2019-12

USING MACHINE LEARNING TO PREDICT EARLY SERVICE SEPARATION OF TECHNICAL AND NON-TECHNICAL SAILORS

Cole, Stephen

Monterey, CA; Naval Postgraduate School

NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

USING MACHINE LEARNING TO PREDICT EARLY SERVICE SEPARATION OF TECHNICAL AND NON-TECHNICAL SAILORS

by

Stephen Cole

December 2019

Thesis Advisor: Sae Young Ahn
Co-Advisor: James J. Fan

Approved for public release. Distribution is unlimited.

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE			*Form Approved OMB No. 0704-0188*
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington, DC 20503.			
1. AGENCY USE ONLY ***(Leave blank)***	**2. REPORT DATE** December 2019	**3. REPORT TYPE AND DATES COVERED** Master's thesis	
4. TITLE AND SUBTITLE USING MACHINE LEARNING TO PREDICT EARLY SERVICE SEPARATION OF TECHNICAL AND NON-TECHNICAL SAILORS			**5. FUNDING NUMBERS**
6. AUTHOR(S) Stephen Cole			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000			**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) N/A			**10. SPONSORING / MONITORING AGENCY REPORT NUMBER**
11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.			
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release. Distribution is unlimited.			**12b. DISTRIBUTION CODE** A

13. ABSTRACT (maximum 200 words)

Sailors are difficult to recruit, expensive to train, and hard to retain. This is particularly true in the technical sailor community. Retention of both technical and non-technical sailors is critical to future manning continuity and capability within the Royal Australian Navy. This research employs machine learning to analyze Royal Australian Navy exit survey data collected between 1999 and 2018 to better predict the attitudes and behaviors of a sailor voluntarily separating between four and eight years of service. Furthermore, this study analyzes in particular whether technical sailors behave differently compared to non-technical sailors.

In comparison to traditional modeling techniques, the analysis finds that machine learning can more accurately detect differences in the attitudes and behaviors of technical and non-technical sailors when they are deciding to voluntarily separate from service. Furthermore, the analysis can identify differences in sentiment across periods of time covering key career milestones. This analysis and its findings may now be employed to analyze specific critical target groups in both the Royal Australian Navy technical and non-technical sailor communities to understand their attitudes and behaviors, and help support current and future sailor retention policy initiatives.

14. SUBJECT TERMS machine learning, regression, retention, sailor, technical sailor, non-technical sailor			**15. NUMBER OF PAGES** 111
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT Unclassified	**18. SECURITY CLASSIFICATION OF THIS PAGE** Unclassified	**19. SECURITY CLASSIFICATION OF ABSTRACT** Unclassified	**20. LIMITATION OF ABSTRACT** UU

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release. Distribution is unlimited.

USING MACHINE LEARNING TO PREDICT EARLY SERVICE SEPARATION OF TECHNICAL AND NON-TECHNICAL SAILORS

Stephen Cole
Lieutenant Commander, Royal Australian Navy
BSE, University of Portsmouth, UK, 2004
MSE, Kingston University, UK, 2009

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN MANAGEMENT

from the

NAVAL POSTGRADUATE SCHOOL
December 2019

Approved by: Sae Young Ahn
Advisor

James J. Fan
Co-Advisor

Marigee Bacolod
Academic Associate, Graduate School of Defense Management

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

Sailors are difficult to recruit, expensive to train, and hard to retain. This is particularly true in the technical sailor community. Retention of both technical and non-technical sailors is critical to future manning continuity and capability within the Royal Australian Navy. This research employs machine learning to analyze Royal Australian Navy exit survey data collected between 1999 and 2018 to better predict the attitudes and behaviors of a sailor voluntarily separating between four and eight years of service. Furthermore, this study analyzes in particular whether technical sailors behave differently compared to non-technical sailors.

In comparison to traditional modeling techniques, the analysis finds that machine learning can more accurately detect differences in the attitudes and behaviors of technical and non-technical sailors when they are deciding to voluntarily separate from service. Furthermore, the analysis can identify differences in sentiment across periods of time covering key career milestones. This analysis and its findings may now be employed to analyze specific critical target groups in both the Royal Australian Navy technical and non-technical sailor communities to understand their attitudes and behaviors, and help support current and future sailor retention policy initiatives.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

LIST OF FIGURES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF ACRONYMS AND ABBREVIATIONS.

AB	Able Seaman (E3)
ADF	Australian Defence Force
AQF	Australian Qualification Framework
ARA	Australian Regular Army
ATA	Aviation Technician Aircraft
ATV	Aviation Technician Avionics
CPO	Chief Petty Officer (E8)
CSO	Combat Systems Operator
CHAID	Chi-Square Automatic Interaction Detection
DCN	Deputy Chief of Navy
DFR	Defence Force Recruiting
DFRT	Defence Force Remuneration Tribunal
DNWR	Director Navy Workforce Requirements
DSPRR	Directorate of Strategic Personnel Planning & Research
ET	Electronics Technician
GE	General Entry
HR	Human Resources
IMPS	Initial Minimum Period of Service
KNN	K-Nearest Neighbor
LOS	Length of Service
LS	Leading Seaman (E5)
ML	Machine Learning
ML-SC	Maritime Logistics – Supply Chain
MSBS	Military Superannuation Benefit Scheme
MT	Marine Technician
NATO	North Atlantic Treaty Organization
NN	Neural Network
NPS	Naval Postgraduate School
NRI	Navy Retention Initiative
OLS	Ordinary Least Squares

PCA	Principle Component Analysis
PO	Petty Officer (E6)
RAN	Royal Australian Navy
RAAF	Royal Australian Air Force
SIV	Silhouette Index Value
SLG	Senior Leadership Group
SMN	Seaman (E2)
SVM	Support Vector Machine
SWPA(N)	Strategic Workforce Planning & Analysis (Navy)
TWM	Total Workforce Model
USN	United States Navy
WO	Warrant Officer
YOS	Years of Service
YOU	Your Opportunities Unlimited

EXECUTIVE SUMMARY

Although this thesis develops a methodology for using machine learning (ML) models specifically to analyze manpower retention issues, this methodology may be applicable across other disciplines. This thesis aims to empirically validate the benefits of machine learning and its applicability as an alternative to traditional regression modeling methods in our context. The outcomes may direct efforts by the Royal Australian Navy (RAN) to invest and further explore the use of machine learning in its manpower modeling systems.

In this thesis machine learning models identify key variables affecting the behaviors of sailors when they decide to separate from the RAN. Furthermore, the models compare the relative behaviors of sailors from technical and non-technical branches and over the periods of time, covering career milestones of the initial minimum period of service (IMPS) and potential promotion.

This thesis provides a methodology for applying machine learning to analyze the potential behavioral differences between technical and non-technical sailors. The analysis assists in identifying key sentiments that may contribute to and drive a particular behavior in the decision to separate from the service. Consequently, this analysis may help shape policy and actions to effect retention of both technical and non-technical sailors.

A. KEY FINDINGS

The analysis was able to identify clear differences in the sentiments, attitudes, and behaviors of different groups of sailors at critical career milestones such as IMPS. When analyzing both communities at around the point of reaching IMPS and the period just after I was able to identify the key differences. Within the non-technical sailor community, prominent concerns include pay, sea service, and a lack of recognition for effort or commitment. High predictors of separation within the technical sailor community include a lack of utilization of skills often related to civilianization of the workforce and a clear understanding of the value of their skills in the civilian workforce.

The analysis also revealed a clear shift in separation sentiment and attitudes when sailors transition from the periods of four to five years of service and from years six to seven. During the former, the shift in sentiment centers on the high predictors of pay and postings. In the later period, concern shifts to a more family-related set of predictors along with the impacts of service life on those families.

ML has been proven as an effective method to analyze data containing responses to questions that provide a gauge of sentiment, feeling, or an emotion, which in turn can describe a behavior. By a process of test and down selection, this thesis identifies four ML models considered appropriate for analyzing the ADF Exit Survey data including: Linear Support Vector Machine (SVM), Chi-square Automatic Interaction Detector, Random Tree, and K-Means. Following test and validation, Linear SVM was chosen as the preferred model for final analysis due to this model's consistently high prediction accuracies and continuity in outputs.

The outcomes of this analysis may lend support to the Deputy Chief of Navy (DCN) released signal detailing the vision for supporting the efforts of the Navy in increasing its future numbers. This vision is to be achieved by not only effective recruitment strategies but with a greater individual support and personal focus within the divisional system to reduce separation (RAN, 2018). The recommendation from this analysis centers on the individual level and the inclusion of questions at divisional interviews that could identify at-risk individuals and the identification of higher-level workgroup attitudinal and behavior separation predictors.

B. FUTURE RESEARCH RECOMMENDATIONS

First, free text answers within the dataset in this instance were not used to support the analysis as the dataset contained only approximately 10% of personnel who actually took the time to write some free text feedback associated with questions such as why are you leaving? As such these answers were not used in this analysis. There is much value in the inclusion of free text answers in this type of analysis, applying some method of sentiment analysis. This would allow quantification of the sentiment, negative or otherwise, to be factored into the analysis and separate those leaving with positive feedback.

