0.5	0.001091	0.5	50	50
142	dataset_numeric_pca_best	0.2	0.031059	0.5	50	50
143	dataset_performance_pca_best	0.9	0.080985	0.5	50	50
144	dataset_psych_pca_best	0.7	0.000629	0.52	52	48
146	dataset_psych_JPI_R_pca_best	0.7	0.001935	0.54	54	46
147	dataset_psych_MMPI_pca_best	0.9	0.007201	0.52	52	48
148	dataset_imputed_performance_pca_best	0.4	0.024162	0.54	54	46
149	dataset_imputed_psych_pca_best	0.1	0.001625	0.5	50	50
150	dataset_imputed_psych_JPI_R_pca_best	0.5	0.034816	0.52	52	48
151	dataset_imputed_psych_IQ_pca_best	0.9	0.003190	0.52	52	48
152	dataset_imputed_psych_MMPI_pca_best	1	0.003136	0.48	48	52
154	dataset_log_psych_pca_best	0.1	0.001470	0.5	50	50
155	dataset_log_psych_JPI_R_pca_best	0.8	0.004371	0.52	52	48
157	dataset_log_psych_MMPI_pca_best	1	0.001365	0.52	52	48
158	dataset_binned_performance_pca_best	0.8	0.006825	0.5	50	50
159	dataset_binned_psych_pca_best	0.2	0.004356	0.5	50	50

Table B.1 continued from previous page

number	dataset	alpha	lambda	threshold	FP_Cost	FN_Cost
160	dataset_binned_psych_JPI_R_pca_best	0.3	0.080723	0.52	52	48
162	dataset_binned_psych_MMPI_pca_best	0.3	0.052389	0.46	46	54
223	dataset_apft_minimum_all_std_one_hot_all	0.3	0.049894	0.48	48	52
230	dataset_apft_minimum_admin_std_one_hot_all	0.6	0.030798	0.5	50	50
232	dataset_apft_minimum_performance_std	0.7	0.008478	0.46	46	54
234	dataset_apft_minimum_psych_std	0.1	0.073382	0.5	50	50
236	dataset_apft_minimum_psych_JPI_R_std	0.6	0.000862	0.52	52	48
238	dataset_apft_minimum_psych_IQ_std	1	0.019931	0.5	50	50
240	dataset_apft_minimum_psych_MMPI_std	0.4	0.029778	0.5	50	50
269	dataset_best_10_std_one_hot_all	1	0.015999	0.48	48	52
270	dataset_best_20_std_one_hot_all	1	0.006945	0.5	50	50

B.1.2 LDA

TABLE B.2: LDA Optimal Thresholds

number	dataset	threshold	FP_Cost	FN_Cost
16	dataset_numeric	0.58	58	42
18	dataset_performance	0.54	54	46
31	dataset_psych	0.56	56	44
32	dataset_psych_IQ	0.5	50	50
34	dataset_psych_JPI_R	0.54	54	46
36	dataset_psych_MMPI	0.54	54	46
49	dataset_imputed_numeric	0.52	52	48
56	dataset_imputed_performance	0.54	54	46
58	dataset_imputed_psych	0.52	52	48
60	dataset_imputed_psych_JPI_R	0.52	52	48
62	dataset_imputed_psych_IQ	0.54	54	46
64	dataset_imputed_psych_MMPI	0.54	54	46
75	dataset_log_numeric	0.48	48	52
82	dataset_log_performance	0.52	52	48
84	dataset_log_psych	0.56	56	44
86	dataset_log_psych_JPI_R	0.54	54	46
88	dataset_log_psych_IQ	0.5	50	50
90	dataset_log_psych_MMPI	0.54	54	46
101	dataset_binned_numeric	0.54	54	46
108	dataset_binned_performance	0.46	46	54
110	dataset_binned_psych	0.5	50	50
112	dataset_binned_psych_JPI_R	0.56	56	44
114	dataset_binned_psych_IQ	0.5	50	50
116	dataset_binned_psych_MMPI	0.48	48	52
142	dataset_numeric_pca_best	0.46	46	54
143	dataset_performance_pca_best	0.5	50	50
144	dataset_psych_pca_best	0.52	52	48
145	dataset_psych_IQ_pca_best	0.5	50	50
146	dataset_psych_JPI_R_pca_best	0.5	50	50
147	dataset_psych_MMPI_pca_best	0.54	54	46
148	dataset_imputed_performance_pca_best	0.54	54	46
149	dataset_imputed_psych_pca_best	0.48	48	52
150	dataset_imputed_psych_JPI_R_pca_best	0.52	52	48
151	dataset_imputed_psych_IQ_pca_best	0.52	52	48
152	dataset_imputed_psych_MMPI_pca_best	0.46	46	54
153	dataset_log_performance_pca_best	0.52	52	48
154	dataset_log_psych_pca_best	0.52	52	48
155	dataset_log_psych_JPI_R_pca_best	0.52	52	48
156	dataset_log_psych_IQ_pca_best	0.5	50	50
157	dataset_log_psych_MMPI_pca_best	0.54	54	46
158	dataset_binned_performance_pca_best	0.5	50	50
159	dataset_binned_psych_pca_best	0.5	50	50
160	dataset_binned_psych_JPI_R_pca_best	0.52	52	48
161	dataset_binned_psych_IQ_pca_best	0.5	50	50
162	dataset_binned_psych_MMPI_pca_best	0.44	44	56
224	dataset_apft_minimum_numeric	0.4	40	60
231	dataset_apft_minimum_performance	0.46	46	54
233	dataset_apft_minimum_psych	0.44	44	56
235	dataset_apft_minimum_psych_JPI_R	0.5	50	50

Table B.2 continued from previous page

number	dataset	threshold	FP_Cost	FN_Cost
237	dataset_apft_minimum_psych_IQ	0.52	52	48
239	dataset_apft_minimum_psych_MMPI	0.5	50	50

B.1.3 QDA

TABLE B.3: QDA Optimal Thresholds

number	dataset	threshold	FP_Cost	FN_Cost
16	dataset_numeric	0.7	70	30
18	dataset_performance	0.56	56	44
31	dataset_psych	0.9	90	10
32	dataset_psych_IQ	0.52	52	48
34	dataset_psych_JPI_R	0.64	64	36
36	dataset_psych_MMPI	0.98	98	2
49	dataset_imputed_numeric	0.5	50	50
56	dataset_imputed_performance	0.66	66	34
58	dataset_imputed_psych	0.18	18	82
60	dataset_imputed_psych_JPI_R	0.4	40	60
62	dataset_imputed_psych_IQ	0.56	56	44
64	dataset_imputed_psych_MMPI	0.26	26	74
75	dataset_log_numeric	0.56	56	44
82	dataset_log_performance	0.68	68	32
84	dataset_log_psych	0.84	84	16
86	dataset_log_psych_JPI_R	0.62	62	38
88	dataset_log_psych_IQ	0.52	52	48
90	dataset_log_psych_MMPI	0.94	94	6
101	dataset_binned_numeric	0.12	12	88
108	dataset_binned_performance	0.46	46	54
110	dataset_binned_psych	0.3	30	70
112	dataset_binned_psych_JPI_R	0.54	54	46
114	dataset_binned_psych_IQ	0.48	48	52
116	dataset_binned_psych_MMPI	0.34	34	66
142	dataset_numeric_pca_best	0.54	54	46
143	dataset_performance_pca_best	0.6	60	40
144	dataset_psych_pca_best	0.76	76	24
145	dataset_psych_IQ_pca_best	0.52	52	48
146	dataset_psych_JPI_R_pca_best	0.62	62	38
147	dataset_psych_MMPI_pca_best	0.7	70	30
148	dataset_imputed_performance_pca_best	0.66	66	34
149	dataset_imputed_psych_pca_best	0.36	36	64
150	dataset_imputed_psych_JPI_R_pca_best	0.46	46	54
151	dataset_imputed_psych_IQ_pca_best	0.52	52	48
152	dataset_imputed_psych_MMPI_pca_best	0.54	54	46
153	dataset_log_performance_pca_best	0.62	62	38
154	dataset_log_psych_pca_best	0.8	80	20
155	dataset_log_psych_JPI_R_pca_best	0.56	56	44
156	dataset_log_psych_IQ_pca_best	0.48	48	52
157	dataset_log_psych_MMPI_pca_best	0.68	68	32
158	dataset_binned_performance_pca_best	0.56	56	44
159	dataset_binned_psych_pca_best	0.48	48	52
160	dataset_binned_psych_JPI_R_pca_best	0.56	56	44
161	dataset_binned_psych_IQ_pca_best	0.5	50	50
162	dataset_binned_psych_MMPI_pca_best	0.5	50	50
224	dataset_apft_minimum_numeric	0.4	40	60
231	dataset_apft_minimum_performance	0.46	46	54
233	dataset_apft_minimum_psych	0.5	50	50
235	dataset_apft_minimum_psych_JPI_R	0.5	50	50
237	dataset_apft_minimum_psych_IQ	0.5	50	50
239	dataset_apft_minimum_psych_MMPI	0.46	46	54

B.1.4 SVM

TABLE B.4: SVM Optimal Tuning Parameters and Thresholds

number	dataset	cost	threshold	FP_Cost	FN_Cost
6	dataset_admin_std_one_hot_all	1	0.56	56	44
12	dataset_all_std_one_hot_all	1	0.52	52	48
17	dataset_numeric_std	1	0.44	44	56
19	dataset_performance_std	1	0.52	52	48
33	dataset_psych_IQ_std	1	0.5	50	50
35	dataset_psych_JPI_R_std	1	0.54	54	46
37	dataset_psych_MMPI_std	1	0.54	54	46
38	dataset_psych_std	1	0.56	56	44
48	dataset_imputed_all_std_one_hot_all	1	0.56	56	44
55	dataset_imputed_admin_std_one_hot_all	1	0.58	58	42
57	dataset_imputed_performance_std	1	0.56	56	44
59	dataset_imputed_psych_std	1	0.52	52	48
63	dataset_imputed_psych_IQ_std	1	0.52	52	48
65	dataset_imputed_psych_MMPI_std	1	0.52	52	48
74	dataset_log_all_std_one_hot_all	1	0.52	52	48
81	dataset_log_admin_std_one_hot_all	1	0.6	60	40
83	dataset_log_performance_std	1	0.54	54	46
85	dataset_log_psych_std	1	0.54	54	46
87	dataset_log_psych_JPI_R_std	1	0.54	54	46
89	dataset_log_psych_IQ_std	1	0.5	50	50
91	dataset_log_psych_MMPI_std	1	0.54	54	46
100	dataset_binned_all_std_one_hot_all	1	0.5	50	50
107	dataset_binned_admin_std_one_hot_all	1	0.58	58	42
109	dataset_binned_performance_std	1	0.5	50	50
111	dataset_binned_psych_std	1	0.5	50	50
113	dataset_binned_psych_JPI_R_std	1	0.56	56	44
115	dataset_binned_psych_IQ_std	1	0.5	50	50
117	dataset_binned_psych_MMPI_std	1	0.48	48	52
141	dataset_all_pca_best	1	0.46	46	54
142	dataset_numeric_pca_best	1	0.44	44	56
143	dataset_performance_pca_best	1	0.48	48	52
144	dataset_psych_pca_best	1	0.54	54	46
145	dataset_psych_IQ_pca_best	1	0.5	50	50
146	dataset_psych_JPI_R_pca_best	1	0.5	50	50
147	dataset_psych_MMPI_pca_best	1	0.54	54	46
148	dataset_imputed_performance_pca_best	1	0.56	56	44
149	dataset_imputed_psych_pca_best	1	0.5	50	50
151	dataset_imputed_psych_IQ_pca_best	1	0.48	48	52
152	dataset_imputed_psych_MMPI_pca_best	1	0.44	44	56
153	dataset_log_performance_pca_best	1	0.48	48	52
154	dataset_log_psych_pca_best	1	0.54	54	46
155	dataset_log_psych_JPI_R_pca_best	1	0.52	52	48
156	dataset_log_psych_IQ_pca_best	1	0.5	50	50
157	dataset_log_psych_MMPI_pca_best	1	0.52	52	48
159	dataset_binned_psych_pca_best	1	0.48	48	52
160	dataset_binned_psych_JPI_R_pca_best	1	0.52	52	48
161	dataset_binned_psych_IQ_pca_best	1	0.5	50	50
162	dataset_binned_psych_MMPI_pca_best	1	0.44	44	56
223	dataset_apft_minimum_all_std_one_hot_all	1	0.52	52	48
230	dataset_apft_minimum_admin_std_one_hot_all	1	0.5	50	50
232	dataset_apft_minimum_performance_std	1	0.46	46	54
234	dataset_apft_minimum_psych_std	1	0.48	48	52
236	dataset_apft_minimum_psych_JPI_R_std	1	0.5	50	50
238	dataset_apft_minimum_psych_IQ_std	1	0.5	50	50
240	dataset_apft_minimum_psych_MMPI_std	1	0.5	50	50
269	dataset_best_10_std_one_hot_all	1	0.46	46	54
270	dataset_best_20_std_one_hot_all	1	0.56	56	44

B.1.5 CART

TABLE B.5: CART Optimal Tuning Parameters and Thresholds

number	dataset	cp	threshold	FP_Cost	FN_Cost
1	dataset_admin	0.001753	0.58	58	42
7	dataset_all	0.008249	0.42	42	58

Table B.5 continued from previous page

number	dataset	cp	threshold	FP_Cost	FN_Cost
16	dataset_numeric	0.008883	0.54	54	46
18	dataset_performance	0.002552	0.64	64	36
31	dataset_psych	0.010064	0.5	50	50
32	dataset_psych_IQ	0.001256	0.44	44	56
34	dataset_psych_JPI_R	0.005012	0.48	48	52
36	dataset_psych_MMPI	0.007155	0.46	46	54
43	dataset_imputed_all	0.003887	0.66	66	34
49	dataset_imputed_numeric	0.003158	0.64	64	36
50	dataset_imputed_admin	0.002268	0.58	58	42
56	dataset_imputed_performance	0.002106	0.64	64	36
58	dataset_imputed_psych	0.004130	0.6	60	40
60	dataset_imputed_psych_JPI_R	0.001984	0.58	58	42
62	dataset_imputed_psych_IQ	0.000292	0.5	50	50
64	dataset_imputed_psych_MMPI	0.004252	0.58	58	42
69	dataset_log_all	0.008249	0.42	42	58
75	dataset_log_numeric	0.008883	0.54	54	46
76	dataset_log_admin	0.001753	0.58	58	42
82	dataset_log_performance	0.002552	0.64	64	36
84	dataset_log_psych	0.010064	0.5	50	50
86	dataset_log_psych_JPI_R	0.005012	0.48	48	52
88	dataset_log_psych_IQ	0.001256	0.44	44	56
90	dataset_log_psych_MMPI	0.007155	0.46	46	54
95	dataset_binned_all	0.005922	0.62	62	38
101	dataset_binned_numeric	0.006345	0.58	58	42
102	dataset_binned_admin	0.001502	0.58	58	42
108	dataset_binned_performance	0.139357	0.5	50	50
110	dataset_binned_psych	0.007649	0.56	56	44
112	dataset_binned_psych_JPI_R	0.006816	0.56	56	44
114	dataset_binned_psych_IQ	0.000105	0.48	48	52
116	dataset_binned_psych_MMPI	0.003695	0.58	58	42
141	dataset_all_pca_best	0.004442	0.56	56	44
142	dataset_numeric_pca_best	0.004442	0.56	56	44
143	dataset_performance_pca_best	0.004594	0.58	58	42
144	dataset_psych_pca_best	0.010870	0.44	44	56
145	dataset_psych_IQ_pca_best	0.001884	0.5	50	50
146	dataset_psych_JPI_R_pca_best	0.005079	0.58	58	42
147	dataset_psych_MMPI_pca_best	0.009442	0.5	50	50
148	dataset_imputed_performance_pca_best	0.001458	0.62	62	38
149	dataset_imputed_psych_pca_best	0.002672	0.58	58	42
150	dataset_imputed_psych_JPI_R_pca_best	0.002915	0.56	56	44
151	dataset_imputed_psych_IQ_pca_best	0.001361	0.54	54	46
152	dataset_imputed_psych_MMPI_pca_best	0.001701	0.56	56	44
153	dataset_log_performance_pca_best	0.002188	0.56	56	44
154	dataset_log_psych_pca_best	0.013285	0.5	50	50
155	dataset_log_psych_JPI_R_pca_best	0.004010	0.52	52	48
156	dataset_log_psych_IQ_pca_best	0.002513	0.5	50	50
157	dataset_log_psych_MMPI_pca_best	0.009442	0.42	42	58
158	dataset_binned_performance_pca_best	0.002042	0.58	58	42
159	dataset_binned_psych_pca_best	0.009662	0.66	66	34
160	dataset_binned_psych_JPI_R_pca_best	0.005613	0.56	56	44
161	dataset_binned_psych_IQ_pca_best	0.000126	0.48	48	52
162	dataset_binned_psych_MMPI_pca_best	0.002258	0.64	64	36
218	dataset_apft_minimum_all	0.007092	0.6	60	40
224	dataset_apft_minimum_numeric	0.018237	0.44	44	56
225	dataset_apft_minimum_admin	0.002071	0.5	50	50
231	dataset_apft_minimum_performance	0.005233	0.5	50	50
233	dataset_apft_minimum_psych	0.019727	0.56	56	44
235	dataset_apft_minimum_psych_JPI_R	0.084592	0.5	50	50
237	dataset_apft_minimum_psych_IQ	0.071807	0.5	50	50
239	dataset_apft_minimum_psych_MMPI	0.007837	0.54	54	46
267	dataset_best_10	0.004646	0.6	60	40
268	dataset_best_20	0.006844	0.5	50	50

B.1.6 KNN

TABLE B.6: KNN Optimal Tuning Parameters and Thresholds

number	dataset	k	threshold	FP_Cost	FN_Cost
6	dataset_admin_std_one_hot_all	50	0.6	60	40
12	dataset_all_std_one_hot_all	50	0.6	60	40
17	dataset_numeric_std	45	0.56	56	44
19	dataset_performance_std	30	0.56	56	44
33	dataset_psych_IQ_std	60	0.5	50	50
35	dataset_psych_JPI_R_std	60	0.54	54	46
37	dataset_psych_MMPI_std	60	0.58	58	42
38	dataset_psych_std	60	0.58	58	42
48	dataset_imputed_all_std_one_hot_all	60	0.54	54	46
55	dataset_imputed_admin_std_one_hot_all	55	0.6	60	40
57	dataset_imputed_performance_std	40	0.58	58	42
59	dataset_imputed_psych_std	25	0.54	54	46
61	dataset_imputed_psych_JPI_R_std	60	0.54	54	46
63	dataset_imputed_psych_IQ_std	30	0.52	52	48
74	dataset_log_all_std_one_hot_all	45	0.58	58	42
81	dataset_log_admin_std_one_hot_all	50	0.62	62	38
83	dataset_log_performance_std	25	0.5	50	50
85	dataset_log_psych_std	40	0.62	62	38
87	dataset_log_psych_JPI_R_std	45	0.52	52	48
89	dataset_log_psych_IQ_std	60	0.5	50	50
91	dataset_log_psych_MMPI_std	60	0.6	60	40
100	dataset_binned_all_std_one_hot_all	30	0.5	50	50
107	dataset_binned_admin_std_one_hot_all	40	0.6	60	40
109	dataset_binned_performance_std	60	0.54	54	46
111	dataset_binned_psych_std	60	0.54	54	46
113	dataset_binned_psych_JPI_R_std	60	0.54	54	46
117	dataset_binned_psych_MMPI_std	60	0.52	52	48
141	dataset_all_pca_best	60	0.56	56	44
142	dataset_numeric_pca_best	60	0.5	50	50
143	dataset_performance_pca_best	55	0.56	56	44
144	dataset_psych_pca_best	50	0.56	56	44
145	dataset_psych_IQ_pca_best	25	0.5	50	50
146	dataset_psych_JPI_R_pca_best	60	0.54	54	46
147	dataset_psych_MMPI_pca_best	60	0.54	54	46
148	dataset_imputed_performance_pca_best	60	0.62	62	38
149	dataset_imputed_psych_pca_best	25	0.58	58	42
150	dataset_imputed_psych_JPI_R_pca_best	20	0.54	54	46
154	dataset_log_psych_pca_best	50	0.56	56	44
155	dataset_log_psych_JPI_R_pca_best	45	0.52	52	48
156	dataset_log_psych_IQ_pca_best	5	0.46	46	54
157	dataset_log_psych_MMPI_pca_best	60	0.56	56	44
158	dataset_binned_performance_pca_best	60	0.54	54	46
159	dataset_binned_psych_pca_best	60	0.5	50	50
160	dataset_binned_psych_JPI_R_pca_best	60	0.54	54	46
161	dataset_binned_psych_IQ_pca_best	55	0.48	48	52
162	dataset_binned_psych_MMPI_pca_best	60	0.48	48	52
223	dataset_apft_minimum_all_std_one_hot_all	30	0.48	48	52
230	dataset_apft_minimum_admin_std_one_hot_all	40	0.52	52	48
232	dataset_apft_minimum_performance_std	60	0.5	50	50
234	dataset_apft_minimum_psych_std	55	0.5	50	50
236	dataset_apft_minimum_psych_JPI_R_std	60	0.5	50	50
238	dataset_apft_minimum_psych_IQ_std	60	0.48	48	52
240	dataset_apft_minimum_psych_MMPI_std	35	0.52	52	48
269	dataset_best_10_std_one_hot_all	60	0.54	54	46
270	dataset_best_20_std_one_hot_all	60	0.58	58	42

B.1.7 Random Forest

TABLE B.7: Random Forest Optimal Tuning Parameters and Thresholds

number	dataset	mtry	ntree	threshold	FP_Cost	FN_Cost
1	dataset_admin	3	500	0.64	64	36
7	dataset_all	13	1500	0.44	44	56
16	dataset_numeric	7	1500	0.42	42	58

Table B.7 continued from previous page

number	dataset	mtry	ntree	threshold	FP_Cost	FN_Cost
18	dataset_performance	3	500	0.38	38	62
31	dataset_psych	13	1500	0.44	44	56
32	dataset_psych_IQ	9	1000	0.24	24	76
34	dataset_psych_JPI_R	5	500	0.42	42	58
36	dataset_psych_MMPI	13	500	0.42	42	58
43	dataset_imputed_all	9	1000	0.42	42	58
49	dataset_imputed_numeric	13	1000	0.46	46	54
50	dataset_imputed_admin	3	500	0.7	70	30
56	dataset_imputed_performance	3	500	0.36	36	64
58	dataset_imputed_psych	3	1500	0.42	42	58
60	dataset_imputed_psych_JPI_R	11	500	0.34	34	66
62	dataset_imputed_psych_IQ	13	1500	0.52	52	48
64	dataset_imputed_psych_MMPI	3	500	0.42	42	58
69	dataset_log_all	13	1000	0.44	44	56
75	dataset_log_numeric	9	1500	0.44	44	56
76	dataset_log_admin	3	1500	0.7	70	30
82	dataset_log_performance	3	1000	0.42	42	58
84	dataset_log_psych	5	1000	0.44	44	56
86	dataset_log_psych_JPI_R	5	1000	0.4	40	60
88	dataset_log_psych_IQ	11	1500	0.38	38	62
90	dataset_log_psych_MMPI	7	500	0.42	42	58
95	dataset_binned_all	13	1000	0.46	46	54
101	dataset_binned_numeric	9	1000	0.46	46	54
102	dataset_binned_admin	3	1500	0.56	56	44
108	dataset_binned_performance	3	500	0.5	50	50
110	dataset_binned_psych	11	1000	0.46	46	54
112	dataset_binned_psych_JPI_R	13	1500	0.46	46	54
114	dataset_binned_psych_IQ	9	1000	0.18	18	82
116	dataset_binned_psych_MMPI	3	1500	0.44	44	56
141	dataset_all_pca_best	13	1500	0.52	52	48
142	dataset_numeric_pca_best	3	500	0.44	44	56
143	dataset_performance_pca_best	11	1000	0.32	32	68
144	dataset_psych_pca_best	11	500	0.42	42	58
145	dataset_psych_IQ_pca_best	5	1500	0.32	32	68
146	dataset_psych_JPI_R_pca_best	9	1500	0.4	40	60
147	dataset_psych_MMPI_pca_best	3	1000	0.48	48	52
148	dataset_imputed_performance_pca_best	3	500	0.38	38	62
149	dataset_imputed_psych_pca_best	5	1500	0.38	38	62
150	dataset_imputed_psych_JPI_R_pca_best	11	1000	0.34	34	66
151	dataset_imputed_psych_IQ_pca_best	13	1500	0.52	52	48
152	dataset_imputed_psych_MMPI_pca_best	9	500	0.34	34	66
153	dataset_log_performance_pca_best	3	500	0.14	14	86
154	dataset_log_psych_pca_best	11	1000	0.42	42	58
155	dataset_log_psych_JPI_R_pca_best	11	1500	0.48	48	52
156	dataset_log_psych_IQ_pca_best	13	1500	0.56	56	44
157	dataset_log_psych_MMPI_pca_best	3	500	0.44	44	56
158	dataset_binned_performance_pca_best	11	1500	0.48	48	52
159	dataset_binned_psych_pca_best	7	500	0.42	42	58
160	dataset_binned_psych_JPI_R_pca_best	9	500	0.4	40	60
161	dataset_binned_psych_IQ_pca_best	9	500	0.34	34	66
162	dataset_binned_psych_MMPI_pca_best	9	1500	0.36	36	64
218	dataset_apft_minimum_all	7	1000	0.5	50	50
224	dataset_apft_minimum_numeric	13	500	0.48	48	52
225	dataset_apft_minimum_admin	3	500	0.5	50	50
231	dataset_apft_minimum_performance	9	500	0.44	44	56
233	dataset_apft_minimum_psych	5	1000	0.5	50	50
235	dataset_apft_minimum_psych_JPI_R	9	500	0.48	48	52
237	dataset_apft_minimum_psych_IQ	5	1000	0.28	28	72
239	dataset_apft_minimum_psych_MMPI	13	1000	0.5	50	50
267	dataset_best_10	3	500	0.42	42	58
268	dataset_best_20	5	1000	0.42	42	58

B.1.8 xgboost

TABLE B.8: xgboost Optimal Tuning Parameters and Thresholds

number	dataset	nrounds	max_depth	eta	gamma	colsample_bytree	min child weight	subsample	threshold	FP_Cost	FN_Cost
1	dataset_admin	100	3	0.3	0	0.8	1	1	0.58	58	42
7	dataset_all	50	3	0.3	0	0.6	1	1	0.5	50	50
16	dataset_numeric	50	2	0.3	0	0.8	1	0.75	0.5	50	50
18	dataset_performance	50	1	0.3	0	0.6	1	0.5	0.58	58	42
31	dataset_psych	100	1	0.4	0	0.8	1	0.5	0.54	54	46
32	dataset_psych_IQ	50	3	0.3	0	0.6	1	0.75	0.52	52	48
34	dataset_psych_JPI_R	50	2	0.3	0	0.8	1	0.75	0.54	54	46
36	dataset_psych_MMPI	50	3	0.3	0	0.8	1	1	0.5	50	50
43	dataset_imputed_all	100	3	0.3	0	0.8	1	1	0.5	50	50
49	dataset_imputed_numeric	100	3	0.3	0	0.6	1	1	0.52	52	48
50	dataset_imputed_admin	50	3	0.3	0	0.6	1	1	0.58	58	42
56	dataset_imputed_performance	50	2	0.4	0	0.6	1	1	0.5	50	50
58	dataset_imputed_psych	150	2	0.3	0	0.6	1	1	0.5	50	50
60	dataset_imputed_psych_JPI_R	100	1	0.3	0	0.6	1	0.5	0.5	50	50
62	dataset_imputed_psych_IQ	150	1	0.4	0	0.8	1	1	0.5	50	50
64	dataset_imputed_psych_MMPI	50	1	0.4	0	0.6	1	1	0.5	50	50
69	dataset_log_all	50	3	0.3	0	0.6	1	1	0.5	50	50
75	dataset_log_numeric	50	2	0.3	0	0.8	1	0.75	0.5	50	50
76	dataset_log_admin	100	3	0.3	0	0.8	1	1	0.58	58	42
82	dataset_log_performance	50	1	0.3	0	0.6	1	0.5	0.58	58	42
84	dataset_log_psych	100	1	0.4	0	0.8	1	0.5	0.54	54	46
86	dataset_log_psych_JPI_R	50	2	0.3	0	0.8	1	0.75	0.54	54	46
88	dataset_log_psych_IQ	50	3	0.3	0	0.6	1	0.75	0.52	52	48
90	dataset_log_psych_MMPI	50	3	0.3	0	0.8	1	1	0.5	50	50
95	dataset_binned_all	50	3	0.3	0	0.6	1	0.75	0.5	50	50
101	dataset_binned_numeric	100	1	0.3	0	0.6	1	0.75	0.52	52	48
102	dataset_binned_admin	150	2	0.4	0	0.8	1	1	0.58	58	42
108	dataset_binned_performance	100	1	0.3	0	0.8	1	0.75	0.6	60	40
110	dataset_binned_psych	150	1	0.4	0	0.8	1	1	0.52	52	48
112	dataset_binned_psych_JPI_R	50	1	0.4	0	0.6	1	0.5	0.54	54	46
114	dataset_binned_psych_IQ	50	2	0.3	0	0.8	1	1	0.5	50	50
116	dataset_binned_psych_MMPI	50	2	0.4	0	0.6	1	0.75	0.48	48	52
141	dataset_all_pca_best	150	1	0.4	0	0.6	1	1	0.48	48	52
142	dataset_numeric_pca_best	50	2	0.3	0	0.6	1	1	0.52	52	48
143	dataset_performance_pca_best	150	1	0.3	0	0.6	1	1	0.54	54	46
144	dataset_psych_pca_best	100	1	0.3	0	0.6	1	0.75	0.5	50	50
145	dataset_psych_IQ_pca_best	50	1	0.4	0	0.6	1	0.75	0.5	50	50
146	dataset_psych_JPI_R_pca_best	150	1	0.3	0	0.8	1	0.75	0.5	50	50
147	dataset_psych_MMPI_pca_best	100	1	0.3	0	0.6	1	0.5	0.5	50	50
148	dataset_imputed_performance_pca_best	150	2	0.3	0	0.8	1	0.75	0.52	52	48
149	dataset_imputed_psych_pca_best	50	3	0.3	0	0.6	1	0.5	0.48	48	52
150	dataset_imputed_psych_JPI_R_pca_best	100	1	0.3	0	0.8	1	0.5	0.5	50	50
151	dataset_imputed_psych_IQ_pca_best	100	1	0.4	0	0.6	1	1	0.5	50	50
152	dataset_imputed_psych_MMPI_pca_best	150	2	0.4	0	0.8	1	0.75	0.54	54	46
153	dataset_log_performance_pca_best	50	1	0.4	0	0.8	1	0.5	0.56	56	44
154	dataset_log_psych_pca_best	100	1	0.3	0	0.8	1	0.75	0.54	54	46
155	dataset_log_psych_JPI_R_pca_best	150	1	0.3	0	0.6	1	0.5	0.52	52	48
156	dataset_log_psych_IQ_pca_best	50	1	0.4	0	0.8	1	0.5	0.52	52	48
157	dataset_log_psych_MMPI_pca_best	50	1	0.4	0	0.6	1	0.75	0.56	56	44
158	dataset_binned_performance_pca_best	100	2	0.4	0	0.8	1	1	0.5	50	50
159	dataset_binned_psych_pca_best	50	1	0.3	0	0.6	1	1	0.5	50	50
160	dataset_binned_psych_JPI_R_pca_best	50	1	0.3	0	0.6	1	0.75	0.52	52	48
161	dataset_binned_psych_IQ_pca_best	100	1	0.3	0	0.6	1	0.75	0.5	50	50
162	dataset_binned_psych_MMPI_pca_best	50	2	0.4	0	0.8	1	1	0.54	54	46
218	dataset_apft_minimum_all	100	1	0.3	0	0.8	1	1	0.5	50	50
224	dataset_apft_minimum_numeric	50	1	0.4	0	0.8	1	1	0.48	48	52
225	dataset_apft_minimum_admin	50	1	0.3	0	0.6	1	0.5	0.5	50	50
231	dataset_apft_minimum_performance	50	1	0.3	0	0.8	1	0.75	0.5	50	50
233	dataset_apft_minimum_psych	50	1	0.3	0	0.8	1	0.75	0.5	50	50
235	dataset_apft_minimum_psych_JPI_R	100	1	0.3	0	0.6	1	0.5	0.48	48	52
237	dataset_apft_minimum_psych_IQ	50	1	0.3	0	0.8	1	0.5	0.5	50	50
239	dataset_apft_minimum_psych_MMPI	100	1	0.3	0	0.6	1	0.5	0.5	50	50

B.1.9 Stack

Stack GLM

TABLE B.9: Stack GLM Model Weights for each Data Subset

number	dataset	(Intercept)	lda	svm_lin	pen_log_reg	CART	knn	rf	xgb
17	dataset_numeric_std	3.123	-0.28	-1.172	-1.941	0.169	-0.444	-3.37	0.499
19	dataset_performance_std	3.111	-6.172	-2.516	5.888	0.194	-0.906	-0.149	-2.066
33	dataset_psych_IQ_std	2.077	-0.469	-1.069	-2.262	-0.288	0.969	-0.125	-0.958
35	dataset_psych_JPI_R_std	2.104	0.761	-2.714	-2.448	0.295	-0.242	-1.58	1.422
37	dataset_psych_MMPI_std	1.96	0.954	-0.994	-1.363	-0.489	0.466	-2.696	-0.461
38	dataset_psych_std	2.004	0.846	0.208	-2.426	0.11	-0.269	-2.349	-0.678
57	dataset_imputed_performance_std	2.732	0.528	0.079	-1.163	-0.817	-2.062	-0.508	-1.498
59	dataset_imputed_psych_std	2.019	-0.866	1.596	-1.524	-0.233	-0.863	-1.232	-1.51
83	dataset_log_performance_std	2.825	-4.049	-2.073	3.346	0.398	-0.223	-0.247	-2.488
85	dataset_log_psych_std	2.23	0.888	-0.044	-3.007	0.243	-0.684	-2.478	0.185
87	dataset_log_psych_JPI_R_std	2.368	0.685	-0.291	-3.992	0.328	0.025	-1.394	-0.397
89	dataset_log_psych_IQ_std	2.105	-1.086	-0.728	-1.788	-0.288	0.64	-0.132	-0.871
91	dataset_log_psych_MMPI_std	1.982	0.533	-0.545	-1.324	-0.516	0.501	-2.828	-0.453
109	dataset_binned_performance_std	2.654	0.89	-0.286	-0.753	0.473	-0.018	0.148	-5.601
111	dataset_binned_psych_std	2.536	2.33	-0.462	-3.235	0.335	0.996	-4.268	-1.598
113	dataset_binned_psych_JPI_R_std	2.344	-2.321	-0.953	2.348	-0.439	-0.625	-1.611	-1.409
117	dataset_binned_psych_MMPI_std	2.248	0.864	1.207	-2.041	-0.115	-0.392	-3.765	-0.992
142	dataset_numeric_pca_best	2.853	-6.415	1.964	0.895	-0.118	-0.365	-1.614	0.063
143	dataset_performance_pca_best	4.794	-31.414	32.267	-6.171	-0.14	-0.362	0.096	-2.417
144	dataset_psych_pca_best	2.383	-1.18	1.816	-3.746	-0.117	-0.234	-1.307	-0.215
146	dataset_psych_JPI_R_pca_best	2.328	1.724	-4.795	0.545	-0.366	-0.645	-0.842	-0.441
147	dataset_psych_MMPI_pca_best	2.006	3.259	-1.038	-5.398	-0.156	0.634	-1.163	-0.428
148	dataset_imputed_performance_pca_best	3.042	2.754	-2.626	-1.252	-0.211	-2.998	-0.331	-1.257
149	dataset_imputed_psych_pca_best	2.169	-0.742	2.078	-1.869	-0.24	-1.28	-0.731	-1.845
154	dataset_log_psych_pca_best	2.193	-0.966	0.324	-2.551	0.3	-0.498	-1.641	0.414
155	dataset_log_psych_JPI_R_pca_best	2.605	2.024	-1.103	-3.183	0.347	-0.639	-1.342	-1.593
157	dataset_log_psych_MMPI_pca_best	1.94	2.802	-2.173	-3.293	0.316	0.4	-0.717	-1.425
159	dataset_binned_psych_pca_best	2.253	-1.787	2.403	-3.354	-0.461	-0.022	-2.632	0.898
160	dataset_binned_psych_JPI_R_pca_best	2.11	0.172	-4.067	2.977	-0.114	-0.914	-0.888	-1.565
162	dataset_binned_psych_MMPI_pca_best	2.051	2.362	-7.211	4.335	0.204	-2.061	0.195	-1.89
234	dataset_apft_minimum_psych_std	2.815	-0.463	-0.642	-3.115	-0.376	1.164	-2.292	0.123
236	dataset_apft_minimum_psych_JPI_R_std	2	-3.28	3.067	-1.262	-0.811	-0.657	-0.561	-0.441
238	dataset_apft_minimum_psych_IQ_std	2.474	19.438	-17.089	-6.806	1.607	-0.93	0.123	-1.311
240	dataset_apft_minimum_psych_MMPI_std	1.904	0.224	1.739	-3.516	0.227	0.647	-2.532	-0.549

Stack Random Forest

TABLE B.10: Stack Random Forest Optimal Tuning Parameters

number	dataset	mtry
17	dataset_numeric_std	2
19	dataset_performance_std	2
33	dataset_psych_IQ_std	2
35	dataset_psych_JPI_R_std	2
37	dataset_psych_MMPI_std	4
38	dataset_psych_std	2
57	dataset_imputed_performance_std	2
59	dataset_imputed_psych_std	2
83	dataset_log_performance_std	2
85	dataset_log_psych_std	2
87	dataset_log_psych_JPI_R_std	4
89	dataset_log_psych_IQ_std	4
91	dataset_log_psych_MMPI_std	7
109	dataset_binned_performance_std	2
111	dataset_binned_psych_std	2
113	dataset_binned_psych_JPI_R_std	2
117	dataset_binned_psych_MMPI_std	2
142	dataset_numeric_pca_best	2
143	dataset_performance_pca_best	2
144	dataset_psych_pca_best	2
146	dataset_psych_JPI_R_pca_best	2
147	dataset_psych_MMPI_pca_best	2

Table B.10 continued from previous page

number	dataset	mtry
148	dataset_imputed_performance_pca_best	2
149	dataset_imputed_psych_pca_best	2
154	dataset_log_psych_pca_best	4
155	dataset_log_psych_JPI_R_pca_best	7
157	dataset_log_psych_MMPI_pca_best	4
159	dataset_binned_psych_pca_best	2
160	dataset_binned_psych_JPI_R_pca_best	2
162	dataset_binned_psych_MMPI_pca_best	2
234	dataset_apft_minimum_psych_std	4
236	dataset_apft_minimum_psych_JPI_R_std	4
238	dataset_apft_minimum_psych_IQ_std	7
240	dataset_apft_minimum_psych_MMPI_std	2

B.2 Feature Importance

To be concise, but also show the holistic feature importance measures, we will share the feature importance of only the data sets using all features with no feature engineering, for both all candidates and candidates who scores above the minimum fitness screening criteria. We will first provide a plot of the top 20 most important features for each respective model and data set for an intuitive visualization, and then the complete feature importance results using a table.

B.2.1 Penalized Logistic Regression (using Elastic Net)

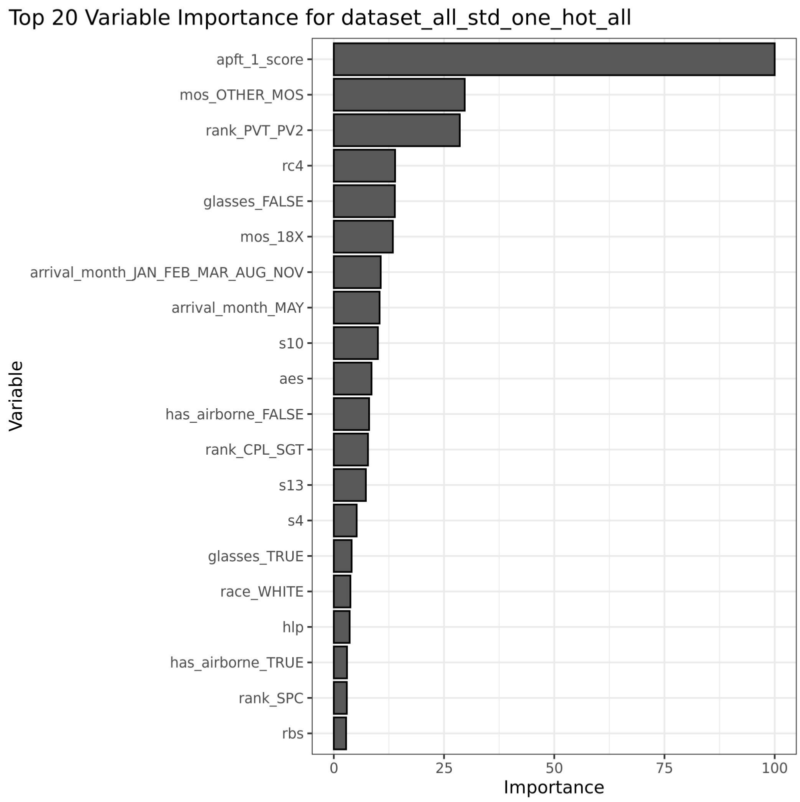


FIGURE B.1: Top 20 Most Important Features with all candidates for dataset_all using Penalized Logistic Regression

TABLE B.11: Feature Importance with all candidates for dataset_all using Penalized Logistic Regression

Feature	Importance
apft_1_score	100
mos_OTHER_MOS	29.69
rank_PVT_PV2	28.57
rc4	13.9
glasses_FALSE	13.84
mos_18X	13.39
arrival_month_JAN_FEB_MAR_AUG_NOV	10.64
arrival_month_MAY	10.39
s10	10
aes	8.56
has_airborne_FALSE	8.01
rank_CPL_SGT	7.75
s13	7.27
s4	5.19
glasses_TRUE	4.05
race_WHITE	3.77
hlp	3.6
has_airborne_TRUE	2.99
rank_SPC	2.95
rbs	2.79
hpc	2.74
performance_iq	2.54
cog	2.45
sub	1.88
‘civilian_education_certification_HIGH SCHOOL DIPLOMA‘	1.28
apft_1_run	1.25
mls	0.7
sfd	0.39
rc2	0.2
s1	0.12
sui	0.04
mos_19D_68W_11C_13F_11B	0
mos_35G_M_N_P	0
rank_PFC	0
race_OTHER	0
race_UNKNOWN	0
arrival_month_APR_JUN_DEC	0
arrival_month_JUL_SEP_OCT	0
arrival_month_UNKNOWN	0
tis_at_arrival	0
parents_together_FALSE	0
parents_together_TRUE	0
‘civilian_education_certification_ALTERNATIVE HIGH SCHOOL/EQUIVALENCY CERTIFICATE‘	0
‘civilian_education_certification_BACHELORS DEGREE/MASTERS DEGREE‘	0
‘civilian_education_certification_SOME COLLEGE/ASSOCIATES DEGREE‘	0
civilian_education_certification_UNKNOWN	0
‘age_at_arrival_<=20‘	0
‘age_at_arrival_>20‘	0
age_at_arrival_UNKNOWN	0
gt_score	0
apft_1_pu	0
apft_1_su	0
s2	0
s3	0
s5	0
s6	0
s7	0
s8	0
s9	0
s11	0
s12	0
s14	0
s15	0
rci	0
inf	0

Table B.11 continued from previous page

Feature	Importance
mis	0
verbal_iq	0
full_scale_iq	0
vri_nr	0
tri_nr	0
fr	0
fpr	0
fs	0
fb_sr	0
lr	0
kr	0
eid	0
thd	0
bxd	0
r_cd	0
rc1	0
rc3	0
rc6	0
rc7	0
rc8	0
rc9	0
nuc	0
gic	0
nfc	0
stw	0
axy	0
anp	0
brf	0
msf	0
jcp	0
agg	0
act	0
fml	0
ipp	0
sav	0
shy	0
dsf	0
mec	0
agg_rr	0
psy_cr	0
dis_cr	0
neg_er	0
int_rr	0
cannot_say	0
pct_true	0

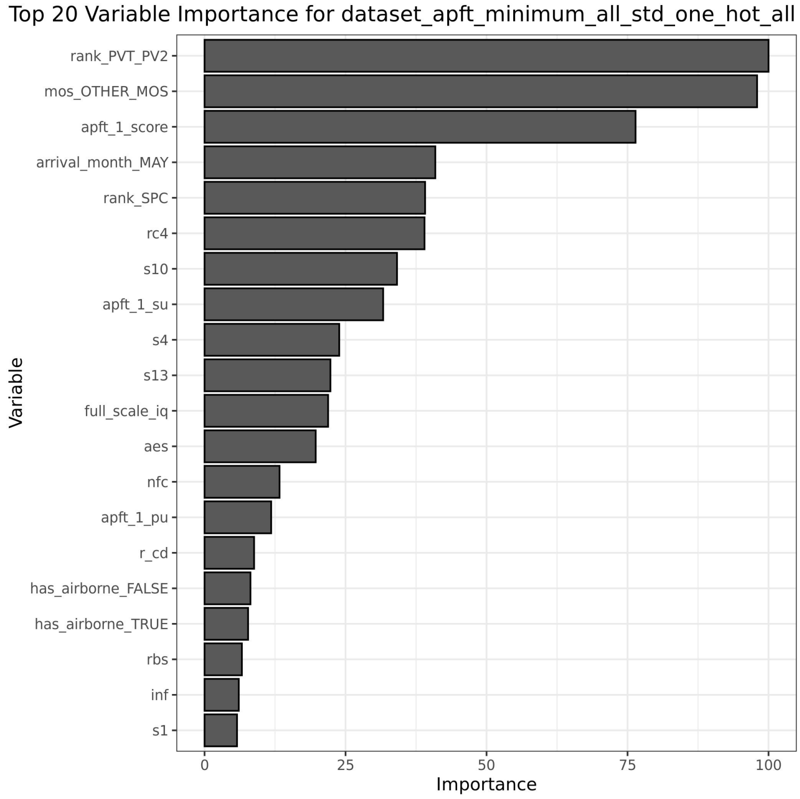


FIGURE B.2: Top 20 Most Important Features with candidates who scored above the minimum fitness test score for dataset_all using Penalized Logistic Regression

TABLE B.12: Feature Importance with candidates who scored above the minimum fitness test score for dataset_all using Penalized Logistic Regression

Feature	Importance
rank_PVT_PV2	100
mos_OTHER_MOS	97.97
apft_1_score	76.42
arrival_month_MAY	40.92
rank_SPC	39.1
rc4	39
s10	34.12
apft_1_su	31.67
s4	23.88
s13	22.32
full_scale_iq	21.91
aes	19.71
nfc	13.3
apft_1_pu	11.81
r_cd	8.78
has_airborne_FALSE	8.14
has_airborne_TRUE	7.72
rbs	6.62
inf	6.07
s1	5.76
performance_iq	5.18
parents_together_FALSE	4.91
sub	4.82
parents_together_TRUE	4.34
verbal_iq	3.47
rc2	3.18

Table B.12 continued from previous page

Feature	Importance
dsf	2.23
fs	0.92
tri_nr	0.17
mos_18X	0
mos_19D_68W_11C_13F_11B	0
mos_35G_M_N_P	0
rank_CPL_SGT	0
rank_PFC	0
race_OTHER	0
race_UNKNOWN	0
race_WHITE	0
arrival_month_APR_JUN_DEC	0
arrival_month_JAN_FEB_MAR_AUG_NOV	0
arrival_month_JUL_SEP_OCT	0
arrival_month_UNKNOWN	0
tis_at_arrival	0
glasses_FALSE	0
glasses_TRUE	0
‘civilian_education_certification_ALTERNATIVE HIGH SCHOOL/EQUIVALENCY CERTIFICATE‘	0
‘civilian_education_certification_BACHELORS DEGREE/MASTERS DEGREE‘	0
‘civilian_education_certification_HIGH SCHOOL DIPLOMA‘	0
‘civilian_education_certification_SOME COLLEGE/ASSOCIATES DEGREE‘	0
civilian_education_certification_UNKNOWN	0
‘age_at_arrival_<=20‘	0
‘age_at_arrival_>20‘	0
age_at_arrival_UNKNOWN	0
gt_score	0
apft_1_run	0
s2	0
s3	0
s5	0
s6	0
s7	0
s8	0
s9	0
s11	0
s12	0
s14	0
s15	0
rci	0
mis	0
vri_nr	0
fr	0
fpr	0
fb_sr	0
lr	0
kr	0
eid	0
thd	0
bxd	0
rc1	0
rc3	0
rc6	0
rc7	0
rc8	0
rc9	0
mls	0
hpc	0
nuc	0
gic	0
sui	0
hlp	0
sfd	0
cog	0
stw	0
axy	0
anp	0
brf	0

Table B.12 continued from previous page

Feature	Importance
msf	0
jcp	0
agg	0
act	0
fml	0
ipp	0
sav	0
shy	0
mec	0
agg_rr	0
psy_cr	0
dis_cr	0
neg_er	0
int_rr	0
cannot_say	0
pct_true	0

B.2.2 LDA

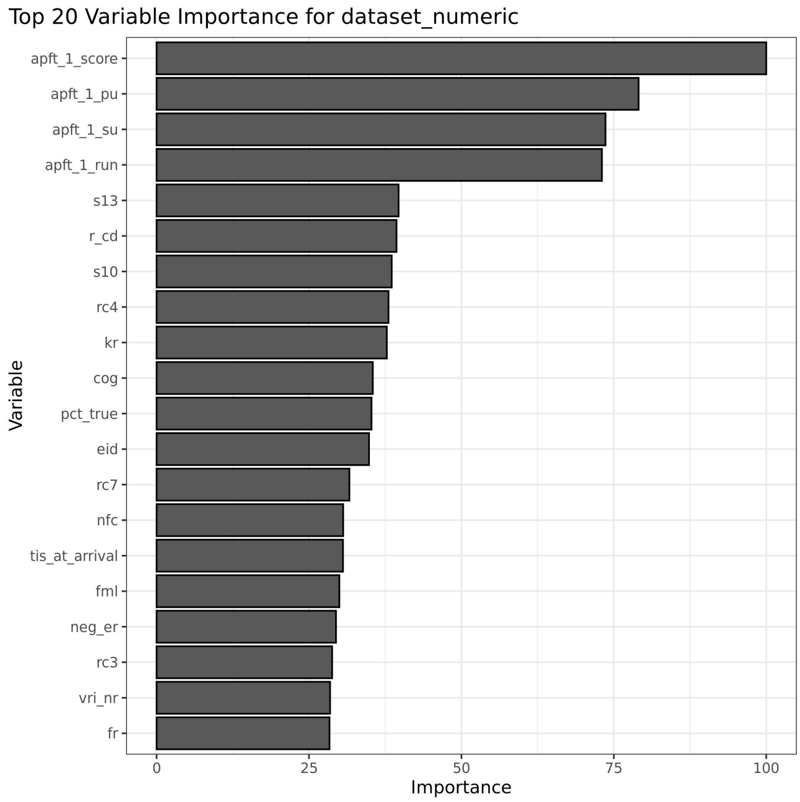


FIGURE B.3: Top 20 Most Important Features with all candidates for dataset_numeric using LDA

TABLE B.13: Feature Importance with all candidates for dataset_numeric using LDA

Feature	Importance
apft_1_score	100
apft_1_pu	79.06
apft_1_su	73.63
apft_1_run	73.05
s13	39.69
r_cd	39.36

Table B.13 continued from previous page

Feature	Importance
s10	38.57
rc4	38.04
kr	37.76
cog	35.46
pct_true	35.25
eid	34.84
rc7	31.62
nfc	30.6
tis_at_arrival	30.57
fml	29.98
neg_er	29.42
rc3	28.79
vri_nr	28.46
fr	28.36
mls	27.49
sfd	27.17
rc2	26.31
jcp	26.23
psy_cr	26.08
bxd	26.06
rc8	26.01
thd	24.76
shy	24.43
dis_cr	23.16
fs	22.93
rc1	22.48
fpr	22.27
rc6	22.21
anp	22.03
sub	21.49
stw	21.37
s9	21.21
act	20.94
agg	20.17
lr	20.12
hlp	19.28
full_scale_iq	19.06
rc9	18.99
aes	17.44
nuc	17.41
performance_iq	17.14
s14	16.92
rbs	16.62
s15	16.22
s6	15.77
verbal_iq	15.5
rci	15.25
gt_score	15.09
sav	14.02
s8	12.28
int_rr	11.59
s1	11.45
brf	10.32
hpc	10.19
dsf	9.79
axy	8.17
s2	8.13
s12	7.2
msf	6.89
mec	6.63
tri_nr	4.84
inf	4.69
agg_rr	4.63
s11	4.62
gic	4.47
sui	4.34
s3	2.5
fb_sr	2.1

Table B.13 continued from previous page

Feature	Importance
mis	1.95
s4	0.65
s5	0.56
ipp	0.33
s7	0.03
cannot_say	0

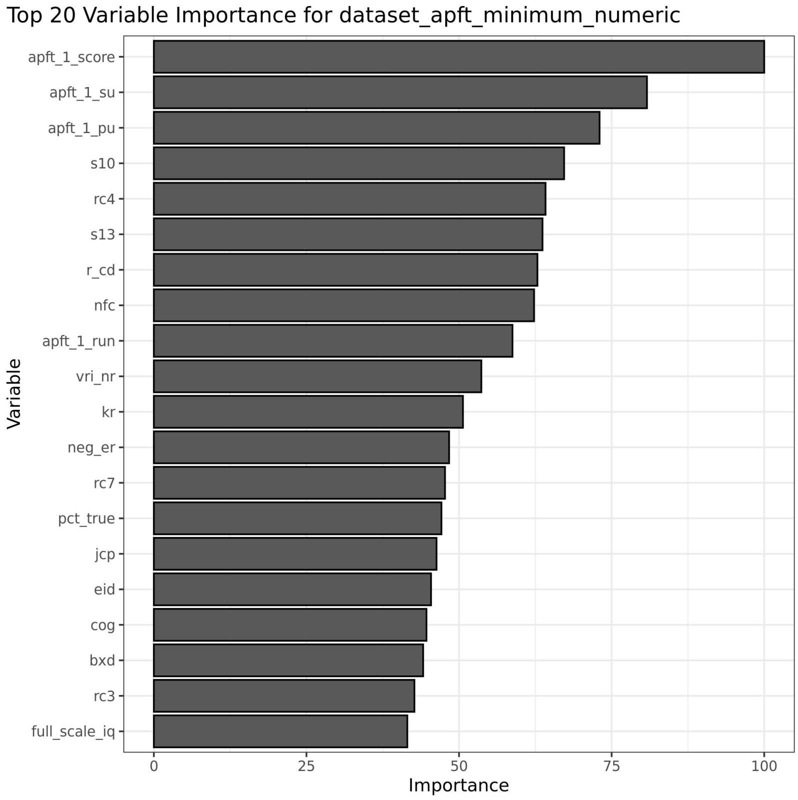


FIGURE B.4: Top 20 Most Important Features with candidates who scored above the minimum fitness test score for dataset_numeric using LDA

TABLE B.14: Feature Importance with candidates who scored above the minimum fitness test score for dataset_numeric using LDA

Feature	Importance
apft_1_score	100
apft_1_su	80.79
apft_1_pu	73.02
s10	67.22
rc4	64.17
s13	63.67
r_cd	62.84
nfc	62.29
apft_1_run	58.75
vri_nr	53.65
kr	50.64
neg_er	48.37
rc7	47.7
pct_true	47.12
jcp	46.31
eid	45.41
cog	44.66

Table B.14 continued from previous page

Feature	Importance
bxd	44.12
rc3	42.69
full_scale_iq	41.52
fr	41.07
fs	39.78
gt_score	39.76
psy_cr	39.1
thd	39.01
fml	39
rc1	38.49
dis_cr	37.89
rc8	37.4
sub	36.38
fpr	36.29
verbal_iq	36.22
rci	36.19
rc6	35.56
performance_iq	35.44
stw	35.41
mls	34.59
aes	34.2
rc2	33.89
shy	32.73
sfd	31.86
rbs	30.3
s15	29.35
s6	29.05
tis_at_arrival	28.46
act	25.98
rc9	25.22
nuc	25.21
sav	23.65
s9	22.92
agg	22.4
anp	22.1
tri_nr	19.41
s8	18.83
int_rr	18.09
brf	17.43
s1	17.22
lr	15.34
hlp	14.19
hpc	14.18
axy	13.92
msf	12.37
s4	10.63
s12	9.79
s14	9.78
inf	9.55
dsf	9.11
cannot_say	8.77
agg_rr	8.04
gic	7.58
s2	6.15
mec	3.84
s7	3.69
sui	3.56
s5	3.11
ipp	1.73
s11	1.32
fb_sr	1.03
s3	0.31
mis	0

B.2.3 QDA

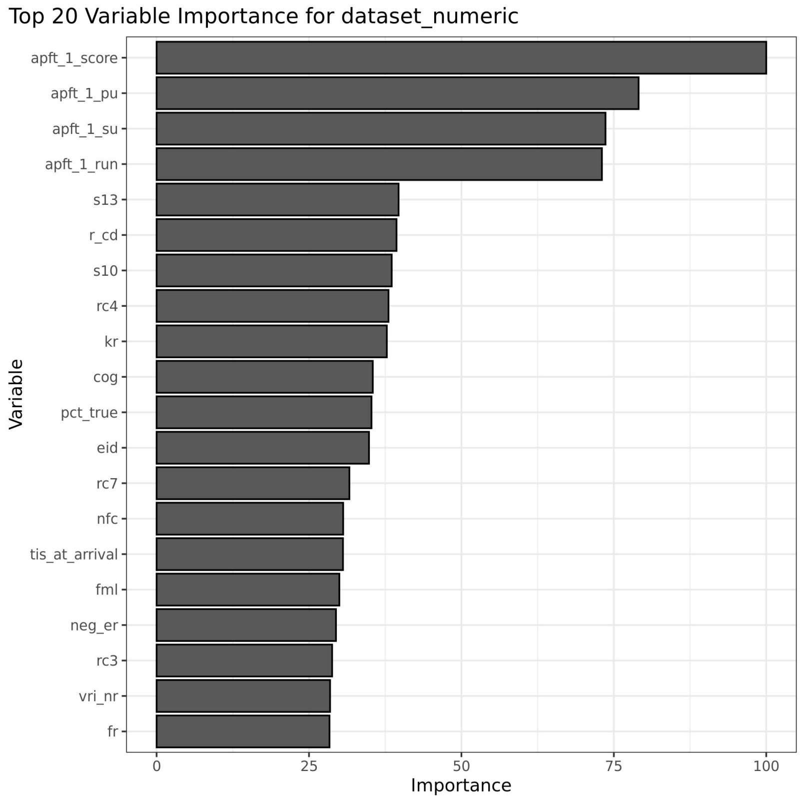


FIGURE B.5: Top 20 Most Important Features with all candidates for dataset_numeric using QDA

TABLE B.15: Feature Importance with all candidates for dataset_numeric using QDA

Feature	Importance
apft_1_score	100
apft_1_pu	79.06
apft_1_su	73.63
apft_1_run	73.05
s13	39.69
r_cd	39.36
s10	38.57
rc4	38.04
kr	37.76
cog	35.46
pct_true	35.25
eid	34.84
rc7	31.62
nfc	30.6
tis_at_arrival	30.57
fml	29.98
neg_er	29.42
rc3	28.79
vri_nr	28.46
fr	28.36
mls	27.49
sfd	27.17
rc2	26.31
jcp	26.23
psy_cr	26.08
bxd	26.06

Table B.15 continued from previous page

Feature	Importance
rc8	26.01
thd	24.76
shy	24.43
dis_cr	23.16
fs	22.93
rc1	22.48
fpr	22.27
rc6	22.21
anp	22.03
sub	21.49
stw	21.37
s9	21.21
act	20.94
agg	20.17
lr	20.12
hlp	19.28
full_scale_iq	19.06
rc9	18.99
aes	17.44
nuc	17.41
performance_iq	17.14
s14	16.92
rbs	16.62
s15	16.22
s6	15.77
verbal_iq	15.5
rci	15.25
gt_score	15.09
sav	14.02
s8	12.28
int_rr	11.59
s1	11.45
brf	10.32
hpc	10.19
dsf	9.79
axy	8.17
s2	8.13
s12	7.2
msf	6.89
mec	6.63
tri_nr	4.84
inf	4.69
agg_rr	4.63
s11	4.62
gic	4.47
sui	4.34
s3	2.5
fb_sr	2.1
mis	1.95
s4	0.65
s5	0.56
ipp	0.33
s7	0.03
cannot_say	0

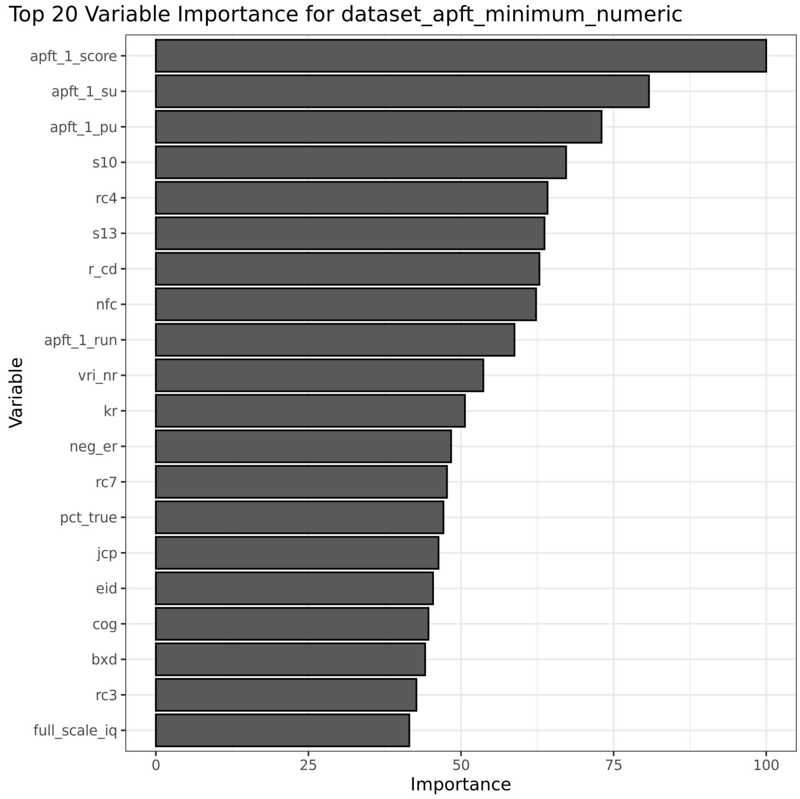


FIGURE B.6: Top 20 Most Important Features with candidates who scored above the minimum fitness test score for dataset_numeric using QDA

TABLE B.16: Feature Importance with candidates who scored above the minimum fitness test score for dataset_numeric using QDA

Feature	**Importance**
apft_1_score	100
apft_1_su	80.79
apft_1_pu	73.02
s10	67.22
rc4	64.17
s13	63.67
r_cd	62.84
nfc	62.29
apft_1_run	58.75
vri_nr	53.65
kr	50.64
neg_er	48.37
rc7	47.7
pct_true	47.12
jcp	46.31
eid	45.41
cog	44.66
bxd	44.12
rc3	42.69
full_scale_iq	41.52
fr	41.07
fs	39.78
gt_score	39.76
psy_cr	39.1
thd	39.01
fml	39
rc1	38.49
dis_cr	37.89

Table B.16 continued from previous page

Feature	Importance
rc8	37.4
sub	36.38
fpr	36.29
verbal_iq	36.22
rci	36.19
rc6	35.56
performance_iq	35.44
stw	35.41
mls	34.59
aes	34.2
rc2	33.89
shy	32.73
sfd	31.86
rbs	30.3
s15	29.35
s6	29.05
tis_at_arrival	28.46
act	25.98
rc9	25.22
nuc	25.21
sav	23.65
s9	22.92
agg	22.4
anp	22.1
tri_nr	19.41
s8	18.83
int_rr	18.09
brf	17.43
s1	17.22
lr	15.34
hlp	14.19
hpc	14.18
axy	13.92
msf	12.37
s4	10.63
s12	9.79
s14	9.78
inf	9.55
dsf	9.11
cannot_say	8.77
agg_rr	8.04
gic	7.58
s2	6.15
mec	3.84
s7	3.69
sui	3.56
s5	3.11
ipp	1.73
s11	1.32
fb_sr	1.03
s3	0.31
mis	0

B.2.4 SVM

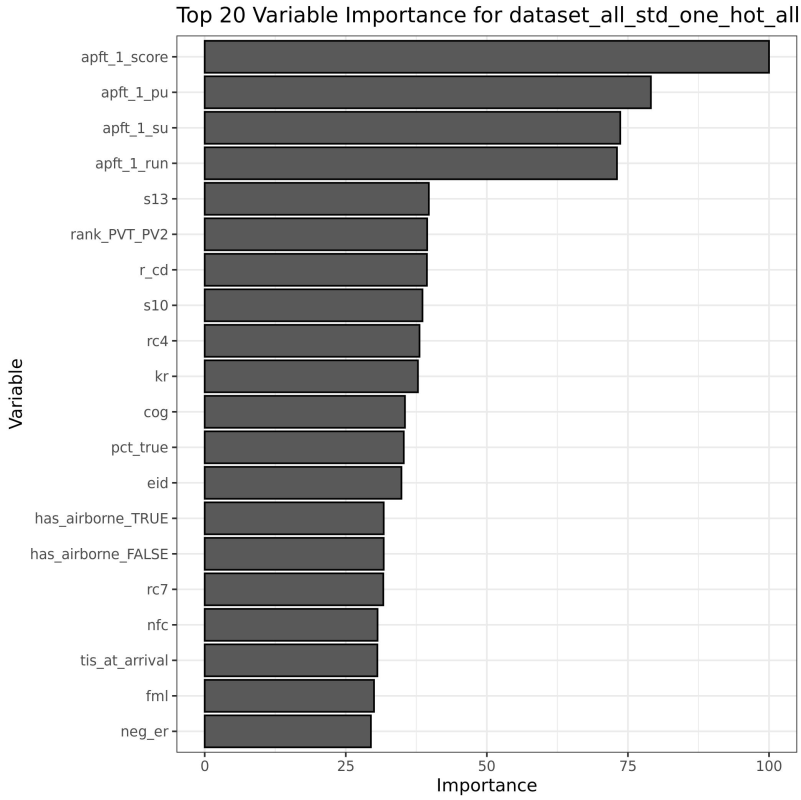


FIGURE B.7: Top 20 Most Important Features with all candidates for dataset_all using SVM

TABLE B.17: Feature Importance with all candidates for dataset_all using SVM

Feature	**Importance**
apft_1_score	100
apft_1_pu	79.07
apft_1_su	73.64
apft_1_run	73.06
s13	39.72
rank_PVT_PV2	39.42
r_cd	39.38
s10	38.59
rc4	38.07
kr	37.79
cog	35.49
pct_true	35.28
eid	34.87
has_airborne_TRUE	31.72
has_airborne_FALSE	31.72
rc7	31.65
nfc	30.63
tis_at_arrival	30.6
fml	30.01
neg_er	29.45
rc3	28.82
vri_nr	28.49
fr	28.39
mls	27.52
sfd	27.2
rc2	26.34

Table B.17 continued from previous page

Feature	Importance
jcp	26.26
psy_cr	26.11
bxd	26.1
rc8	26.04
thd	24.8
shy	24.46
dis_cr	23.19
fs	22.97
rc1	22.51
fpr	22.3
rank_SPC	22.27
rc6	22.25
anp	22.07
sub	21.52
stw	21.41
s9	21.24
act	20.98
agg	20.21
lr	20.15
hlp	19.31
full_scale_iq	19.09
rc9	19.03
aes	17.48
nuc	17.44
performance_iq	17.17
s14	16.95
rbs	16.66
s15	16.25
age_at_arrival_<=20	16.17
age_at_arrival_>20	15.94
s6	15.81
verbal_iq	15.54
rci	15.29
mos_18X	15.17
gt_score	15.13
mos_OTHER_MOS	14.73
sav	14.05
s8	12.31
rank_PFC	11.92
int_rr	11.62
s1	11.48
civilian_education_certification_BACHELORS DEGREE/MASTERS DEGREE	10.38
brf	10.36
hpc	10.23
dsf	9.83
arrival_month_JAN_FEB_MAR_AUG_NOV	9.23
axy	8.21
s2	8.17
s12	7.24
msf	6.93
mec	6.67
race_WHITE	6.55
glasses_TRUE	5.94
glasses_FALSE	5.94
civilian_education_certification_HIGH SCHOOL DIPLOMA	5.78
rank_CPL_SGT	5.24
tri_nr	4.89
inf	4.73
agg_rr	4.68
s11	4.66
gic	4.51
sui	4.38
arrival_month_MAY	4.03
race_OTHER	3.9
civilian_education_certification_UNKNOWN	3.8
mos_19D_68W_11C_13F_11B	3.23
civilian_education_certification_ALTERNATIVE HIGH SCHOOL/EQUIVALENCY CERTIFICATE	3
parents_together_FALSE	2.88

Table B.17 continued from previous page

Feature	Importance
parents_together_TRUE	2.88
mos_35G_M_N_P	2.78
arrival_month_APR_JUN_DEC	2.75
race_UNKNOWN	2.65
s3	2.54
arrival_month_JUL_SEP_OCT	2.46
civilian_education_certification_SOME COLLEGE/ASSOCIATES DEGREE	2.2
fb_sr	2.15
mis	1.99
s4	0.69
s5	0.6
ipp	0.37
age_at_arrival_UNKNOWN	0.22
s7	0.07
cannot_say	0.04
arrival_month_UNKNOWN	0

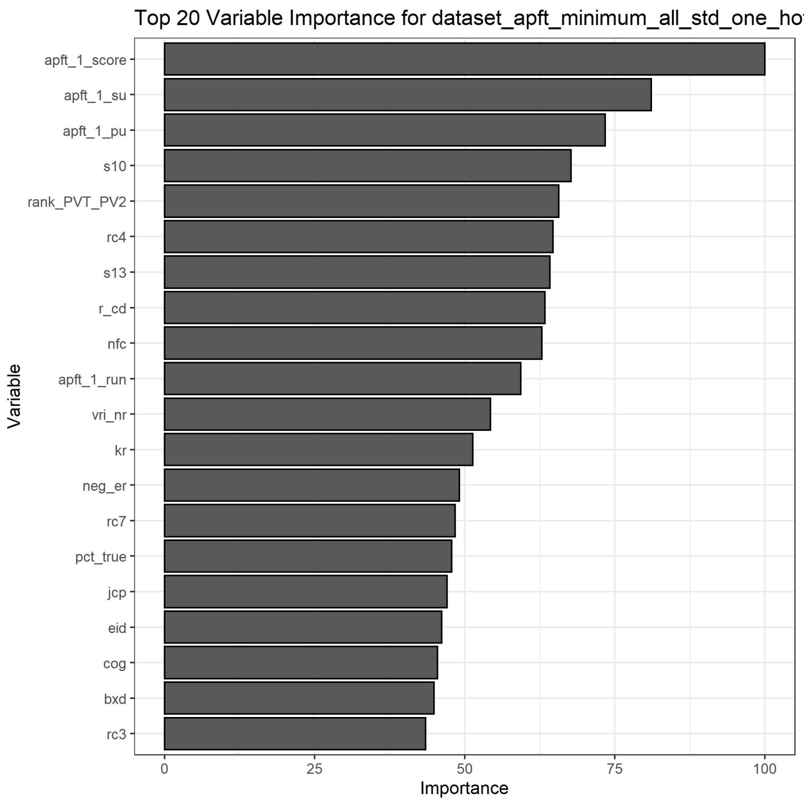


FIGURE B.8: Top 20 Most Important Features with candidates who scored above the minimum fitness test score for dataset_all using SVM

TABLE B.18: Feature Importance with candidates who scored above the minimum fitness test score for dataset_all using SVM

Feature	Importance
apft_1_score	100
apft_1_su	81.06
apft_1_pu	73.4
s10	67.67
rank_PVT_PV2	65.65
rc4	64.67
s13	64.18

Table B.18 continued from previous page

Feature	Importance
r_cd	63.36
nfc	62.81
apft_1_run	59.33
vri_nr	54.3
kr	51.33
neg_er	49.09
rc7	48.43
pct_true	47.85
jcp	47.06
eid	46.17
cog	45.43
bxd	44.9
rc3	43.49
full_scale_iq	42.34
fr	41.89
fs	40.62
gt_score	40.59
psy_cr	39.95
thd	39.85
fml	39.85
rc1	39.35
dis_cr	38.75
rc8	38.27
rank_SPC	38.22
sub	37.27
fpr	37.17
verbal_iq	37.11
rci	37.08
rc6	36.45
performance_iq	36.33
stw	36.31
mls	35.5
aes	35.12
rc2	34.81
shy	33.66
sfd	32.81
rbs	31.27
mos_OTHER_MOS	30.96
has_airborne_FALSE	30.67
has_airborne_TRUE	30.67
s15	30.33
s6	30.04
tis_at_arrival	29.45
act	27.01
age_at_arrival_<=20	26.99
age_at_arrival_>20	26.51
rc9	26.26
nuc	26.25
sav	24.71
s9	23.99
agg	23.48
anp	23.19
rank_PFC	20.86
civilian_education_certification_BACHELORS DEGREE/MASTERS DEGREE	20.8
tri_nr	20.54
mos_18X	20.31
s8	19.96
int_rr	19.23
brf	18.57
s1	18.37
lr	16.52
hlp	15.38
hpc	15.38
race_WHITE	15.32
axy	15.12
msf	13.59
civilian_education_certification_HIGH SCHOOL DIPLOMA	12.37
s4	11.87

Table B.18 continued from previous page

Feature	Importance
parents_together_TRUE	11.07
parents_together_FALSE	11.07
s12	11.04
s14	11.04
inf	10.81
dsf	10.37
cannot_say	10.04
agg_rr	9.32
gic	8.86
arrival_month_JAN_FEB_MAR_AUG_NOV	8.83
race_OTHER	8.04
mos_19D_68W_11C_13F_11B	7.57
s2	7.45
race_UNKNOWN	7.29
glasses_FALSE	7.1
glasses_TRUE	7.1
arrival_month_MAY	7.05
rank_CPL_SGT	6.57
civilian_education_certification_UNKNOWN	6.11
mec	5.18
s7	5.03
sui	4.9
arrival_month_JUL_SEP_OCT	4.6
s5	4.46
civilian_education_certification_ALTERNATIVE HIGH SCHOOL/EQUIVALENCY CERTIFICATE	3.16
ipp	3.09
mos_35G_M_N_P	3.07
arrival_month_APR_JUN_DEC	2.81
s11	2.69
fb_sr	2.41
s3	1.7
mis	1.39
civilian_education_certification_SOME COLLEGE/ASSOCIATES DEGREE	0.85
age_at_arrival_UNKNOWN	0.48
arrival_month_UNKNOWN	0

B.2.5 CART

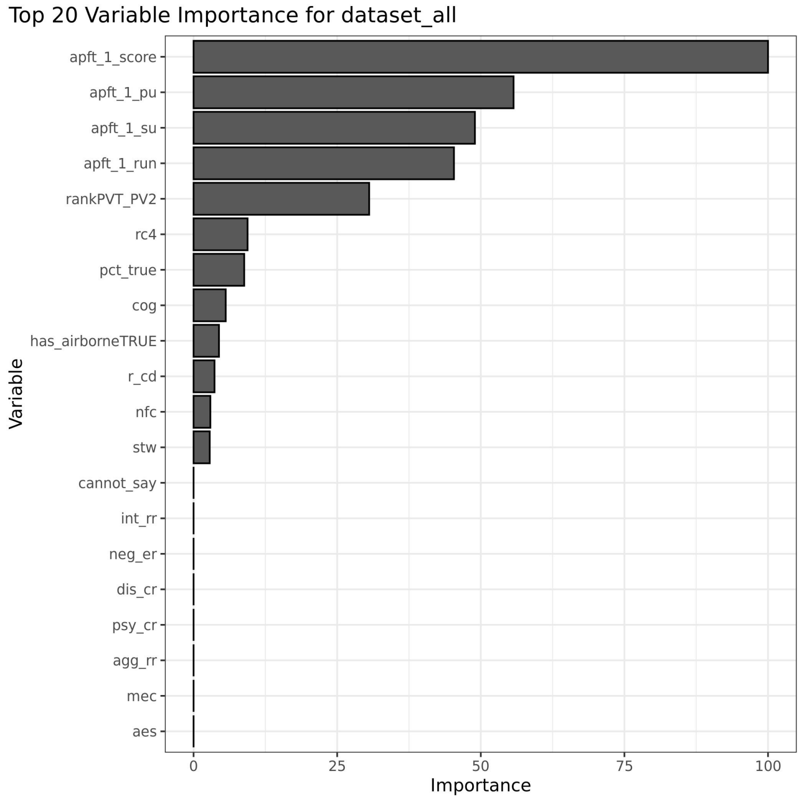


FIGURE B.9: Top 20 Most Important Features with all candidates for dataset_all using CART

TABLE B.19: Feature Importance with all candidates for dataset_all using CART

Feature	Importance
apft_1_score	100
apft_1_pu	55.7
apft_1_su	48.98
apft_1_run	45.33
rankPVT_PV2	30.57
rc4	9.4
pct_true	8.81
cog	5.6
has_airborneTRUE	4.42
r_cd	3.65
nfc	2.92
stw	2.81
mos19D_68W_11C_13F_11B	0
mos35G_M_N_P	0
mosOTHER_MOS	0
rankPFC	0
rankSPC	0
raceUNKNOWN	0
raceWHITE	0
arrival_monthJAN_FEB_MAR_AUG_NOV	0
arrival_monthJUL_SEP_OCT	0
arrival_monthMAY	0
arrival_monthUNKNOWN	0
tis_at_arrival	0
parents_togetherTRUE	0
glassesTRUE	0

Table B.19 continued from previous page

Feature	Importance
‘civilian_education_certificationBACHELORS DEGREE/MASTERS DEGREE‘	0
‘civilian_education_certificationHIGH SCHOOL DIPLOMA‘	0
‘civilian_education_certificationSOME COLLEGE/ASSOCIATES DEGREE‘	0
civilian_education_certificationUNKNOWN	0
‘age_at_arrival>20‘	0
age_at_arrivalUNKNOWN	0
gt_score	0
s1	0
s2	0
s3	0
s4	0
s5	0
s6	0
s7	0
s8	0
s9	0
s10	0
s11	0
s12	0
s13	0
s14	0
s15	0
rci	0
inf	0
mis	0
verbal_iq	0
performance_iq	0
full_scale_iq	0
vri_nr	0
tri_nr	0
fr	0
fpr	0
fs	0
fb_sr	0
rbs	0
lr	0
kr	0
eid	0
thd	0
bxd	0
rc1	0
rc2	0
rc3	0
rc6	0
rc7	0
rc8	0
rc9	0
mls	0
hpc	0
nuc	0
gic	0
sui	0
hlp	0
sfd	0
axy	0
anp	0
brf	0
msf	0
jcp	0
sub	0
agg	0
act	0
fml	0
ipp	0
sav	0
shy	0
dsf	0
aes	0

Table B.19 continued from previous page

Feature	Importance
mec	0
agg_rr	0
psy_cr	0
dis_cr	0
neg_er	0
int_rr	0
cannot_say	0

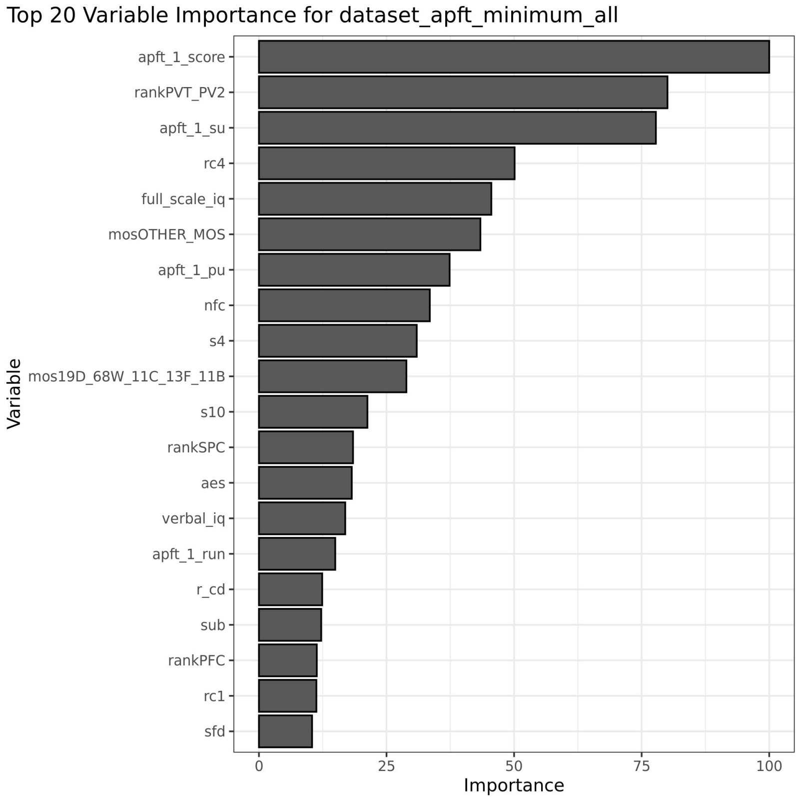


FIGURE B.10: Top 20 Most Important Features with candidates who scored above the minimum fitness test score for dataset_all using CART

TABLE B.20: Feature Importance with candidates who scored above the minimum fitness test score for dataset_all using CART