Second, the exit survey data used in this analysis included only RAN sailor exit survey responses to meet the thesis objective. I recommend expanding the focus to a more holistic analysis to implement a larger survey dataset and better exploit the ML applications in predicting separation behaviors of a wider set of personnel.

Third, a research topic that specifically looks in more detail at a comparison of the output of ML versus traditional modeling techniques would be worthwhile. Such a study could select a specific critical branch in service and apply both ML and traditional regression modeling to determine any correlation or disparity between the two methods and identify the value of ML beyond or in complementing traditional modeling methods.

Last, the analysis found differences in attitudes, sentiments, or behaviors at the level of technical versus non-technical sailor. As such further analysis by specific technical category, or more refined demographic data, using the same methodology may identify specific behaviors affecting branches or groups of personnel with acute retention issues and identify findings that may support highly focused retention policy.

Reference

Royal Australian Navy (RAN). (2018, July 7). Deputy Chief of Navy, Signal: *Retention of Navy People*. Retrieved from https://www.forcenet.gov.au/community/viewgroup?Id=09a8177c-8f45-4bca-84d1-c37e6c739430

THIS PAGE INTENTIONALLY LEFT BLANK

ACKNOWLEDGMENTS

I would like to give thanks to my thesis advisors, Dr. Sae Young (Tom) Ahn and Dr. James Fan, for their input and mentoring during this demanding process. I must thank Dr. Christopher Griffin of Penn State University for his expert machine learning advice and support. I also would like to thank the Graduate School of Defense Management, NPS staff, and my colleagues for creating a unique and powerful learning experience.

I must thank the RAN and my sponsor organization, Director Navy Workforce Requirements (DNWR), for affording me this opportunity. I truly hope to repay the investment.

Finally, I would like to thank my wife, Suzanne, and children, Rees and Elliot, for their unwavering support in me and for following me continuously in my career and particularly during this recent challenge at NPS.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

Sailors are difficult and expensive to recruit, train, and retain. This is particularly relevant in the technical sailor community whose skills are in high demand in the civilian sector and sensitive to flux in civilian technical industry demand. Consequently, the Royal Australian Navy (RAN) faces challenges retaining the right technical skillsets at the right levels.

Current anecdotal evidence from career managers would suggest that technical sailors are separating at a much higher rate than non-technical sailors. Nonetheless, the current suite of statistical data published by Strategic Workforce Planning & Analysis (Navy) (SWPA(N)) does not support the hypothesis that technical sailors across all technical disciplines suffer from greater separation rates than the non-technical community.

Director Navy Workforce Requirements (DNWR) is the lead department in the RAN for workforce requirements analysis. DNWR, as sponsor of RAN students at the Naval Postgraduate School (NPS) in the manpower analysis curriculum, has a high level of interest in the capacity for high quality research that can be delivered by the graduates during the thesis process. As such, DNWR requires more research to better understand the power of machine learning (ML) and how it can be exploited to support analysis of manpower and workforce issues and, in this case, sailor retention.

The purpose of this thesis is two-fold: first, to deliver analysis into the separation behaviors of technical and non-technical sailors and identify any discernable differences that can be detected and used to provide retention policy guidance. Second, the methodology centers around the application of ML technologies to analyze RAN exit survey data and prove the effectiveness of ML in this type of analysis.

The analysis and subsequent comparison of the attitudes, sentiment, and behaviors displayed by both technical and non-technical sailors can be critical to long-term efforts to reduce separation. If these attitudes and behaviors were clearly understood, this understanding could support efforts to reduce sailor separation in particular, as well as the risk to capability associated with their loss and critical gap shortages in workgroups. This understanding is also particularly useful when economics and industry predictions show a rise in the demand for technical and non-technical skills in the civilian sector. The understanding of these behaviors may allow the prescient application of policy to reduce the impact to the service of the loss of skills, which traditionally spikes during civilian industry boom.

The findings of the analysis and in particular the outcomes of the Linear Support Vector Machine (SVM) ML model, identify key differences between technical and non-technical sailors, and these differences may be taken into consideration when shaping both technical and non-technical sailor retention policy. The more prominent attitudinal differences among technical sailors as opposed to non-technical sailors are associated with a desire for more challenging work, a lack of utilization of skills, and the lack of opportunities to use those skills due to civilianization of technical positions. Conversely, non-technical sailors' attitudes identified that differentiate them from technical sailors center around pay, sea service, and recognition. These include: "lack of recognition or credit for work done," "a more attractive salary package in the civilian sector," and a "feeling of little financial reward for what would be considered overtime in the civilian workplace."

The application of ML technologies in the analysis and in particular the application of the ML model, Linear SVM, was proven to be effective in analyzing the Australian Defence Force (ADF) Exit Survey data. As such ML technologies are suitable for consideration as a modeling technique in solving future workforce issues, particularly in analyzing survey data.

The approach to the analysis was first to set up the exit survey data ready for analysis in several configurations centered to provide the broadest output across selected variable parameters. Second, the data was analyzed through a number of ML algorithms to ascertain a short list of four candidate models that provided outputs that could be interpreted to gauge model performance. This process of down selection led to the preferred Linear SVM model. The data was then set up to analyze specifically the attitudes and behaviors of technical and non-technical sailors at the rank of Able Seaman (AB) and

Leading Seaman (LS) and across two key career periods: the combination of four and five years of service and that of six and seven years of service.

The key career periods for analysis were identified and decided upon using two methods. First, the study used some outline analysis of the SWPA(N) organization published Length of Service Profiles for all sailors and those experienced by the technical and non-technical communities. Second, the general technical and non-technical sailor training and employment profiles over the first four and six years were selected. These periods lead up to completion of the initial minimum period of service (IMPS) at four and six years as the first opportunity to separate for technical sailors and non-technical sailors, respectively. The outputs of these parameters modeled within the Linear SVM were then used to carry out a side-by-side comparison of the technical and non-technical sailor's key behavior predictors and across both periods.

The rest of this thesis explains in detail the background of Royal Australian Navy Sailor recruitment and retention, examines associated academic literature, defines the analysis process, presents its findings, and ultimately makes recommendations for future research and policy.

THIS PAGE INTENTIONALLY LEFT BLANK

II. ROYAL AUSTRALIAN NAVY BACKGROUND

The RAN is manned by 14,094 officers and sailors, including 10,609 Sailors as of 01 May 2019 (SWPA(N), 2019a). Traditionally the RAN as the maritime arm of the ADF has suffered from shortages of personnel in specialist skills branches. A study by the Australian National Audit Office in 2014 on the *Recruitment and Retention of Specialist Skills for the RAN* (ANAO, 2014) found that shortfalls in a number of critical areas within both officer and sailor communities had persisted over the previous 15 years and had largely never been resolved. Over the last ten years there have been various technical and non-technical sailor branches that have suffered critical shortages (ANAO, 2014). The RAN, via Defence Force Recruiting (DFR), executes intensive recruitment campaigns to target any particular shortage branch and also implements retention strategies. These retention strategies can be incentive based to relieve the issue on a short-term basis or as part of a career continuum initiative that will attempt to inculcate longer term behavioral changes in separation decisions.

This chapter establishes the background understanding of the RAN Workforce structure, qualifications, methods of entry, incentives, and overall manpower planning philosophies. This background supports the reader in understanding the recommendations and outputs from this thesis. Furthermore, the chapter describes the underlying issues and challenges for the ADF in retaining technically skilled personnel, providing the impetus for research into the understanding of technical sailor separation behaviors.

A. SAILOR RATES

Sailor rates range from Seaman to Warrant Officer (WO). This study only deals with the data associated with the responses from Seaman (SMN) to Chief Petty Officer (CPO) Rates. WOs are excluded for a couple of reasons. First, due to lack of data and the very minimal numbers effecting the analysis. Second, it is due to the majority of WO operating within roles outside of their core branch and as such having little or no effect on branch skill retention issues. In total 164 WO were excluded for the purposes of the analysis. Sailors join the navy as recruits and once they have graduated from recruit school are promoted to Seaman, when they have the opportunity to be promoted to E9/10, equivalent to WO. Figure 1 depicts the rates and their United States Navy (USN) equivalent rates up to and including E8 or CPO. As can be seen the RAN has no equivalent rate to E4 and E7. The role and responsibilities that would be filled and accepted by the E4 and E7 would ordinarily be taken by the rates above and below. In this case, the AB and LS would deliver the outputs of the E4. The Petty Officer (PO) and CPO would deliver the output of the E7.

Figure 1. Royal Australian Navy Rates and USN Equivalents. Adapted from Dodds (2018).

B. STRUCTURE

Sailors workgroups are separated into communities based upon their particular trade or skill. The higher-level family functions for sailors are Engineering, Aviation, Health services, Logistics & Administration, and Warfare. Each family function is then further separated into sub-functions. An example particularly relevant to this thesis is the sailor engineering community, which is then further broken down into Marine Technician, Electronics Technician, and Aviation Technician. Each family may then be further broken down by particular skills. Aviation Technician, for example, is further separated into Aviation Technician Aircraft and Aviation Technician Avionics. This type of breakdown within workgroups is common across all sailor workgroups. The breakdown of sailor family functions and sub-functions is depicted in Tables 1 and 2.