Feature	Importance
apft_1_score	100
rankPVT_PV2	80.06
apft_1_su	77.79
rc4	50.11
full_scale_iq	45.54
mosOTHER_MOS	43.4
apft_1_pu	37.37
nfc	33.48
s4	30.92
mos19D_68W_11C_13F_11B	28.9
s10	21.26
rankSPC	18.44
aes	18.19
verbal_iq	16.91
apft_1_run	14.95
r_cd	12.4

Table B.20 continued from previous page

Feature	Importance
sub	12.2
rankPFC	11.34
rc1	11.26
sfd	10.41
parents_togetherTRUE	8.2
fpr	5.34
mos35G_M_N_P	0
raceUNKNOWN	0
raceWHITE	0
arrival_monthJAN_FEB_MAR_AUG_NOV	0
arrival_monthJUL_SEP_OCT	0
arrival_monthMAY	0
arrival_monthUNKNOWN	0
tis_at_arrival	0
has_airborneTRUE	0
glassesTRUE	0
‘civilian_education_certificationBACHELORS DEGREE/MASTERS DEGREE‘	0
‘civilian_education_certificationHIGH SCHOOL DIPLOMA‘	0
‘civilian_education_certificationSOME COLLEGE/ASSOCIATES DEGREE‘	0
civilian_education_certificationUNKNOWN	0
‘age_at_arrival>20‘	0
age_at_arrivalUNKNOWN	0
gt_score	0
s1	0
s2	0
s3	0
s5	0
s6	0
s7	0
s8	0
s9	0
s11	0
s12	0
s13	0
s14	0
s15	0
rci	0
inf	0
mis	0
performance_iq	0
vri_nr	0
tri_nr	0
fr	0
fs	0
fb_sr	0
rbs	0
lr	0
kr	0
eid	0
thd	0
bxd	0
rc2	0
rc3	0
rc6	0
rc7	0
rc8	0
rc9	0
mls	0
hpc	0
nuc	0
gic	0
sui	0
hlp	0
cog	0
stw	0
axy	0
anp	0
brf	0

Table B.20 continued from previous page

Feature	Importance
msf	0
jcp	0
agg	0
act	0
fml	0
ipp	0
sav	0
shy	0
dsf	0
mec	0
agg_rr	0
psy_cr	0
dis_cr	0
neg_er	0
int_rr	0
cannot_say	0
pct_true	0

B.2.6 KNN

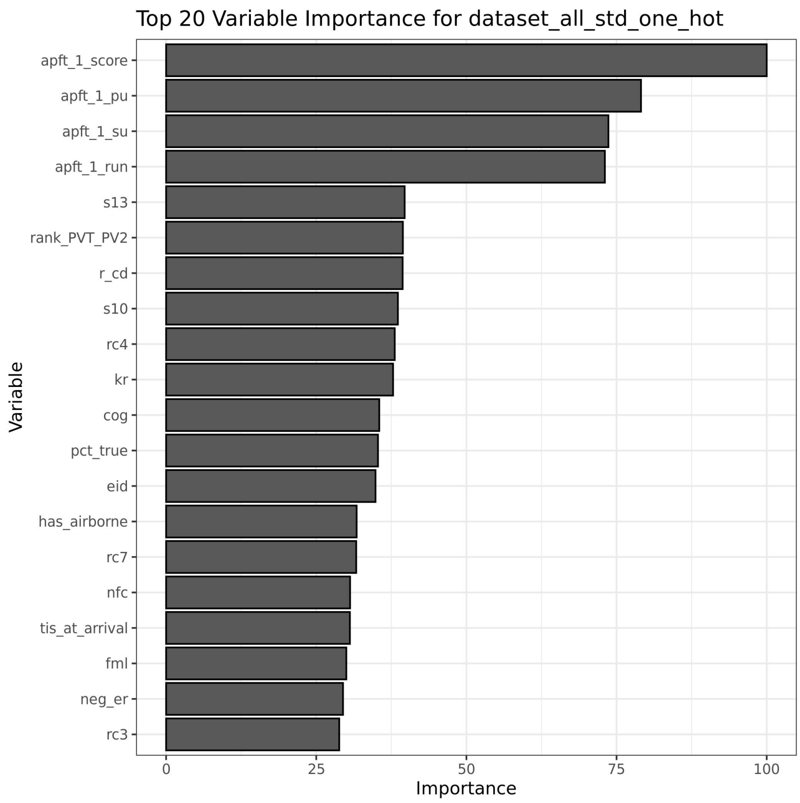


FIGURE B.11: Top 20 Most Important Features with all candidates for dataset_all using KNN

TABLE B.21: Feature Importance with all candidates for dataset_all using KNN

Feature	Importance
apft_1_score	100
apft_1_pu	79.07
apft_1_su	73.64
apft_1_run	73.06
s13	39.72

Table B.21 continued from previous page

Feature	Importance
rank_PVT_PV2	39.42
r_cd	39.38
s10	38.59
rc4	38.07
kr	37.79
cog	35.49
pct_true	35.28
eid	34.87
has_airborne_TRUE	31.72
has_airborne_FALSE	31.72
rc7	31.65
nfc	30.63
tis_at_arrival	30.6
fml	30.01
neg_er	29.45
rc3	28.82
vri_nr	28.49
fr	28.39
mls	27.52
sfd	27.2
rc2	26.34
jcp	26.26
psy_cr	26.11
bxd	26.1
rc8	26.04
thd	24.8
shy	24.46
dis_cr	23.19
fs	22.97
rc1	22.51
fpr	22.3
rank_SPC	22.27
rc6	22.25
anp	22.07
sub	21.52
stw	21.41
s9	21.24
act	20.98
agg	20.21
lr	20.15
hlp	19.31
full_scale_iq	19.09
rc9	19.03
aes	17.48
nuc	17.44
performance_iq	17.17
s14	16.95
rbs	16.66
s15	16.25
age_at_arrival_<=20	16.17
age_at_arrival_>20	15.94
s6	15.81
verbal_iq	15.54
rci	15.29
mos_18X	15.17
gt_score	15.13
mos_OTHER_MOS	14.73
sav	14.05
s8	12.31
rank_PFC	11.92
int_rr	11.62
s1	11.48
civilian_education_certification_BACHELORS DEGREE/MASTERS DEGREE	10.38
brf	10.36
hpc	10.23
dsf	9.83
arrival_month_JAN_FEB_MAR_AUG_NOV	9.23
axy	8.21

Table B.21 continued from previous page

Feature	Importance
s2	8.17
s12	7.24
msf	6.93
mec	6.67
race_WHITE	6.55
glasses_TRUE	5.94
glasses_FALSE	5.94
civilian_education_certification_HIGH SCHOOL DIPLOMA	5.78
rank_CPL_SGT	5.24
tri_nr	4.89
inf	4.73
agg_rr	4.68
s11	4.66
gic	4.51
sui	4.38
arrival_month_MAY	4.03
race_OTHER	3.9
civilian_education_certification_UNKNOWN	3.8
mos_19D_68W_11C_13F_11B	3.23
civilian_education_certification_ALTERNATIVE HIGH SCHOOL/EQUIVALENCY CERTIFICATE	3
parents_together_FALSE	2.88
parents_together_TRUE	2.88
mos_35G_M_N_P	2.78
arrival_month_APR_JUN_DEC	2.75
race_UNKNOWN	2.65
s3	2.54
arrival_month_JUL_SEP_OCT	2.46
civilian_education_certification_SOME COLLEGE/ASSOCIATES DEGREE	2.2
fb_sr	2.15
mis	1.99
s4	0.69
s5	0.6
ipp	0.37
age_at_arrival_UNKNOWN	0.22
s7	0.07
cannot_say	0.04
arrival_month_UNKNOWN	0

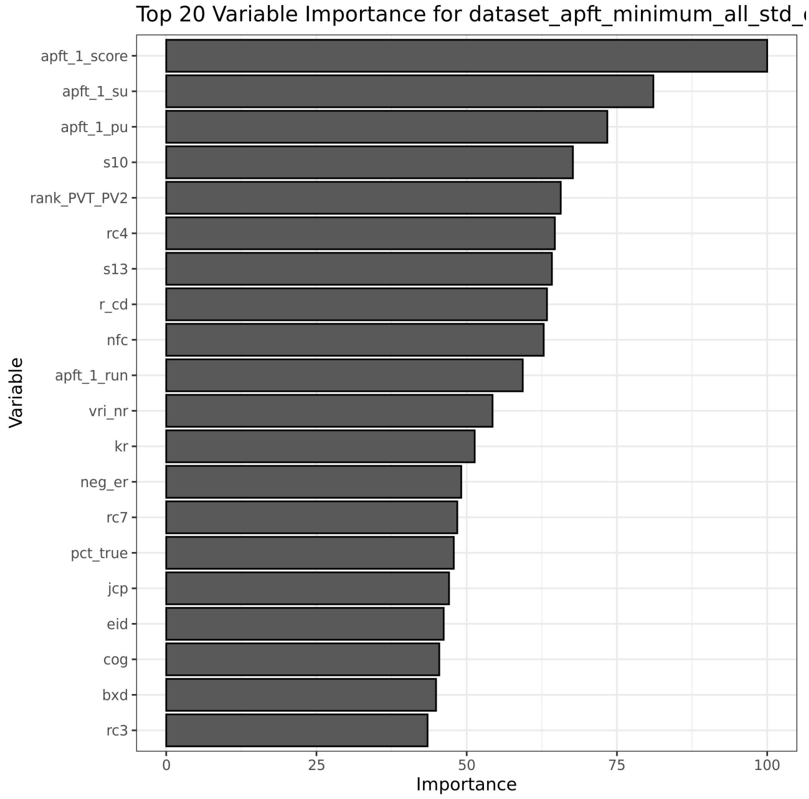


FIGURE B.12: Top 20 Most Important Features with candidates who scored above the minimum fitness test score for dataset_all using KNN

TABLE B.22: Feature Importance with candidates who scored above the minimum fitness test score for dataset_all using KNN

Feature	**Importance**
apft_1_score	100
apft_1_su	81.06
apft_1_pu	73.4
s10	67.67
rank_PVT_PV2	65.65
rc4	64.67
s13	64.18
r_cd	63.36
nfc	62.81
apft_1_run	59.33
vri_nr	54.3
kr	51.33
neg_er	49.09
rc7	48.43
pct_true	47.85
jcp	47.06
eid	46.17
cog	45.43
bxd	44.9
rc3	43.49
full_scale_iq	42.34
fr	41.89
fs	40.62
gt_score	40.59
psy_cr	39.95
thd	39.85
fml	39.85
rc1	39.35

Table B.22 continued from previous page

Feature	Importance
dis_cr	38.75
rc8	38.27
rank_SPC	38.22
sub	37.27
fpr	37.17
verbal_iq	37.11
rci	37.08
rc6	36.45
performance_iq	36.33
stw	36.31
mls	35.5
aes	35.12
rc2	34.81
shy	33.66
sfd	32.81
rbs	31.27
mos_OTHER_MOS	30.96
has_airborne_FALSE	30.67
has_airborne_TRUE	30.67
s15	30.33
s6	30.04
tis_at_arrival	29.45
act	27.01
age_at_arrival_<=20	26.99
age_at_arrival_>20	26.51
rc9	26.26
nuc	26.25
sav	24.71
s9	23.99
agg	23.48
anp	23.19
rank_PFC	20.86
civilian_education_certification_BACHELORS DEGREE/MASTERS DEGREE	20.8
tri_nr	20.54
mos_18X	20.31
s8	19.96
int_rr	19.23
brf	18.57
s1	18.37
lr	16.52
hlp	15.38
hpc	15.38
race_WHITE	15.32
axy	15.12
msf	13.59
civilian_education_certification_HIGH SCHOOL DIPLOMA	12.37
s4	11.87
parents_together_TRUE	11.07
parents_together_FALSE	11.07
s12	11.04
s14	11.04
inf	10.81
dsf	10.37
cannot_say	10.04
agg_rr	9.32
gic	8.86
arrival_month_JAN_FEB_MAR_AUG_NOV	8.83
race_OTHER	8.04
mos_19D_68W_11C_13F_11B	7.57
s2	7.45
race_UNKNOWN	7.29
glasses_FALSE	7.1
glasses_TRUE	7.1
arrival_month_MAY	7.05
rank_CPL_SGT	6.57
civilian_education_certification_UNKNOWN	6.11
mec	5.18
s7	5.03

Table B.22 continued from previous page

Feature	Importance
sui	4.9
arrival_month_JUL_SEP_OCT	4.6
s5	4.46
civilian_education_certification_ALTERNATIVE HIGH SCHOOL/EQUIVALENCY CERTIFICATE	3.16
ipp	3.09
mos_35G_M_N_P	3.07
arrival_month_APR_JUN_DEC	2.81
s11	2.69
fb_sr	2.41
s3	1.7
mis	1.39
civilian_education_certification_SOME COLLEGE/ASSOCIATES DEGREE	0.85
age_at_arrival_UNKNOWN	0.48
arrival_month_UNKNOWN	0

B.2.7 Random Forest

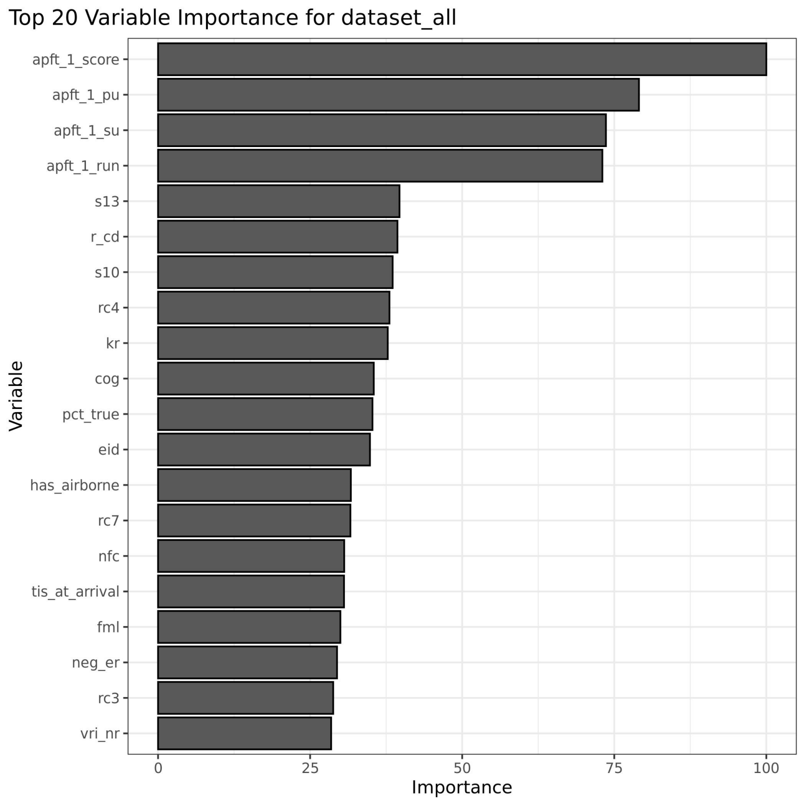


FIGURE B.13: Top 20 Most Important Features with all candidates for dataset_all using Random Forest

TABLE B.23: Feature Importance with all candidates for dataset_all using Random Forest

Feature	Importance
apft_1_score	100
apft_1_pu	79.06
apft_1_su	73.63
apft_1_run	73.05
s13	39.69
r_cd	39.36
s10	38.57
rc4	38.04

Table B.23 continued from previous page

Feature	Importance
kr	37.76
cog	35.46
pct_true	35.25
eid	34.84
has_airborne	31.69
rc7	31.62
nfc	30.6
tis_at_arrival	30.57
fml	29.98
neg_er	29.42
rc3	28.79
vri_nr	28.46
fr	28.36
mls	27.49
sfd	27.17
rc2	26.31
jcp	26.23
psy_cr	26.08
bxd	26.06
rc8	26.01
thd	24.76
shy	24.43
mos	23.9
dis_cr	23.16
fs	22.93
rc1	22.48
fpr	22.27
rc6	22.21
anp	22.03
sub	21.49
stw	21.37
s9	21.21
act	20.94
agg	20.17
lr	20.12
hlp	19.28
full_scale_iq	19.06
rc9	18.99
aes	17.44
nuc	17.41
performance_iq	17.14
s14	16.92
rbs	16.62
age_at_arrival	16.22
s15	16.22
s6	15.77
verbal_iq	15.5
rci	15.25
gt_score	15.09
sav	14.02
s8	12.28
int_rr	11.59
s1	11.45
brf	10.32
hpc	10.19
dsf	9.79
axy	8.17
s2	8.13
s12	7.2
civilian_education_certification	7.07
msf	6.89
mec	6.63
race	6.62
glasses	5.9
tri_nr	4.84
inf	4.69
agg_rr	4.63
s11	4.62

Table B.23 continued from previous page

Feature	Importance
gic	4.47
sui	4.34
arrival_month	3.8
parents_together	2.83
s3	2.5
fb_sr	2.1
mis	1.95
rank	1.53
s4	0.65
s5	0.56
ipp	0.33
s7	0.03
cannot_say	0

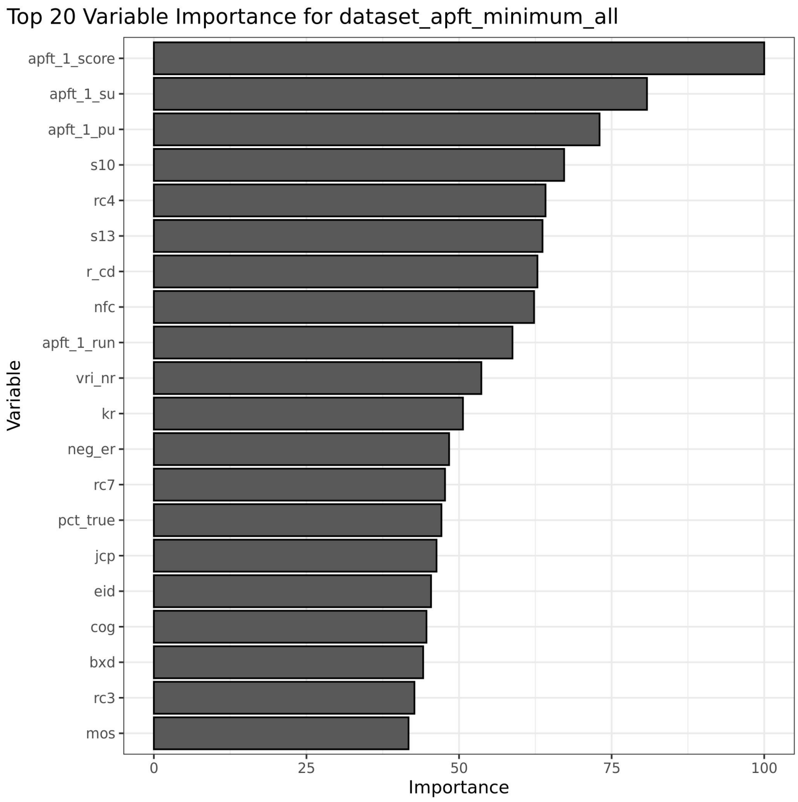


FIGURE B.14: Top 20 Most Important Features with candidates who scored above the minimum fitness test score for dataset_all using Random Forest

TABLE B.24: Feature Importance with candidates who scored above the minimum fitness test score for dataset_all using Random Forest

Feature	Importance
apft_1_score	100
apft_1_su	80.79
apft_1_pu	73.02
s10	67.22
rc4	64.17
s13	63.67
r_cd	62.84
nfc	62.29
apft_1_run	58.75
vri_nr	53.65

Table B.24 continued from previous page

Feature	Importance
kr	50.64
neg_er	48.37
rc7	47.7
pct_true	47.12
jcp	46.31
eid	45.41
cog	44.66
bxd	44.12
rc3	42.69
mos	41.72
full_scale_iq	41.52
fr	41.07
fs	39.78
gt_score	39.76
psy_cr	39.1
thd	39.01
fml	39
rc1	38.49
dis_cr	37.89
rc8	37.4
sub	36.38
fpr	36.29
verbal_iq	36.22
rci	36.19
rc6	35.56
performance_iq	35.44
stw	35.41
mls	34.59
aes	34.2
rc2	33.89
shy	32.73
sfd	31.86
rbs	30.3
has_airborne	29.69
s15	29.35
s6	29.05
tis_at_arrival	28.46
age_at_arrival	26.17
act	25.98
rc9	25.22
nuc	25.21
sav	23.65
s9	22.92
agg	22.4
anp	22.1
tri_nr	19.41
s8	18.83
int_rr	18.09
brf	17.43
s1	17.22
civilian_education_certification	16.8
lr	15.34
hlp	14.19
hpc	14.18
race	14.1
axy	13.92
msf	12.37
arrival_month	11.23
s4	10.63
parents_together	9.82
s12	9.79
s14	9.78
inf	9.55
dsf	9.11
cannot_say	8.77
agg_rr	8.04
gic	7.58
s2	6.15

Table B.24 continued from previous page

Feature	Importance
glasses	5.79
rank	4.27
mec	3.84
s7	3.69
sui	3.56
s5	3.11
ipp	1.73
s11	1.32
fb_sr	1.03
s3	0.31
mis	0

B.2.8 xgboost

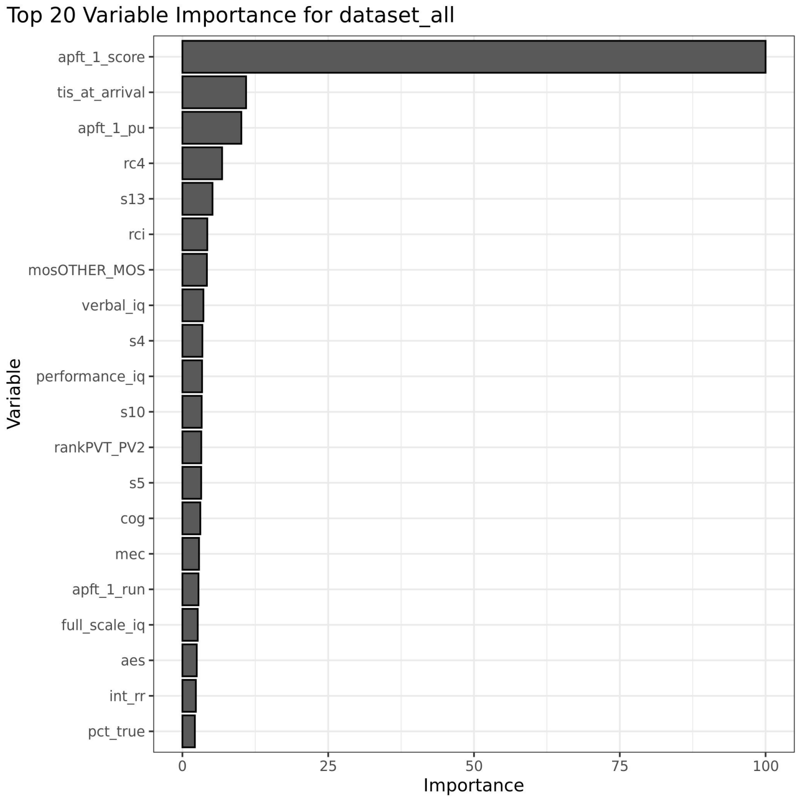


FIGURE B.15: Top 20 Most Important Features with all candidates for dataset_all using xgboost

TABLE B.25: Feature Importance with all candidates for dataset_all using xgboost

Feature	Importance
apft_1_score	100
tis_at_arrival	10.92
apft_1_pu	10.1
rc4	6.81
s13	5.16
rci	4.27
mosOTHER_MOS	4.2
verbal_iq	3.6
s4	3.42
performance_iq	3.37
s10	3.31

Table B.25 continued from previous page

Feature	Importance
rankPVT_PV2	3.24
s5	3.22
cog	3.07
mec	2.85
apft_1_run	2.76
full_scale_iq	2.63
aes	2.46
int_rr	2.31
pct_true	2.13
act	1.96
s15	1.93
jcp	1.87
r_cd	1.8
inf	1.77
gt_score	1.75
shy	1.72
s12	1.69
kr	1.55
agg	1.47
hpc	1.39
dsf	1.39
rbs	1.33
s7	1.33
ipp	1.29
s11	1.27
mis	1.22
psy_cr	1.2
arrival_monthJAN_FEB_MAR_AUG_NOV	1.17
has_airborneTRUE	1.13
s8	1.01
arrival_monthMAY	1
rc3	0.99
rc2	0.95
mls	0.93
apft_1_su	0.86
rankSPC	0.84
sub	0.83
glassesFALSE	0.81
s1	0.81
fr	0.77
thd	0.76
stw	0.75
msf	0.73
rc8	0.71
rc7	0.68
agg_rr	0.66
s3	0.66
rc6	0.63
rc1	0.61
eid	0.6
fpr	0.59
s14	0.56
has_airborneFALSE	0.55
rc9	0.54
raceOTHER	0.52
s6	0.5
neg_er	0.47
lr	0.46
fb_sr	0.45
sav	0.43
fs	0.43
vri_nr	0.4
s2	0.39
hlp	0.33
bxd	0.29
sfd	0.25
nuc	0.22
fml	0.2

Table B.25 continued from previous page

Feature	Importance
sui	0.18
age_at_arrival<=20	0.16
tri_nr	0.01
mos18X	0
mos19D_68W_11C_13F_11B	0
mos35G_M_N_P	0
rankCPL_SGT	0
rankPFC	0
raceUNKNOWN	0
raceWHITE	0
arrival_monthAPR_JUN_DEC	0
arrival_monthJUL_SEP_OCT	0
arrival_monthUNKNOWN	0
parents_togetherFALSE	0
parents_togetherTRUE	0
glassesTRUE	0
civilian_education_certificationALTERNATIVE HIGH SCHOOL/EQUIVALENCY CERTIFICATE	0
civilian_education_certificationBACHELORS DEGREE/MASTERS DEGREE	0
civilian_education_certificationHIGH SCHOOL DIPLOMA	0
civilian_education_certificationSOME COLLEGE/ASSOCIATES DEGREE	0
civilian_education_certificationUNKNOWN	0
age_at_arrival>20	0
age_at_arrivalUNKNOWN	0
s9	0
gic	0
nfc	0
axy	0
anp	0
brf	0
dis_cr	0
cannot_say	0

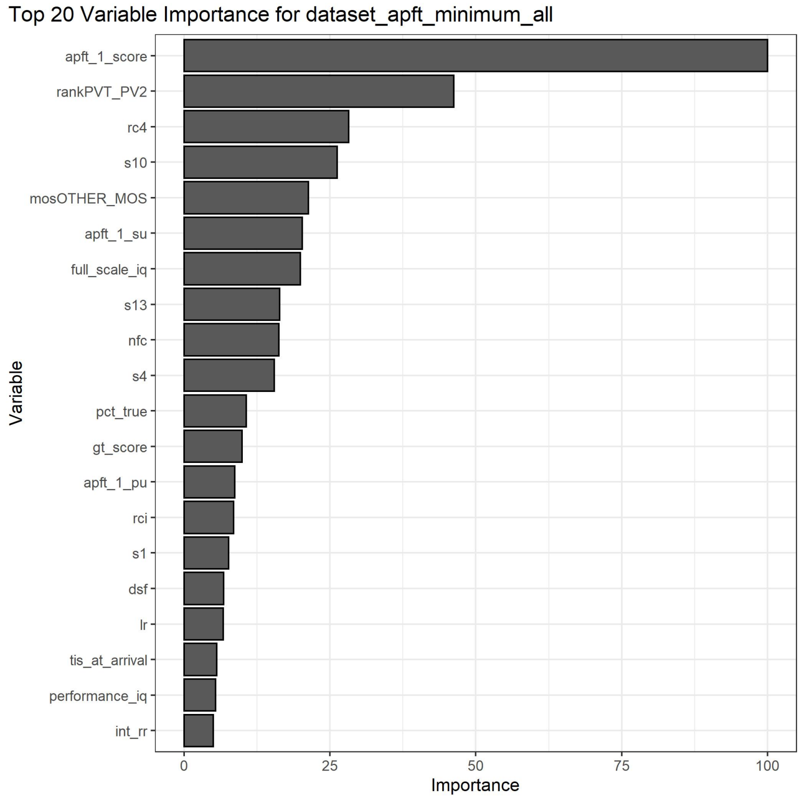


FIGURE B.16: Top 20 Most Important Features with candidates who scored above the minimum fitness test score for dataset_all using xgboost

TABLE B.26: Feature Importance with candidates who scored above the minimum fitness test score for dataset_all using xgboost

Feature	Importance
apft_1_score	100
rankPVT_PV2	46.19
rc4	28.21
s10	26.22
mosOTHER_MOS	21.28
apft_1_su	20.21
full_scale_iq	19.88
s13	16.33
nfc	16.2
s4	15.42
pct_true	10.65
gt_score	9.93
apft_1_pu	8.69
rci	8.46
s1	7.62
dsf	6.78
lr	6.71
tis_at_arrival	5.56
performance_iq	5.4
int_rr	4.97
inf	4.28
rc1	4.06
apft_1_run	4.02
s9	3.86
s15	3.46
eid	3.16
psy_cr	2.69
rbs	2.35

Table B.26 continued from previous page

Feature	Importance
shy	2.25
mls	2.14
verbal_iq	2.09
rc9	1.96
s11	1.79
ipp	1.73
mis	1.64
jcp	1.62
cannot_say	1.62
fr	1.61
cog	1.56
hpc	1.52
fml	1.5
sub	1.48
msf	1.44
sfd	1.39
s2	1.38
s5	1.37
rc3	1.32
mos19D_68W_11C_13F_11B	0
mos35G_M_N_P	0
rankPFC	0
rankSPC	0
raceUNKNOWN	0
raceWHITE	0
arrival_monthJAN_FEB_MAR_AUG_NOV	0
arrival_monthJUL_SEP_OCT	0
arrival_monthMAY	0
arrival_monthUNKNOWN	0
parents_togetherTRUE	0
has_airborneTRUE	0
glassesTRUE	0
civilian_education_certificationBACHELORS DEGREE/MASTERS DEGREE	0
civilian_education_certificationHIGH SCHOOL DIPLOMA	0
civilian_education_certificationSOME COLLEGE/ASSOCIATES DEGREE	0
civilian_education_certificationUNKNOWN	0
age_at_arrival>20	0
age_at_arrivalUNKNOWN	0
s3	0
s6	0
s7	0
s8	0
s12	0
s14	0
vri_nr	0
tri_nr	0
fpr	0
fs	0
fb_sr	0
kr	0
thd	0
bxd	0
r_cd	0
rc2	0
rc6	0
rc7	0
rc8	0
nuc	0
gic	0
sui	0
hlp	0
stw	0
axy	0
anp	0
brf	0
agg	0
act	0
sav	0

Table B.26 continued from previous page

Feature	Importance
aes	0
mec	0
agg_rr	0
dis_cr	0
neg_er	0

B.3 Model Coefficients

The only models that used coefficients were Penalized Logistic Regression and LDA. To be concise, but also show the holistic feature importance measures, we will share the feature importance of only the data sets using all features with no feature engineering, for both all candidates and candidates who scores above the minimum fitness screening criteria.

The coefficients are ordered by the maximum absolute value to least. The larger the coefficient, the more important it is in prediction. Also, coefficients have an inverse relationship to the response feature. If the coefficient is **negative**, it means that it impacted selection **positively** (e.g. apft_1_score is negative so it means candidates with high fitness test scores were more likely to be selected). If the coefficient is **positive**, it means that it impacted selection **negatively** (e.g. if a candidate is rank_PVT_PVT, then it means they were less likely to be selected).

B.3.1 Penalized Logistic Regression (using Elastic Net)

TABLE B.27: Model Coefficients with all candidates for dataset_all using Penalized Logistic Regression

Feature	Coefficient
apft_1_score	-1.39455
(Intercept)	0.41729
mos_OTHER_MOS	0.41406
rank_PVT_PV2	0.39843
rc4	0.19387
glasses_FALSE	-0.19306
mos_18X	-0.18679
arrival_month_JAN_FEB_MAR_AUG_NOV	-0.14838
arrival_month_MAY	0.14485
s10	-0.13951
aes	0.11933
has_airborne_FALSE	0.11169
rank_CPL_SGT	-0.10809
s13	-0.10137
s4	0.07238
glasses_TRUE	0.05641
race_WHITE	-0.05255
hlp	0.05015
has_airborne_TRUE	-0.0417
rank_SPC	-0.04118
rbs	0.0389
hpc	-0.03819
performance_iq	-0.03542
cog	0.0342
sub	0.02624
'civilian_education_certification_HIGH SCHOOL DIPLOMA'	-0.01791
apft_1_run	0.01742
mls	0.00978
sfd	0.00544
rc2	0.00281
s1	0.00165
sui	5.80E-04
mos_19D_68W_11C_13F_11B	0

Table B.27 continued from previous page

Feature	Coefficient
mos_35G_M_N_P	0
rank_PFC	0
race_OTHER	0
race_UNKNOWN	0
arrival_month_APR_JUN_DEC	0
arrival_month_JUL_SEP_OCT	0
arrival_month_UNKNOWN	0
tis_at_arrival	0
parents_together_FALSE	0
parents_together_TRUE	0
‘civilian_education_certification_ALTERNATIVE HIGH SCHOOL/EQUIVALENCY CERTIFICATE‘	0
‘civilian_education_certification_BACHELORS DEGREE/MASTERS DEGREE‘	0
‘civilian_education_certification_SOME COLLEGE/ASSOCIATES DEGREE‘	0
civilian_education_certification_UNKNOWN	0
‘age_at_arrival_<=20‘	0
‘age_at_arrival_>20‘	0
age_at_arrival_UNKNOWN	0
gt_score	0
apft_1_pu	0
apft_1_su	0
s2	0
s3	0
s5	0
s6	0
s7	0
s8	0
s9	0
s11	0
s12	0
s14	0
s15	0
rci	0
inf	0
mis	0
verbal_iq	0
full_scale_iq	0
vri_nr	0
tri_nr	0
fr	0
fpr	0
fs	0
fb_sr	0
lr	0
kr	0
eid	0
thd	0
bxd	0
r_cd	0
rc1	0
rc3	0
rc6	0
rc7	0
rc8	0
rc9	0
nuc	0
gic	0
nfc	0
stw	0
axy	0
anp	0
brf	0
msf	0
jcp	0
agg	0
act	0
fml	0
ipp	0
sav	0

Table B.27 continued from previous page

Feature	Coefficient
shy	0
dsf	0
mec	0
agg_rr	0
psy_cr	0
dis_cr	0
neg_er	0
int_rr	0
cannot_say	0
pct_true	0

TABLE B.28: Model Coefficients with candidates who scored above the minimum fitness test score for dataset_all using Penalized Logistic Regression

Feature	Coefficient
rank_PVT_PV2	0.38842
mos_OTHER_MOS	0.38054
apft_1_score	-0.29684
(Intercept)	-0.2826
arrival_month_MAY	0.15892
rank_SPC	-0.15189
rc4	0.15148
s10	-0.13254
apft_1_su	-0.12301
s4	0.09276
s13	-0.08668
full_scale_iq	-0.08509
aes	0.07655
nfc	0.05165
apft_1_pu	-0.04588
r_cd	0.03412
has_airborne_FALSE	0.03162
has_airborne_TRUE	-0.02998
rbs	0.02571
inf	-0.02357
s1	0.02236
performance_iq	-0.02013
parents_together_FALSE	0.01906
sub	0.01873
parents_together_TRUE	-0.01684
verbal_iq	-0.01347
rc2	0.01234
dsf	-0.00868
fs	0.00359
tri_nr	6.60E-04
mos_18X	0
mos_19D_68W_11C_13F_11B	0
mos_35G_M_N_P	0
rank_CPL_SGT	0
rank_PFC	0
race_OTHER	0
race_UNKNOWN	0
race_WHITE	0
arrival_month_APR_JUN_DEC	0
arrival_month_JAN_FEB_MAR_AUG_NOV	0
arrival_month_JUL_SEP_OCT	0
arrival_month_UNKNOWN	0
tis_at_arrival	0
glasses_FALSE	0
glasses_TRUE	0
‘civilian_education_certification_ALTERNATIVE HIGH SCHOOL/EQUIVALENCY CERTIFICATE‘	0
‘civilian_education_certification_BACHELORS DEGREE/MASTERS DEGREE‘	0
‘civilian_education_certification_HIGH SCHOOL DIPLOMA‘	0
‘civilian_education_certification_SOME COLLEGE/ASSOCIATES DEGREE‘	0
civilian_education_certification_UNKNOWN	0

Table B.28 continued from previous page

Feature	Coefficient
‘age_at_arrival_<=20‘	0
‘age_at_arrival_>20‘	0
age_at_arrival_UNKNOWN	0
gt_score	0
apft_1_run	0
s2	0
s3	0
s5	0
s6	0
s7	0
s8	0
s9	0
s11	0
s12	0
s14	0
s15	0
rci	0
mis	0
vri_nr	0
fr	0
fpr	0
fb_sr	0
lr	0
kr	0
eid	0
thd	0
bxd	0
rc1	0
rc3	0
rc6	0
rc7	0
rc8	0
rc9	0
mls	0
hpc	0
nuc	0
gic	0
sui	0
hlp	0
sfd	0
cog	0
stw	0
axy	0
anp	0
brf	0
msf	0
jcp	0
agg	0
act	0
fml	0
ipp	0
sav	0
shy	0
mec	0
agg_rr	0
psy_cr	0
dis_cr	0
neg_er	0
int_rr	0
cannot_say	0
pct_true	0

B.3.2 LDA

TABLE B.29: Model Coefficients with all candidates for dataset_numeric using LDA

Feature	Coefficient
rci	-0.40045
full_scale_iq	-0.18929
verbal_iq	0.10102
performance_iq	0.08463
rc4	0.04826
eid	-0.04676
s10	-0.03658
s13	-0.03252
s15	0.03036
r_cd	0.03
s4	0.02957
psy_cr	0.02889
s9	0.02695
aes	0.02683
apft_1_score	-0.02607
anp	0.02462
hpc	-0.02361
s1	0.02283
rc2	0.02156
bxd	-0.02084
s2	-0.02012
s5	-0.01888
thd	-0.01758
s11	-0.01583
rc8	-0.01554
rc9	-0.01518
s6	-0.01499
agg_rr	0.01495
shy	0.01413
cog	0.01396
s12	0.01363
apft_1_pu	-0.01286
hlp	0.01186
sui	0.01133
dsf	-0.01078
stw	0.00976
rbs	0.0092
jcp	-0.00893
mis	-0.00888
dis_cr	0.00848
s14	0.00802
rc7	-0.00778
vri_nr	-0.00741
agg	-0.00738
act	0.00709
rc6	0.00696
kr	-0.00694
sav	0.0067
mec	-0.00668
fb_sr	0.006
tri_nr	-0.00589
fs	0.00586
fr	-0.00582
rc1	-0.00528
rc3	0.00525
s3	0.00524
apft_1_su	-0.00487
inf	-0.00487
axy	-0.00465
s7	-0.00452
sub	0.00444
msf	-0.00384
mls	0.00363
fpr	0.00344
pct_true	0.00324

Table B.29 continued from previous page

Feature	Coefficient
fml	-0.00242
nuc	-0.0019
neg_er	-0.00188
nfc	-0.00185
ipp	-0.00172
brf	0.00155
apft_1_run	0.00139
int_rr	0.00123
gt_score	-0.0012
s8	0
lr	0
sfd	0
tis_at_arrival	0
gic	0
cannot_say	0

TABLE B.30: Model Coefficients with candidates who scored above the minimum fitness test score for dataset_numeric using LDA

Feature	Coefficient
rci	-0.32582
s10	-0.09178
full_scale_iq	-0.08013
s4	0.0725
rc4	0.07101
s9	0.06379
s13	-0.05606
eid	-0.05096
s1	0.04434
s5	-0.0413
verbal_iq	0.03961
dis_cr	-0.03775
s11	-0.03347
apft_1_score	-0.0307
agg	-0.03038
rc2	0.02955
pct_true	0.02769
r_cd	0.02499
rc8	-0.02475
performance_iq	0.02197
cannot_say	0.02184
aes	0.02098
hpc	-0.01997
sfd	0.01986
s14	0.01941
apft_1_su	-0.01917
inf	-0.01842
sui	0.01838
s15	0.01814
dsf	-0.0172
tri_nr	0.01639
psy_cr	0.01627
bxd	0.01626
apft_1_pu	-0.01476
nfc	0.01461
msf	-0.01449
int_rr	0.01438
fml	-0.01355
shy	0.01298
rc7	0.01259
mis	0.01207
nuc	-0.01177
rc9	0.01132
thd	0.01117
rc3	0.01106
kr	0.01048

Table B.30 continued from previous page

Feature	Coefficient
agg_rr	0.0098
jcp	-0.00891
rbs	0.00799
fs	0.00795
fr	-0.00777
ipp	-0.00744
s8	-0.00743
lr	0.00724
anp	0.00677
sav	0.00675
gt_score	-0.00614
cog	0.00508
gic	0.0049
brf	0.00487
mec	0.00467
vri_nr	0.00419
s6	-0.00406
rc1	0.00404
s7	-0.0037
fpr	-0.00308
fb_sr	0.00301
neg_er	-0.00286
stw	0.00284
s2	-0.00278
mls	0.00274
rc6	0.00244
s3	-0.0021
sub	-0.0019
act	-0.00176
s12	-0.00107
apft_1_run	0
hlp	0
axy	0
tis_at_arrival	0

B.4 CART Trees

This section has all the CART trees for all 9 data subsets using all candidates and candidates who scored above the minimum fitness screening criteria.

B.4.1 All candidates

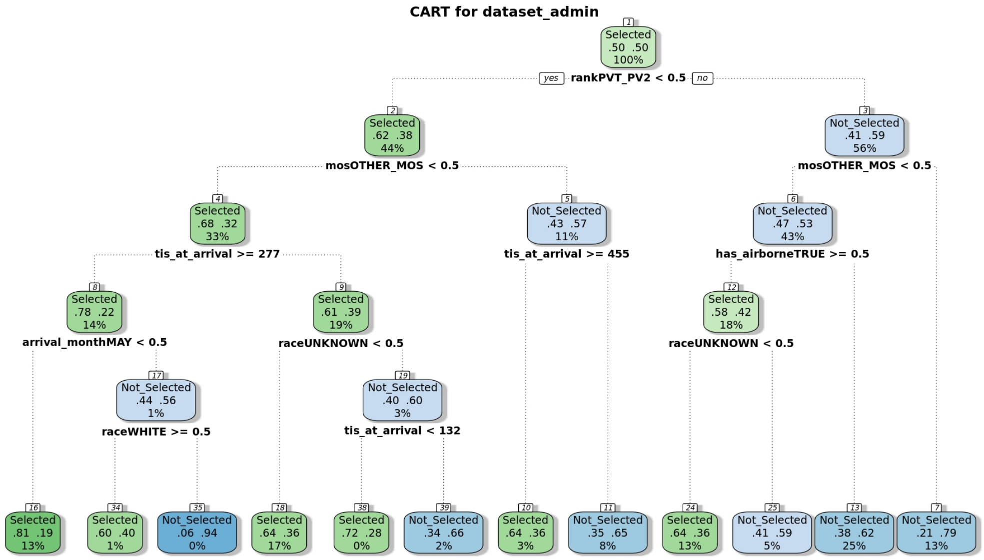


FIGURE B.17: CART Tree with all candidates for dataset_admin

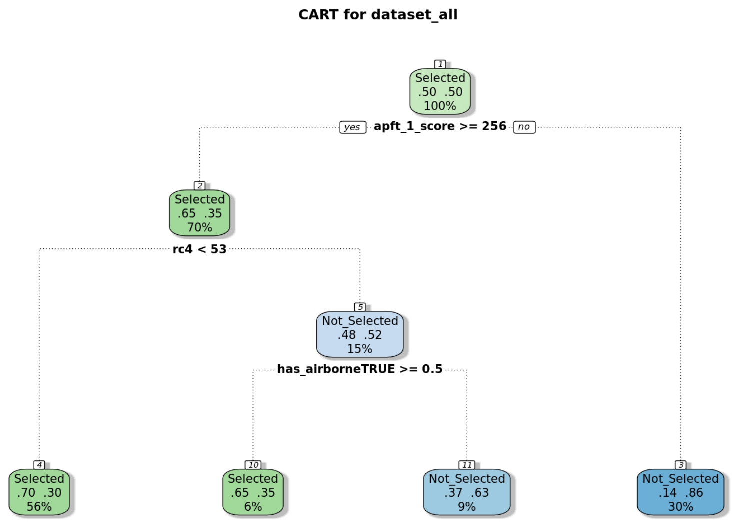


FIGURE B.18: CART Tree with all candidates for dataset_all

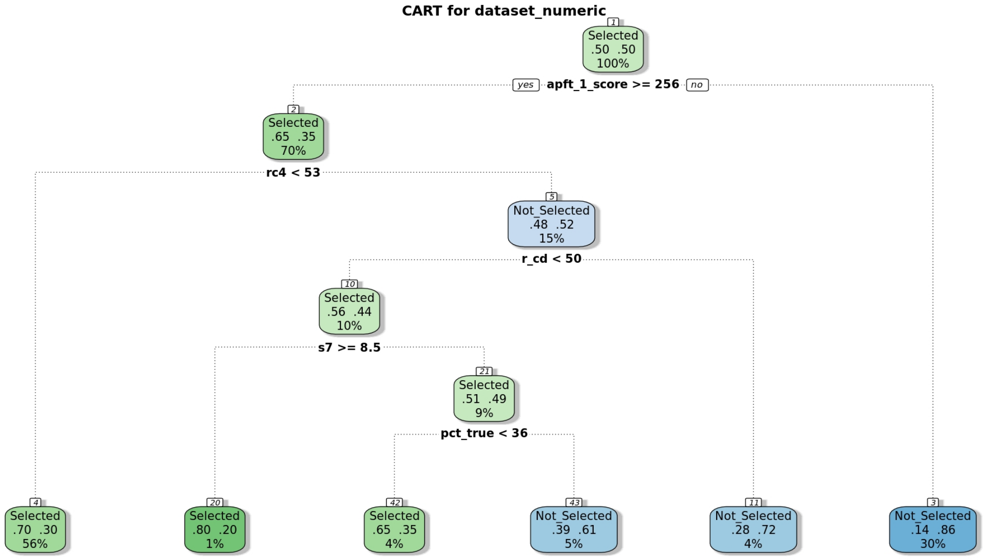


FIGURE B.19: CART Tree with all candidates for dataset_numeric

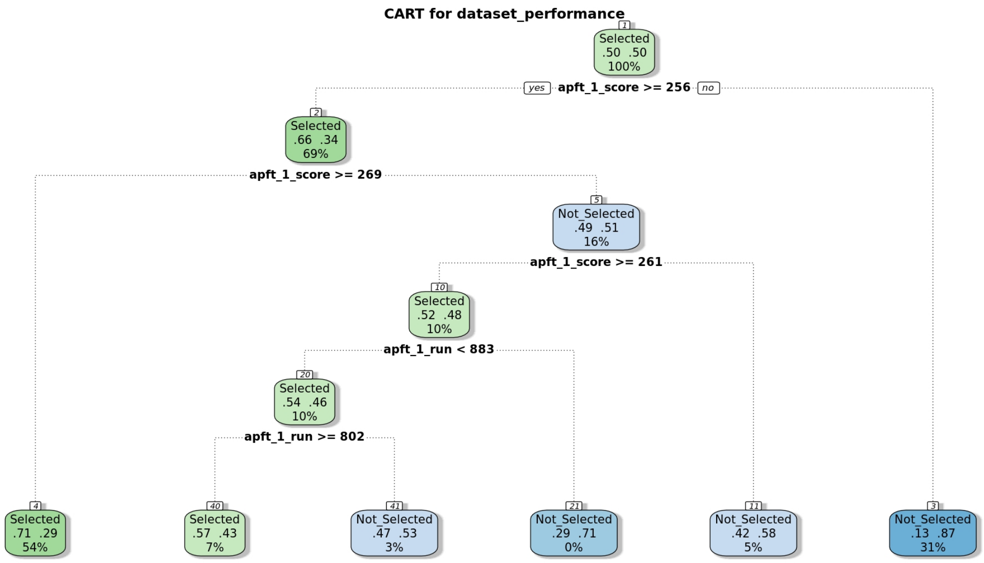


FIGURE B.20: CART Tree with all candidates for dataset_performance

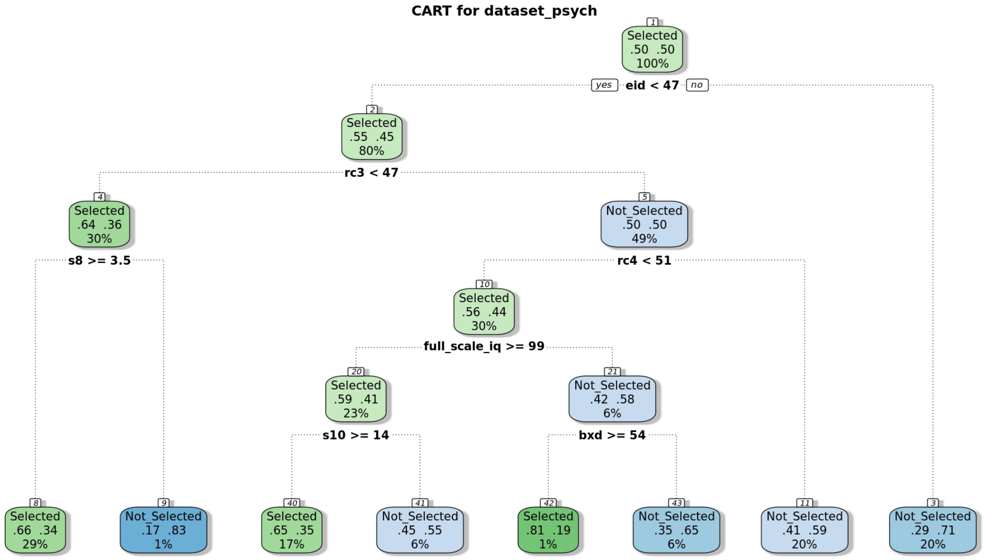


FIGURE B.21: CART Tree with all candidates for dataset_psych

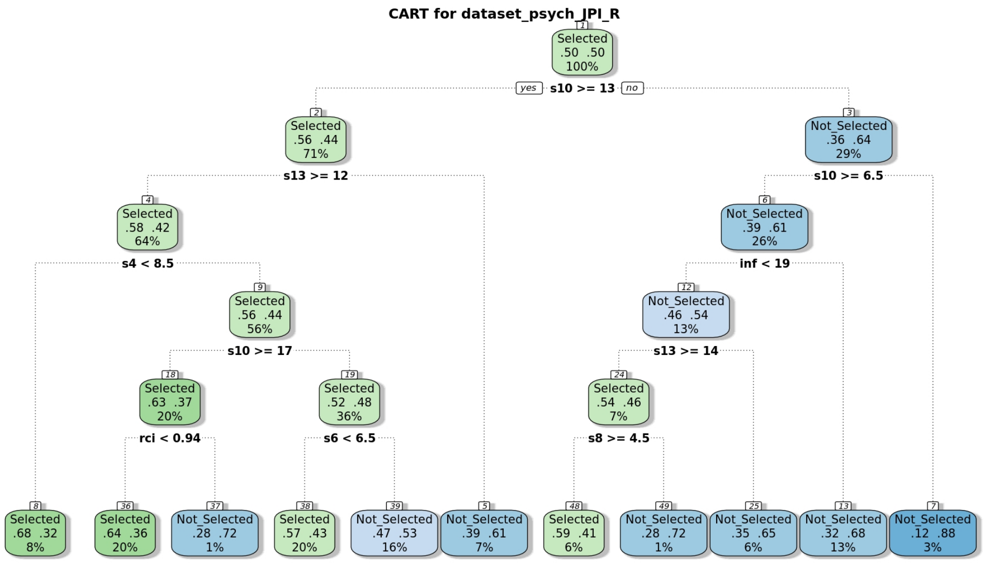


FIGURE B.22: CART Tree with all candidates for dataset_psych_JPI_R

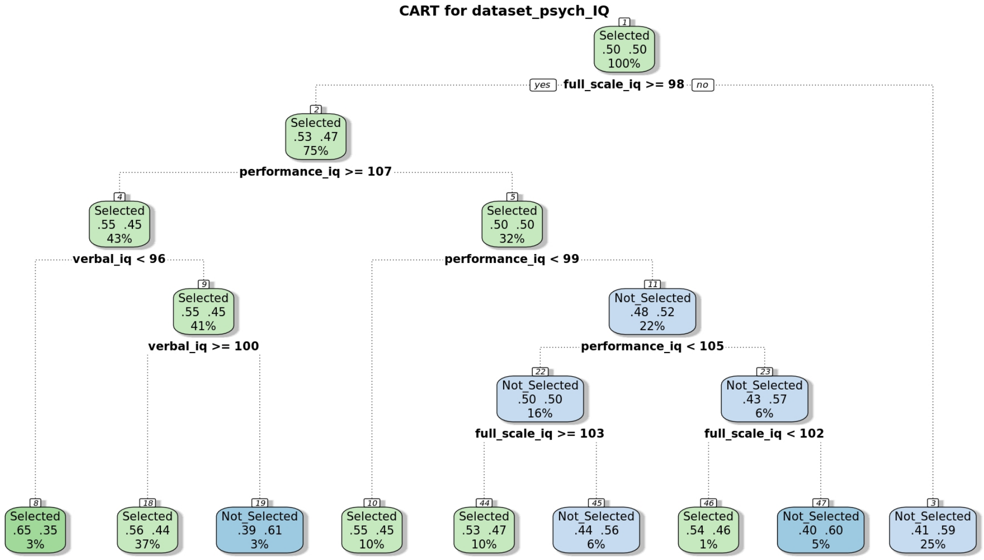


FIGURE B.23: CART Tree with all candidates for dataset_psych_IQ

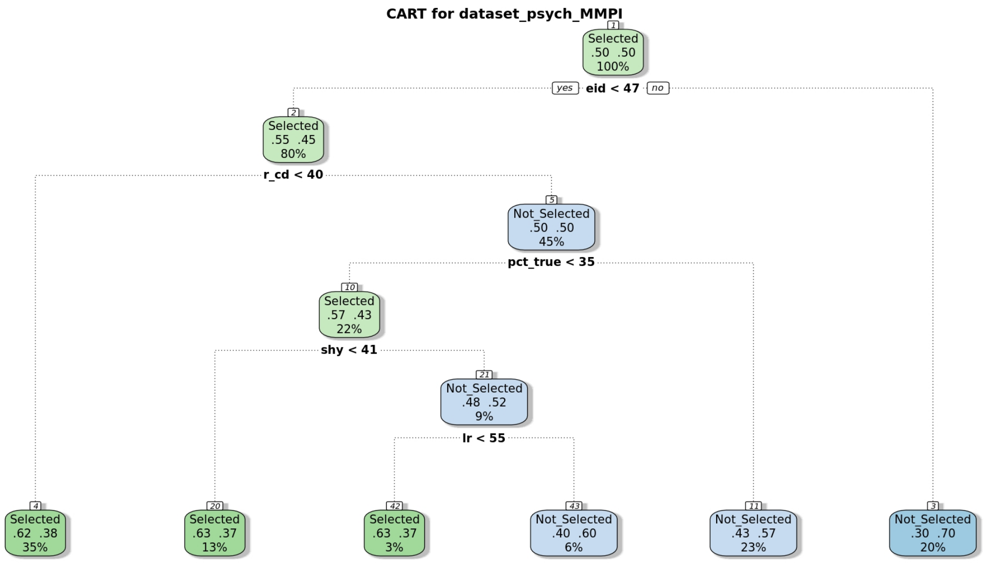


FIGURE B.24: CART Tree with all candidates for dataset_psych_MMPI

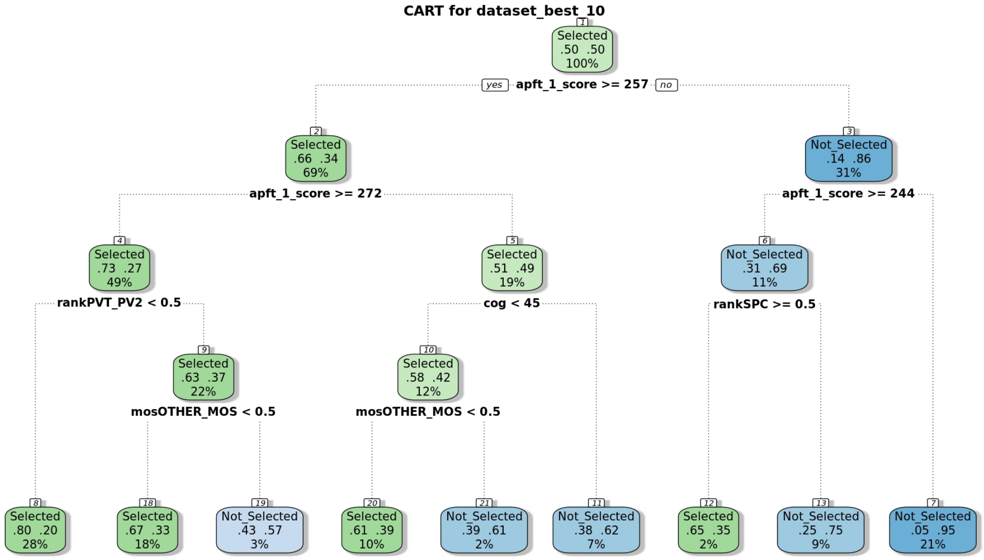


FIGURE B.25: CART Tree with all candidates for dataset_best_10

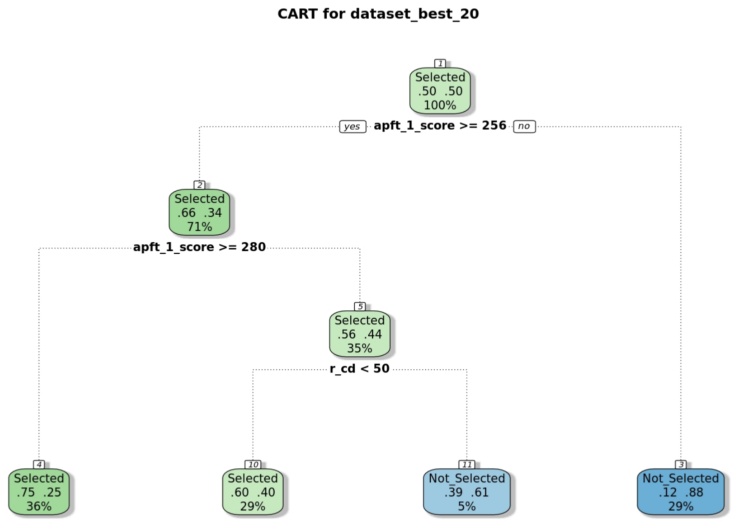


FIGURE B.26: CART Tree with all candidates for dataset_best_20

B.4.2 Candidates who met minimum fitness screening criteria

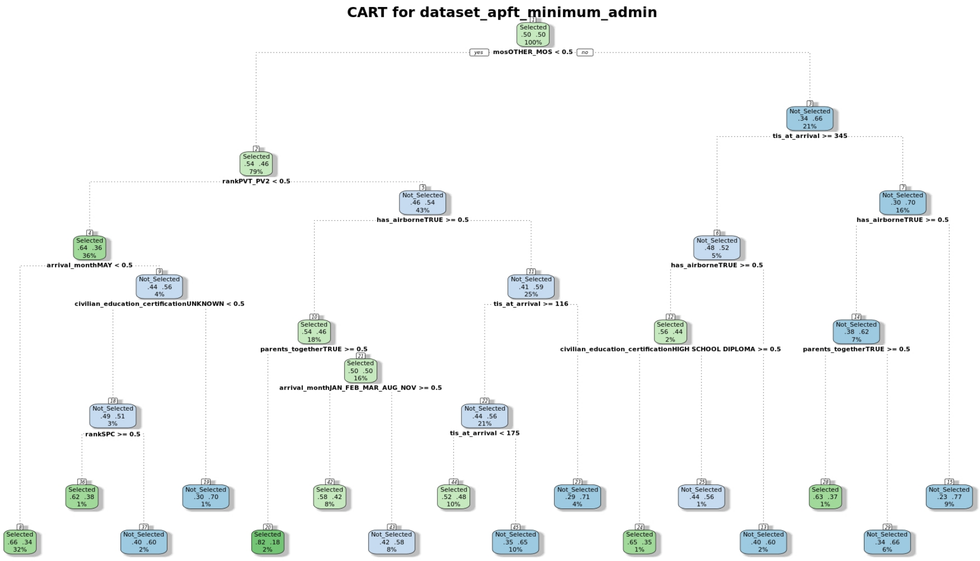


FIGURE B.27: CART Tree with candidates who scored above the minimum fitness test score for dataset_admin

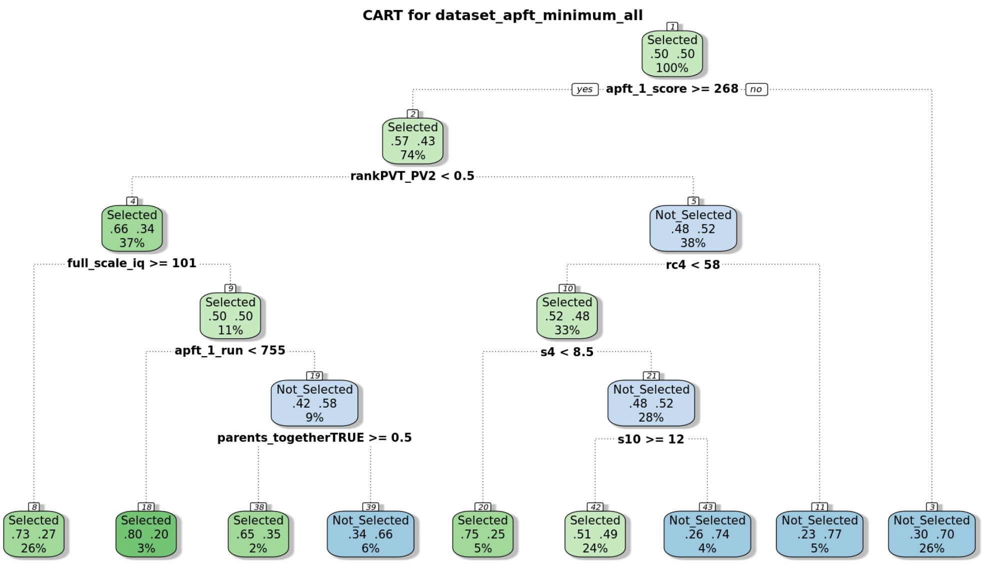


FIGURE B.28: CART Tree with candidates who scored above the minimum fitness test score for dataset_all

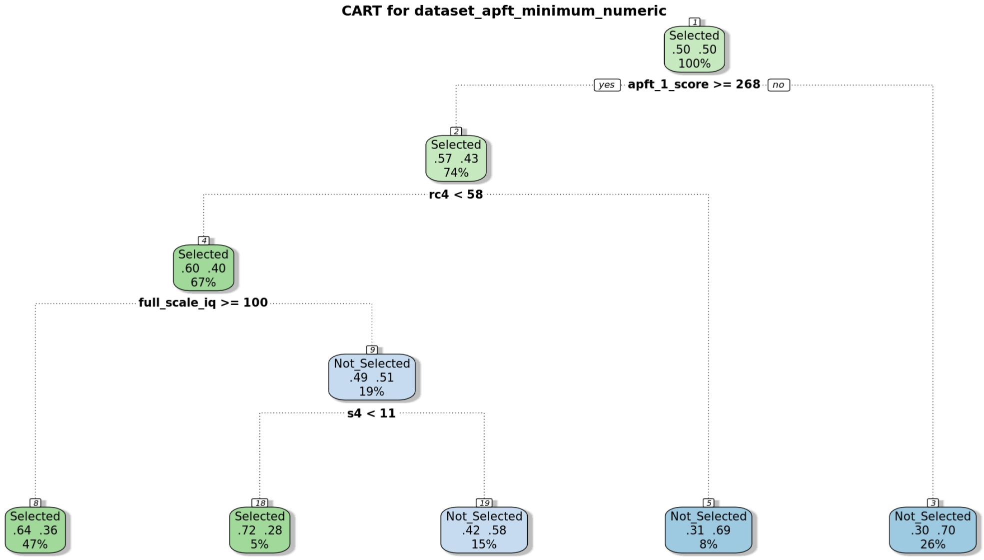


FIGURE B.29: CART Tree with candidates who scored above the minimum fitness test score for dataset_numeric

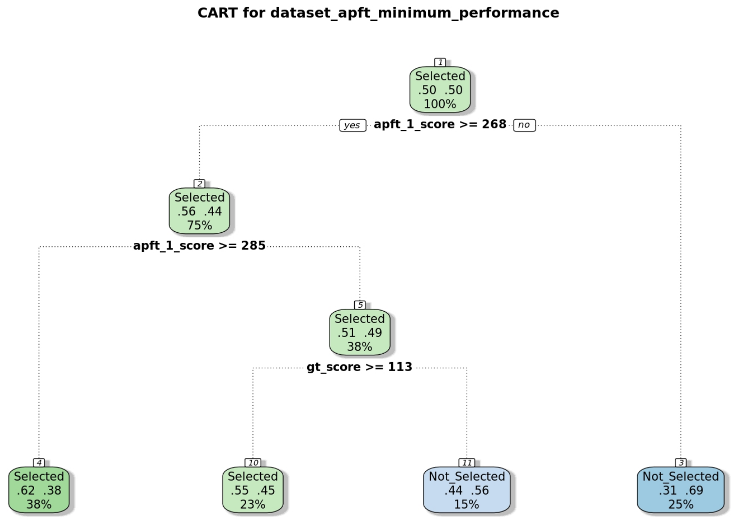


FIGURE B.30: CART Tree with candidates who scored above the minimum fitness test score for dataset_performance

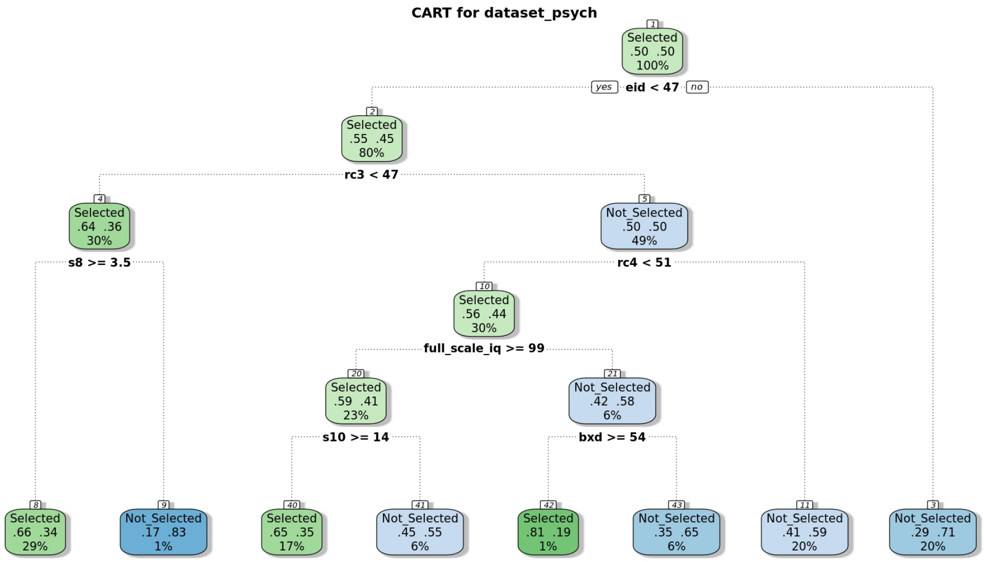


FIGURE B.31: CART Tree with candidates who scored above the minimum fitness test score for dataset_psych

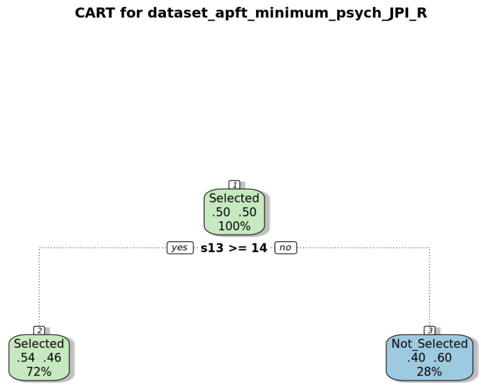


FIGURE B.32: CART Tree with candidates who scored above the minimum fitness test score for dataset_psych_JPI_R

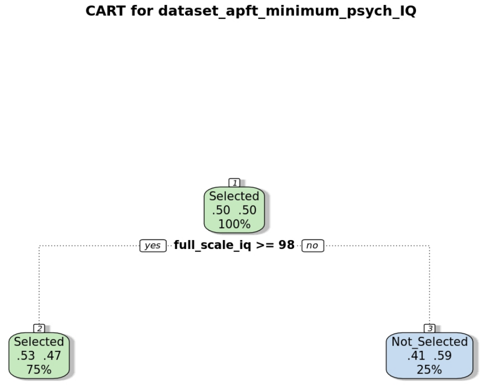


FIGURE B.33: CART Tree with candidates who scored above the minimum fitness test score for dataset_psych_IQ

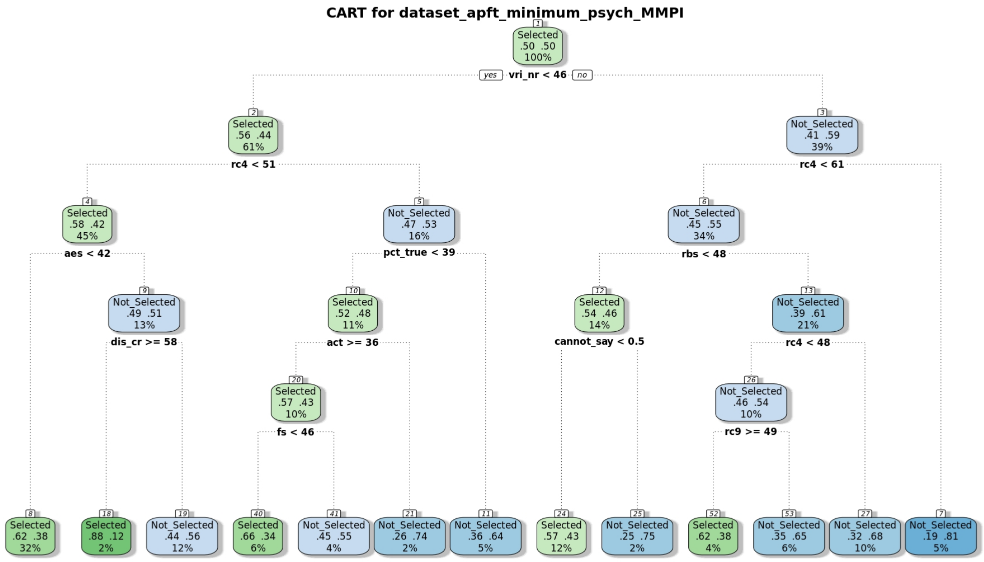


FIGURE B.34: CART Tree with candidates who scored above the minimum fitness test score for dataset_psych_MMPI

Appendix C

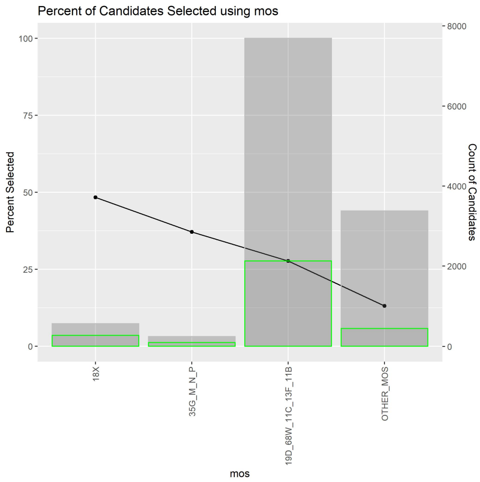
Tables and Figures

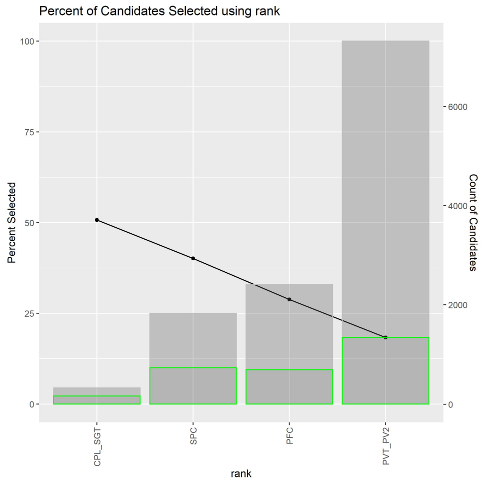
C.1 PCA Analysis Results

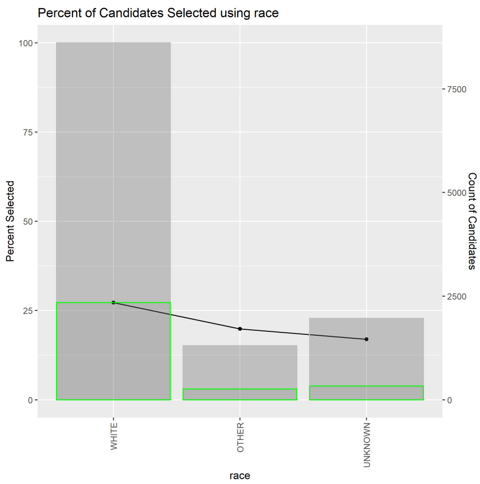
TABLE C.1: PCA Analysis Results Summary

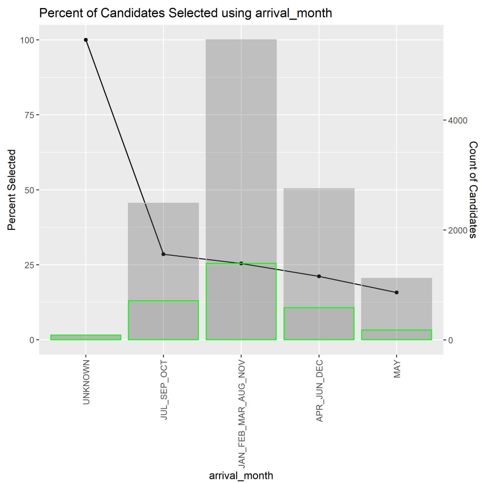
number	dataset	PCs	Cumulative Variance	accuracy	precision_PPV	NPV	sensitivity	specificity	kappa	f1
1	dataset_all_pca	6	0.47	0.72	0.58	0.87	0.83	0.65	0.44	0.68
2	dataset_numeric_pca	6	0.47	0.72	0.58	0.87	0.83	0.65	0.44	0.68
3	dataset_performance_pca	2	0.77	0.7	0.54	0.85	0.78	0.66	0.39	0.64
4	dataset_psych_pca	24	0.76	0.61	0.44	0.77	0.63	0.6	0.21	0.52
5	dataset_psych_IQ_pca	1	0.86	0.53	0.34	0.71	0.53	0.53	0.05	0.41
6	dataset_psych_JPI_R_pca	14	0.91	0.58	0.4	0.75	0.62	0.55	0.15	0.49
7	dataset_psych_MMPI_pca	7	0.6	0.62	0.44	0.77	0.61	0.62	0.21	0.51
8	dataset_imputed_performance_pca	5	1	0.78	0.54	0.87	0.63	0.83	0.43	0.58
9	dataset_imputed_psych_pca	19	0.72	0.63	0.36	0.83	0.6	0.65	0.2	0.45
10	dataset_imputed_psych_JPI_R_pca	17	0.98	0.7	0.39	0.8	0.38	0.81	0.19	0.39
11	dataset_imputed_psych_IQ_pca	2	0.99	0.62	0.29	0.77	0.39	0.69	0.07	0.33
12	dataset_imputed_psych_MMPI_pca	12	0.72	0.65	0.36	0.81	0.51	0.7	0.19	0.42
13	dataset_log_performance_pca	1	0.56	0.7	0.54	0.85	0.77	0.66	0.39	0.64
14	dataset_log_psych_pca	22	0.73	0.6	0.43	0.75	0.62	0.59	0.18	0.51
15	dataset_log_psych_JPI_R_pca	11	0.82	0.59	0.41	0.76	0.61	0.58	0.17	0.49
16	dataset_log_psych_IQ_pca	1	0.86	0.52	0.34	0.72	0.56	0.5	0.05	0.42
17	dataset_log_psych_MMPI_pca	10	0.65	0.62	0.44	0.77	0.63	0.61	0.22	0.52
18	dataset_binned_performance_pca	4	0.98	0.7	0.55	0.83	0.74	0.68	0.39	0.63
19	dataset_binned_psych_pca	18	0.7	0.65	0.48	0.8	0.68	0.63	0.28	0.56
20	dataset_binned_psych_JPI_R_pca	6	0.6	0.64	0.46	0.77	0.59	0.67	0.24	0.52
21	dataset_binned_psych_IQ_pca	1	0.81	0.52	0.33	0.71	0.54	0.51	0.04	0.41
22	dataset_binned_psych_MMPI_pca	2	0.43	0.6	0.44	0.83	0.79	0.52	0.25	0.57

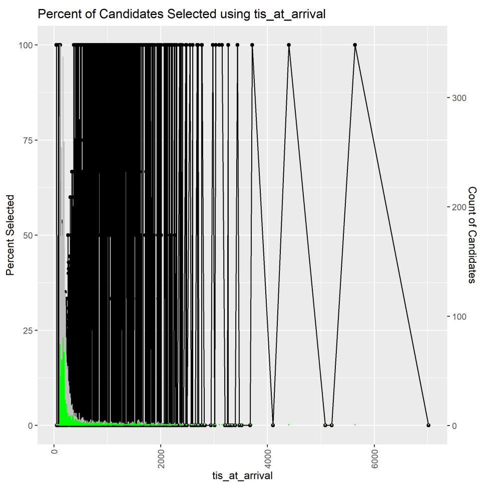
C.2 Contingency Table Plots for All Features

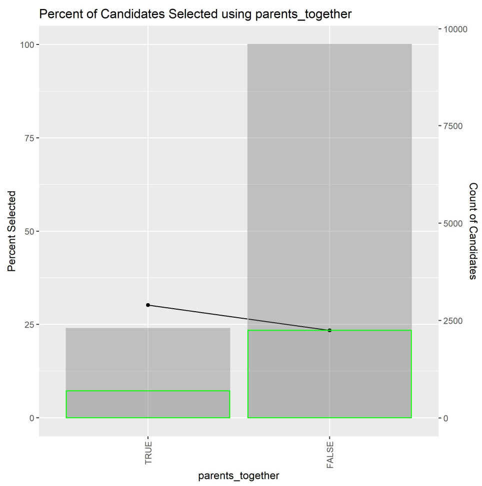


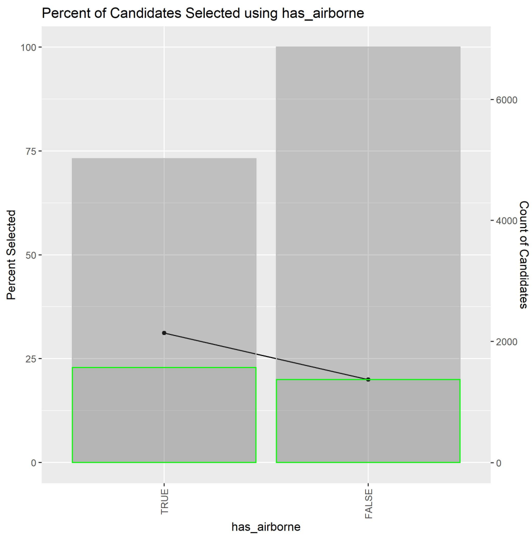


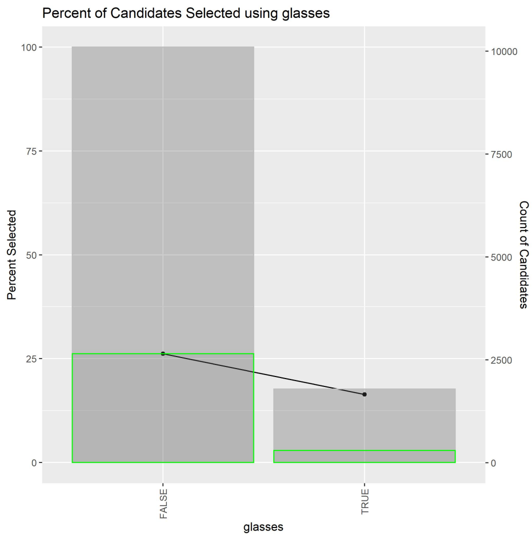


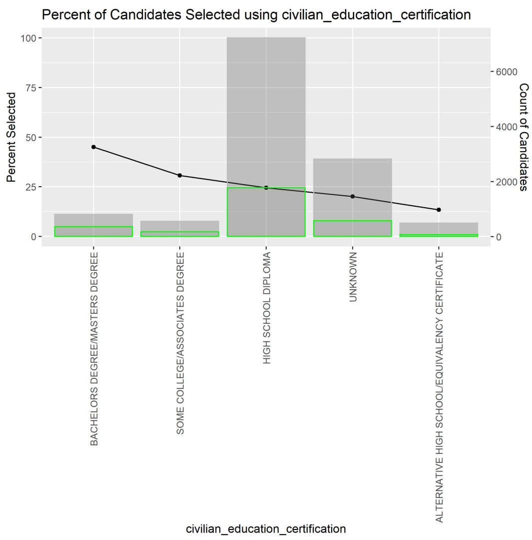


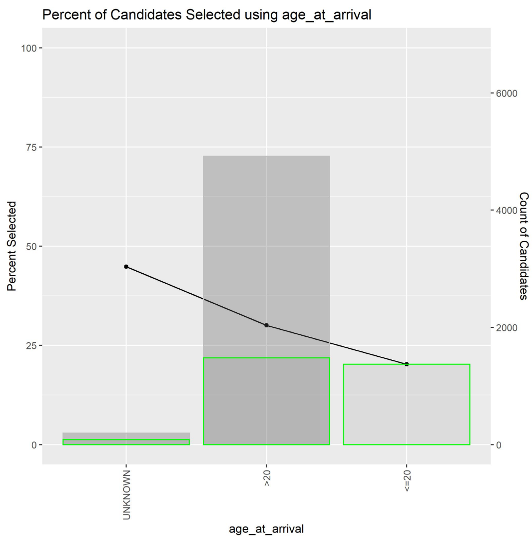


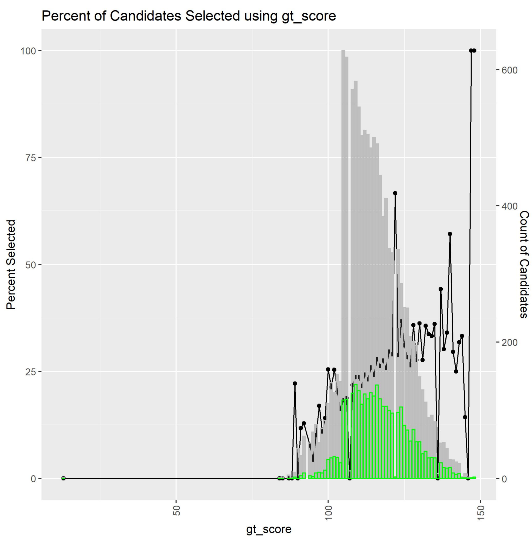


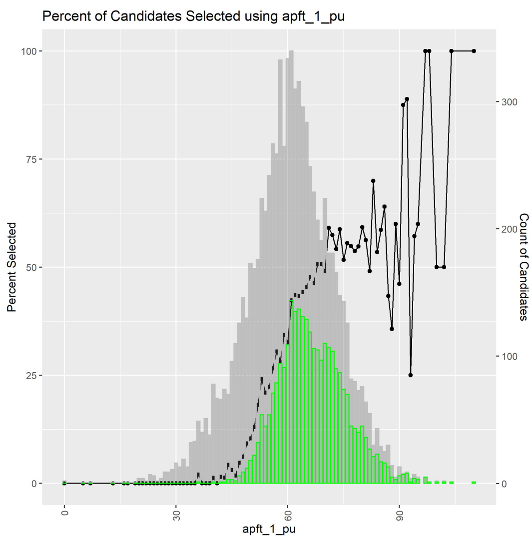

Percent of Candidates Selected using has_airborne
Percent Selected
Count of Candidates
TRUE
FALSE
has_airborne