Table 1. Non-technical Branches of the Royal Australian Navy. Adapted from ADF (n.d.-a).

Branch	**Sub Branch (as applicable)**	**Qualification / AQF Level**
Cryptologic Linguist		None
Cryptologic Systems		None
Cryptologic Networks		None
Acoustic Analyst		None
Aviation Ground Crew		None
Boatswains mate		None
Boatswains mate	Submariner	None
Maritime Logistics	Chef	Certificate III (Level 3)
Maritime Logistics	Chef Submariner	Certificate III (Level 3)
Maritime Logistics	Hospitality & Logistics	Certificate III (Level 3)
Maritime Logistics	Hospitality & Logistics	Certificate III (Level 3)
Maritime Logistics	Personnel Administration	None
Maritime Logistics	Warehouse Store-person	None
Maritime Logistics	Warehouse Store-person	None
Combat Systems	Operator	None
Combat Systems Operator	Operator-Mine Warfare	None
Communications &	Operator	None
Communications &	Submariner	None
Electronic Warfare	Operator	None
Electronic Warfare	Submariner	None
Hydrographic Surveyor		None
Medic		None
Medic	Submariner	None
Dental Assistant		N/K
Musician		N/K
Naval Police Coxswain		None
Clearance Diver		None

Table 1 also details the current complement of non-technical sailor branches, along with the relevant qualifications received post initial trade training, as described in the next section. The technical branches are detailed in Table 2, again along with the relevant qualifications received post initial trade training, as described in the next section.

Table 2. Technical Branches of the Royal Australian Navy. Adapted from ADF (n.d.-a).

Branch	**Sub Branch (as applicable)**	**Qualification / AQF Level**
Marine Technician		Certificate III/IV (Level 3/4)
Marine Technician	Submarines	Certificate III/IV (Level 3/4)
Electronics Technician		Certificate III (Level 3)
Electronics Technician	Submarines	Certificate III (Level 3)
Aviation Technician	Aircraft	Certificate IV (Level 4)
Aviation Technician	Avionics	Certificate IV (Level 4)

C. QUALIFICATIONS

Traditionally, technical trades receive a higher level of qualification following completion of their in-service trade training than their non-technical counterparts. The RAN utilizes the Australian Qualifications Framework, AQF (n.d.) to set the qualification standard. Not all non-technical trades will receive any AQF recognized qualification during their initial trade training but they receive RAN delivered competency-based training to the required skill level. That is, the RAN training received by some non-technical sailors may not be in line with any specific civilian-based or recognized curriculum. That is to be expected given the often-unique skillsets utilized by sailors in the military context that are simply not used outside of the RAN or in the military in general. In some branches, there may be opportunities to receive AQF recognized qualifications based upon further skills training and education. The various qualifications range from a "no accredited" qualification to a Certificate IV. The AQF Framework, detailing the levels of AQF in context of the overall education system and highlighting the level achieved by technical sailors, is presented in Table 3.

Table 3. Australian Qualifications Framework. Source: AQF (n.d.).

HIGH SCHOOL	TAFE NSW	UNIVERSITY	AQF Level
		Doctoral Degree	Level 10
		Masters Degree	Level 9
	Graduate Diploma Graduate Certificate Bachelor Honours Degree		Level 8
	Bachelor Degree		Level 7
	Associate Degree Advanced Diploma		Level 6
	Diploma		Level 5
	Certificate IV		Level 4
	Certificate III		Level 3
Certificate II			Level 2
Certificate I			Level 1
Senior Secondary Certificates of Education (HSC in NSW)			

D. CURRENT RETENTION INCENTIVES FOR BOTH TECHNICAL AND NON-TECHNICAL SAILOR BRANCHES

There are differences in the service incentives associated with technical and non-technical sailors. These incentives are largely based upon three areas: pay, retention bonuses, and qualifications.

First, the RAN operates a graded pay system whereby the allocated pay grade is assessed on skills, working environment, and type of service. The Defense Force Remuneration Tribunal (DFRT) approves the recommended pay grades on an annual basis. As can be seen from Table 4, a differential of one and two pay grades exists between most non-technical sailor and technical branches. There are exceptions to this rule, however; on the large part, non-technical sailors are paid a lower rate across equivalent career progression points. A typical technical to non-technical paygrade differential taken for the ADF pay rates table as of 01 November 2018 is $3,389AU per annum between paygrades two and three at the AB level (ADF, n.d.-e).

Table 4. Royal Australian Navy Sailor Pay Groups. Adapted from ADF (n.d.-b).

		Pay Group							
	Category	1	2	3	4	5	6	7	8
Tech / Eng	DMSA-P		Tech / Eng						
	ATA				ATA1	ATA2	ATA3	ATA4 ATA5	FSMS
	ATV				ATV1	ATV2	ATV3	ATV4 ATA5	FSMS
	ET			ET1	ET2	ET3 ET2 FCO	ET4 ET3 FCO	ET5 ET4 FCO	
	MT			MT1	MT2	MT3	MT4	MT5	MT5 STO
Combat 1	DMSA-P	Combat 1 Family							
	BM		BM1	BM1NavYeo BM1SLSE BM2	BM2 LNWC Bosun	Bosun NWC			
	CSO			CSO1	CSO2	CSS	CSM		
	CSO Air Controller					ASAC	AIC		
	CSO MW			CSO MW1	CSO MW2	CSS MW CSM MW	CSS MW NWC CSM MW NWC		
	Diver (Res)		DVR1	DVR2 DVR3					
	HSO			HSO1	HSO2	HSM			
	HSO NAV					HSO2 LNWC	HSM NWC		
	NPC			NPC1	NPC2	INV/LNWC			
Support	DMSA-P	Support Family							
	CK		CK1	CK2 CK3 CK4	CK5				
	MUSN			MUSN1	MUSN2	MUSN3			
	SN		SN1	SN2 SN3 SN4	SN5				
	STD		STD1 STD2 STD3	STD4 STD5					
	WTR		WTR1	WTR2 WTR3 WTR4	WTR5				
Health	DMSA-P	Health Family							
	DEN		DA1	DA2	DAP3 DS	DM			
	MED			MED1	MED2	UMC	MED3		
	PTI			PT1	PT2 PT3				

Second, the RAN employs retention bonuses in an effort to reduce attrition within a particularly critical branch or part of service. Examples include retention bonuses awarded to personnel in the submarine service and the Marine Technician workforce. Alongside these are more generic retention incentives paid to all qualifying personnel across all branches. One such is the Navy Retention Initiative (NRI), which is a financial payment to incentivize retention and address workforce hollowness at specific ranks and experience levels (RAN, n.d.). The NRI is paid to eligible individuals at key career decision points to incentivize a continued 12–24 months of service. The payment is a fixed amount of $20,000AU gross payable on the anniversary of seven, eight, and 12 years' completed service (RAN, n.d.).

Another retention benefit is the Military Superannuation Benefit Scheme (MSBS) retention benefit (ADF, n.d.-c). The MSBS retention ceased to be open to personnel who joined the RAN after 01 Jul 2007. Those personnel who joined prior to the cut-off date and reach E6 or O4 and a total of 15 years' service can sign on for a commitment of a further five years' service for the equivalent of one year of pay (ADF, n.d.-c).

Third, the civilian-recognized qualifications at the Certificate III and IV levels awarded post technical branch trade training can be considered as an incentive for retention, certainly from an outside perspective. Yet, it is no secret that the issue of civilian-accredited qualifications can also encourage separation based upon the economic climate in the civilian workforce, and this is particularly prevalent in the engineering industry. Given that all technical trades within the RAN receive up to a Certificate III/IV on completion of their trade training it can be argued that this is an incentive for retention and separation then based upon the individual's behaviors and assessment by the individual member.

E. OTHER RETENTION INITIATIVES

Recently Deputy Chief of Navy (DCN) released a signal detailing their vision for supporting the efforts of the RAN in increasing its numbers. This is to be achieved by not only effective recruitment strategies but increase support of and personal focus on the individual within the divisional system to reduce separation (RAN, 2018).

The signal stated, "I require all Divisional Staff to use the divisional system to engage and support their (our) people. Your approach should be highly engaged in nature. Including monthly divisional meetings and regular divisional interviews with each and every member of your division" (RAN, 2018, p. 5). The analysis approach in this thesis is able to identify key attitudes, sentiments, and behaviors within sailors. If these factors can be captured effectively and analyzed, there may be scope to apply specific methods to increase retention at the individual level.

F. METHOD OF ENTRY AND MINIMUM SERVICE

All sailors are recruited by the DFR. DFR is a tri-service organization responsible for assessing suitability of all applicants and determining their suitability for particular workgroups. All recruits will receive an assessment day called a Your Opportunities Unlimited (YOU) session (ADF, n.d.-a). The result of the YOU session will determine the workgroups and specific trades that may be offered to applicants should they decide to join. Better performance of applicants during aptitude testing will result in a wider variety of jobs being made available to them as part of their application and this may include access to technical trades. The reasoning for this process of assessment and offer is an attempt to minimize risk of attrition during training and better ensure success during the more academically intensive technical training programs.