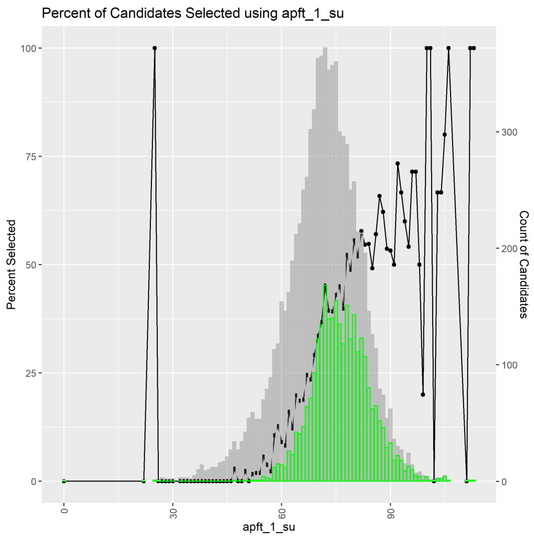

Percent of Candidates Selected using glasses
Percent Selected
Count of Candidates
FALSE
TRUE
glasses

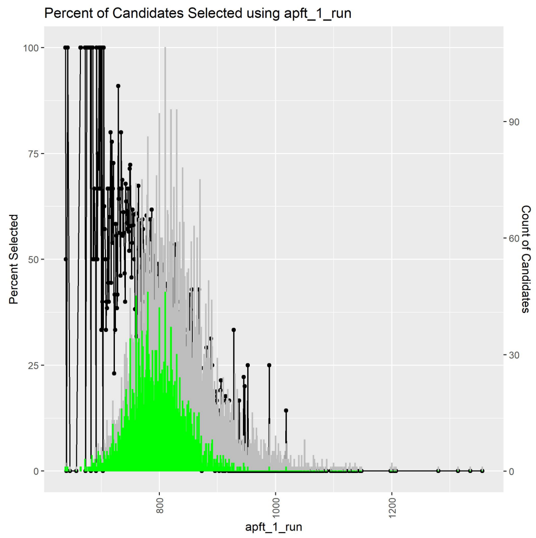

Percent of Candidates Selected using civilian_education_certification
Percent Selected
Count of Candidates
BACHELORS DEGREE/MASTERS DEGREE
SOME COLLEGE/ASSOCIATES DEGREE
HIGH SCHOOL DIPLOMA
UNKNOWN
ALTERNATIVE HIGH SCHOOL/EQUIVALENCY CERTIFICATE
civilian_education_certification

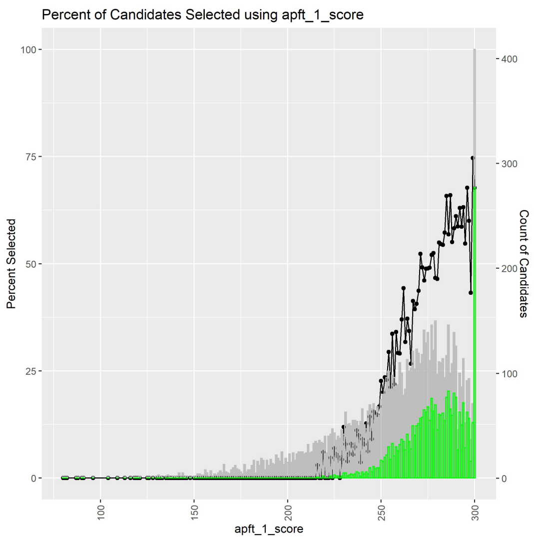

Percent of Candidates Selected using age_at_arrival
Percent Selected
Count of Candidates
UNKNOWN
>20
<=20
age_at_arrival

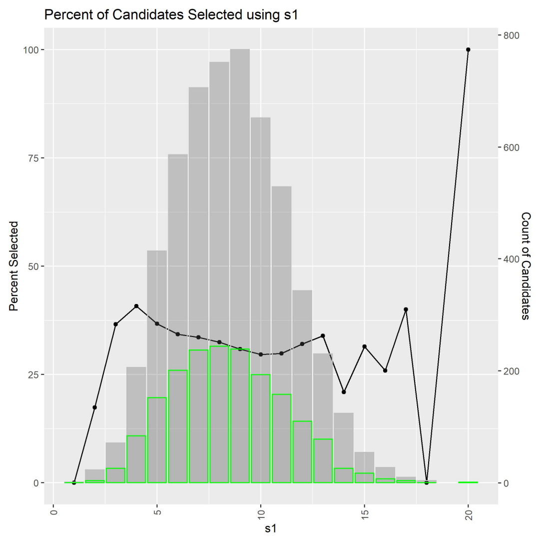

Percent of Candidates Selected using gt_score
Percent Selected
Count of Candidates
gt_score

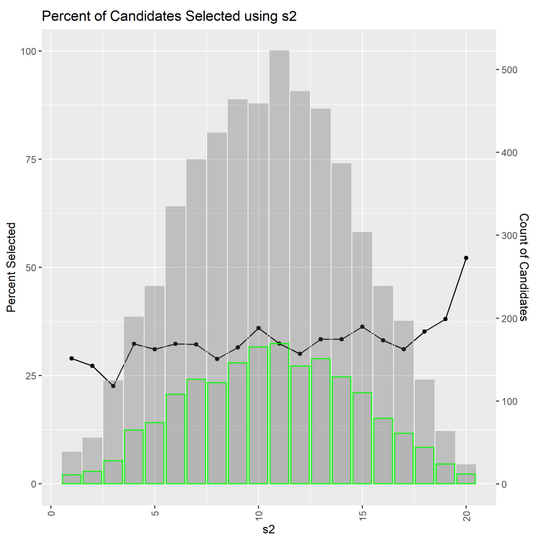

Percent of Candidates Selected using apft_1_pu
Percent Selected
Count of Candidates
apft_1_pu

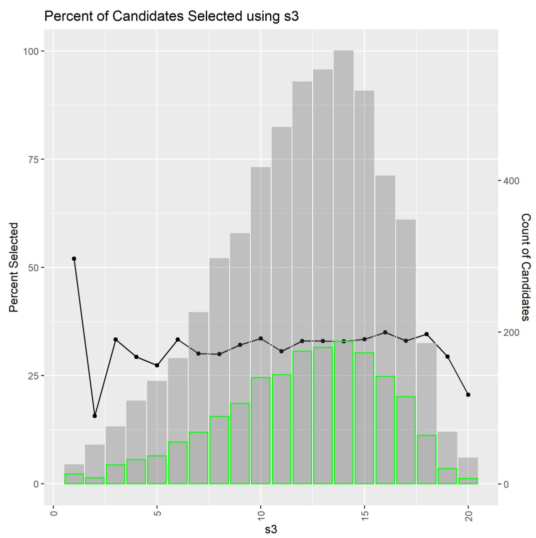

Percent of Candidates Selected using apft_1_su
Percent Selected
0
25
50
75
100
Count of Candidates
0
100
200
300
0
30
60
90
apft_1_su

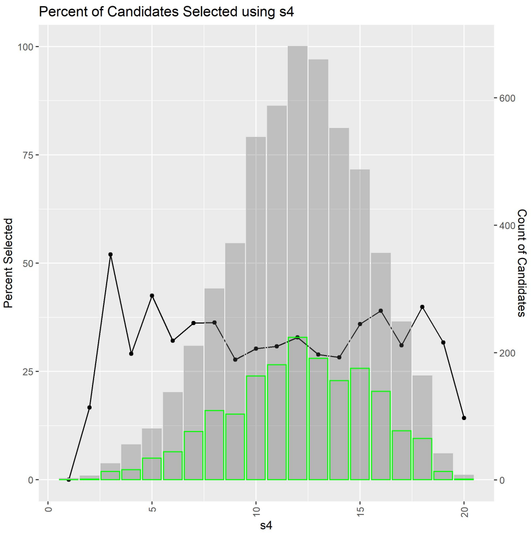

Percent of Candidates Selected using apft_1_run
Percent Selected
0
25
50
75
100
Count of Candidates
0
30
60
90
800
1000
1200
apft_1_run

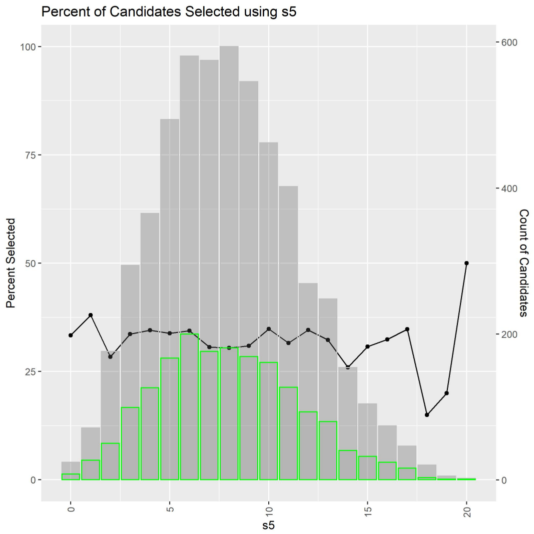

Percent of Candidates Selected using apft_1_score
Percent Selected
0
25
50
75
100
Count of Candidates
0
100
200
300
400
100
150
200
250
300
apft_1_score

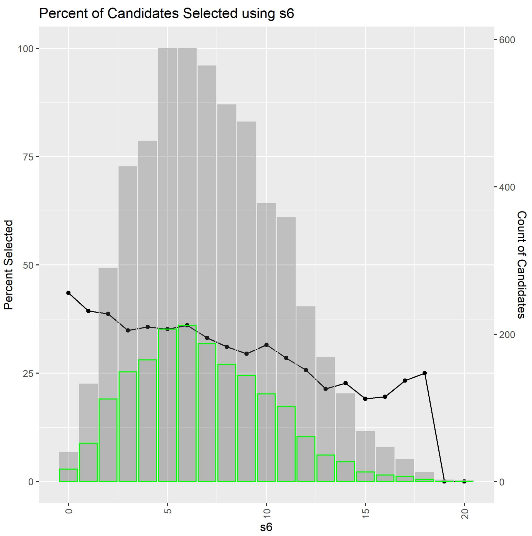

Percent of Candidates Selected using s1
Percent Selected
0
25
50
75
100
Count of Candidates
0
200
400
600
800
0
5
10
15
20
s1

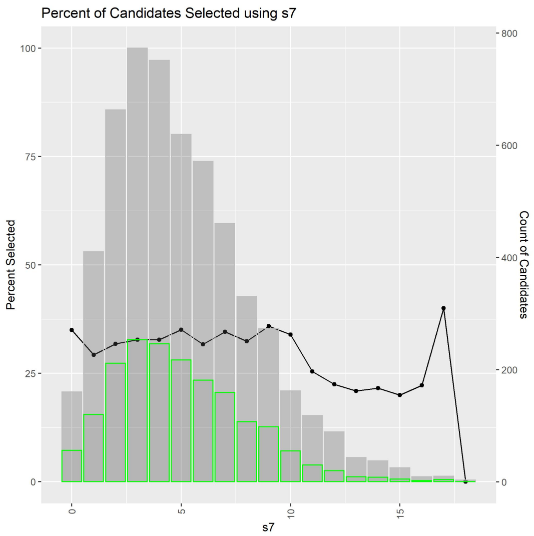

Percent of Candidates Selected using s2
Percent Selected
0
25
50
75
100
Count of Candidates
0
100
200
300
400
500
0
5
10
15
20
s2

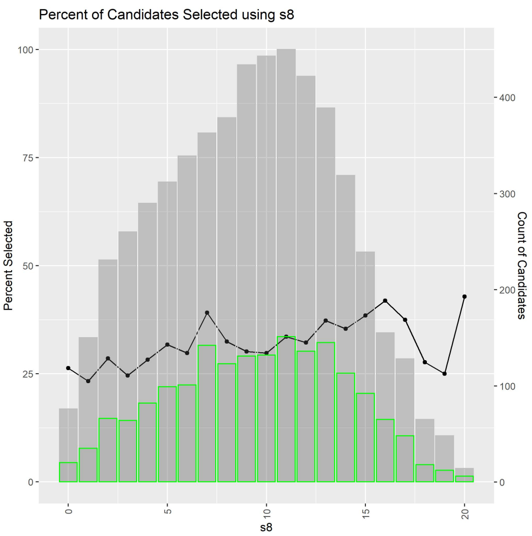

Percent of Candidates Selected using s3
Percent Selected
0
25
50
75
100
Count of Candidates
0
200
400
0
5
10
15
20
s3

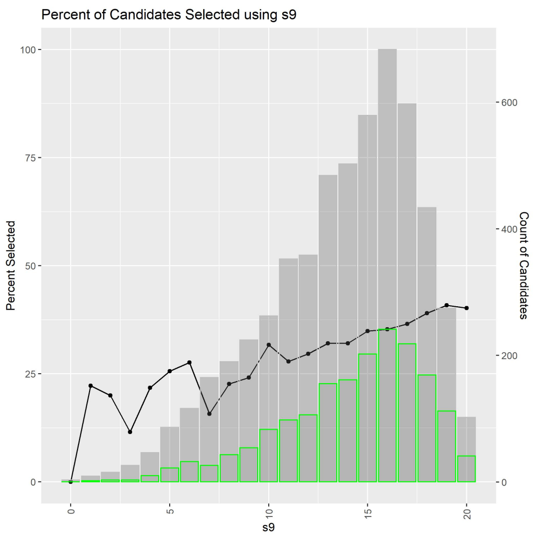

Percent of Candidates Selected using s4
Percent Selected
Count of Candidates
s4

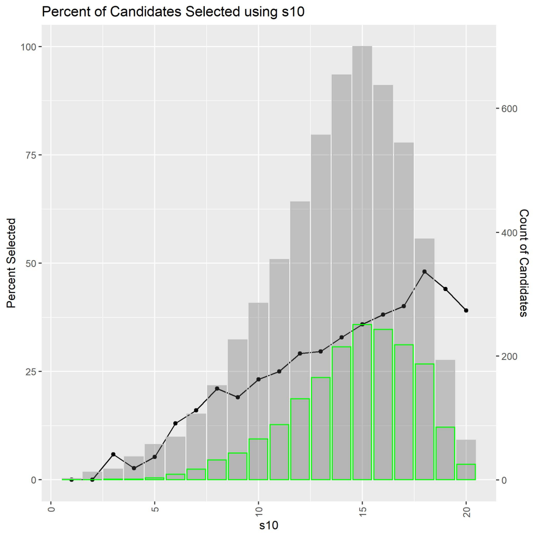

Percent of Candidates Selected using s5
Percent Selected
Count of Candidates
s5

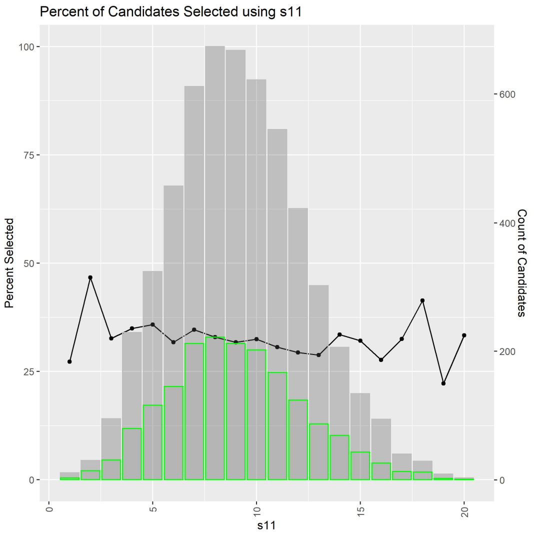

Percent of Candidates Selected using s6
Percent Selected
Count of Candidates
s6

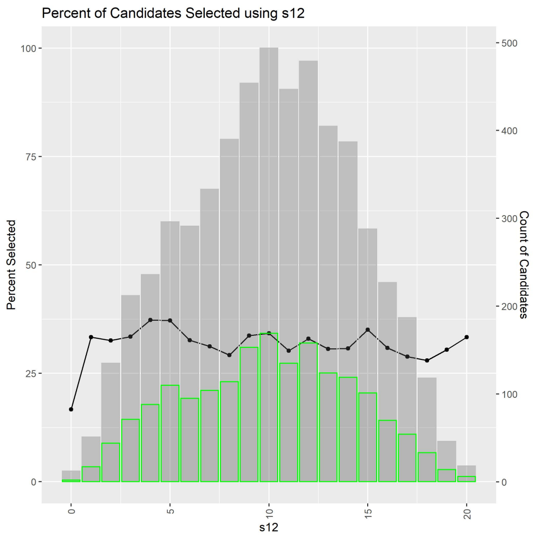

Percent of Candidates Selected using s7
Percent Selected
Count of Candidates
s7

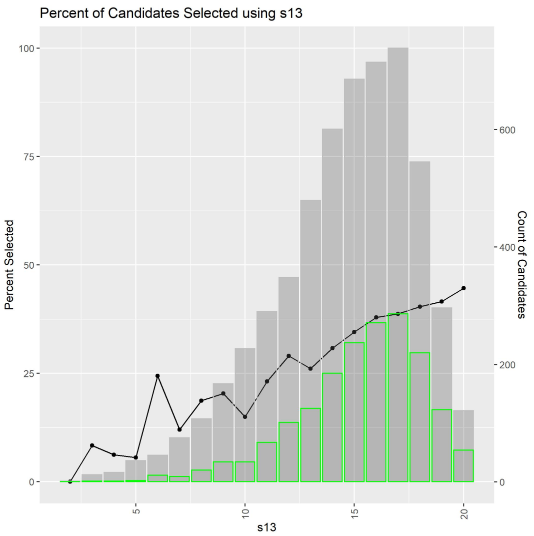

Percent of Candidates Selected using s8
Percent Selected
Count of Candidates
s8

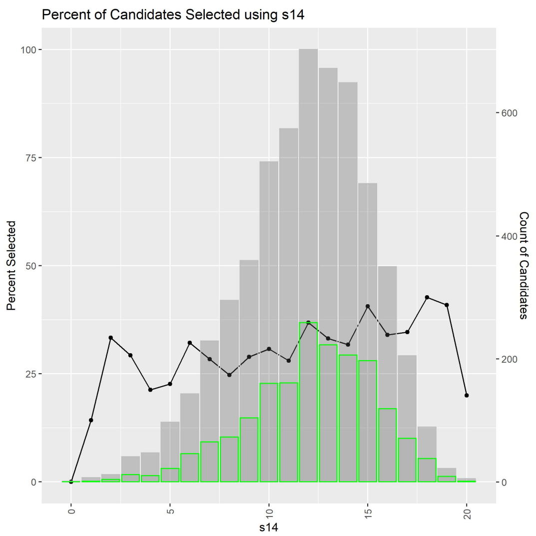

Percent of Candidates Selected using s9
Percent Selected
Count of Candidates
s9

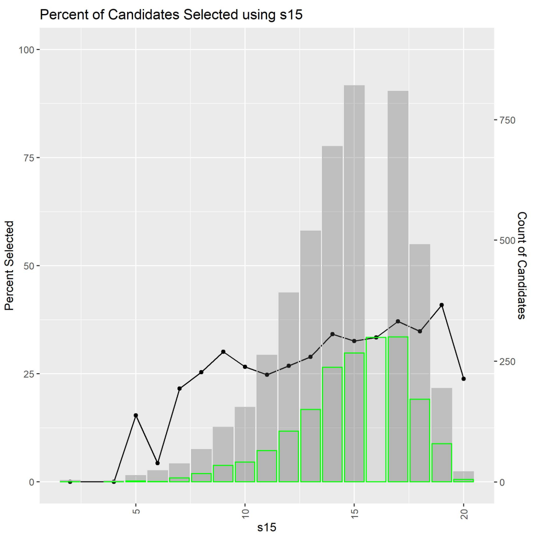

Percent of Candidates Selected using s10
Percent Selected
Count of Candidates
s10

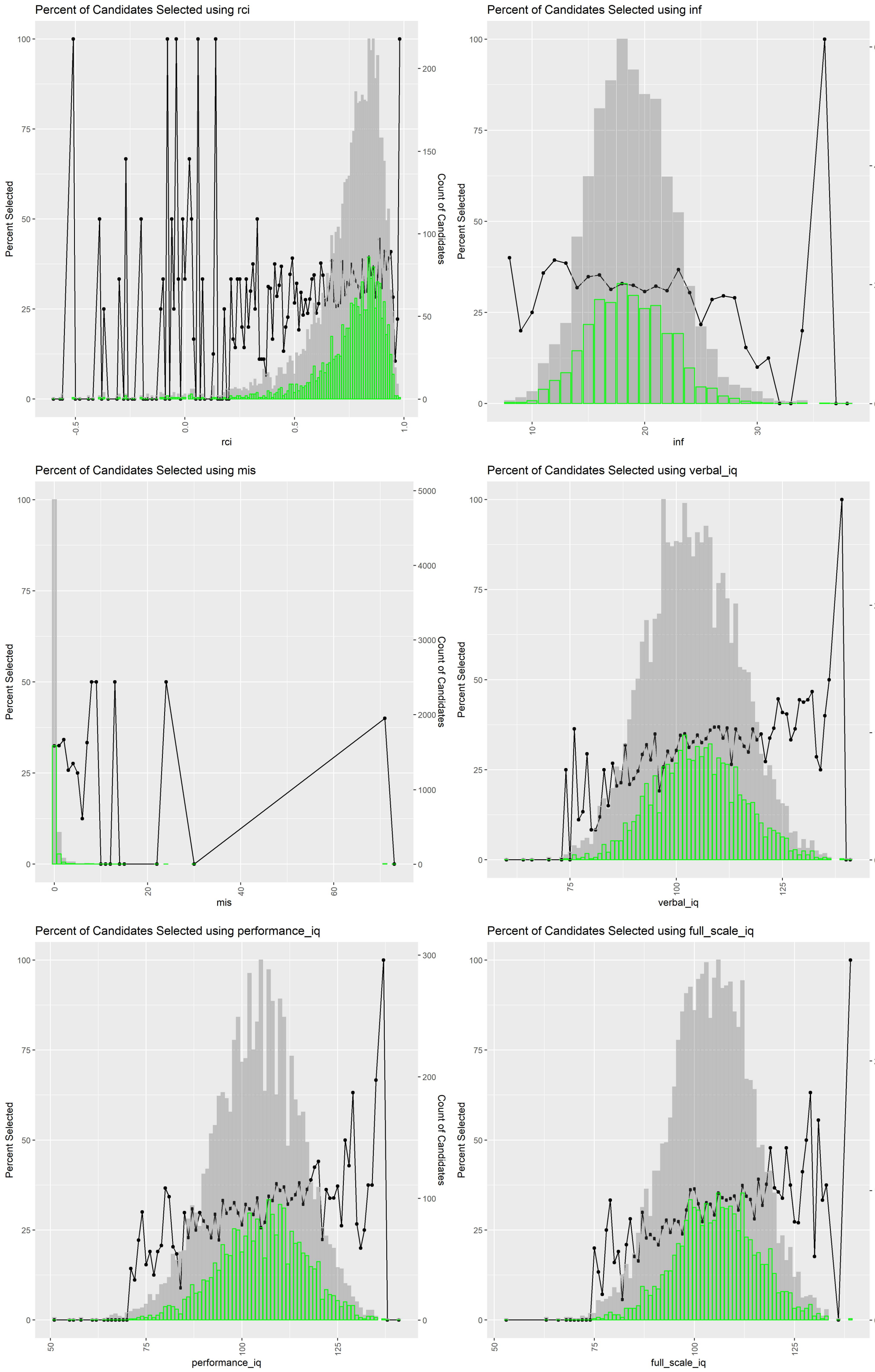

Percent of Candidates Selected using s11
Percent Selected
Count of Candidates
s11

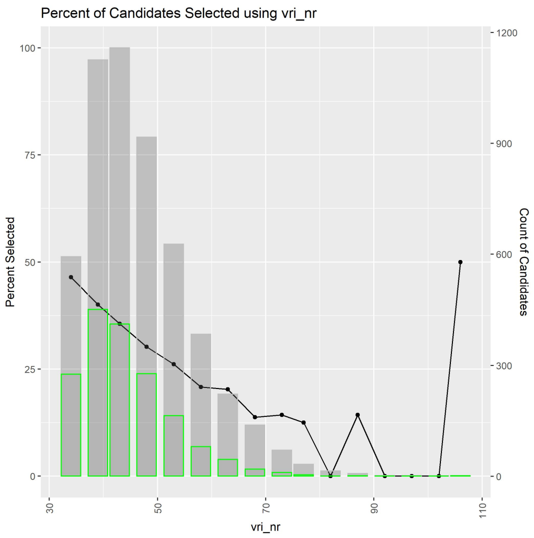

Percent of Candidates Selected using s12
Percent Selected
Count of Candidates
s12

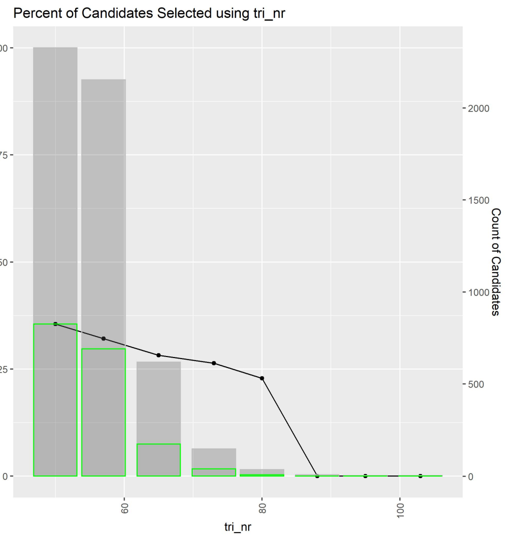

Percent of Candidates Selected using s13
Percent Selected
Count of Candidates
s13

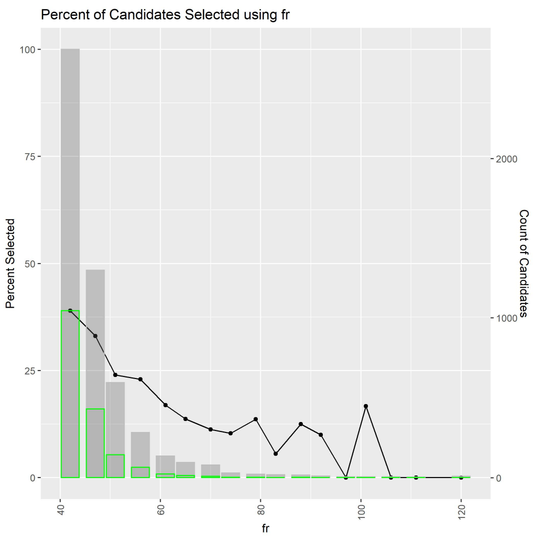

Percent of Candidates Selected using s14
Percent Selected
Count of Candidates
s14

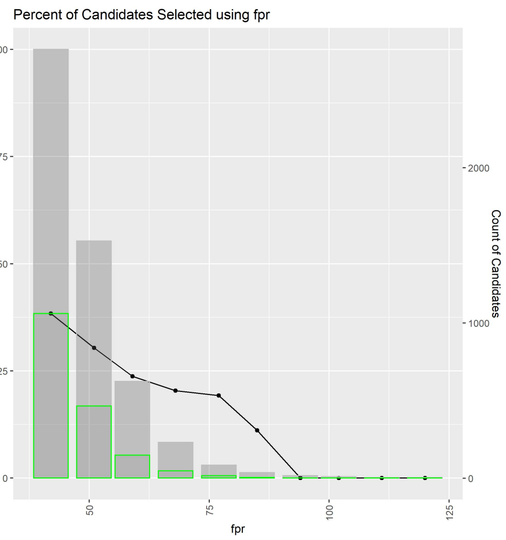

Percent of Candidates Selected using s15
Percent Selected
Count of Candidates
s15

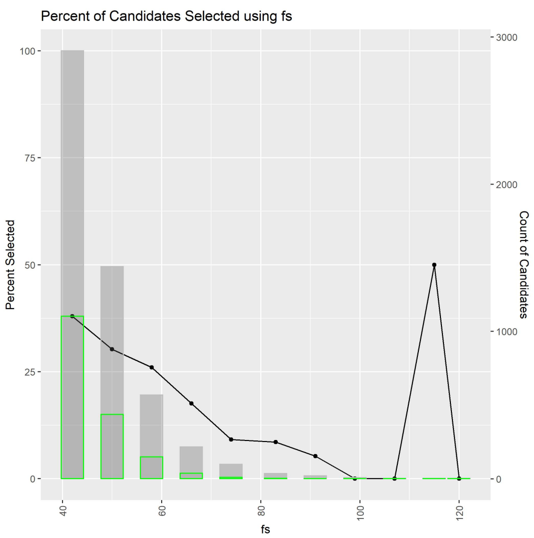

Percent of Candidates Selected using rci
Percent Selected
Count of Candidates
rci
Percent of Candidates Selected using inf
Percent Selected
Count of Candidates
inf
Percent of Candidates Selected using mis
Percent Selected
Count of Candidates
mis
Percent of Candidates Selected using verbal_iq
Percent Selected
Count of Candidates
verbal_iq
Percent of Candidates Selected using performance_iq
Percent Selected
Count of Candidates
performance_iq
Percent of Candidates Selected using full_scale_iq
Percent Selected
Count of Candidates
full_scale_iq

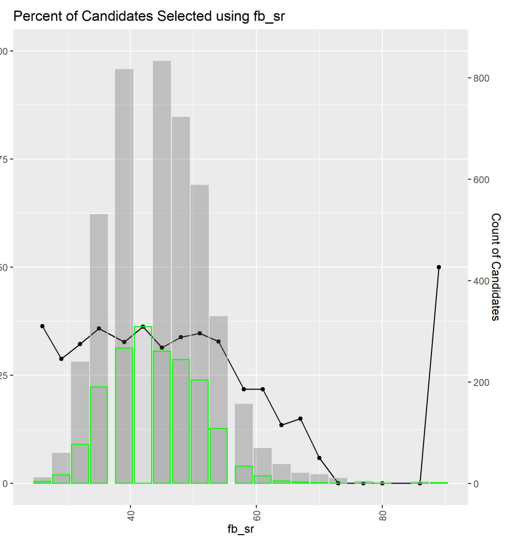

Percent of Candidates Selected using vri_nr
Percent Selected
Count of Candidates
vri_nr

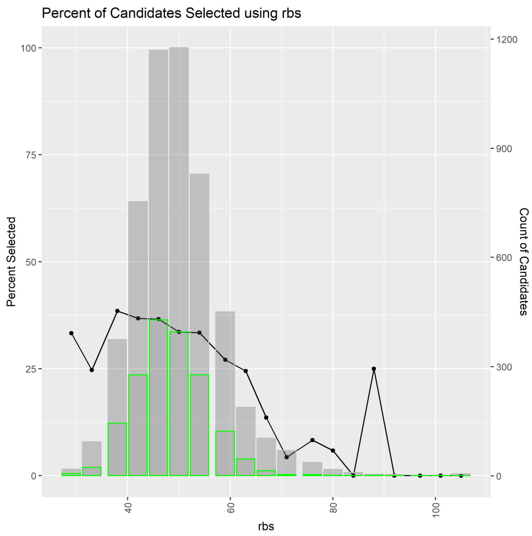

Percent of Candidates Selected using tri_nr
Percent Selected
Count of Candidates
tri_nr

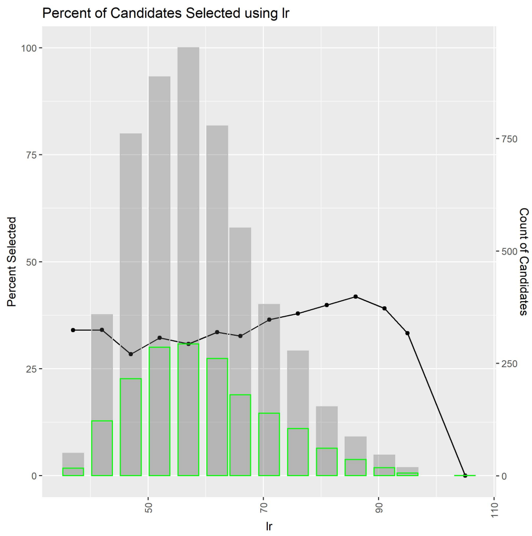

Percent of Candidates Selected using fr
Percent Selected
Count of Candidates
fr

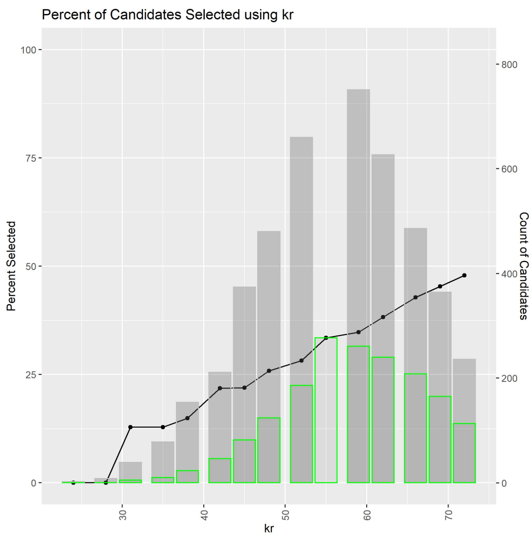

Percent of Candidates Selected using fpr
Percent Selected
Count of Candidates
fpr

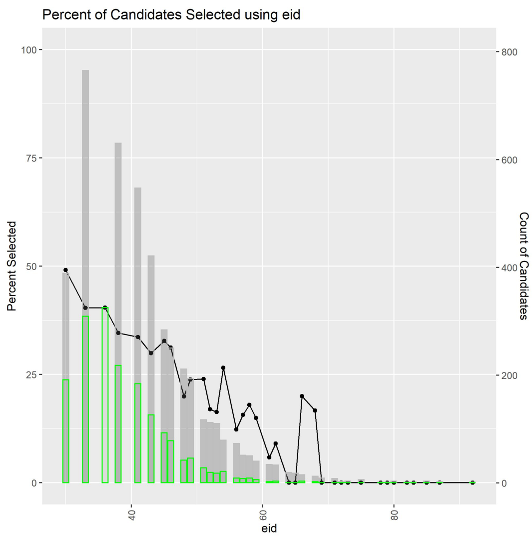

Percent of Candidates Selected using fs
Percent Selected
Count of Candidates
fs

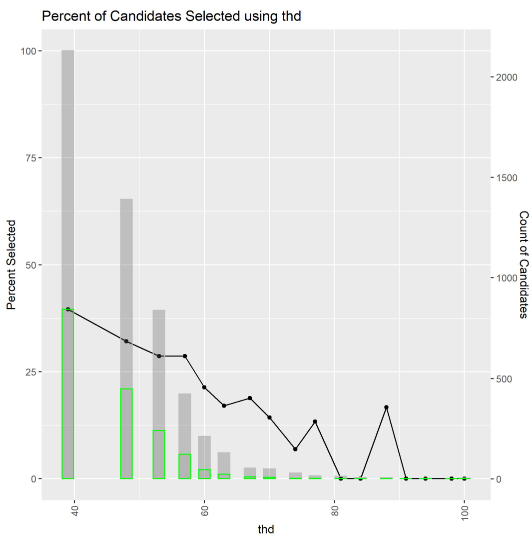

Percent of Candidates Selected using fb_sr
Percent Selected
Count of Candidates
fb_sr

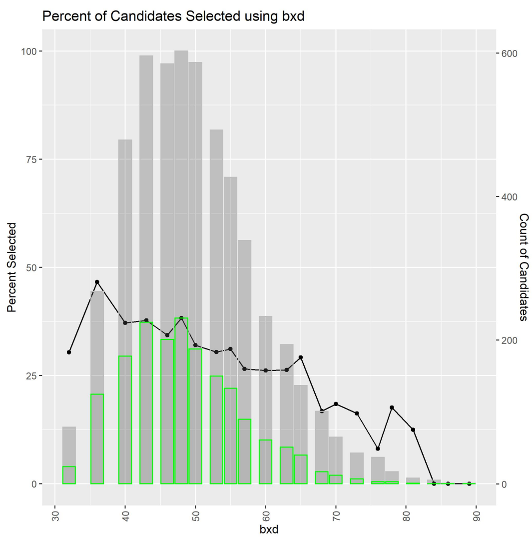

Percent of Candidates Selected using rbs
Percent Selected
Count of Candidates
rbs

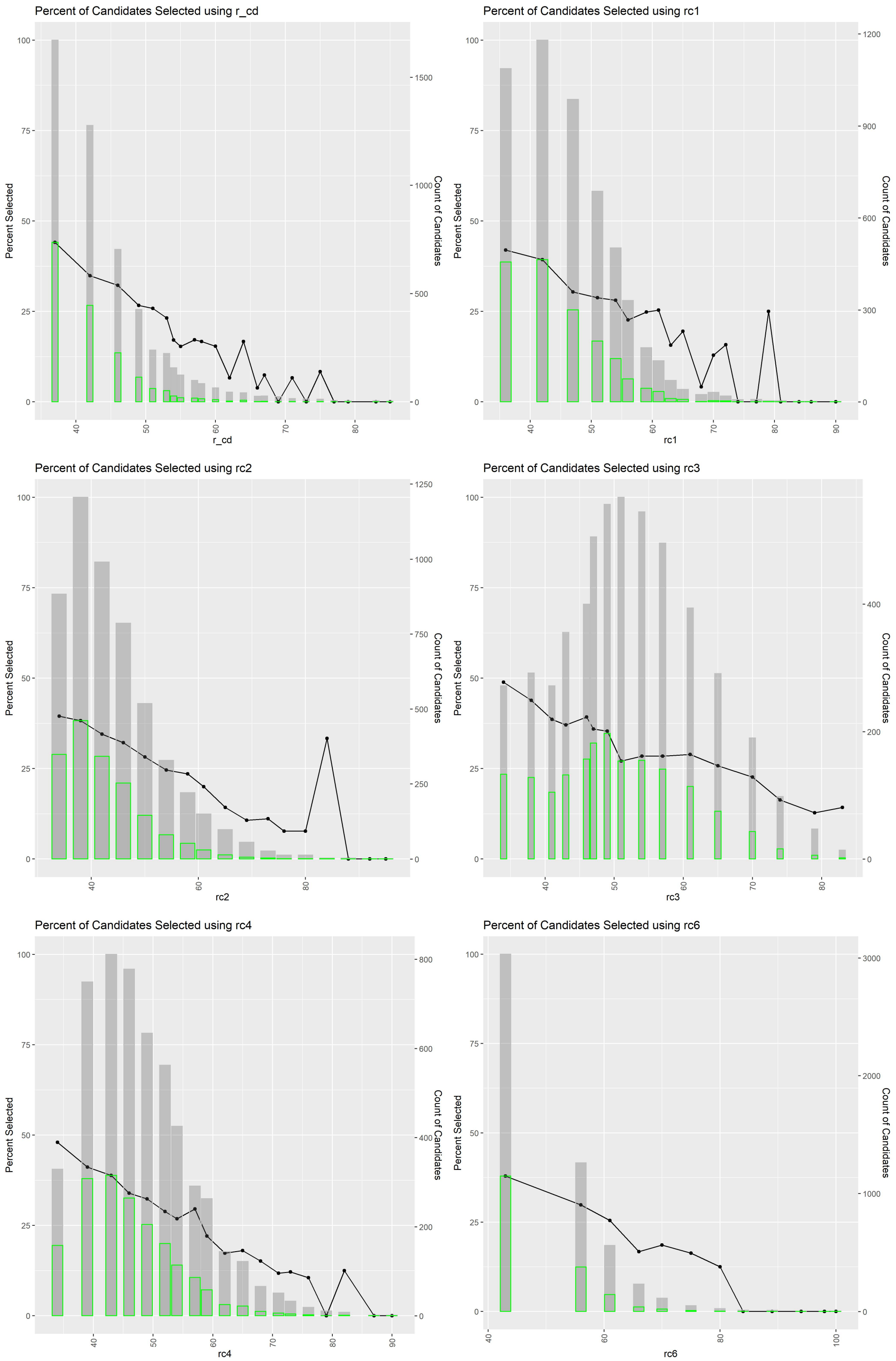

Percent of Candidates Selected using lr
Percent Selected
Count of Candidates
lr

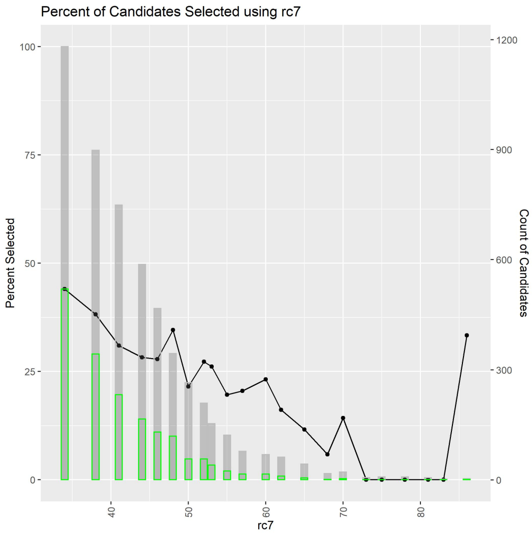

Percent of Candidates Selected using kr
Percent Selected
Count of Candidates
kr

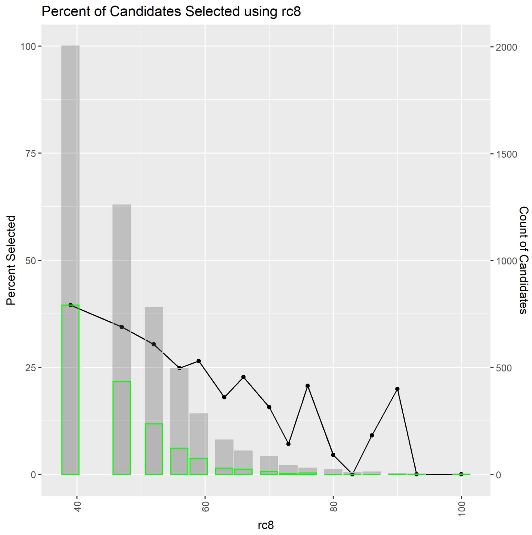

Percent of Candidates Selected using eid
Percent Selected
Count of Candidates
eid

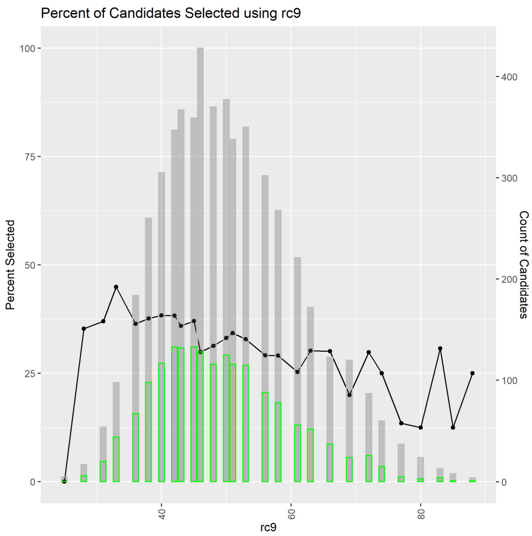

Percent of Candidates Selected using thd
Percent Selected
Count of Candidates
thd

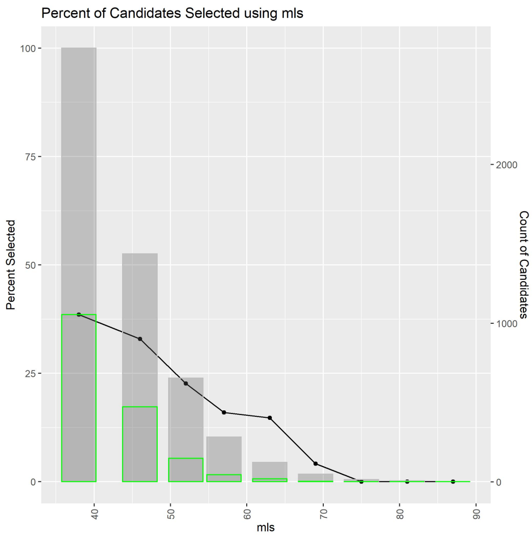

Percent of Candidates Selected using bxd
Percent Selected
Count of Candidates
bxd

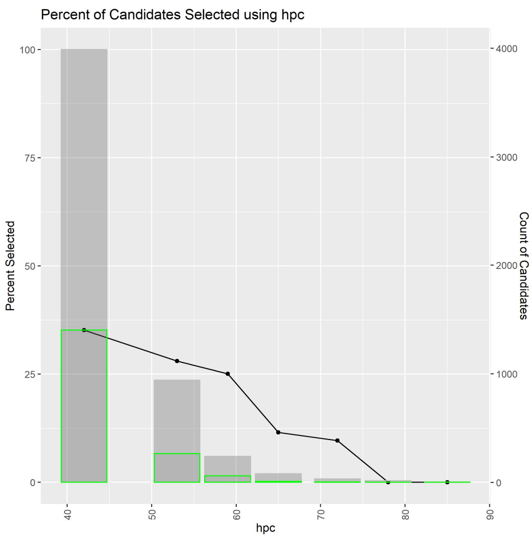

Percent of Candidates Selected using r_cd
Percent Selected
Count of Candidates
r_cd
Percent of Candidates Selected using rc1
Percent Selected
Count of Candidates
rc1
Percent of Candidates Selected using rc2
Percent Selected
Count of Candidates
rc2
Percent of Candidates Selected using rc3
Percent Selected
Count of Candidates
rc3
Percent of Candidates Selected using rc4
Percent Selected
Count of Candidates
rc4
Percent of Candidates Selected using rc6
Percent Selected
Count of Candidates
rc6

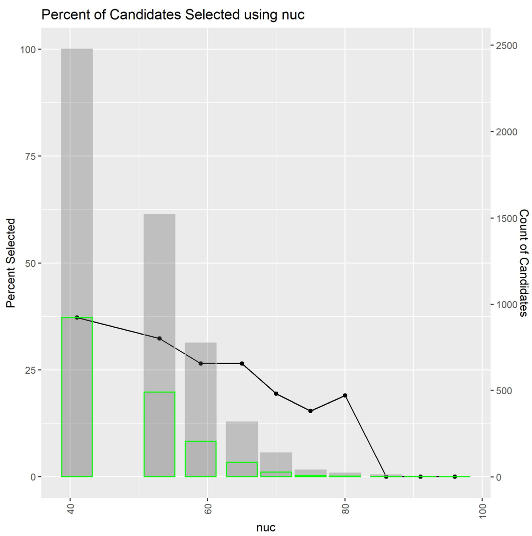

Percent of Candidates Selected using rc7
Percent Selected
Count of Candidates
rc7
0
25
50
75
100
300
600
900
1200
40
50
60
70
80

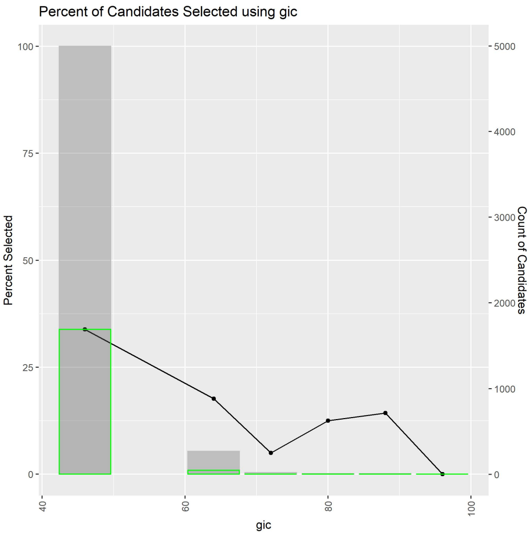

Percent of Candidates Selected using rc8
Percent Selected
Count of Candidates
rc8
0
25
50
75
100
500
1000
1500
2000
40
60
80
100

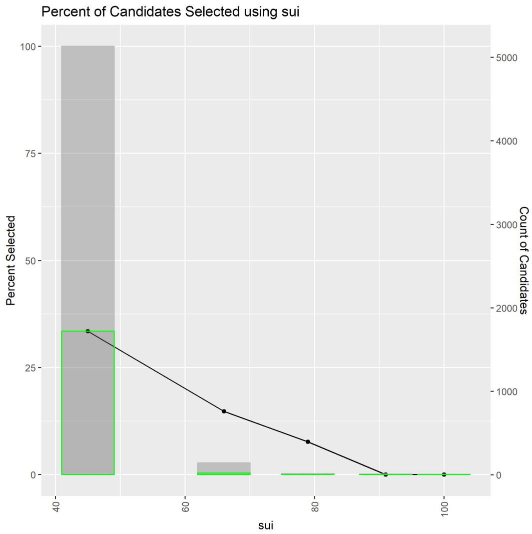

Percent of Candidates Selected using rc9
Percent Selected
Count of Candidates
rc9
0
25
50
75
100
200
300
400
40
60
80

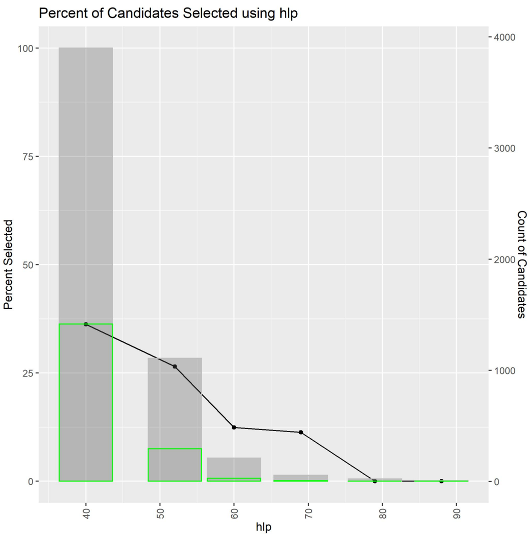

Percent of Candidates Selected using mls
Percent Selected
Count of Candidates
mls
0
25
50
75
100
1000
2000
40
50
60
70
80
90

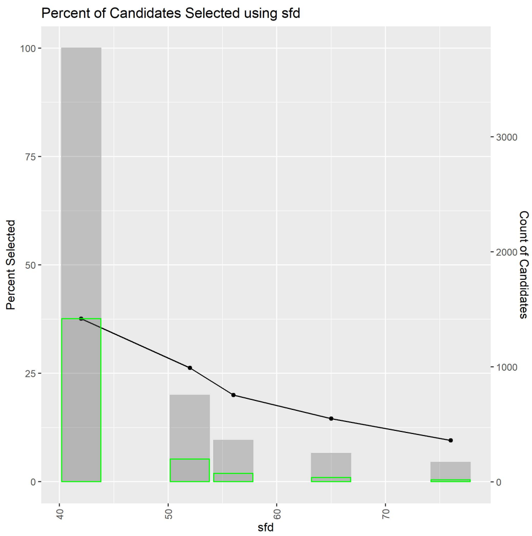

Percent of Candidates Selected using hpc
Percent Selected
Count of Candidates
hpc
0
25
50
75
100
1000
2000
3000
4000
40
50
60
70
80
90

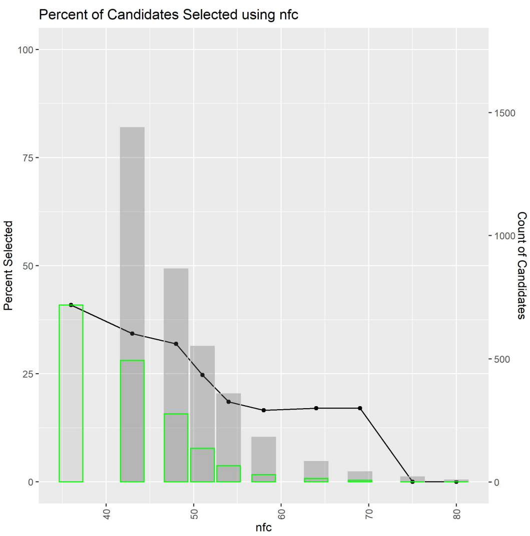

Percent of Candidates Selected using nuc
Percent Selected
Count of Candidates
nuc
0
25
50
75
100
500
1000
1500
2000
2500
40
60
80
100

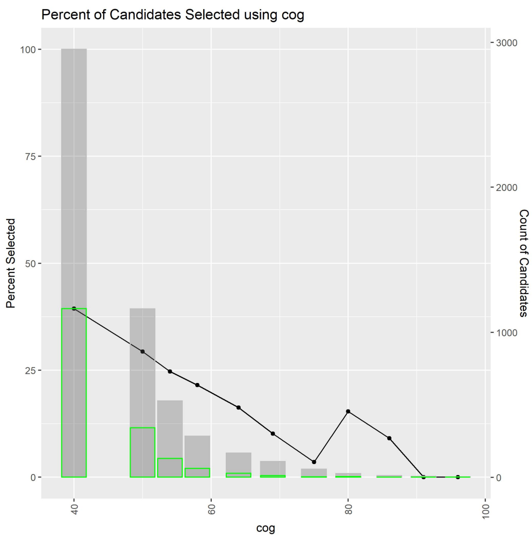

Percent of Candidates Selected using gic
Percent Selected
Count of Candidates
gic

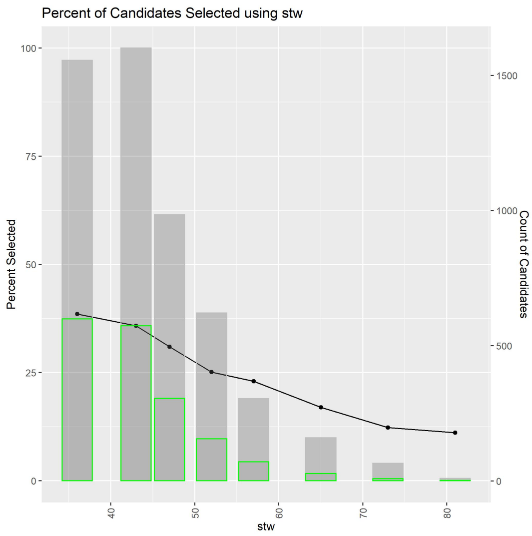

Percent of Candidates Selected using sui
Percent Selected
Count of Candidates
sui

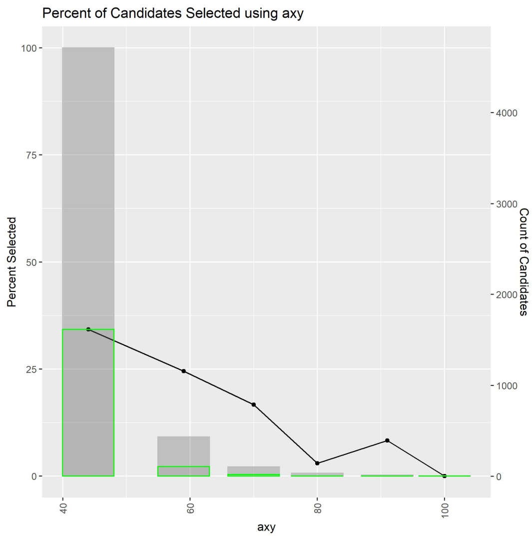

Percent of Candidates Selected using hlp
Percent Selected
Count of Candidates
hlp

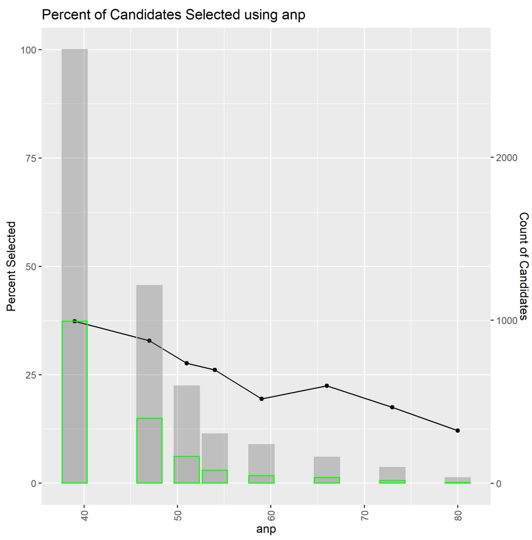

Percent of Candidates Selected using sfd
Percent Selected
Count of Candidates
sfd

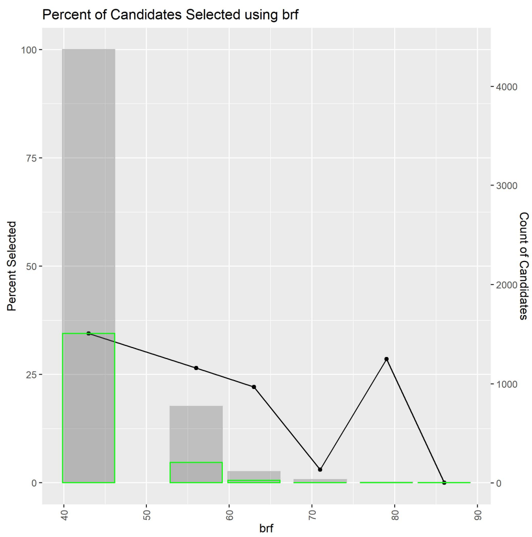

Percent of Candidates Selected using nfc
Percent Selected
Count of Candidates
nfc

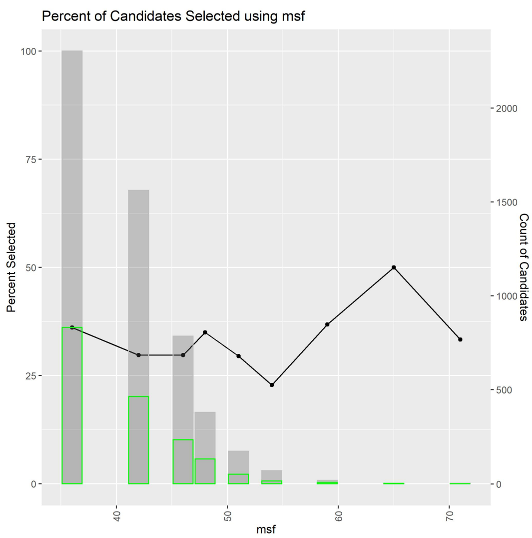

Percent of Candidates Selected using cog
Percent Selected
Count of Candidates
cog

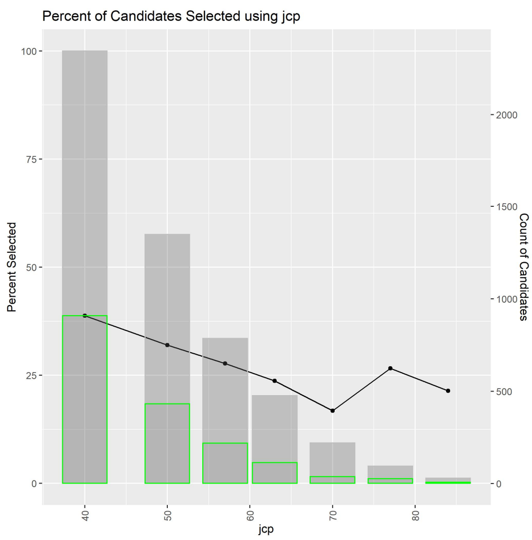

Percent of Candidates Selected using stw
Percent Selected
Count of Candidates
stw

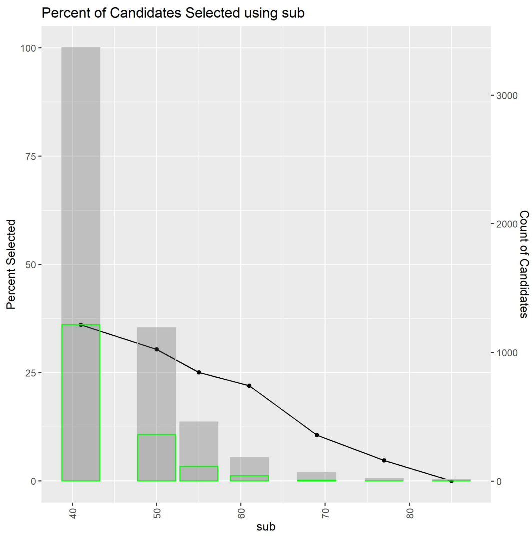

Percent of Candidates Selected using axy
Percent Selected
Count of Candidates
axy

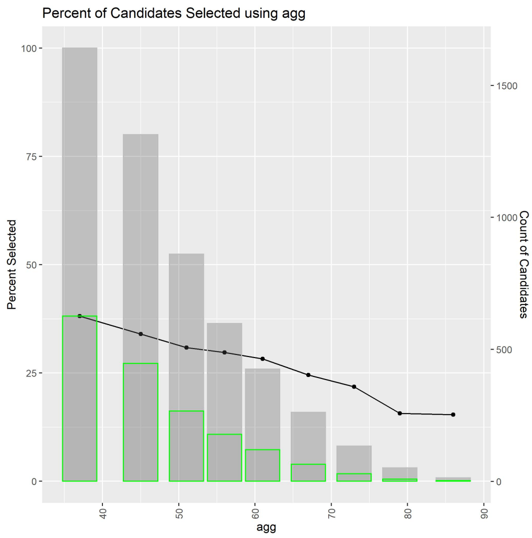

Percent of Candidates Selected using anp
Percent Selected
Count of Candidates
anp

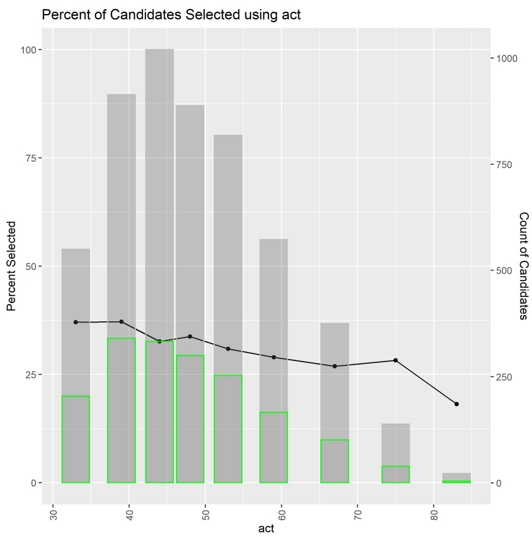

Percent of Candidates Selected using brf
Percent Selected
Count of Candidates
brf

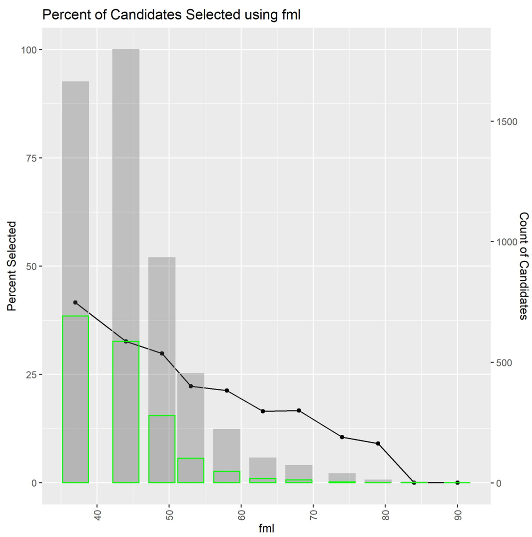

Percent of Candidates Selected using msf
Percent Selected
Count of Candidates
msf

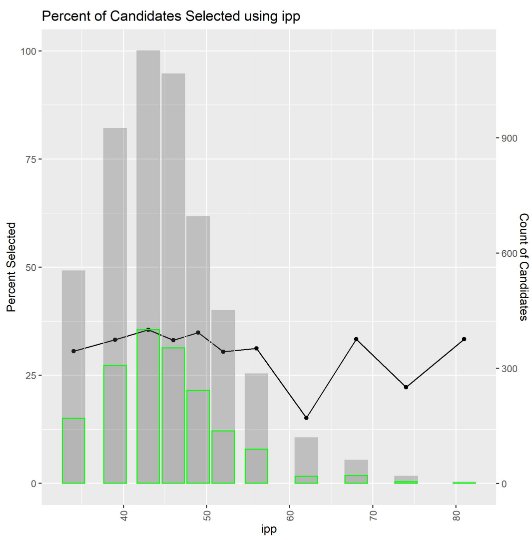

Percent of Candidates Selected using jcp
Percent Selected
Count of Candidates
jcp

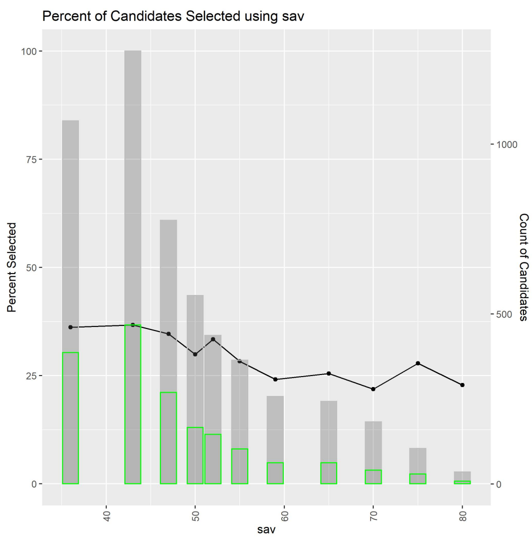

Percent of Candidates Selected using sub
Percent Selected
Count of Candidates
sub

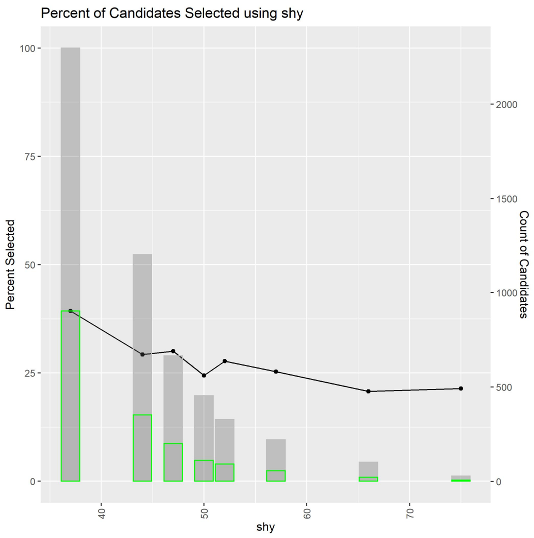

Percent of Candidates Selected using agg
Percent Selected
Count of Candidates
agg

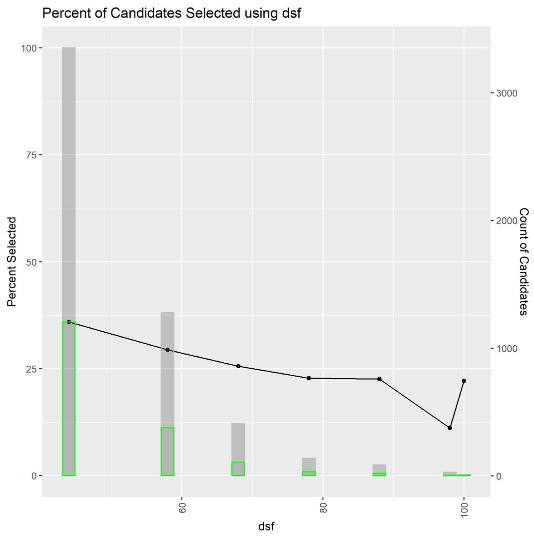

Percent of Candidates Selected using act
Percent Selected
Count of Candidates
act

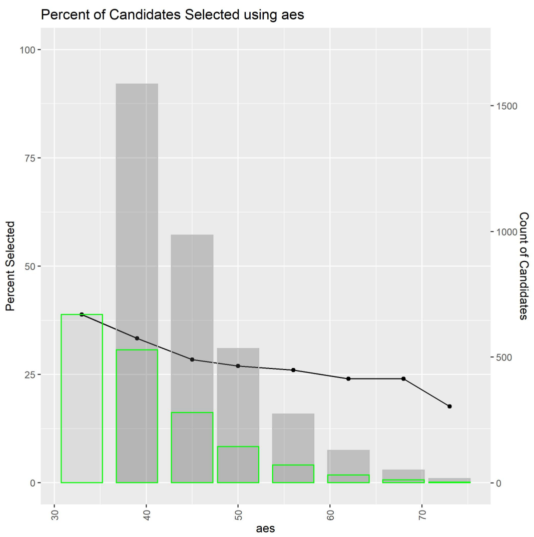

Percent of Candidates Selected using fml
Percent Selected
Count of Candidates
fml

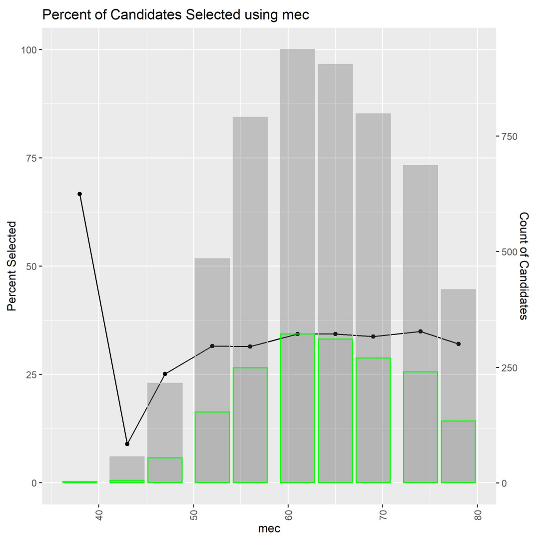

Percent of Candidates Selected using ipp
Percent Selected
Count of Candidates
ipp

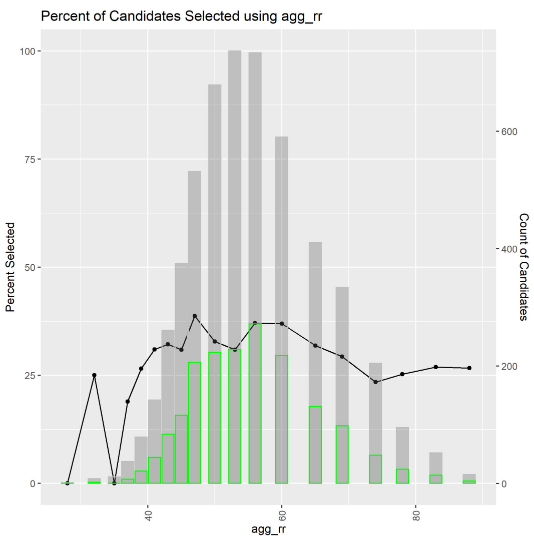

Percent of Candidates Selected using sav
Percent Selected
Count of Candidates
sav

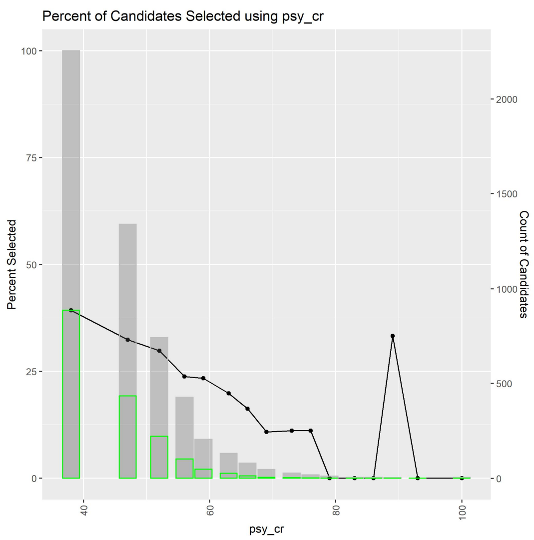

Percent of Candidates Selected using shy
Percent Selected
Count of Candidates
shy

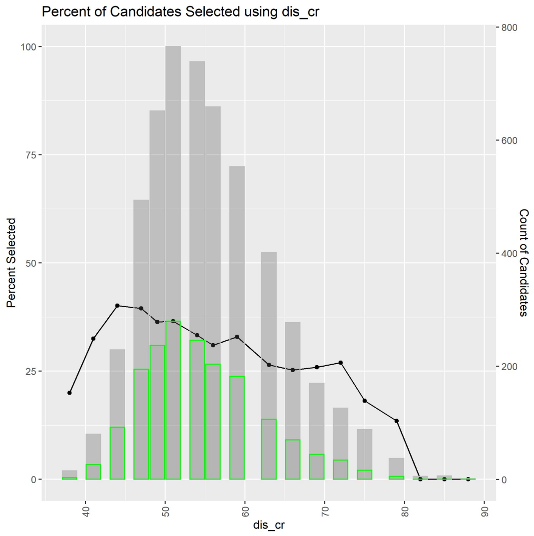

Percent of Candidates Selected using dsf
Percent Selected
Count of Candidates
dsf

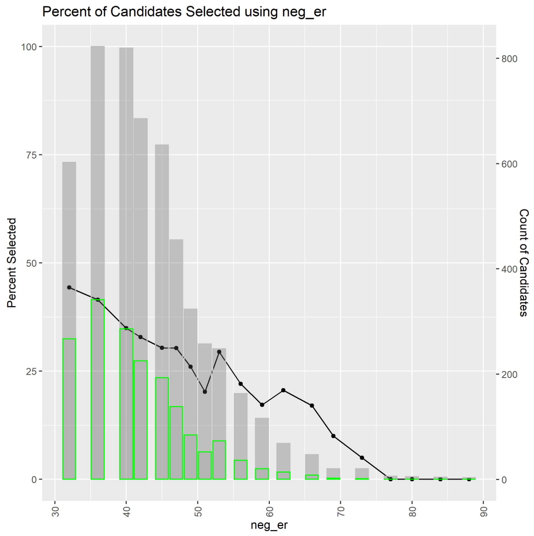

Percent of Candidates Selected using aes
Percent Selected
Count of Candidates
aes

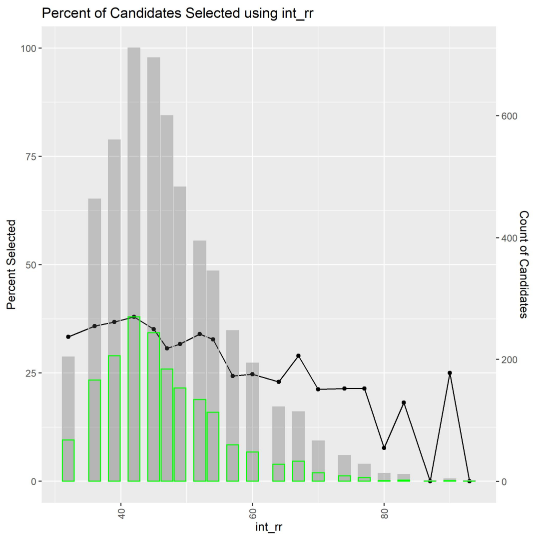

Percent of Candidates Selected using mec
Percent Selected
Count of Candidates
mec

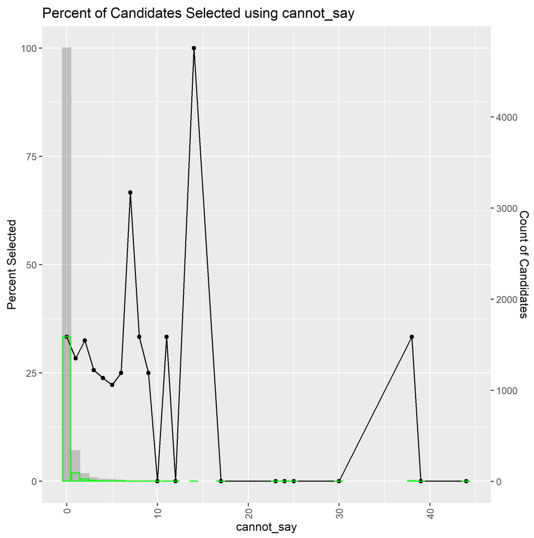

Percent of Candidates Selected using agg_rr
Percent Selected
Count of Candidates
agg_rr

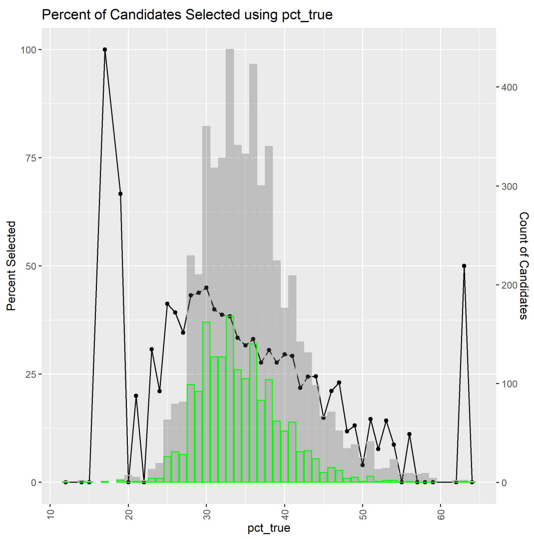

Percent of Candidates Selected using psy_cr
Percent Selected
Count of Candidates
psy_cr