New sailors entering the navy enter under the General Entry (GE) scheme, and this is the only method of entry for new sailors (ADF, n.d.-a). The GE scheme further delineates among GE Non-Technical, GE Technical, and GE Qualified. The GE Technical entry scheme recruits sailors for the RAN technical branches, considering such candidates suitable to pass technical training in service (ADF, n.d.-a). The GE Qualified Scheme recruits personnel with prior skills and qualifications into the RAN technical branches. The advantage of recruiting personnel with prior skills above GE Technical is a reduced training time, leveraging of prior qualifications. The GE Non-technical entry scheme recruits personnel into all other branches (ADF, n.d.-a).

On enlistment, the sailor signs a contract for the IMPS and may be offered an option to re-enlist for a further contract period at the end of the IMPS based upon the workgroup and workforce needs (ADF, n.d.-a). IMPS can vary between workgroups largely due to the type of GE entry and the differences in the length of training and subsequent return on investment. Sailors joining under the GE Technical scheme ordinarily serve an IMPS of six years while sailors on the GE Non-technical scheme serve an IMPS of four years and sailors entering under the GE Qualified scheme serve a four-year IMPS (ADF, n.d.-a).

The RAN also runs an ADF Gap Year scheme to encourage school leavers and undergraduates to try the service of their choice for one year, again with the aim to encourage personnel to sign on for further service beyond the first year. Since its inception in 2015 the ADF Gap Year program has returned 54% of Gap Year Applicants, who transfer to permanent service (ADF, n.d.-a).

G. TOTAL WORKFORCE MODEL

At the beginning of July 2016, the ADF, including the RAN, began to implement a Total Workforce Model (TWM) (ADF, n.d.-f). The TWM is largely based on the understanding that the decision to depart from the ADF or continue in service, for a significant portion of the workforce, is influenced by the lack of flexibility in ADF working arrangements (ADF, n.d.-f). Permanent members want more flexibility in the service options available to better balance career and personal commitments while reservists want more opportunity to serve. Given the persistent skills shortage in the Australian workforce, including the ADF, the TWM was introduced and is being developed to combat personnel attrition within the ADF (ADF, n.d.-f).

As can be seen in Figure 3, the framework allows for more options when deciding how and when to serve and in what capacity, which is highlighted in red in the model in Figure 2. The TWM is relatively young in its implementation, and as such, there is currently little output data that may be used to support the analysis in this thesis. It is important, however, to understand the concept of TWM and how it may be a supporting mechanism in increasing retention in sailors and thus support any recommendations made in later chapters.

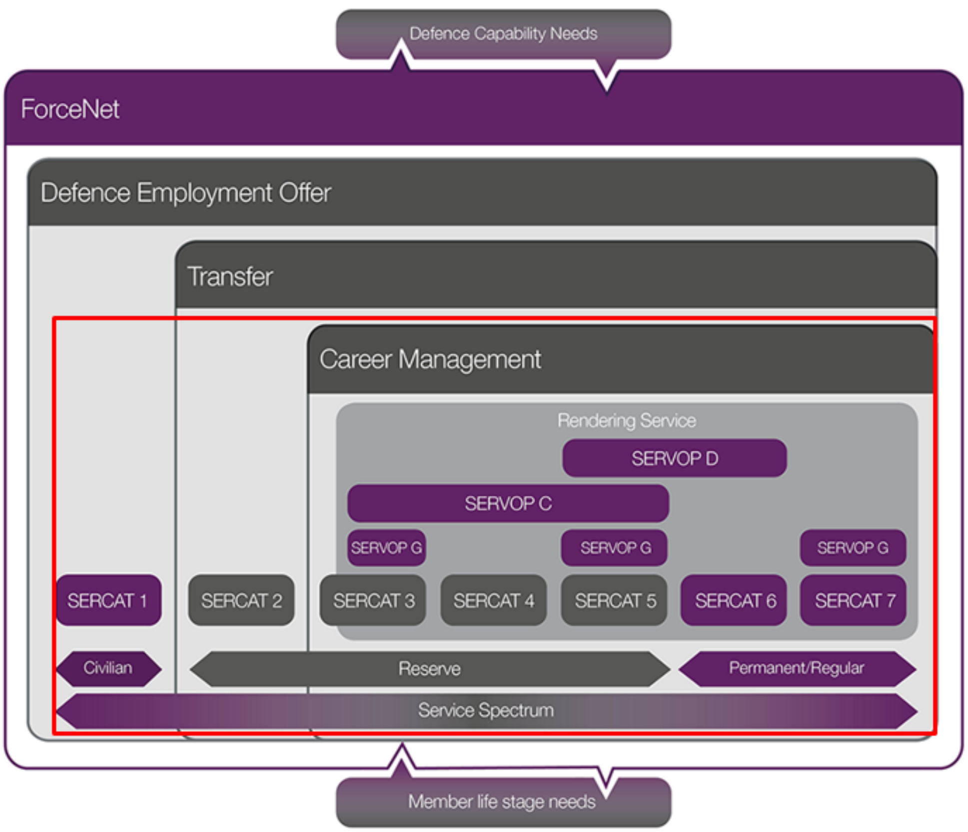

Figure 2. ADF Total Workforce Model. Source ADF (n.d.-f).

THIS PAGE INTENTIONALLY LEFT BLANK

III. REVIEW OF ALL CURRENT SAILOR SEPARATION STATISTICS

This chapter provides some understanding of and background into the current statistics associated with technical and non-technical sailor workgroups. Moreover, this chapter establishes the need for this research. The findings in this chapter lead us to surmise that there is little data or analysis that supports the hypothesis that technical sailors have greater separation rates than their non-technical counterparts. In fact, separation rates vary quite markedly among categories in both the technical and non-technical sailor communities. That said, between technical and non-technical sailors there is likely a difference in the attitudes and behaviors leading to separation. Given the value we place in the skills and expertise retained within our technical sailor force, an understanding of these attitudes and behaviors may be critical to long-term retention of these key skills.

A. SEPARATION STATISTICS FOR CURRENT SAILOR CATEGORIES

Overall, the statistics detailed in Figure 4 show that the whole sailor population has a separation rate of 8.7%, which is slightly higher than the whole of RAN average of 8.1% (SWPA(N), 2019b). We can further drill down into the separation by rate and the separation profile by years of service (YOS) for all sailors. The separation rates as of 01 May 2019 by the relative rank are 3% for Seaman (SMN), 9.9% for AB, 12.1% for LS, 7.6% for PO, and 5.7% for CPO (SWPA(N), 2019i). This spike at the earlier rates of AB and LS is a concern for the navy and typically reflects the notion that separation in the first ten years is high in relation to the majority of personnel serving ten years or under predominantly holding those two rates. This is again reflected in the separation profile in Figure 3 related to YOS, with large increases in separation between five and eight years. This spike is likely related to the completion of the IMPS for most sailors, but given the opportunity for continued service it does not tell the whole story; more likely, it explains one part of the decision behavior of the whole sailor community.

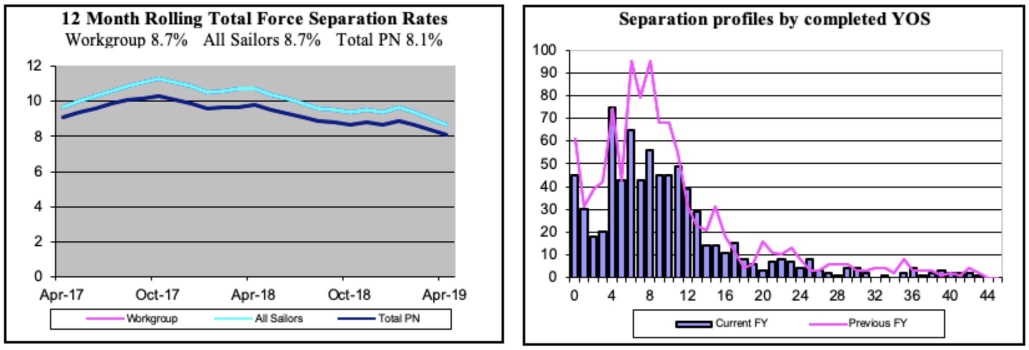

Figure 3. All Sailor 12-Month Rolling Separation Rates and Separation Profile by YOS. Source: SWPA(N) (2019b).

B. CURRENT SAILOR SEPARATION STATISTICS BY TECHNICAL TRADES

Figures 4, 5, 6, and 7 illustrate the current separation rates for all of the technical branches and contain varied separation rates and YOS profiles across all technical trades. The technical branch that currently is suffering the highest separation rates is the Aviation Technician Aircraft (ATA) branch. The ATA branch suffers a particularly high overall separation rate at 10.3% compared to the other technical branches, where only MT is close with a separation rate of 9.8%. The ET branch and Aviation Technician Avionics (ATV) branch record separation rates of 7.7% and 7.4%, respectively. The ATA branch separation rate is 13.4% for the AB and 12.6% for LS. These values are significantly higher than the overall sailor separation rates of 10.2% and 10.7%, respectively (SWPA(N), 2019i).

In the technical branch statistics for separation by rate we can see variation again between branches. An example is the separation rate for LS compared to AB. The ET branch has a very high separation rate of LS compared to AB, with rates at 11.8% and 8.2% for LS and AB, respectively. This is in contrast to the ATA and ATV branches, which conversely have higher rates at the AB level—13.4% and 10.3%, respectively—compared to the LS level at 12.6% and 7.5%, respectively (SWPA(N), 2019i).

A review of the YOS profiles for all of the technical branches SWPA(N) (2019c) shows a steady trend across the branches, with higher separations by YOS in the period between four to eight years; however, there are some particularly high spikes in separations in MT and ATV branches (SWPA(N), 2019c). This could again be attributed to the IMPS profile of these branches. Given the branches all operate the same IMPS, however, there may be other explanations for the high spikes in separations, such as the point at which full qualifications are received. Also, given the comparison of the YOS separation profiles by current year and previous year we can see marked variation in separation by year.

In essence we can see that technical branches do suffer from high separation rates in general, but certainly the data does not support the hypothesis that the technical branches suffer any worse attrition in comparison to the overall sailor population. Also, given the variation between the technical branches for separation rates, no real trend in separations rates by rank and YOS separation profile exists that would differentiate the technical branches from the rest of the sailor population.

What may help better understand any real specific technical sailor separation trends is a comparison of a sample of non-technical sailor branches to the overall sailor community and then to the non-technical branches.

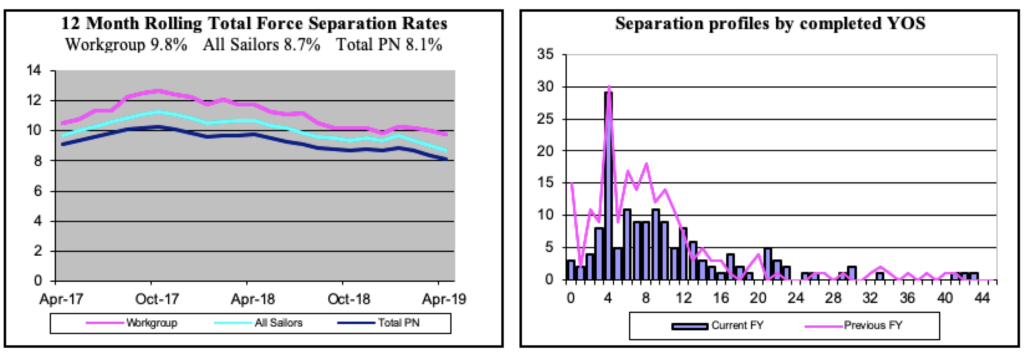

Figure 4. MT Branch 12-Month Rolling Separation Rates and Separation Profile by YOS. Source: SWPA(N) (2019c).

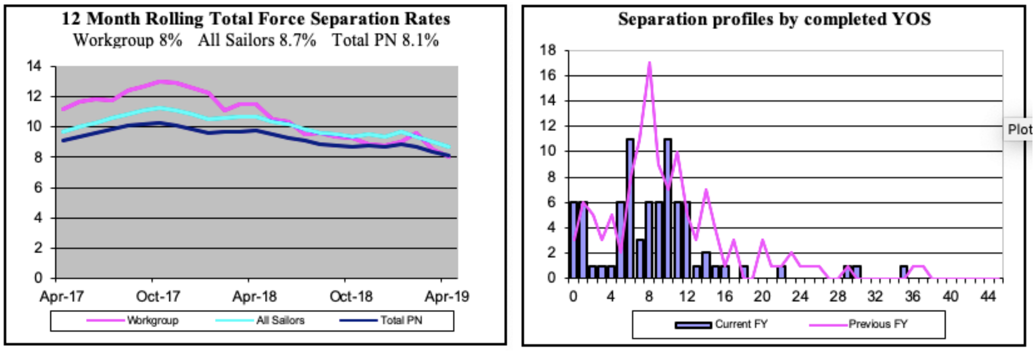

Figure 5. ET Branch 12-Month Rolling Separation Rates and Separation Profile by YOS. Source: SWPA(N) (2019d).

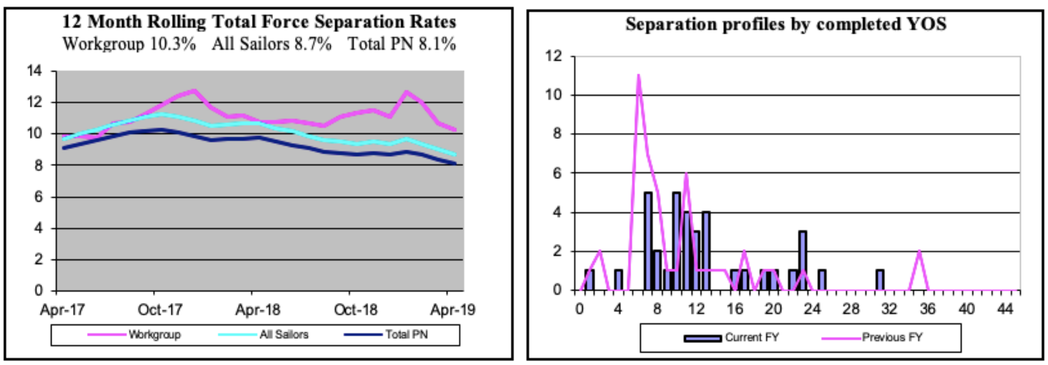

Figure 6. ATA Branch 12-Month Rolling Separation Rates and Separation Profile by YOS. Source: SWPA(N) (2019e).

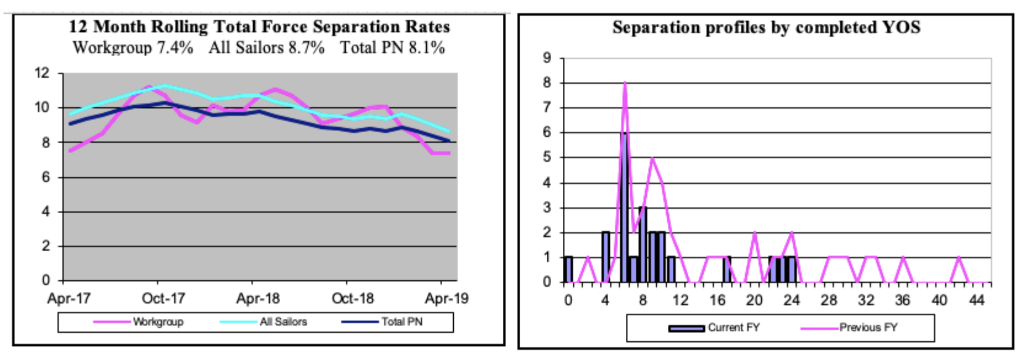

Figure 7. ATV Branch 12-Month Rolling Separation Rates and Separation Profile by YOS. Source: SWPA(N) (2019f).

C. CURRENT SAILOR SEPARATION STATISTICS BY NON-TECHNICAL TRADES

These following sections analyze the current branch statistics for the Maritime Logistics-Supply Chain (ML-SC) and Combat Systems Operator, analyze statistical differences between technical and non-technical sailors, and examine technical sailor retention historical issues.

1. Comparison to Non-technical Branches

For comparison, I have used the Maritime Logistics-Supply Chain (ML-SC) Sailor data and the Combat Systems Operator (CSO) sailor data.

As can be seen from Figures 8 and 9, the current ML-SC workgroup rolling separation rate is 7.4%, which is 1.3% less than the overall sailor average, with a separation rate by rank of 13.9% for AB, 8.1% for LS, 4.7% for PO, and 2.3% for CPO (SWPA(N), 2019i).

Another comparison group of non-technical sailors is the CSO Branch. As can be seen from Figure 9, the current CSO workgroup rolling separation rate is 10.1%, which is 1.4% greater than the overall sailor average, with a separation rate by rank of 12.4% for AB, 12.4% for LS, 10.2% for PO, and 3.4% for CPO (SWPA(N), 2019i). Thus, even a comparison of various non-technical sailor branches can show much variation in branch statistics. These examples show a large disparity between workgroup rolling separation rates, 7.4% and 10.1%, which are less than and greater than the whole of navy sailor average, respectively. Neither this comparison nor the subsequent comparison of non-technical branches to technical branches supports the hypothesis that technical sailors have a higher separation rate than non-technical sailors. Although no obvious difference emerges in retention levels across non-technical and technical branches due to the broad range of roles and size of workgroups, there is still a possibility of key behavioral differences between those branches within non-technical and technical workgroups.

2. Current ML-SC Family Statistics

The graphs in Figure 8 show the ML-SC branch 12-month rolling separation rates and separation profile by YOS.