C.3 Complete Results

TABLE C.2: Complete Results for all models and data subsets

Model	Data Subset	n	Selected	Not Selected	Percent Selected	Accuracy	PPV	NPV	Sensivitiy	Specificity	Kappa	F1
LDA	dataset_numeric	3095	1125	1970	36.35	73.06	62.25	79.76	65.58	77.33	42.42	63.87
LDA	dataset_performance	8156	2798	5358	34.31	70.77	55.46	84.05	75.09	68.51	40.2	63.8
LDA	dataset_psych	3544	1182	2362	33.35	63.09	45.52	74.76	54.52	67.37	20.86	49.61
LDA	dataset_psych_IQ	7286	2273	5013	31.2	52.98	33.92	71.48	53.6	52.69	5.43	41.55
LDA	dataset_psych_JPI_R	5506	1781	3725	32.35	61.66	42.28	73.96	50.75	66.88	16.75	46.13
LDA	dataset_psych_MMPI	5298	1740	3558	32.84	60.29	42.18	74.44	56.32	62.23	17.09	48.24
LDA	dataset_imputed_numeric	11885	2939	8946	24.73	77.86	53.99	89.26	70.6	80.25	46.09	61.19
LDA	dataset_imputed_performance	11885	2939	8946	24.73	77.64	54.11	87.1	62.77	82.52	42.98	58.12
LDA	dataset_imputed_psych	11885	2939	8946	24.73	68.57	39.15	81.73	48.92	75.03	22.1	43.49
LDA	dataset_imputed_psych_JPI_R	11885	2939	8946	24.73	70.31	39.72	80.06	38.82	80.66	19.63	39.27
LDA	dataset_imputed_psych_IQ	11885	2939	8946	24.73	67.54	31.4	77.04	26.45	81.03	7.89	28.71
LDA	dataset_imputed_psych_MMPI	11885	2939	8946	24.73	69.39	39.22	80.73	43.36	77.94	20.57	41.19
LDA	dataset_log_numeric	3095	1125	1970	36.35	71.55	57.95	84.86	78.93	67.34	42.94	66.83
LDA	dataset_log_performance	8156	2798	5358	34.31	70.4	54.87	84.9	77.23	66.83	40.16	64.16
LDA	dataset_log_psych	3544	1182	2362	33.35	62.99	45.41	74.73	54.52	67.23	20.71	49.55
LDA	dataset_log_psych_JPI_R	5506	1781	3725	32.35	60.99	41.67	73.86	51.5	65.53	16.04	46.06
LDA	dataset_log_psych_IQ	7286	2273	5013	31.2	52.29	33.87	71.64	55.65	50.77	5.46	42.11
LDA	dataset_log_psych_MMPI	5298	1740	3558	32.84	60.67	42.44	74.34	55.36	63.26	17.28	48.05
LDA	dataset_binned_numeric	3095	1125	1970	36.35	73.17	61.11	82.14	71.81	73.94	44.09	66.03
LDA	dataset_binned_performance	8156	2798	5358	34.31	69.91	54.3	84.92	77.59	65.9	39.45	63.89
LDA	dataset_binned_psych	3544	1182	2362	33.35	63.37	46.49	78.33	65.54	62.29	25.24	54.4
LDA	dataset_binned_psych_JPI_R	5506	1781	3725	32.35	66.87	48.91	76.99	54.49	72.78	26.49	51.55
LDA	dataset_binned_psych_IQ	7286	2273	5013	31.2	52.06	33.27	70.92	53.45	51.43	4.19	41.01
LDA	dataset_binned_psych_MMPI	5298	1740	3558	32.84	62.68	45.76	81.52	73.37	57.45	26.71	56.36
LDA	dataset_numeric_pca_best	3095	1125	1970	36.35	71.66	57.64	86.94	82.79	65.31	43.98	67.97
LDA	dataset_performance_pca_best	8156	2798	5358	34.31	69.79	54.15	85	77.83	65.59	39.31	63.86
LDA	dataset_psych_pca_best	3544	1182	2362	33.35	61.11	44.18	76.58	63.28	60.03	21.03	52.03
LDA	dataset_psych_IQ_pca_best	7286	2273	5013	31.2	52.93	33.86	71.42	53.45	52.69	5.31	41.46
LDA	dataset_psych_JPI_R_pca_best	5506	1781	3725	32.35	57.54	39.98	75.43	62.36	55.24	15.35	48.72
LDA	dataset_psych_MMPI_pca_best	5298	1740	3558	32.84	61.55	43.9	76.51	61.3	61.67	20.86	51.16
LDA	dataset_imputed_performance_pca_best	11885	2939	8946	24.73	77.64	54.11	87.1	62.77	82.52	42.98	58.12
LDA	dataset_imputed_psych_pca_best	11885	2939	8946	24.73	63.47	35.73	83.05	59.82	64.67	19.96	44.74
LDA	dataset_imputed_psych_JPI_R_pca_best	11885	2939	8946	24.73	70.26	39.43	79.87	37.91	80.88	19.04	38.66
LDA	dataset_imputed_psych_IQ_pca_best	11885	2939	8946	24.73	61.64	29.16	77.44	38.59	69.21	7.04	33.22
LDA	dataset_imputed_psych_MMPI_pca_best	11885	2939	8946	24.73	65.46	36.04	81.43	51.31	70.11	18.75	42.34
LDA	dataset_log_performance_pca_best	8156	2798	5358	34.31	69.87	54.29	84.58	76.88	66.21	39.19	63.64
LDA	dataset_log_psych_pca_best	3544	1182	2362	33.35	59.79	42.83	75.41	61.58	58.9	18.46	50.52
LDA	dataset_log_psych_JPI_R_pca_best	5506	1781	3725	32.35	59.06	41.01	75.61	60.67	58.28	16.85	48.94
LDA	dataset_log_psych_IQ_pca_best	7286	2273	5013	31.2	52.06	33.81	71.63	56.09	50.23	5.36	42.19
LDA	dataset_log_psych_MMPI_pca_best	5298	1740	3558	32.84	61.93	44.41	77.3	63.22	61.29	22.12	52.17
LDA	dataset_binned_performance_pca_best	8156	2798	5358	34.31	70.28	54.96	83.38	73.9	68.39	39.07	63.04
LDA	dataset_binned_psych_pca_best	3544	1182	2362	33.35	64.78	48	79.72	67.8	63.28	28.17	56.21
LDA	dataset_binned_psych_JPI_R_pca_best	5506	1781	3725	32.35	64.2	45.85	77.28	58.99	66.7	23.9	51.6
LDA	dataset_binned_psych_IQ_pca_best	7286	2273	5013	31.2	52.06	33.3	70.96	53.6	51.36	4.25	41.08
LDA	dataset_binned_psych_MMPI_pca_best	5298	1740	3558	32.84	60.42	44.23	83.08	78.54	51.55	25.11	56.59
LDA	dataset_apft_minimum_numeric	1938	998	940	51.5	64.54	61.6	71.11	82.61	45.39	28.29	70.57
LDA	dataset_apft_minimum_performance	4927	2470	2457	50.13	62.58	60.15	66.67	75.17	49.93	25.12	66.83
LDA	dataset_apft_minimum_psych	1994	1053	941	52.81	60.64	60.26	61.35	74.6	45.04	19.92	66.67
LDA	dataset_apft_minimum_psych_JPI_R	3026	1608	1418	53.14	55.57	58.35	52.53	57.26	53.65	10.89	57.8
LDA	dataset_apft_minimum_psych_IQ	3917	2028	1889	51.77	54.6	58.31	52.28	43.26	66.78	9.95	49.67
LDA	dataset_apft_minimum_psych_MMPI	2887	1521	1366	52.68	58.5	60.48	56.22	61.4	55.26	16.67	60.94
QDA	dataset_numeric	3095	1125	1970	36.35	69.83	58.77	75.79	56.68	77.33	34.27	57.7
QDA	dataset_performance	8156	2798	5358	34.31	68.97	53.13	86.3	80.93	62.73	38.8	64.15
QDA	dataset_psych	3544	1182	2362	33.35	61.02	42.54	72.27	48.31	67.37	15.16	45.24
QDA	dataset_psych_IQ	7286	2273	5013	31.2	53.21	33.78	71.23	52.13	53.69	5.07	40.99
QDA	dataset_psych_JPI_R	5506	1781	3725	32.35	56.75	39	74.19	59.74	55.33	13.23	47.19
QDA	dataset_psych_MMPI	5298	1740	3558	32.84	64.07	45.33	73.31	45.59	73.1	18.67	45.46
QDA	dataset_imputed_numeric	11885	2939	8946	24.73	71.66	44.86	86.23	63.9	74.21	33.36	52.72
QDA	dataset_imputed_performance	11885	2939	8946	24.73	78.48	54.98	89.59	71.4	80.81	47.44	62.12
QDA	dataset_imputed_psych	11885	2939	8946	24.73	62.88	36.14	84.52	65.38	62.06	21.57	46.55
QDA	dataset_imputed_psych_JPI_R	11885	2939	8946	24.73	62.65	33.67	80.92	52.67	65.93	15.64	41.08
QDA	dataset_imputed_psych_IQ	11885	2939	8946	24.73	62.57	31.72	79.01	44.61	68.47	11.51	37.08
QDA	dataset_imputed_psych_MMPI	11885	2939	8946	24.73	65.18	35.71	81.29	51.08	69.81	18.25	42.04
QDA	dataset_log_numeric	3095	1125	1970	36.35	69.18	57.43	76.07	58.46	75.3	33.62	57.94
QDA	dataset_log_performance	8156	2798	5358	34.31	69.83	54.07	86.06	79.98	64.53	39.93	64.52
QDA	dataset_log_psych	3544	1182	2362	33.35	62.24	44.14	73.22	50	68.36	17.77	46.89

Table C.2 continued from previous page

Model	Data Subset	n	Selected	Not Selected	Percent Selected	Accuracy	PPV	NPV	Sensivitiy	Specificity	Kappa	F1
QDA	dataset_log_psych_JPI_R	5506	1781	3725	32.35	54.82	37.95	74.06	62.55	51.12	11.69	47.24
QDA	dataset_log_psych_IQ	7286	2273	5013	31.2	49.73	33.06	71.25	59.77	45.18	4.05	42.57
QDA	dataset_log_psych_MMPI	5298	1740	3558	32.84	63.31	44.46	73.31	46.93	71.32	18	45.67
QDA	dataset_binned_numeric	3095	1125	1970	36.35	66.59	53.22	77.6	66.17	66.84	31.37	58.99
QDA	dataset_binned_performance	8156	2798	5358	34.31	70.07	54.48	84.92	77.47	66.21	39.68	63.98
QDA	dataset_binned_psych	3544	1182	2362	33.35	59.98	41.49	71.94	48.87	65.54	13.79	44.88
QDA	dataset_binned_psych_JPI_R	5506	1781	3725	32.35	64.99	46.59	76.82	56.37	69.11	24.16	51.02
QDA	dataset_binned_psych_IQ	7286	2273	5013	31.2	47.3	32.7	71.31	65.2	39.19	3.45	43.55
QDA	dataset_binned_psych_MMPI	5298	1740	3558	32.84	59.41	42.56	77.69	67.43	55.48	19.94	52.19
QDA	dataset_numeric_pca_best	3095	1125	1970	36.35	72.2	58.76	84.91	78.64	68.53	43.97	67.26
QDA	dataset_performance_pca_best	8156	2798	5358	34.31	69.71	54.25	83.46	74.49	67.21	38.29	62.78
QDA	dataset_psych_pca_best	3544	1182	2362	33.35	60.45	43.29	75.26	60.17	60.59	18.92	50.35
QDA	dataset_psych_IQ_pca_best	7286	2273	5013	31.2	52.29	33.64	71.32	54.48	51.3	4.94	41.59
QDA	dataset_psych_JPI_R_pca_best	5506	1781	3725	32.35	54.75	38.07	74.41	63.67	50.49	12.05	47.65
QDA	dataset_psych_MMPI_pca_best	5298	1740	3558	32.84	61.74	41.79	71.55	41.95	71.42	13.36	41.87
QDA	dataset_imputed_performance_pca_best	11885	2939	8946	24.73	78.48	54.98	89.59	71.4	80.81	47.44	62.12
QDA	dataset_imputed_psych_pca_best	11885	2939	8946	24.73	64.34	36.81	83.84	61.75	65.19	21.94	46.12
QDA	dataset_imputed_psych_JPI_R_pca_best	11885	2939	8946	24.73	63.47	33.92	80.62	50.4	67.76	15.61	40.55
QDA	dataset_imputed_psych_IQ_pca_best	11885	2939	8946	24.73	63.95	31.6	78.35	39.39	72.01	10.53	35.07
QDA	dataset_imputed_psych_MMPI_pca_best	11885	2939	8946	24.73	67.31	36.9	80.6	45.4	74.51	18.48	40.71
QDA	dataset_log_performance_pca_best	8156	2798	5358	34.31	69.46	53.9	83.97	75.8	66.15	38.24	63
QDA	dataset_log_psych_pca_best	3544	1182	2362	33.35	61.49	43.71	73.92	53.95	65.25	18.15	48.29
QDA	dataset_log_psych_JPI_R_pca_best	5506	1781	3725	32.35	54.57	38.51	75.81	67.79	48.25	13.39	49.12
QDA	dataset_log_psych_IQ_pca_best	7286	2273	5013	31.2	44.37	33.6	75.9	80.32	28.08	6.09	47.38
QDA	dataset_log_psych_MMPI_pca_best	5298	1740	3558	32.84	59.03	39.94	71.94	49.04	63.92	12.25	44.02
QDA	dataset_binned_performance_pca_best	8156	2798	5358	34.31	70.73	55.73	82.45	71.28	70.44	39.11	62.55
QDA	dataset_binned_psych_pca_best	3544	1182	2362	33.35	63.37	46.72	80.15	70.34	59.89	26.83	56.14
QDA	dataset_binned_psych_JPI_R_pca_best	5506	1781	3725	32.35	64.93	46.31	75.82	52.81	70.73	22.7	49.34
QDA	dataset_binned_psych_IQ_pca_best	7286	2273	5013	31.2	52.06	33.3	70.96	53.6	51.36	4.25	41.08
QDA	dataset_binned_psych_MMPI_pca_best	5298	1740	3558	32.84	61.36	44.64	80.98	73.37	55.48	24.78	55.51
QDA	dataset_apft_minimum_numeric	1938	998	940	51.5	57.31	56.56	58.85	73.58	40.07	13.77	63.95
QDA	dataset_apft_minimum_performance	4927	2470	2457	50.13	63.13	60.56	67.45	75.84	50.34	26.2	67.35
QDA	dataset_apft_minimum_psych	1994	1053	941	52.81	57.12	57.66	56.13	70.48	42.2	12.84	63.43
QDA	dataset_apft_minimum_psych_JPI_R	3026	1608	1418	53.14	54.69	55.45	52.73	74.9	31.76	6.82	63.72
QDA	dataset_apft_minimum_psych_IQ	3917	2028	1889	51.77	58.52	58.94	57.95	65.62	50.88	16.58	62.1
QDA	dataset_apft_minimum_psych_MMPI	2887	1521	1366	52.68	60.23	59.27	62.45	78.51	39.85	18.71	67.55
pen_log_reg	dataset_admin	11689	2852	8837	24.4	69.22	40	82.94	52.4	74.65	24.48	45.37
pen_log_reg	dataset_all	3095	1125	1970	36.35	74.89	62.56	84.82	76.85	73.77	48.26	68.97
pen_log_reg	dataset_numeric	3095	1125	1970	36.35	73.17	59.95	85.19	78.64	70.05	45.63	68.04
pen_log_reg	dataset_performance	8156	2798	5358	34.31	70.52	55.03	84.77	76.88	67.21	40.26	64.15
pen_log_reg	dataset_psych_IQ	7286	2273	5013	31.2	53.21	34.17	71.73	54.04	52.83	5.92	41.87
pen_log_reg	dataset_psych_JPI_R	5506	1781	3725	32.35	59.54	40.64	74.01	54.49	61.95	15.1	46.56
pen_log_reg	dataset_psych_MMPI	5298	1740	3558	32.84	60.79	42.78	74.94	57.28	62.51	18.21	48.98
pen_log_reg	dataset_psych	3544	1182	2362	33.35	62.81	45.2	74.65	54.52	66.95	20.42	49.42
pen_log_reg	dataset_imputed_all	11885	2939	8946	24.73	80.53	59.51	88.53	66.4	85.17	49.64	62.77
pen_log_reg	dataset_imputed_admin	11885	2939	8946	24.73	72.67	45.06	82.58	48.13	80.73	28.21	46.54
pen_log_reg	dataset_imputed_performance	11885	2939	8946	24.73	77.61	54.07	87.03	62.54	82.56	42.85	58
pen_log_reg	dataset_imputed_psych	11885	2939	8946	24.73	67.26	38.56	82.76	54.71	71.38	22.87	45.24
pen_log_reg	dataset_imputed_psych_JPI_R	11885	2939	8946	24.73	70.29	39.55	79.94	38.25	80.81	19.27	38.89
pen_log_reg	dataset_imputed_psych_IQ	11885	2939	8946	24.73	61.34	28.71	77.22	38.02	68.99	6.33	32.71
pen_log_reg	dataset_imputed_psych_MMPI	11885	2939	8946	24.73	68.24	38.58	81.46	48.13	74.84	21.21	42.83
pen_log_reg	dataset_log_all	3095	1125	1970	36.35	74.78	61.78	86.35	80.12	71.74	48.75	69.77
pen_log_reg	dataset_log_admin	11689	2852	8837	24.4	69.88	40.53	82.62	50.29	76.2	24.49	44.89
pen_log_reg	dataset_log_performance	8156	2798	5358	34.31	70.32	54.83	84.49	76.4	67.14	39.8	63.84
pen_log_reg	dataset_log_psych	3544	1182	2362	33.35	61.86	44.68	75.99	60.45	62.57	21.16	51.38
pen_log_reg	dataset_log_psych_JPI_R	5506	1781	3725	32.35	60.75	41.31	73.57	50.75	65.53	15.37	45.55
pen_log_reg	dataset_log_psych_IQ	7286	2273	5013	31.2	52.75	34.32	72.11	56.39	51.1	6.37	42.67
pen_log_reg	dataset_log_psych_MMPI	5298	1740	3558	32.84	60.54	43.19	76.89	63.79	58.95	20.27	51.51
pen_log_reg	dataset_binned_all	3095	1125	1970	36.35	76.83	64.95	86.15	78.64	75.8	52.08	71.14
pen_log_reg	dataset_binned_admin	11689	2852	8837	24.4	68.48	39.39	83.2	54.27	73.07	24.23	45.65
pen_log_reg	dataset_binned_performance	8156	2798	5358	34.31	70.11	54.75	83.44	74.14	68.01	38.86	62.99
pen_log_reg	dataset_binned_psych	3544	1182	2362	33.35	62.62	45.95	79.28	68.93	59.46	25.24	55.14
pen_log_reg	dataset_binned_psych_JPI_R	5506	1781	3725	32.35	67.05	49.21	78.17	58.43	71.17	28.22	53.42
pen_log_reg	dataset_binned_psych_IQ	7286	2273	5013	31.2	52.06	33.27	70.92	53.45	51.43	4.19	41.01
pen_log_reg	dataset_binned_psych_MMPI	5298	1740	3558	32.84	61.8	44.82	79.95	70.5	57.54	24.46	54.8
pen_log_reg	dataset_all_pca_best	3095	1125	1970	36.35	75	62.9	84.45	75.96	74.45	48.26	68.82
pen_log_reg	dataset_numeric_pca_best	3095	1125	1970	36.35	73.92	61.07	84.97	77.74	71.74	46.75	68.41

Table C.2 continued from previous page

Model	Data Subset	n	Selected	Not Selected	Percent Selected	Accuracy	PPV	NPV	Sensivitiy	Specificity	Kappa	F1
pen_log_reg	dataset_performance_pca_best	8156	2798	5358	34.31	70.03	54.51	84.41	76.4	66.71	39.33	63.62
pen_log_reg	dataset_psych_pca_best	3544	1182	2362	33.35	61.49	44.58	76.94	63.84	60.31	21.8	52.5
pen_log_reg	dataset_psych_JPI_R_pca_best	5506	1781	3725	32.35	61.54	41.82	73.31	48.31	67.86	15.54	44.83
pen_log_reg	dataset_psych_MMPI_pca_best	5298	1740	3558	32.84	59.72	42.8	77.76	67.24	56.04	20.32	52.31
pen_log_reg	dataset_imputed_performance_pca_best	11885	2939	8946	24.73	77.5	53.81	87.22	63.34	82.15	42.93	58.19
pen_log_reg	dataset_imputed_psych_pca_best	11885	2939	8946	24.73	65.54	36.73	82.23	54.48	69.18	20.35	43.88
pen_log_reg	dataset_imputed_psych_JPI_R_pca_best	11885	2939	8946	24.73	71.02	38.69	78.53	29.51	84.64	15.38	33.48
pen_log_reg	dataset_imputed_psych_IQ_pca_best	11885	2939	8946	24.73	61.5	28.78	77.23	37.8	69.29	6.41	32.68
pen_log_reg	dataset_imputed_psych_MMPI_pca_best	11885	2939	8946	24.73	66.33	36.61	81.25	49.49	71.86	19.09	42.08
pen_log_reg	dataset_log_psych_pca_best	3544	1182	2362	33.35	59.32	43.01	77.38	67.8	55.08	20	52.63
pen_log_reg	dataset_log_psych_JPI_R_pca_best	5506	1781	3725	32.35	58.87	40.86	75.52	60.67	58.01	16.59	48.83
pen_log_reg	dataset_log_psych_MMPI_pca_best	5298	1740	3558	32.84	59.47	42.65	77.87	67.82	55.39	20.17	52.37
pen_log_reg	dataset_binned_performance_pca_best	8156	2798	5358	34.31	70.03	54.73	82.96	73.06	68.45	38.44	62.58
pen_log_reg	dataset_binned_psych_pca_best	3544	1182	2362	33.35	65.25	48.48	79.89	67.8	63.98	28.9	56.54
pen_log_reg	dataset_binned_psych_JPI_R_pca_best	5506	1781	3725	32.35	64.99	46.54	76.55	55.43	69.56	23.81	50.6
pen_log_reg	dataset_binned_psych_MMPI_pca_best	5298	1740	3558	32.84	60.79	44.44	82.65	77.39	52.67	25.28	56.46
pen_log_reg	dataset_apft_minimum_all	1938	998	940	51.5	64.03	63.8	64.31	69.57	58.16	27.8	66.56
pen_log_reg	dataset_apft_minimum_admin	4888	2414	2474	49.39	62.69	62.38	62.98	61.6	63.75	25.35	61.99
pen_log_reg	dataset_apft_minimum_performance	4927	2470	2457	50.13	62.72	60.22	66.97	75.57	49.8	25.39	67.03
pen_log_reg	dataset_apft_minimum_psych	1994	1053	941	52.81	62.14	63.78	60.22	65.4	58.51	23.94	64.58
pen_log_reg	dataset_apft_minimum_psych_JPI_R	3026	1608	1418	53.14	52.59	56.74	49.52	45.44	60.71	6.06	50.46
pen_log_reg	dataset_apft_minimum_psych_IQ	3917	2028	1889	51.77	56.56	58.11	54.91	57.73	55.3	13.03	57.92
pen_log_reg	dataset_apft_minimum_psych_MMPI	2887	1521	1366	52.68	58.61	60.21	56.62	63.38	53.3	16.73	61.75
pen_log_reg	dataset_best_10	3295	1229	2066	37.3	72.14	59.28	85.39	80.71	67.04	44.49	68.35
pen_log_reg	dataset_best_20	3155	1182	1973	37.46	72.59	60.53	83.6	77.12	69.88	44.55	67.83
svm_lin	dataset_admin	11689	2852	8837	24.4	69.68	39.62	81.71	46.43	77.18	22.31	42.76
svm_lin	dataset_all	3095	1125	1970	36.35	74.78	63.04	83.49	73.89	75.3	47.43	68.03
svm_lin	dataset_numeric	3095	1125	1970	36.35	71.66	57.58	87.27	83.38	64.97	44.11	68.12
svm_lin	dataset_performance	8156	2798	5358	34.31	70.52	55.12	84.23	75.69	67.83	39.95	63.79
svm_lin	dataset_psych_IQ	7286	2273	5013	31.2	52.93	33.83	71.38	53.3	52.76	5.24	41.39
svm_lin	dataset_psych_JPI_R	5506	1781	3725	32.35	62.39	43.13	74.36	51.12	67.77	18.02	46.79
svm_lin	dataset_psych_MMPI	5298	1740	3558	32.84	60.48	42.23	74.2	55.17	63.07	16.92	47.84
svm_lin	dataset_psych	3544	1182	2362	33.35	63.94	46.35	74.44	51.98	69.92	21.25	49
svm_lin	dataset_imputed_all	11885	2939	8946	24.73	79.38	56.65	89.5	70.6	82.26	48.82	62.86
svm_lin	dataset_imputed_admin	11885	2939	8946	24.73	72.17	44.62	83.38	52.21	78.72	29.26	48.12
svm_lin	dataset_imputed_performance	11885	2939	8946	24.73	77.69	54.17	87.25	63.34	82.41	43.29	58.4
svm_lin	dataset_imputed_psych	11885	2939	8946	24.73	67.45	38.21	81.98	51.31	72.75	21.58	43.8
svm_lin	dataset_imputed_psych_IQ	11885	2939	8946	24.73	62.29	29.24	77.34	37	70.59	6.97	32.67
svm_lin	dataset_imputed_psych_MMPI	11885	2939	8946	24.73	67.12	37.57	81.57	49.94	72.75	20.44	42.88
svm_lin	dataset_log_all	3095	1125	1970	36.35	75.75	63.93	84.79	76.26	75.47	49.67	69.55
svm_lin	dataset_log_admin	11689	2852	8837	24.4	70.42	40.38	81.52	44.68	78.73	22.59	42.42
svm_lin	dataset_log_performance	8156	2798	5358	34.31	70.97	55.78	83.7	74.14	69.32	40.29	63.66
svm_lin	dataset_log_psych	3544	1182	2362	33.35	63.18	45.75	75.28	56.21	66.67	21.64	50.44
svm_lin	dataset_log_psych_JPI_R	5506	1781	3725	32.35	60.93	41.42	73.51	50.19	66.07	15.41	45.39
svm_lin	dataset_log_psych_IQ	7286	2273	5013	31.2	52.11	33.69	71.46	55.36	50.63	5.1	41.89
svm_lin	dataset_log_psych_MMPI	5298	1740	3558	32.84	61.42	42.95	74.05	53.07	65.51	17.52	47.47
svm_lin	dataset_binned_all	3095	1125	1970	36.35	75	63.57	83.18	73	76.14	47.62	67.96
svm_lin	dataset_binned_admin	11689	2852	8837	24.4	69.79	39.66	81.59	45.73	77.56	22.14	42.48
svm_lin	dataset_binned_performance	8156	2798	5358	34.31	69.71	54.06	84.98	77.83	65.46	39.18	63.8
svm_lin	dataset_binned_psych	3544	1182	2362	33.35	63.75	46.8	77.99	64.12	63.56	25.34	54.11
svm_lin	dataset_binned_psych_JPI_R	5506	1781	3725	32.35	67.35	49.57	77.11	54.31	73.59	27.22	51.83
svm_lin	dataset_binned_psych_IQ	7286	2273	5013	31.2	52.06	33.27	70.92	53.45	51.43	4.19	41.01
svm_lin	dataset_binned_psych_MMPI	5298	1740	3558	32.84	61.99	45.06	80.5	71.65	57.26	25.12	55.33
svm_lin	dataset_all_pca_best	3095	1125	1970	36.35	73.71	60.4	86.07	80.12	70.05	46.88	68.88
svm_lin	dataset_numeric_pca_best	3095	1125	1970	36.35	71.77	57.83	86.64	82.2	65.82	44.03	67.89
svm_lin	dataset_performance_pca_best	8156	2798	5358	34.31	69.75	54.16	84.49	76.76	66.09	38.96	63.51
svm_lin	dataset_psych_pca_best	3544	1182	2362	33.35	62.43	44.83	74.64	55.08	66.1	20.04	49.43
svm_lin	dataset_psych_IQ_pca_best	7286	2273	5013	31.2	52.98	33.86	71.4	53.3	52.83	5.3	41.41
svm_lin	dataset_psych_JPI_R_pca_best	5506	1781	3725	32.35	58.39	40.73	76.03	62.92	56.22	16.76	49.45
svm_lin	dataset_psych_MMPI_pca_best	5298	1740	3558	32.84	61.36	43.8	76.74	62.26	60.92	20.93	51.42
svm_lin	dataset_imputed_performance_pca_best	11885	2939	8946	24.73	77.67	54.12	87.25	63.34	82.37	43.24	58.37
svm_lin	dataset_imputed_psych_pca_best	11885	2939	8946	24.73	65.97	37.05	82.21	53.92	69.92	20.68	43.92
svm_lin	dataset_imputed_psych_IQ_pca_best	11885	2939	8946	24.73	51.49	27.68	78.67	59.7	48.79	6.12	37.83
svm_lin	dataset_imputed_psych_MMPI_pca_best	11885	2939	8946	24.73	62.99	34.49	81.69	55.28	65.52	17.3	42.48
svm_lin	dataset_log_performance_pca_best	8156	2798	5358	34.31	69.71	54.06	84.92	77.71	65.53	39.15	63.77
svm_lin	dataset_log_psych_pca_best	3544	1182	2362	33.35	60.92	42.96	73.29	52.54	65.11	16.72	47.27
svm_lin	dataset_log_psych_JPI_R_pca_best	5506	1781	3725	32.35	59.48	41.36	75.75	60.49	59	17.39	49.13

Table C.2 continued from previous page

Model	Data Subset	n	Selected	Not Selected	Percent Selected	Accuracy	PPV	NPV	Sensivitiy	Specificity	Kappa	F1
svm_lin	dataset_log_psych_IQ_pca_best	7286	2273	5013	31.2	51.97	33.6	71.37	55.36	50.43	4.92	41.82
svm_lin	dataset_log_psych_MMPI_pca_best	5298	1740	3558	32.84	59.35	42.55	77.81	67.82	55.2	19.99	52.29
svm_lin	dataset_binned_psych_pca_best	3544	1182	2362	33.35	63.84	47.15	80.22	70.06	60.73	27.46	56.36
svm_lin	dataset_binned_psych_JPI_R_pca_best	5506	1781	3725	32.35	63.9	45.52	77.16	58.99	66.25	23.43	51.39
svm_lin	dataset_binned_psych_IQ_pca_best	7286	2273	5013	31.2	52.06	33.3	70.96	53.6	51.36	4.25	41.08
svm_lin	dataset_binned_psych_MMPI_pca_best	5298	1740	3558	32.84	60.35	44.22	83.36	79.12	51.17	25.21	56.73
svm_lin	dataset_apft_minimum_all	1938	998	940	51.5	67.13	71.09	64	60.87	73.76	34.48	65.59
svm_lin	dataset_apft_minimum_admin	4888	2414	2474	49.39	63.37	62.72	64.02	63.67	63.07	26.74	63.19
svm_lin	dataset_apft_minimum_performance	4927	2470	2457	50.13	63.19	60.45	68.04	76.92	49.39	26.33	67.7
svm_lin	dataset_apft_minimum_psych	1994	1053	941	52.81	61.64	61.88	61.28	71.11	51.06	22.37	66.17
svm_lin	dataset_apft_minimum_psych_JPI_R	3026	1608	1418	53.14	54.69	57.7	51.57	55.19	54.12	9.28	56.42
svm_lin	dataset_apft_minimum_psych_IQ	3917	2028	1889	51.77	56.98	58.43	55.42	58.72	55.12	13.84	58.57
svm_lin	dataset_apft_minimum_psych_MMPI	2887	1521	1366	52.68	58.61	60.43	56.46	62.28	54.52	16.83	61.34
svm_lin	dataset_best_10	3295	1229	2066	37.3	72.75	59.57	87.23	83.7	66.24	46.14	69.6
svm_lin	dataset_best_20	3155	1182	1973	37.46	72.59	62.03	80.18	69.21	74.62	42.83	65.42
CART	dataset_admin	11689	2852	8837	24.4	67.11	38.88	84.58	60.94	69.11	25.2	47.47
CART	dataset_all	3095	1125	1970	36.35	69.5	55.6	84.53	79.53	63.79	39.65	65.45
CART	dataset_numeric	3095	1125	1970	36.35	68.43	54.51	83.86	78.93	62.44	37.73	64.48
CART	dataset_performance	8156	2798	5358	34.31	70.36	55.14	83.01	72.94	69.01	38.95	62.8
CART	dataset_psych	3544	1182	2362	33.35	60.26	42.54	73.6	54.8	62.99	16.6	47.9
CART	dataset_psych_IQ	7286	2273	5013	31.2	47.71	33.96	74.16	71.66	36.86	6.54	46.08
CART	dataset_psych_JPI_R	5506	1781	3725	32.35	56.03	38.06	73.08	57.3	55.42	11.24	45.74
CART	dataset_psych_MMPI	5298	1740	3558	32.84	58.28	40.11	73.06	54.79	59.98	13.51	46.32
CART	dataset_imputed_all	11885	2939	8946	24.73	75.36	50.12	88.42	69.13	77.41	41.28	58.11
CART	dataset_imputed_numeric	11885	2939	8946	24.73	74.97	49.57	89.09	71.62	76.07	41.49	58.59
CART	dataset_imputed_admin	11885	2939	8946	24.73	67.76	40.3	85.15	63.22	69.25	27.27	49.23
CART	dataset_imputed_performance	11885	2939	8946	24.73	75.9	50.89	89.11	71.17	77.45	42.88	59.35
CART	dataset_imputed_psych	11885	2939	8946	24.73	68.24	37.54	80.33	42.91	76.56	18.57	40.04
CART	dataset_imputed_psych_JPI_R	11885	2939	8946	24.73	62.71	34.44	81.87	56.3	64.82	17.41	42.74
CART	dataset_imputed_psych_IQ	11885	2939	8946	24.73	57.04	30.63	80.54	58.34	56.62	11.47	40.17
CART	dataset_imputed_psych_MMPI	11885	2939	8946	24.73	64.96	35.71	81.5	52.21	69.14	18.48	42.42
CART	dataset_log_all	3095	1125	1970	36.35	69.5	55.6	84.53	79.53	63.79	39.65	65.45
CART	dataset_log_numeric	3095	1125	1970	36.35	68.43	54.51	83.86	78.93	62.44	37.73	64.48
CART	dataset_log_admin	11689	2852	8837	24.4	67.11	38.88	84.58	60.94	69.11	25.2	47.47
CART	dataset_log_performance	8156	2798	5358	34.31	70.36	55.14	83.01	72.94	69.01	38.95	62.8
CART	dataset_log_psych	3544	1182	2362	33.35	60.26	42.54	73.6	54.8	62.99	16.6	47.9
CART	dataset_log_psych_JPI_R	5506	1781	3725	32.35	56.03	38.06	73.08	57.3	55.42	11.24	45.74
CART	dataset_log_psych_IQ	7286	2273	5013	31.2	47.71	33.96	74.16	71.66	36.86	6.54	46.08
CART	dataset_log_psych_MMPI	5298	1740	3558	32.84	58.28	40.11	73.06	54.79	59.98	13.51	46.32
CART	dataset_binned_all	3095	1125	1970	36.35	72.63	60.95	80.69	68.55	74.96	42.37	64.53
CART	dataset_binned_numeric	3095	1125	1970	36.35	72.41	59.14	84.54	77.74	69.37	44.14	67.18
CART	dataset_binned_admin	11689	2852	8837	24.4	67.68	38.9	83.67	56.96	71.14	24.28	46.23
CART	dataset_binned_performance	8156	2798	5358	34.31	69.71	54.06	84.98	77.83	65.46	39.18	63.8
CART	dataset_binned_psych	3544	1182	2362	33.35	60.36	43.12	74.96	59.32	60.88	18.46	49.94
CART	dataset_binned_psych_JPI_R	5506	1781	3725	32.35	65.96	47.89	78.12	59.55	69.02	26.87	53.09
CART	dataset_binned_psych_IQ	7286	2273	5013	31.2	48.26	32.69	71.03	62.26	41.92	3.35	42.87
CART	dataset_binned_psych_MMPI	5298	1740	3558	32.84	62.62	45.29	78.67	66.28	60.82	24.24	53.81
CART	dataset_all_pca_best	3095	1125	1970	36.35	69.72	56.33	81.89	73.89	67.34	38.64	63.93
CART	dataset_numeric_pca_best	3095	1125	1970	36.35	69.83	56.38	82.33	74.78	67.01	39.04	64.29
CART	dataset_performance_pca_best	8156	2798	5358	34.31	67.5	51.59	88.16	84.98	58.37	37.55	64.21
CART	dataset_psych_pca_best	3544	1182	2362	33.35	57.91	39.36	70.88	48.59	62.57	10.54	43.49
CART	dataset_psych_IQ_pca_best	7286	2273	5013	31.2	50.73	34.63	73.75	65.35	44.11	7.61	45.27
CART	dataset_psych_JPI_R_pca_best	5506	1781	3725	32.35	60.87	37.16	69.38	30.34	75.47	6.1	33.4
CART	dataset_psych_MMPI_pca_best	5298	1740	3558	32.84	60.1	41.36	73.01	51.34	64.39	14.81	45.81
CART	dataset_imputed_performance_pca_best	11885	2939	8946	24.73	73.82	48.31	93.19	84.34	70.37	43.75	61.43
CART	dataset_imputed_psych_pca_best	11885	2939	8946	24.73	69.87	40	80.95	43.81	78.42	21.54	41.82
CART	dataset_imputed_psych_JPI_R_pca_best	11885	2939	8946	24.73	62.12	34.19	82.03	57.55	63.62	17.22	42.89
CART	dataset_imputed_psych_IQ_pca_best	11885	2939	8946	24.73	55.58	30.97	81.98	64.81	52.55	12.7	41.91
CART	dataset_imputed_psych_MMPI_pca_best	11885	2939	8946	24.73	64.37	35.79	82.19	55.62	67.24	19.27	43.56
CART	dataset_log_performance_pca_best	8156	2798	5358	34.31	69.87	54.34	84.25	76.16	66.58	39	63.42
CART	dataset_log_psych_pca_best	3544	1182	2362	33.35	58	41.19	74.26	60.73	56.64	15.53	49.09
CART	dataset_log_psych_JPI_R_pca_best	5506	1781	3725	32.35	51.61	36.74	74.39	68.73	43.42	9.9	47.88
CART	dataset_log_psych_IQ_pca_best	7286	2273	5013	31.2	48.95	33.56	72.45	65.05	41.65	5.33	44.28
CART	dataset_log_psych_MMPI_pca_best	5298	1740	3558	32.84	59.97	41.2	72.9	51.15	64.29	14.55	45.64
CART	dataset_binned_performance_pca_best	8156	2798	5358	34.31	70.52	55.39	82.77	72.23	69.63	39.03	62.7
CART	dataset_binned_psych_pca_best	3544	1182	2362	33.35	63.47	45.87	74.62	53.39	68.5	21.03	49.35
CART	dataset_binned_psych_JPI_R_pca_best	5506	1781	3725	32.35	63.9	44.38	73.7	45.88	72.52	18.24	45.12