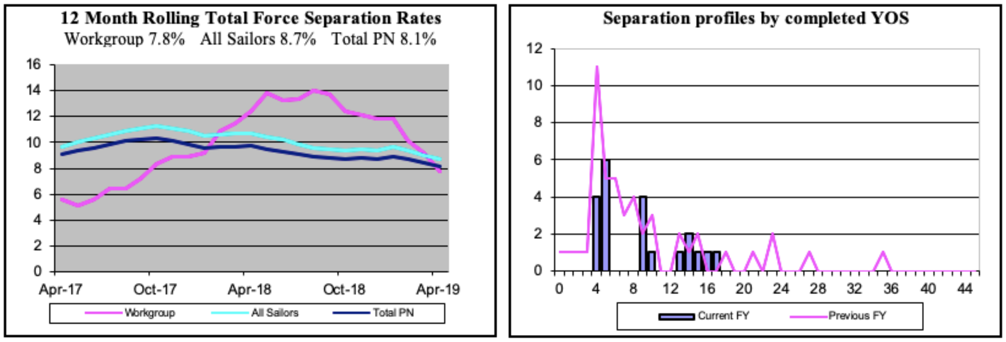

Figure 8. ML-SC Branch 12-Month Rolling Separation Rates and Separation Profile by YOS. Source: SWPA(N) (2019g).

3. Current CSO Family Statistics

The graphs in Figure 9 show the CSO branch 12-month rolling separation rates and separation profile by YOS.

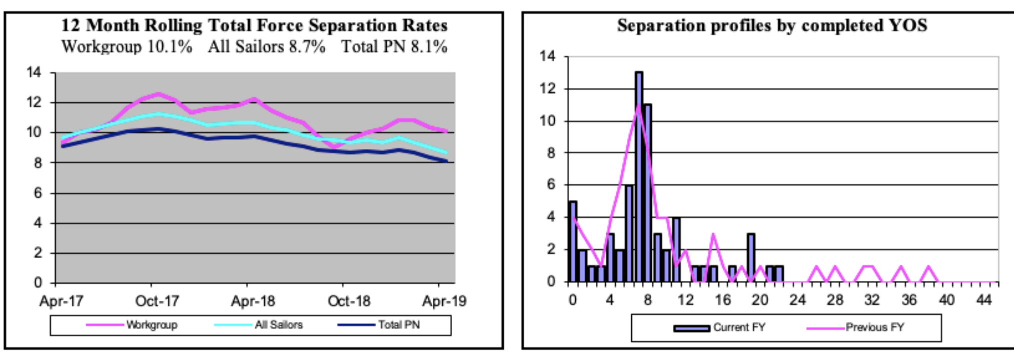

Figure 9. CSO Branch 12-Month Rolling Separation Rates and Separation Profile by YOS. Source: SWPA(N) (2019h).

D. TECHNICAL VERSUS NON-TECHNICAL CURRENT STATISTICS

As can be seen from the statistical data presented previously, despite a belief that there is a clear empirical difference in separation, both technical and non-technical branches do not suffer from specifically high or low attrition rates compared to the overall sailor separation rate of 8.7%. In fact, a few factors contribute to a general misconception that the technical branches suffer greater attrition. First, because it is considered more critical when we lose a technical sailor, there may be more effort to resolve a loss in a technical branch, and therefore, the issue receives more focus. Second, the ability to grow a replacement technical sailor takes longer based upon a longer training period. Also, a ship or unit may accept the loss of a non-technical sailor over technical sailor more readily, given they may be able to share the workload loss more effectively. This overall misconception may lead technical branch managers to believe they are seeing a greater attrition rate among technical sailors than that of their non-technical sailor equivalents.

This attempt to compare non-technical branches to technical branches statistically on raw numbers is certainly not telling the whole story. As such we cannot say that technical or non-technical sailors, in particular, markedly suffer from greater attrition than the other. Thus, although no obvious difference emerges in retention levels across non-technical and technical branches due to the broad range of roles and size of workgroups, there is still a possibility of key behavioral differences between those branches within non-technical and technical workgroups. This thesis aims to identify those possible differences.

E. TECHNICAL SAILOR RETENTION HISTORY

Without a clear difference between the separation rates of technical and non-technical trades, should we still be concerned with specifically understanding their behaviors? The technical trades have traditionally suffered from shortages as a part of the total workgroup numbers as opposed to experiencing mere fluctuations in retention levels. This is due to the higher sensitivity of the technical trades, not just in the RAN but in the wider ADF, to economic booms within technically skilled industries. Examples of this would be the mining boom of the early 2000s discussed by Parnell (2012) and the recent expansion in the ship-building industry (ADF, n.d.-d), and these are set against a rapidly expanding and developing technical industry as part of Australia's overall growth.

The Australia Trade and Investment Commission Report for 2017–2018 highlighted the expansion in technical services, noting that

> Australia's strong growth has also been broad based over the past 27 years, with 12 out of 19 major sectors expanding by at least 3% a year…. The average annual growth rate of Australia's Services sector of 3.4% has outpaced growth in non-services industries of 2.1%. In particular, Australia's technology-driven industries, such as Information, Media & Telecommunications; and Professional, Scientific & Technical Services increased by an annual average rate of about 5% over this period. (Australian Trade Commission, 2018)

These findings are depicted in Figure 10.

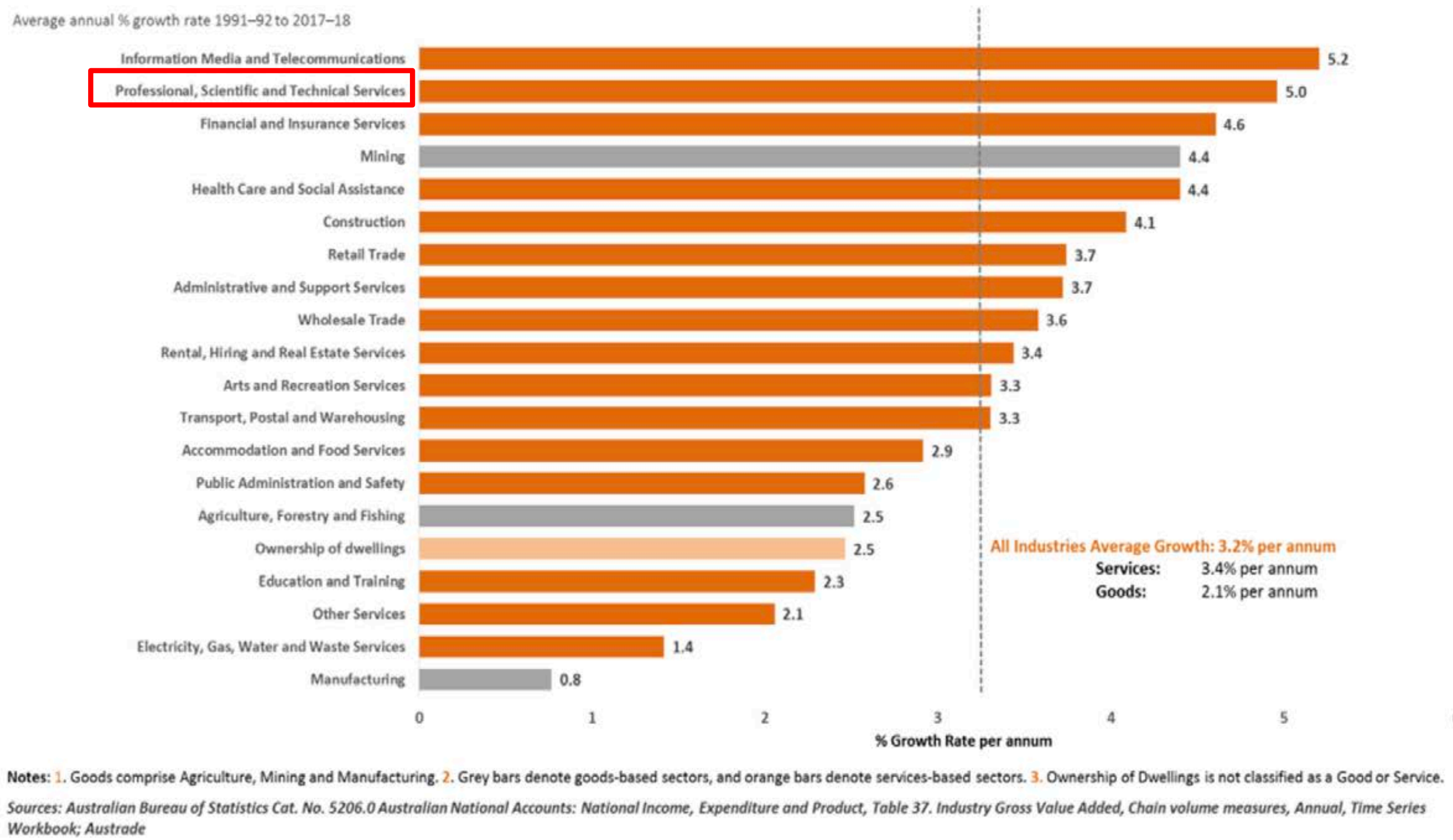

Figure 10. Growth by Industry in Australia 1991/92–2017/18. Source: Australian Trade Commission (n.d.).

The mining boom in Australia was a clear example of industry expansion affecting technical skills retention in the ADF. In 2012, the *Weekend Australian* ran an article detailing the effect of the mining boom on the ADF and concluded "the number of people leaving the Australian Defence Force is rising dramatically, a trend blamed not on dissatisfaction with the military or events in Afghanistan but the lure of the mining dollar" (Parnell, 2012). The rapid expansion of mining and the lucrative nature of the work enticed many technically qualified personnel from the whole of ADF to separate and seek employment in the sector.

Australian Auditor General Report No. 17 of 2014–2015 recognized a specific skill shortage of the Marine Technician sailors within the RAN. The report stated that the navy was experiencing a shortfall among Marine Technician sailors and submariners amounting to some 461 personnel, which is a total 26% fewer than needed. This shortfall presented tangible risk should the RAN experience periods of particularly heavy operational demand (ANO, 2016).

Even as recently as August 2019, reports in Australia have shown a shortage of technically qualified and skilled personnel in the civilian technical industries. The ABC News aired a report 23 August 2019 titled "Tech Jobs are waiting for Australians but not enough people have the skills, companies say" (Robertson, 2019). In the report the Wise Tech Global CEO said "The fundamental issue is we're short, and it's probably short about two hundred thousand people across the whole marketplace in Australia" (Robertson, 2019). It is these types of acute civilian industry shortages that put pressure on the RAN to seek innovative ways to incentivize technical personnel to remain in service.

A major output of the RAN that considered the effects of key engineering and technical workforce shortfalls and sought a plan to overcome those shortfalls was the Naval Engineering Future State Blueprint of August 2013. The Naval Engineering Future State Blueprint espoused a need to sustain the naval technical workforce's mastery. This goal was to be achieved by increasing the retention of people. The report went on to say "In an increasingly competitive market for engineering skills and experience, the Naval Engineering organization needs to offer more flexible employment options, whilst developing core skills over time to rebuild and thence avoid depletion of vulnerable categories” (RAN, 2013, p. 33).

The risks associated with a reduction in the technical workforce, which are associated with overall expansion in technical industries in Australia and more localized expansion such as the recent shipbuilding program in Adelaide, would lead us to want to understand technical sailor separation behaviors. This understanding would allow the RAN and the wider ADF to potentially put in place extra safeguards against increased separations associated with pending industry expansion.

THIS PAGE INTENTIONALLY LEFT BLANK

IV. ACADEMIC LITERATURE REVIEW

This chapter reviews the academic literature associated with the thesis main topic covering labor economics, manpower planning mechanisms, and machine learning methods in manpower modeling.

A. ECONOMICS OF MANPOWER

Over the last 30 years, major events in the world have shaped militaries across the globe, requiring the defense departments of these militaries to seek new policy to overcome the effective change in defense posture due to these events. In 2007, Asch, Hoesk, and Warner identified four major events: The Cold War, increased university attendance, increased operational tempo, and the cost of U.S. Military entitlements as the drivers for shaping the manpower agenda (Asch et al. 2007).

The process of supply and demand blended with other economic theories has been the basis of the method for analyzing and reporting the behavior of those within the labor force. In this case, the military and the RAN are subject to these common labor force demand and supply influences. Sections 1, 2, and 3 cover some key tenets of labor: labor supply, labor demand, and positive economics.

1. Labor Supply

The Standard Occupational Choice Theory from Rosen (1986), and subsequently adapted by Warner and Asch (1995), explains the decision to join the military or to re-enlist. It considers the pay and non-pecuniary benefits in both the military and the civilian sectors and is based on comparing the maximum utility to the available benefit, whereby people will choose the military when the following equalities are true: where W^M and W^C are the wage in the military and civilian sectors, U^M and U^C are the utility function for the military and the civilian sectors, and τ^M and τ^C represent the non-pecuniary benefits in the military and civilian sectors, respectively (Warner & Asch, 1995). These conditions are expressed as

$$U^M = W^M + \tau^M > U^C = W^C + \tau^C \text{ or } W^M - W^C > \tau^C - \tau^M.$$

Non-pecuniary benefits can include the value of serving one's country, the risk associated with losing one's life, and an individual's appetite for a military lifestyle (Warner & Asch, 1995). This simple but effective way of using utility to assess propensity to join or re-enlist is still relevant today and appropriate to be associated with current manpower philosophies of most militaries including the ADF and RAN.

2. Labor Demand

The end of the Cold War brought with it a reduction and downsizing of the North Atlantic Treaty Organization (NATO), as well as Western and European military organizations (Asch, Hosek & Warner. 2007). The United States alone reduced its military by 38% following the Cold War (Asch et al. 2007). This reduction in militaries and subsequent manpower requirements led to reductions in countries operating conscription and some to a completely volunteer force. It should be noted that Australia last operated a conscription manning policy in 1964, under the National Service Act (National Service Act, 1964). The move to an all-volunteer force led to increased challenges for those militaries recruiting and retaining a workforce. Subsequently, the militaries of the world have faced many of the labor supply challenges that virtually all civilian sectors have encountered.

3. Positive Economics

Positive economics and the understanding of behavior is an important consideration. According to Ehrenberg and Smith (2016), positive economics is the process of analyzing the behavior of people to understand how they respond to market incentives. Those scholars also allude to the fact that we live in a complex world, and as such, they consider what it means to understand behavior (Ehrenberg & Smith, 2016).

So, the system we are analyzing within the complex array of considerations afforded to an individual deciding whether to separate from the service is difficult to understand. And, it is even more difficult to predict the individual's behavior in that system. Although complex and in this case exhaustive, the question set contained within the exit survey utilized in this thesis captures some of the less traditional considerations outside of pure economic forces. These may shape behaviors and allow us to predict to some extent the influence of other factors, such as leadership and job satisfaction, in understanding why sailors may separate.

B. MANPOWER PLANNING

This section details key manpower planning influences; specifically, the ADF application of the closed labor market and the utilization of the Length of Service (LOS) profile.

1. Closed Labor Market

Traditionally, militaries around the world operate within a closed labor market; that is, the bulk of personnel enter into service at the basic training stage with little or no skills or experience (Rodney, 2017). The RAN currently operates predominantly within a closed labor market in which entrants to the RAN join under the Sailor GE system described in Chapter II, Section F. This presents challenges to military manpower managers in determining the number of personnel to recruit to meet the long-term workgroup requirements. These predictions are complex and not well supported by the closed labor market model (Rodney, 2017).

There are efforts to introduce personnel at different skill levels, to plug specific skill shortages by method of lateral entry from other ADF services or from other militaries; however, this effort is minimal. These efforts have now been supplemented by the TWM as described in Chapter II, Section G, whereby the model affords much greater flexibility to career managers and service personnel alike in managing careers in and out of various service categories. This flexibility allows the RAN to better respond to fluctuations in recruitment and retention.

2. LOS Profiles

In this case, we would like to understand, given a set of responses to attitudinal questions with a given sentiment, how long someone may serve or how much at risk an individual is for separating. If we were to see this sentiment or attitude across a group of individuals within the same type of workgroup, the LOS profiles for that particular group or groups would help us understand the workgroup retention dynamics at various points within the career timeline. The LOS profiles reviewed and analyzed in Chapter III are demonstrative of the value of understanding critical points for separation by rank and year in an LOS profile. The LOS profiles further support the decision to analyze specific points of time within the survey data utilized in this thesis analysis.

Rodney (2017) proposes a length of service distribution equation by number of personnel within a workgroup, which will indicate total number of personnel at each year of service (*YOS*), where *P* = the number of personnel:

$$P \text{ with } n \text{ } YOS \text{ in year } t \leq P \text{ with } n\text{-}1 \text{ } YOS \text{ in year } t\text{-}1.$$

The LOS profile gives an indication of the experience levels of the workgroup and also gives an insight into fluctuations in recruitment and retention. These fluctuations can be seen traveling through a workgroup over time as a hollowing or widening as a certain cohort transitions through career ranks. The LOS profile for all sailors in the RAN as of 01 May 2019 is shown in Figure 11 (SWPA(N), 2019a).

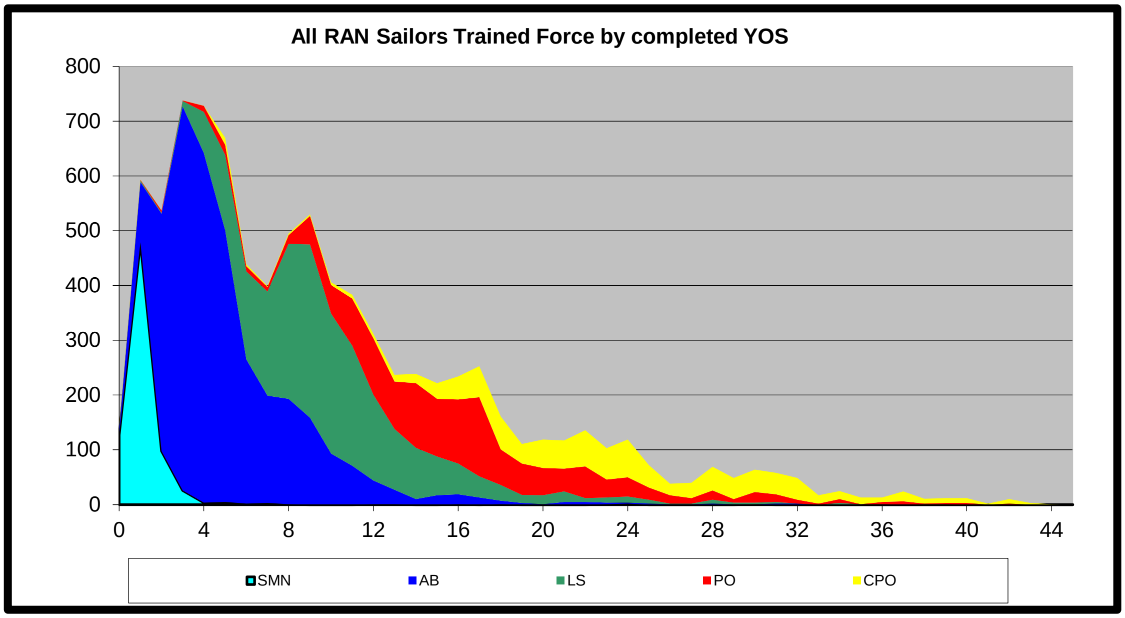

Figure 11. All RAN Sailor Trained Force LOS Profile by YOS.
Source: SWPA(N) (2019a).