Table C.2 continued from previous page

Model	Data Subset	n	Selected	Not Selected	Percent Selected	Accuracy	PPV	NPV	Sensivitiy	Specificity	Kappa	F1
CART	dataset_binned_psych_IQ_pca_best	7286	2273	5013	31.2	45.28	33.38	74.14	75.77	31.47	5.37	46.34
CART	dataset_binned_psych_MMPI_pca_best	5298	1740	3558	32.84	65.2	47.54	76.77	57.28	69.07	25.04	51.95
CART	dataset_apft_minimum_all	1938	998	940	51.5	62.31	68.69	58.58	49.16	76.24	25.19	57.31
CART	dataset_apft_minimum_numeric	1938	998	940	51.5	60.59	61.01	60.08	64.88	56.03	20.95	62.88
CART	dataset_apft_minimum_admin	4888	2414	2474	49.39	62.28	60.75	64.08	66.71	57.95	24.63	63.59
CART	dataset_apft_minimum_performance	4927	2470	2457	50.13	62.25	60.32	65.14	72.2	52.24	24.45	65.72
CART	dataset_apft_minimum_psych	1994	1053	941	52.81	57.62	57.24	58.58	77.78	35.11	13.16	65.95
CART	dataset_apft_minimum_psych_JPI_R	3026	1608	1418	53.14	56.23	56.57	55.38	75.93	33.88	10.05	64.84
CART	dataset_apft_minimum_psych_IQ	3917	2028	1889	51.77	56.98	55.72	61.17	82.57	29.51	12.29	66.53
CART	dataset_apft_minimum_psych_MMPI	2887	1521	1366	52.68	56.42	58.39	54.06	60.31	52.08	12.41	59.33
CART	dataset_best_10	3295	1229	2066	37.3	70.82	57.81	84.84	80.43	65.11	42.19	67.27
CART	dataset_best_20	3155	1182	1973	37.46	69.52	55.94	88.95	87.85	58.54	41.63	68.35
KNN	dataset_admin	11689	2852	8837	24.4	71.11	42.15	82.76	49.59	78.05	26.08	45.57
KNN	dataset_all	3095	1125	1970	36.35	70.15	58.82	76.7	59.35	76.31	35.59	59.08
KNN	dataset_numeric	3095	1125	1970	36.35	68.32	55.21	78.83	67.66	68.7	34.67	60.8
KNN	dataset_performance	8156	2798	5358	34.31	69.01	53.42	83.66	75.45	65.65	37.41	62.55
KNN	dataset_psych_IQ	7286	2273	5013	31.2	49.68	32.73	70.74	58.15	45.84	3.29	41.88
KNN	dataset_psych_JPI_R	5506	1781	3725	32.35	59.12	39.33	72.32	48.69	64.1	12.04	43.51
KNN	dataset_psych_MMPI	5298	1740	3558	32.84	60.04	42.4	75.53	60.34	59.89	18.26	49.8
KNN	dataset_psych	3544	1182	2362	33.35	56.12	37.61	69.84	48.02	60.17	7.66	42.18
KNN	dataset_imputed_all	11885	2939	8946	24.73	74.61	48.83	84.97	56.64	80.51	35.25	52.44
KNN	dataset_imputed_admin	11885	2939	8946	24.73	73.71	47.03	83.32	50.4	81.36	31.02	48.66
KNN	dataset_imputed_performance	11885	2939	8946	24.73	76.04	51	91.25	77.98	75.4	45.33	61.67
KNN	dataset_imputed_psych	11885	2939	8946	24.73	65.38	36.11	81.6	52.1	69.74	19.01	42.66
KNN	dataset_imputed_psych_JPI_R	11885	2939	8946	24.73	65.01	36.01	81.83	53.46	68.8	19.15	43.03
KNN	dataset_imputed_psych_IQ	11885	2939	8946	24.73	60.77	31.31	79.46	49.15	64.59	11.53	38.25
KNN	dataset_log_all	3095	1125	1970	36.35	67.24	54.73	74.78	56.68	73.27	29.71	55.69
KNN	dataset_log_admin	11689	2852	8837	24.4	72.25	43.66	82.57	47.49	80.23	26.92	45.49
KNN	dataset_log_performance	8156	2798	5358	34.31	68.72	52.82	87.19	82.72	61.42	38.88	64.47
KNN	dataset_log_psych	3544	1182	2362	33.35	61.21	41.42	70.44	39.55	72.03	11.71	40.46
KNN	dataset_log_psych_JPI_R	5506	1781	3725	32.35	54.75	37.03	72.29	56.93	53.72	9.33	44.87
KNN	dataset_log_psych_IQ	7286	2273	5013	31.2	49.54	32.73	70.78	58.59	45.44	3.32	42
KNN	dataset_log_psych_MMPI	5298	1740	3558	32.84	61.04	42.6	73.98	53.45	64.76	17.1	47.41
KNN	dataset_binned_all	3095	1125	1970	36.35	66.38	52.63	80.79	74.18	61.93	33.19	61.58
KNN	dataset_binned_admin	11689	2852	8837	24.4	71.71	43	82.83	49.24	78.95	26.87	45.91
KNN	dataset_binned_performance	8156	2798	5358	34.31	69.42	53.77	84.67	77.35	65.28	38.59	63.44
KNN	dataset_binned_psych	3544	1182	2362	33.35	61.68	44.79	77.22	64.41	60.31	22.28	52.84
KNN	dataset_binned_psych_JPI_R	5506	1781	3725	32.35	64.2	45.38	75.44	52.43	69.83	21.4	48.65
KNN	dataset_binned_psych_MMPI	5298	1740	3558	32.84	62.05	45.14	80.69	72.03	57.17	25.34	55.5
KNN	dataset_all_pca_best	3095	1125	1970	36.35	71.66	59.84	79.71	66.77	74.45	40.22	63.11
KNN	dataset_numeric_pca_best	3095	1125	1970	36.35	68.32	54.29	84.78	80.71	61.25	37.99	64.92
KNN	dataset_performance_pca_best	8156	2798	5358	34.31	70.65	55.29	84.11	75.33	68.2	40.06	63.77
KNN	dataset_psych_pca_best	3544	1182	2362	33.35	56.03	40.57	76.03	68.64	49.72	15.65	51
KNN	dataset_psych_IQ_pca_best	7286	2273	5013	31.2	52.75	35.48	74.15	63	48.1	9.16	45.4
KNN	dataset_psych_JPI_R_pca_best	5506	1781	3725	32.35	57.36	38.42	72.52	52.81	59.53	11.24	44.48
KNN	dataset_psych_MMPI_pca_best	5298	1740	3558	32.84	59.41	41.68	74.82	59	59.61	16.82	48.85
KNN	dataset_imputed_performance_pca_best	11885	2939	8946	24.73	77.44	53.02	91.04	76.73	77.67	47.3	62.71
KNN	dataset_imputed_psych_pca_best	11885	2939	8946	24.73	66.02	36.69	81.67	51.65	70.74	19.69	42.9
KNN	dataset_imputed_psych_JPI_R_pca_best	11885	2939	8946	24.73	63.97	35.32	81.93	55.05	66.9	18.49	43.03
KNN	dataset_log_psych_pca_best	3544	1182	2362	33.35	57.44	41.28	75.6	65.54	53.39	16.5	50.66
KNN	dataset_log_psych_JPI_R_pca_best	5506	1781	3725	32.35	53.97	36.12	71.33	55.06	53.45	7.48	43.62
KNN	dataset_log_psych_IQ_pca_best	7286	2273	5013	31.2	48.99	32.58	70.67	59.47	44.24	3.03	42.1
KNN	dataset_log_psych_MMPI_pca_best	5298	1740	3558	32.84	58.78	40.4	72.99	53.64	61.29	13.78	46.09
KNN	dataset_binned_performance_pca_best	8156	2798	5358	34.31	69.75	54.21	84.16	76.04	66.46	38.77	63.29
KNN	dataset_binned_psych_pca_best	3544	1182	2362	33.35	58.95	42.98	78.45	70.9	52.97	20.53	53.52
KNN	dataset_binned_psych_JPI_R_pca_best	5506	1781	3725	32.35	65.17	46.54	75.61	51.69	71.62	22.65	48.98
KNN	dataset_binned_psych_IQ_pca_best	7286	2273	5013	31.2	47.48	33.66	73.48	70.48	37.06	5.81	45.56
KNN	dataset_binned_psych_MMPI_pca_best	5298	1740	3558	32.84	58.09	42.53	82.08	78.54	48.08	21.89	55.18
KNN	dataset_apft_minimum_all	1938	998	940	51.5	58.69	58.04	59.81	71.24	45.39	16.74	63.96
KNN	dataset_apft_minimum_admin	4888	2414	2474	49.39	62.35	61.53	63.19	63.4	61.32	24.71	62.45
KNN	dataset_apft_minimum_performance	4927	2470	2457	50.13	61.64	60.14	63.71	69.64	53.6	23.24	64.54
KNN	dataset_apft_minimum_psych	1994	1053	941	52.81	62.81	61.54	65.46	78.73	45.04	24.16	69.08
KNN	dataset_apft_minimum_psych_JPI_R	3026	1608	1418	53.14	56.01	57.19	53.94	68.46	41.88	10.49	62.32
KNN	dataset_apft_minimum_psych_IQ	3917	2028	1889	51.77	52.47	53.42	50.9	64.14	39.93	4.1	58.3
KNN	dataset_apft_minimum_psych_MMPI	2887	1521	1366	52.68	56.53	59.3	53.79	55.92	57.21	13.09	57.56
KNN	dataset_best_10	3295	1229	2066	37.3	70.92	57.89	85.02	80.71	65.11	42.42	67.42
KNN	dataset_best_20	3155	1182	1973	37.46	71.11	59.14	81.67	74.01	69.37	41.3	65.75

Table C.2 continued from previous page

Model	Data Subset	n	Selected	Not Selected	Percent Selected	Accuracy	PPV	NPV	Sensivitiy	Specificity	Kappa	F1
rf	dataset_admin	11689	2852	8837	24.4	71.51	43.01	83.36	51.81	77.86	27.75	47
rf	dataset_all	3095	1125	1970	36.35	76.19	63.81	86.42	79.53	74.28	51.1	70.81
rf	dataset_numeric	3095	1125	1970	36.35	74.46	61.21	86.72	81.01	70.73	48.37	69.73
rf	dataset_performance	8156	2798	5358	34.31	66.93	51.2	83.36	76.28	62.04	34.31	61.27
rf	dataset_psych	3544	1182	2362	33.35	62.62	44.36	72.83	47.74	70.06	17.46	45.99
rf	dataset_psych_IQ	7286	2273	5013	31.2	47.34	32.25	70.45	62.56	40.45	2.39	42.56
rf	dataset_psych_JPI_R	5506	1781	3725	32.35	61.84	41.43	72.32	43.45	70.64	13.9	42.41
rf	dataset_psych_MMPI	5298	1740	3558	32.84	62.43	44.15	74.79	54.21	66.45	19.53	48.67
rf	dataset_imputed_all	11885	2939	8946	24.73	80.61	58.53	90.65	74.01	82.78	52.16	65.36
rf	dataset_imputed_numeric	11885	2939	8946	24.73	79.41	57.16	88.41	66.63	83.6	47.58	61.53
rf	dataset_imputed_admin	11885	2939	8946	24.73	73.29	46.1	82.62	47.67	81.7	29.04	46.88
rf	dataset_imputed_performance	11885	2939	8946	24.73	75.76	50.64	90.77	76.62	75.48	44.44	60.98
rf	dataset_imputed_psych	11885	2939	8946	24.73	69.14	40.05	82.14	50.06	75.4	23.49	44.5
rf	dataset_imputed_psych_JPI_R	11885	2939	8946	24.73	64.87	36.43	82.57	56.53	67.61	20.36	44.31
rf	dataset_imputed_psych_IQ	11885	2939	8946	24.73	63.24	29.63	77.34	35.41	72.38	7.32	32.26
rf	dataset_imputed_psych_MMPI	11885	2939	8946	24.73	71.02	41.76	81.23	43.7	79.99	23.32	42.71
rf	dataset_log_all	3095	1125	1970	36.35	75.54	63.1	85.83	78.64	73.77	49.77	70.01
rf	dataset_log_numeric	3095	1125	1970	36.35	75.22	62.77	85.46	78.04	73.6	49.08	69.58
rf	dataset_log_admin	11689	2852	8837	24.4	72.73	44.2	82.16	45.03	81.67	26.53	44.61
rf	dataset_log_performance	8156	2798	5358	34.31	68.64	53.09	82.76	73.66	66.02	36.32	61.71
rf	dataset_log_psych	3544	1182	2362	33.35	63.28	45.16	73.04	47.46	71.19	18.41	46.28
rf	dataset_log_psych_JPI_R	5506	1781	3725	32.35	60.45	41.29	73.97	52.81	64.1	15.76	46.34
rf	dataset_log_psych_IQ	7286	2273	5013	31.2	52.56	33.7	71.32	53.89	51.96	5.03	41.47
rf	dataset_log_psych_MMPI	5298	1740	3558	32.84	61.99	43.61	74.45	53.64	66.07	18.62	48.11
rf	dataset_binned_all	3095	1125	1970	36.35	76.62	64.63	86.1	78.64	75.47	51.69	70.95
rf	dataset_binned_numeric	3095	1125	1970	36.35	75.32	63.57	84.15	75.07	75.47	48.65	68.84
rf	dataset_binned_admin	11689	2852	8837	24.4	68.88	40.53	84.51	59.06	72.05	26.94	48.07
rf	dataset_binned_performance	8156	2798	5358	34.31	68.15	52.44	84.1	77	63.53	36.45	62.39
rf	dataset_binned_psych	3544	1182	2362	33.35	66.38	49.64	77.13	58.19	70.48	27.49	53.58
rf	dataset_binned_psych_JPI_R	5506	1781	3725	32.35	68.14	50.73	76.74	51.87	75.92	27.63	51.3
rf	dataset_binned_psych_IQ	7286	2273	5013	31.2	46.79	33.61	73.78	72.39	35.2	5.77	45.9
rf	dataset_binned_psych_MMPI	5298	1740	3558	32.84	64.44	46.89	77.95	62.07	65.6	25.56	53.42
rf	dataset_all_pca_best	3095	1125	1970	36.35	75.22	68.26	78.43	59.35	84.26	44.87	63.49
rf	dataset_numeric_pca_best	3095	1125	1970	36.35	71.77	59.89	79.96	67.36	74.28	40.55	63.41
rf	dataset_performance_pca_best	8156	2798	5358	34.31	66.23	50.51	83.69	77.47	60.36	33.55	61.15
rf	dataset_psych_pca_best	3544	1182	2362	33.35	61.39	43.64	73.95	54.24	64.97	18.11	48.36
rf	dataset_psych_IQ_pca_best	7286	2273	5013	31.2	49.73	32.84	70.9	58.59	45.71	3.54	42.09
rf	dataset_psych_JPI_R_pca_best	5506	1781	3725	32.35	59.3	40.03	73.2	51.87	62.85	13.66	45.19
rf	dataset_psych_MMPI_pca_best	5298	1740	3558	32.84	62.68	42.16	70.86	36.59	75.45	12.46	39.18
rf	dataset_imputed_performance_pca_best	11885	2939	8946	24.73	77.02	52.37	91.28	77.64	76.82	46.86	62.55
rf	dataset_imputed_psych_pca_best	11885	2939	8946	24.73	65.43	36.04	81.45	51.42	70.03	18.76	42.38
rf	dataset_imputed_psych_JPI_R_pca_best	11885	2939	8946	24.73	64.45	35.72	82.01	54.82	67.61	19.02	43.26
rf	dataset_imputed_psych_IQ_pca_best	11885	2939	8946	24.73	63.19	29.53	77.3	35.3	72.34	7.17	32.16
rf	dataset_imputed_psych_MMPI_pca_best	11885	2939	8946	24.73	63.86	36.03	83.1	59.59	65.26	20.38	44.91
rf	dataset_log_performance_pca_best	8156	2798	5358	34.31	63.78	48.11	79.92	71.16	59.93	27.89	57.4
rf	dataset_log_psych_pca_best	3544	1182	2362	33.35	62.05	44.26	74.02	53.39	66.38	18.8	48.4
rf	dataset_log_psych_JPI_R_pca_best	5506	1781	3725	32.35	63.48	41.48	70.63	31.46	78.78	10.93	35.78
rf	dataset_log_psych_IQ_pca_best	7286	2273	5013	31.2	56.09	33.72	70.45	42.29	62.34	4.33	37.52
rf	dataset_log_psych_MMPI_pca_best	5298	1740	3558	32.84	61.36	42.01	72.36	46.36	68.7	14.67	44.08
rf	dataset_binned_performance_pca_best	8156	2798	5358	34.31	67.54	51.81	83.85	76.88	62.66	35.44	61.9
rf	dataset_binned_psych_pca_best	3544	1182	2362	33.35	66.48	49.78	79.33	64.97	67.23	29.92	56.37
rf	dataset_binned_psych_JPI_R_pca_best	5506	1781	3725	32.35	61.48	42.76	75.4	56.37	63.92	18.73	48.63
rf	dataset_binned_psych_IQ_pca_best	7286	2273	5013	31.2	46.98	33.61	73.66	71.81	35.73	5.75	45.79
rf	dataset_binned_psych_MMPI_pca_best	5298	1740	3558	32.84	58.9	41.74	76.01	63.41	56.7	17.76	50.34
rf	dataset_apft_minimum_all	1938	998	940	51.5	65.58	65.81	65.3	68.9	62.06	31	67.32
rf	dataset_apft_minimum_numeric	1938	998	940	51.5	62.13	61.38	63.25	71.24	52.48	23.84	65.94
rf	dataset_apft_minimum_admin	4888	2414	2474	49.39	62.82	62.45	63.19	62.02	63.61	25.63	62.23
rf	dataset_apft_minimum_performance	4927	2470	2457	50.13	59.54	57.92	62.09	70.58	48.44	19.03	63.63
rf	dataset_apft_minimum_psych	1994	1053	941	52.81	60.3	61.54	58.69	66.03	53.9	20.02	63.71
rf	dataset_apft_minimum_psych_JPI_R	3026	1608	1418	53.14	54.69	55.85	52.33	70.33	36.94	7.4	62.26
rf	dataset_apft_minimum_psych_IQ	3917	2028	1889	51.77	52.39	52.82	51.15	75.49	27.56	3.1	62.15
rf	dataset_apft_minimum_psych_MMPI	2887	1521	1366	52.68	58.61	60.08	56.73	64.04	52.57	16.67	62
rf	dataset_best_10	3295	1229	2066	37.3	70.72	57.52	85.71	82.07	63.97	42.37	67.64
rf	dataset_best_20	3155	1182	1973	37.46	71.96	59.21	85.28	80.79	66.67	44.23	68.34
xgb	dataset_admin	11689	2852	8837	24.4	72.22	43.8	82.92	49.12	79.67	27.65	46.31
xgb	dataset_all	3095	1125	1970	36.35	74.25	61.5	85.06	77.74	72.25	47.31	68.68
xgb	dataset_numeric	3095	1125	1970	36.35	73.17	59.87	85.48	79.23	69.71	45.76	68.2

Table C.2 continued from previous page

Model	Data Subset	n	Selected	Not Selected	Percent Selected	Accuracy	PPV	NPV	Sensivitiy	Specificity	Kappa	F1
xgb	dataset_performance	8156	2798	5358	34.31	71.38	56.41	83.32	72.94	70.57	40.66	63.62
xgb	dataset_psych	3544	1182	2362	33.35	62.43	44.55	73.81	51.98	67.66	18.85	47.98
xgb	dataset_psych_IQ	7286	2273	5013	31.2	52.7	31.55	69.11	44.2	56.55	0.68	36.82
xgb	dataset_psych_JPI_R	5506	1781	3725	32.35	61.42	40.95	72.18	43.63	69.92	13.32	42.25
xgb	dataset_psych_MMPI	5298	1740	3558	32.84	58.97	41.4	74.91	59.96	58.48	16.55	48.98
xgb	dataset_imputed_all	11885	2939	8946	24.73	79.21	55.56	92.14	79.46	79.13	51.19	65.39
xgb	dataset_imputed_numeric	11885	2939	8946	24.73	77.89	53.77	90.65	75.26	78.76	47.62	62.72
xgb	dataset_imputed_admin	11885	2939	8946	24.73	74.05	47.57	83.07	48.92	82.3	30.92	48.24
xgb	dataset_imputed_performance	11885	2939	8946	24.73	75.76	50.58	93.17	83.65	73.16	46.59	63.05
xgb	dataset_imputed_psych	11885	2939	8946	24.73	67.06	38.33	82.68	54.6	71.15	22.54	45.04
xgb	dataset_imputed_psych_JPI_R	11885	2939	8946	24.73	64.59	36.44	82.91	58.12	66.72	20.7	44.79
xgb	dataset_imputed_psych_IQ	11885	2939	8946	24.73	59.79	32.01	80.79	55.73	61.13	13.5	40.66
xgb	dataset_imputed_psych_MMPI	11885	2939	8946	24.73	66.89	37.88	82.24	53.01	71.45	21.56	44.18
xgb	dataset_log_all	3095	1125	1970	36.35	74.25	61.5	85.06	77.74	72.25	47.31	68.68
xgb	dataset_log_numeric	3095	1125	1970	36.35	73.17	59.87	85.48	79.23	69.71	45.76	68.2
xgb	dataset_log_admin	11689	2852	8837	24.4	72.22	43.8	82.92	49.12	79.67	27.65	46.31
xgb	dataset_log_performance	8156	2798	5358	34.31	71.38	56.41	83.32	72.94	70.57	40.66	63.62
xgb	dataset_log_psych	3544	1182	2362	33.35	62.43	44.55	73.81	51.98	67.66	18.85	47.98
xgb	dataset_log_psych_JPI_R	5506	1781	3725	32.35	61.42	40.95	72.18	43.63	69.92	13.32	42.25
xgb	dataset_log_psych_IQ	7286	2273	5013	31.2	52.7	31.55	69.11	44.2	56.55	0.68	36.82
xgb	dataset_log_psych_MMPI	5298	1740	3558	32.84	58.97	41.4	74.91	59.96	58.48	16.55	48.98
xgb	dataset_binned_all	3095	1125	1970	36.35	75.11	63.18	84.22	75.37	74.96	48.33	68.74
xgb	dataset_binned_numeric	3095	1125	1970	36.35	74.46	61.68	85.4	78.34	72.25	47.81	69.02
xgb	dataset_binned_admin	11689	2852	8837	24.4	72.5	44.43	83.38	50.88	79.48	28.94	47.44
xgb	dataset_binned_performance	8156	2798	5358	34.31	71.95	57.54	82.18	69.61	73.18	40.74	63
xgb	dataset_binned_psych	3544	1182	2362	33.35	64.78	47.96	79.2	66.38	63.98	27.71	55.69
xgb	dataset_binned_psych_JPI_R	5506	1781	3725	32.35	66.69	48.65	76.77	53.93	72.78	25.98	51.15
xgb	dataset_binned_psych_IQ	7286	2273	5013	31.2	47.34	33.61	73.44	70.63	36.79	5.71	45.55
xgb	dataset_binned_psych_MMPI	5298	1740	3558	32.84	63	46.14	82.59	75.48	56.89	27.84	57.27
xgb	dataset_all_pca_best	3095	1125	1970	36.35	72.95	59.95	84.27	76.85	70.73	44.86	67.36
xgb	dataset_numeric_pca_best	3095	1125	1970	36.35	73.28	61.38	81.94	71.22	74.45	44.15	65.93
xgb	dataset_performance_pca_best	8156	2798	5358	34.31	70.07	54.59	84.15	75.8	67.08	39.24	63.47
xgb	dataset_psych_pca_best	3544	1182	2362	33.35	58.76	41.6	74.02	58.76	58.76	15.88	48.71
xgb	dataset_psych_IQ_pca_best	7286	2273	5013	31.2	47.8	33.96	74.1	71.37	37.13	6.53	46.02
xgb	dataset_psych_JPI_R_pca_best	5506	1781	3725	32.35	55	37.05	72.16	55.99	54.52	9.28	44.59
xgb	dataset_psych_MMPI_pca_best	5298	1740	3558	32.84	60.04	42.54	75.96	61.69	59.23	18.76	50.35
xgb	dataset_imputed_performance_pca_best	11885	2939	8946	24.73	76.54	51.59	92.89	82.63	74.54	47.57	63.53
xgb	dataset_imputed_psych_pca_best	11885	2939	8946	24.73	64.37	36.46	83.22	59.48	65.97	21	45.21
xgb	dataset_imputed_psych_JPI_R_pca_best	11885	2939	8946	24.73	63.41	35.66	83.01	59.7	64.63	19.84	44.65
xgb	dataset_imputed_psych_IQ_pca_best	11885	2939	8946	24.73	58.73	31.54	80.82	57.21	59.22	12.91	40.66
xgb	dataset_imputed_psych_MMPI_pca_best	11885	2939	8946	24.73	68.6	38.73	81.18	46.42	75.89	20.91	42.23
xgb	dataset_log_performance_pca_best	8156	2798	5358	34.31	70.03	54.53	84.25	76.04	66.89	39.24	63.51
xgb	dataset_log_psych_pca_best	3544	1182	2362	33.35	61.39	43.4	73.35	51.98	66.1	17.23	47.3
xgb	dataset_log_psych_JPI_R_pca_best	5506	1781	3725	32.35	56.87	37.87	72.08	52.06	59.18	10.24	43.85
xgb	dataset_log_psych_IQ_pca_best	7286	2273	5013	31.2	53.16	33.71	71.16	51.98	53.69	4.94	40.9
xgb	dataset_log_psych_MMPI_pca_best	5298	1740	3558	32.84	63.06	44.21	73.35	47.51	70.67	17.83	45.8
xgb	dataset_binned_performance_pca_best	8156	2798	5358	34.31	68.89	53.23	84.17	76.64	64.84	37.54	62.82
xgb	dataset_binned_psych_pca_best	3544	1182	2362	33.35	64.03	47.2	79	66.67	62.71	26.63	55.27
xgb	dataset_binned_psych_JPI_R_pca_best	5506	1781	3725	32.35	64.08	45.23	75.39	52.43	69.65	21.2	48.57
xgb	dataset_binned_psych_IQ_pca_best	7286	2273	5013	31.2	48.31	33.63	72.92	67.55	39.59	5.59	44.9
xgb	dataset_binned_psych_MMPI_pca_best	5298	1740	3558	32.84	64.07	46.24	76.44	57.66	67.2	23.38	51.32
xgb	dataset_apft_minimum_all	1938	998	940	51.5	65.23	65.59	64.81	68.23	62.06	30.32	66.89
xgb	dataset_apft_minimum_numeric	1938	998	940	51.5	63.51	63.22	63.89	69.57	57.09	26.74	66.24
xgb	dataset_apft_minimum_admin	4888	2414	2474	49.39	63.71	62.9	64.54	64.64	62.8	27.43	63.76
xgb	dataset_apft_minimum_performance	4927	2470	2457	50.13	62.92	61.21	65.32	71.12	54.68	25.81	65.79
xgb	dataset_apft_minimum_psych	1994	1053	941	52.81	57.45	59.33	55.19	61.59	52.84	14.46	60.44
xgb	dataset_apft_minimum_psych_JPI_R	3026	1608	1418	53.14	53.8	56.21	50.75	59.13	47.76	6.92	57.63
xgb	dataset_apft_minimum_psych_IQ	3917	2028	1889	51.77	55.28	56.01	54.24	63.65	46.29	9.99	59.58
xgb	dataset_apft_minimum_psych_MMPI	2887	1521	1366	52.68	58.38	60.3	56.14	61.62	54.77	16.41	60.95
xgb	dataset_best_10	3295	1229	2066	37.3	71.43	58.57	84.74	79.89	66.4	43.11	67.59
xgb	dataset_best_20	3155	1182	1973	37.46	71.32	58.56	84.78	80.23	65.99	43.02	67.7
stack_glm	dataset_numeric	3095	1125	1970	36.35	74.46	61.85	84.98	77.45	72.76	47.62	68.77
stack_glm	dataset_performance	8156	2798	5358	34.31	69.75	54.03	85.56	79.02	64.9	39.56	64.18
stack_glm	dataset_psych_IQ	7286	2273	5013	31.2	51.1	33.53	71.56	57.86	48.04	4.92	42.46
stack_glm	dataset_psych_JPI_R	5506	1781	3725	32.35	58.33	39.73	73.81	55.81	59.53	13.88	46.42
stack_glm	dataset_psych_MMPI	5298	1740	3558	32.84	59.35	42.09	76.15	63.22	57.45	18.32	50.54
stack_glm	dataset_psych	3544	1182	2362	33.35	61.11	44.32	77.16	64.97	59.18	21.63	52.69

Table C.2 continued from previous page

Model	Data Subset	n	Selected	Not Selected	Percent Selected	Accuracy	PPV	NPV	Sensivitiy	Specificity	Kappa	F1
stack_glm	dataset_imputed_performance	11885	2939	8946	24.73	75.9	50.75	93.47	84.45	73.09	47.05	63.4
stack_glm	dataset_imputed_psych	11885	2939	8946	24.73	64.39	37.17	84.45	63.79	64.59	22.88	46.97
stack_glm	dataset_log_performance	8156	2798	5358	34.31	70.11	54.5	85.15	77.95	66.02	39.87	64.15
stack_glm	dataset_log_psych	3544	1182	2362	33.35	60.92	44.1	76.88	64.41	59.18	21.15	52.35
stack_glm	dataset_log_psych_JPI_R	5506	1781	3725	32.35	57.78	40.12	75.42	61.99	55.77	15.55	48.71
stack_glm	dataset_log_psych_IQ	7286	2273	5013	31.2	51.97	34.62	72.98	60.79	47.97	7.26	44.11
stack_glm	dataset_log_psych_MMPI	5298	1740	3558	32.84	59.6	42.37	76.46	63.79	57.54	18.9	50.92
stack_glm	dataset_binned_performance	8156	2798	5358	34.31	70.28	54.74	84.81	77.12	66.71	39.92	64.03
stack_glm	dataset_binned_psych	3544	1182	2362	33.35	65.16	48.46	81.18	71.19	62.15	29.84	57.67
stack_glm	dataset_binned_psych_JPI_R	5506	1781	3725	32.35	66.02	48.06	79.14	62.73	67.59	28.09	54.43
stack_glm	dataset_binned_psych_MMPI	5298	1740	3558	32.84	62.87	45.8	80.62	71.07	58.86	26.23	55.71
stack_glm	dataset_numeric_pca_best	3095	1125	1970	36.35	72.95	60.14	83.73	75.67	71.4	44.6	67.02
stack_glm	dataset_performance_pca_best	8156	2798	5358	34.31	68.15	52.23	87.64	83.79	59.99	38.25	64.35
stack_glm	dataset_psych_pca_best	3544	1182	2362	33.35	59.89	43.36	77.12	66.38	56.64	20.32	52.46
stack_glm	dataset_psych_JPI_R_pca_best	5506	1781	3725	32.35	58.51	40.57	75.41	60.86	57.39	16.14	48.69
stack_glm	dataset_psych_MMPI_pca_best	5298	1740	3558	32.84	59.35	42.84	79.11	71.07	53.61	21.13	53.46
stack_glm	dataset_imputed_performance_pca_best	11885	2939	8946	24.73	76.18	51.09	93.83	85.36	73.16	47.76	63.92
stack_glm	dataset_imputed_psych_pca_best	11885	2939	8946	24.73	64.23	37.04	84.44	63.9	64.33	22.71	46.9
stack_glm	dataset_log_psych_pca_best	3544	1182	2362	33.35	59.98	43.26	76.45	64.41	57.77	19.76	51.76
stack_glm	dataset_log_psych_JPI_R_pca_best	5506	1781	3725	32.35	58.87	41.34	76.9	64.79	56.04	18.15	50.47
stack_glm	dataset_log_psych_MMPI_pca_best	5298	1740	3558	32.84	57.58	41.4	77.87	70.11	51.45	18.32	52.06
stack_glm	dataset_binned_psych_pca_best	3544	1182	2362	33.35	64.69	47.87	79.26	66.67	63.7	27.65	55.73
stack_glm	dataset_binned_psych_JPI_R_pca_best	5506	1781	3725	32.35	62.99	44.8	77.8	62.17	63.38	23.21	52.08
stack_glm	dataset_binned_psych_MMPI_pca_best	5298	1740	3558	32.84	60.35	43.88	80.91	74.14	53.61	23.59	55.13
stack_glm	dataset_apft_minimum_psych	1994	1053	941	52.81	60.8	63.28	58.22	61.27	60.28	21.51	62.26
stack_glm	dataset_apft_minimum_psych_JPI_R	3026	1608	1418	53.14	55.79	58.32	52.86	58.92	52.24	11.16	58.62
stack_glm	dataset_apft_minimum_psych_IQ	3917	2028	1889	51.77	55.62	56.31	54.64	63.82	46.82	10.69	59.83
stack_glm	dataset_apft_minimum_psych_MMPI	2887	1521	1366	52.68	59.08	61.18	56.72	61.18	56.72	17.91	61.18
stack_rf	dataset_numeric	3095	1125	1970	36.35	72.31	61.05	79.51	65.58	76.14	41.07	63.23
stack_rf	dataset_performance	8156	2798	5358	34.31	70.77	55.98	81.63	69.13	71.62	38.59	61.87
stack_rf	dataset_psych_IQ	7286	2273	5013	31.2	63.78	36.39	70	21.59	82.9	5.05	27.1
stack_rf	dataset_psych_JPI_R	5506	1781	3725	32.35	64.75	43.65	71.01	30.9	80.93	12.81	36.18
stack_rf	dataset_psych_MMPI	5298	1740	3558	32.84	63	42.73	71.1	37.16	75.63	13.24	39.75
stack_rf	dataset_psych	3544	1182	2362	33.35	64.03	45.36	71.45	38.7	76.69	15.98	41.77
stack_rf	dataset_imputed_performance	11885	2939	8946	24.73	77.78	54.43	86.95	62.09	82.93	42.99	58.01
stack_rf	dataset_imputed_psych	11885	2939	8946	24.73	73.15	43.54	79	29.06	87.63	18.78	34.85
stack_rf	dataset_log_performance	8156	2798	5358	34.31	71.18	56.43	82.12	70.08	71.75	39.55	62.52
stack_rf	dataset_log_psych	3544	1182	2362	33.35	65.35	47.71	72.49	41.24	77.4	19.3	44.24
stack_rf	dataset_log_psych_JPI_R	5506	1781	3725	32.35	63.17	41.06	70.57	31.84	78.16	10.61	35.86
stack_rf	dataset_log_psych_IQ	7286	2273	5013	31.2	62.45	33.72	69.44	21.15	81.17	2.58	25.99
stack_rf	dataset_log_psych_MMPI	5298	1740	3558	32.84	64	44.39	71.65	37.93	76.76	15.26	40.91
stack_rf	dataset_binned_performance	8156	2798	5358	34.31	71.3	57.05	80.68	66.03	74.05	38.63	61.22
stack_rf	dataset_binned_psych	3544	1182	2362	33.35	63.84	45.86	73.14	46.89	72.32	19.1	46.37
stack_rf	dataset_binned_psych_JPI_R	5506	1781	3725	32.35	68.5	51.54	74.98	44.01	80.21	25.18	47.47
stack_rf	dataset_binned_psych_MMPI	5298	1740	3558	32.84	64.38	45.94	73.93	47.7	72.54	20.04	46.8
stack_rf	dataset_numeric_pca_best	3095	1125	1970	36.35	73.49	63.11	79.69	64.99	78.34	43.05	64.04
stack_rf	dataset_performance_pca_best	8156	2798	5358	34.31	68.77	53.89	78.46	61.98	72.31	33.1	57.65
stack_rf	dataset_psych_pca_best	3544	1182	2362	33.35	63.47	44.55	71.33	39.27	75.56	15.28	41.74
stack_rf	dataset_psych_JPI_R_pca_best	5506	1781	3725	32.35	61.24	34.77	68.3	22.66	79.68	2.57	27.44
stack_rf	dataset_psych_MMPI_pca_best	5298	1740	3558	32.84	64.95	45.5	71.25	33.91	80.13	15.02	38.86
stack_rf	dataset_imputed_performance_pca_best	11885	2939	8946	24.73	77.41	53.88	86.34	59.93	83.15	41.52	56.74
stack_rf	dataset_imputed_psych_pca_best	11885	2939	8946	24.73	72.31	40.64	78.29	26.11	87.48	15.44	31.79
stack_rf	dataset_log_psych_pca_best	3544	1182	2362	33.35	63.84	45.19	71.6	39.83	75.85	16.16	42.34
stack_rf	dataset_log_psych_JPI_R_pca_best	5506	1781	3725	32.35	62.51	39.45	69.95	29.78	78.16	8.47	33.94
stack_rf	dataset_log_psych_MMPI_pca_best	5298	1740	3558	32.84	64.51	45.33	72.08	39.08	76.94	16.61	41.98
stack_rf	dataset_binned_psych_pca_best	3544	1182	2362	33.35	67.23	51.01	73.5	42.66	79.52	23.12	46.46
stack_rf	dataset_binned_psych_JPI_R_pca_best	5506	1781	3725	32.35	65.66	46.37	72.99	39.51	78.16	18.38	42.67
stack_rf	dataset_binned_psych_MMPI_pca_best	5298	1740	3558	32.84	62.81	43.55	72.58	44.64	71.7	16.23	44.09
stack_rf	dataset_apft_minimum_psych	1994	1053	941	52.81	59.97	61.18	58.37	66.03	53.19	19.31	63.51
stack_rf	dataset_apft_minimum_psych_JPI_R	3026	1608	1418	53.14	50.28	52.91	46.52	58.51	40.94	-0.56	55.57
stack_rf	dataset_apft_minimum_psych_IQ	3917	2028	1889	51.77	55.37	56.95	53.68	56.58	54.06	10.64	56.77
stack_rf	dataset_apft_minimum_psych_MMPI	2887	1521	1366	52.68	58.15	60.31	55.75	60.31	55.75	16.05	60.31

Bibliography

[1] Vasant Dhar. "Data Science and Prediction". In: *Communications of the ACM* 56.12 (Dec. 2013), pp. 64–73. ISSN: 0001-0782, 1557-7317. DOI: 10.1145/2500499.

[2] Longbing Cao. "Data Science: A Comprehensive Overview". In: *ACM Computing Surveys* 50.3 (May 31, 2018), pp. 1–42. ISSN: 0360-0300, 1557-7341. DOI: 10.1145/3076253.

[3] Silvia Chiusano et al. "Breakthroughs on Cross-Cutting Data Management, Data Analytics, and Applied Data Science". In: *Information Systems Frontiers* 23.1 (Feb. 1, 2021), pp. 1–7. ISSN: 1572-9419. DOI: 10.1007/s10796-020-10091-8.

[4] *United States Special Operations Command (USSOCOM)*. URL: https://www.socom.mil (visited on 03/11/2022).

[5] *U.S. Navy SEAL Careers | Navy.Com*. URL: https://www.navy.com/seals (visited on 03/11/2022).

[6] *U.S. Army Special Operations Center of Excellence*. URL: https://www.soc.mil/SWCS/about.html (visited on 03/11/2022).

[7] U.S. Office of Personnel Management. Human Resources Solutions. *75th Ranger Regiment Job Analysis Technical Report*. Apr. 2015.

[8] *Marine Special Operations Forces Publication 1*. May 12, 2011.

[9] Paul Geehreng. "Turning Airmen Into Air Commandos". In: *Air Commando Journal* 1.1 (Aut. 2011), pp. 44–48.

[10] Emily K. Farina et al. "Physical Performance, Demographic, Psychological, and Physiological Predictors of Success in the U.S. Army Special Forces Assessment and Selection Course". In: *Physiology & Behavior* 210 (Oct. 15, 2019), p. 112647. ISSN: 0031-9384. DOI: 10.1016/j.physbeh.2019.112647.

[11] Andrew P. Hunt, Robin M. Orr, and Daniel C. Billing. "Developing Physical Capability Standards That Are Predictive of Success on Special Forces Selection Courses". In: *Military Medicine* 178.6 (June 1, 2013), pp. 619–624. ISSN: 0026-4075. DOI: 10.7205/MILMED-D-12-00347.

[12] Martha L. Teplitzky. *Physical Performance Predictors of Success in Special Forces Assessment and Selection*. U.S. Army Research Institute for the Behavioral and Social Sciences, 1991. 40 pp. Google Books: aaUrAAAAYAAJ.

[13] Michelle Zazanis et al. *Prescreening Methods for Special Forces Assessment and Selection*. U.S. Army Research Institute for the Behavioral and Social Sciences, 1999. 46 pp. Google Books: dK8rAAAAYAAJ.

[14] Scott A. Beal. *The Roles of Perseverance, Cognitive Ability, and Physical Fitness in U.S. Army Special Forces Assessment and Selection*. ARMY RESEARCH INST FOR THE BEHAVIORAL AND SOCIAL SCIENCES FORT BRAGG NC SCIENTIFIC COORDINATION OFFICE, July 1, 2010.

[15] Marcus K. Taylor et al. *Predictors of Success in Basic Underwater Demolition/SEAL (BUD/S) Training - Part 1: What Do We Know and Where Do We Go From Here?* NAVAL HEALTH RESEARCH CENTER SAN DIEGO CA, Oct. 30, 2006.

[16] Aaron Coombs. *Modeling Attrition in a Military Selection Context*. Masters Thesis. Blacksburg, VA: Virginia Polytechnic Institute and State University, Oct. 2, 2020.

[17] A. J. Allsopp and A. Shariff. “Improving the Selection of Candidates for Royal Marine Recruit Training by the Use of a Combination of Performance Tests”. In: *Journal of the Royal Naval Medical Service* 90.3 (2004), pp. 117–124. ISSN: 0035-9033. pmid: 15745257.

[18] D. G. McDonald, J. P. Norton, and J. A. Hodgdon. “Training Success in U.S. Navy Special Forces”. In: *Aviation, Space, and Environmental Medicine* 61.6 (June 1990), pp. 548–554. ISSN: 0095-6562. pmid: 2369395.

[19] Judith E. Brooks and Michelle M. Zazanis. *Enhancing U.S. Army Special Forces: Research and Applications.* ARMY RESEARCH INST FOR THE BEHAVIORAL AND SOCIAL SCIENCES ALEXANDRIA VA, Oct. 1, 1997.

[20] Jeffrey Peter Stolrow. “The Assessment and Selection of Special Forces Qualification Course Candidates with the MMPI”. PhD thesis. United States – California: California School of Professional Psychology - Los Angeles. 235 pp.

[21] Paul T. Bartone et al. “Psychological Hardiness Predicts Success in US Army Special Forces Candidates”. In: *International Journal of Selection and Assessment* 16.1 (2008), pp. 78–81. ISSN: 1468-2389. DOI: 10.1111/j.1468-2389.2008.00412.x.

[22] James Picano et al. “Assessment of Elite Operational Personnel”. In: *Handbook of Military Psychology: Clinical and Organizational Practice*. Dec. 8, 2017, pp. 277–289. ISBN: 978-3-319-66190-2. DOI: 10.1007/978-3-319-66192-6_17.

[23] Anthony D. Smith. *Predicting Ranger Assessment and Selection Program 1 Success and Optimizing Class Composition*. Naval Postgraduate School Monterey United States, June 1, 2017.

[24] Max Kuhn and Kjell Johnson. *Feature Engineering and Selection: A Practical Approach for Predictive Models*.

[25] Max Kuhn. *Caret: Classification and Regression Training*. manual. 2021.

[26] Bradley Boehmke & Brandon Greenwell. *Hands-On Machine Learning with R*.

[27] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. “Regularization Paths for Generalized Linear Models via Coordinate Descent”. In: *Journal of Statistical Software* 33.1 (2010), pp. 1–22.

[28] Trevor Hastie, Janyang Quian, and Kenneth Tay. *An Introduction to ‘glmnet‘*. URL: https://glmnet.stanford.edu/articles/glmnet.html (visited on 02/25/2022).

[29] W. N. Venables and B. D. Ripley. *Modern Applied Statistics with s*. 4th ed. New York: Springer, 2002.

[30] Alexandros Karatzoglou et al. “Kernlab – an S4 Package for Kernel Methods in R”. In: *Journal of Statistical Software* 11.9 (2004), pp. 1–20.

[31] Terry Therneau and Beth Atkinson. *Rpart: Recursive Partitioning and Regression Trees*. manual. 2019.

[32] Gareth James et al. *An Introduction to Statistical Learning: With Applications in R*. Second Edition. Springer, 2021.

[33] Alina Beygelzimer et al. *FNN: Fast Nearest Neighbor Search Algorithms and Applications*. manual. 2019.

[34] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*. Second Edition. Springer Series in Statistics. New York, NY, USA: Springer New York Inc., 2017.

[35] Andy Liaw and Matthew Wiener. "Classification and Regression by randomForest". In: *R News* 2.3 (2002), pp. 18–22.

[36] Tianqi Chen et al. *Xgboost: Extreme Gradient Boosting*. manual. 2021.

[37] Zachary A. Deane-Mayer and Jared E. Knowles. *caretEnsemble: Ensembles of Caret Models*. manual. 2019.

[38] Max Kuhn and Kjell Johnson. *Applied Predictive Modeling*. 2013. URL: http://www.amazon.com/Applied-Predictive-Modeling-Max-Kuhn/dp/1461468485/.

[39] Piotr Romanski, Lars Kotthoff, and Patrick Schratz. *FSelector: Selecting Attributes*. manual. 2021.

[40] Gilles Kratzer and Reinhard Furrer. *Varrank: An R Package for Variable Ranking Based on Mutual Information with Applications to Observed Systemic Datasets*. 2018. arXiv: 1804.07134.