C. MACHINE LEARNING VERSUS TRADITIONAL STATISTICAL MODELING

The initial thought process when developing a methodology to discern behavioral differences between technical and non-technical sailors given the ADF Exit Survey data was based upon understanding which model type would most benefit from the question set responses. A large portion of the question set responses is in a form that expresses a type of sentiment to the question (e.g., Not Applicable, Not Important, Slightly Important, Moderately Important, Very Important, and Extremely Important). Further, the number of these variables is large, in this case more than 80.

Given the number and type of variables within the exit survey data, it was considered that traditional methods of analysis such as Ordinary Least Squares (OLS), Logit, or Probit may give unusable results. The complexity of modeling increases as it transitions through simple OLS, multivariate OLS, logit, and then on to ML models. This increase in complexity, in turn, is more effective at handling more complex data.

The application of machine learning and its subsequent comparison to some more traditional statistical modeling techniques, such as OLS, were considered most appropriate when associated with building a predictive profile that given a set of demographics, question responses and sentiment can predict likelihood of separation. This model or profile can be used to assess an individual's propensity to separate if he or she were to be questioned and respond with a particular set of answers. The connection between our desired prediction model and ML takes the form of understanding which type of ML algorithm will provide us the most effective output, while ensuring we reduce errors, biases, and any other impacts to the validity of results.

D. CAUSALITY VERSUS PREDICTION

In more traditional statistical modeling techniques such as OLS and Multivariate Analysis the aim is to find causality. In machine learning methods, causality cannot be established; instead an inference or a prediction can be made about how an independent variable affects a dependent variable. Arkes (2019) identifies four main objectives for regression analysis: to qualify how one factor causally affects another, to forecast or predict an outcome, to determine the predictors of some factor, and lastly, to adjust an outcome for various factors. The multiple regression model that would allow the regression of many independent variables on a dependent variable is given as

$$Y_i = ß_o + ß_1 X_{1i} + ß_2 X_{2i} + \dots + ß_k K_{ki} + \varepsilon_{I},$$

where K equals the number of explanatory variables. A regression is an equation that represents how a set of factors explains an outcome (Arkes, 2019). In this case, we are operating a large number of variables, and as such, regression analysis may become too busy and the variable valuables predicted may become biased by much correlation, lack of variable independence and multi collinearity. This thesis ran traditional logistic regression models to determine whether the models can make predictions or causal inferences about the data and in particular identify key variables on the decision to separate. This large number of variables does make it difficult to identify the most important predictors, true relationships between variables or make any causal inferences about the target population.

Machine learning, on the other hand, is able to cope with many variables and is able to identify the most important variables and also identify relationships between variables.

In this case ML is able to absorb large quantities of data for analysis without limitation and, as Professor Tom Ahn asserted in an NPS class lecture, “throw the kitchen sink at it.” This application of ML then very well supports our efforts to apply a whole myriad of survey questions and their subsequent responses as variables to first identify the key variables; secondly, to identify trends, and lastly, to uncover inter-variable relationships that may not have been discovered with traditional statistical modeling techniques.

E. MACHINE LEARNING

Machine learning is a method of optimization that uses computer programming to analyze example data or past experience. In employing this method, we essentially use a model, which has some parameters set, and learning is the application of the computer program in optimizing the parameters of the training data. The model can be descriptive or predictive or even both (Alpaydin, 2010). This thesis does both; it uses the power of machine learning to make predictions of individual behaviors and their effect on separation, and it bolsters knowledge by examining relationships between independent variables and between the independent and dependent variables.

F. MANPOWER RESEARCH AND MACHINE LEARNING

Despite much searching, I have found little or no research that directly links military manpower modeling and machine learning. Even more obscure would be any reference to manpower modeling specific to sailors (technical or otherwise) based upon ML methodologies. To that end, I believe there is a real need to explore this type of analysis and the potential expansion of the application of ML in this space in the future within the RAN. This thesis breaks new ground by connecting the need for ML in manpower analysis in general and in particular for looking at behavior and sentiment and their links to separation decision making.

1. Supervised ML versus Unsupervised ML

Machine learning can largely be divided into supervised and unsupervised learning. There are other types less used, such as semi-supervised learning and reinforcement learning, which I will not cover. In supervised learning, the aim is to learn the dependent variable to the target variable relationship, where the value of this relationship is the supervisor (Alpaydin, 2010). In unsupervised learning, a target variable is not set, and as such, we analyze all variables collectively (Alpaydin, 2010). Using our labeled dataset, our supervised learning model algorithm will first learn on a training dataset, and then we evaluate its accuracy in providing the algorithm with which to predict or infer relationships from our data.

2. ML Algorithms

Common ML algorithms include Nearest Neighbor, Naive Bayes, Decision Trees, Linear Regression, SVM, and Neural Networks.

An example of the application of ML algorithms on manning issues not strictly related to the military, but which certainly helped my understanding of ML and its relevance in the manning domain, is the paper *A Comparative Analysis of Machine Learning Techniques for Student Retention Management* (Delen, 2010). The paper promotes comparative analysis by use of data mining and analytical models applied to historical data to better understand freshmen student attrition. The analysis utilized four classification methods: logistic regression, neural network, decision trees, and support vector machines. The paper amplified the ML benefits over those of linear regression. Logistic regression is a widely used statistical tool for classification problems; however, this method has some limitations with regards to assumptions of independence and normality, which have led to a rise in the popularity of ML methodologies to meet the demands of complex real-world issues (Delen, 2010).

SVM and the associated family of SVM models are used for classification or regression utilizing the value of the linear combination of features (Delen, 2010). Decision trees is a classifier model that is built of many decision trees, the output of which is the mode of the classes output broken down by the individual trees. Each tree is a simple function that will give an output when a set of predictor values is entered (Delen, 2010). Neural networks, on the other hand, are designed to mimic biology—in particular, our human brain—to apply analysis. This type of model is capable of analyzing very complex non-linear functions (Delen, 2010). The results of the student retention study showed that the SVM model produced the most effective output, with a prediction rate of 87.23% (Delen, 2010). The conclusion indicated that given the correct data with the correct variables, the ML methods applied could predict student attrition with approximately 80% accuracy (Delen, 2010). It is this sort of analysis that demonstrates that ML can be applied to manning issues.

Random tree algorithms are both a regression and a classifier. Random tree algorithms are quick computationally and are able to process numerous variables with varying measurement units (Mather & Mahesh, 2003). The structure of a random tree model consists of a collection of small trees or networks. In this model, input variable values in each tree are compared to the target variable and classified against the target. This comparison process is carried out across all trees, and the output class is the one that is identified most often; consequently, the response is the average of all the responses (Kalmegh, 2015). Although methods such as boosting, attribute selection, and pruning can improve the accuracy of random tree prediction by between 3% and 6% (Mather & Mahesh, 2003), these methods are not employed in this thesis in the random tree model application.

CHAID (Chi-square automatic interaction detector) is another ML model. Created in 1980, CHAID is a powerful tool for ascertaining relationships and their strengths between variables (Díaz-Pérez & Bethencourt, 2016). CHAID constructs a predictive model or tree system similar to that of the random tree method to ascertain how variables combine and best explain the outcome of the target variable. Like the random tree model, it also copes well with diverse variables having different measures and values such as ordinal or nominal. In CHAID there is no prior assumption of the distribution of the independent variables; thus, it provides a non-parametric statistical application of a free distribution (Díaz-Pérez & Bethencourt, 2016).

K-Means is an unsupervised clustering model and therefore has no target variable. Instead, it more sorts the data into clusters of related data points (Wagstaff, Cardie, Rogers, & Schroedl, 2001). The K-Means clustering method partitions the data in a number of groups (k) and then continues to create clusters and refine the models. During this process, each repetition creates new clusters and each data point is then allotted to its nearest cluster.

As such each cluster center is recalculated as part of the mean of the data points within the cluster. This process of creating clusters continues until there is no change in centers and the model has converged to give the most powerful cluster output (Wagstaff et al., 2001). It should be noted that K-Means method does not provide a model prediction accuracy as it has no target variable to predict. The analyses of cluster size and silhouette index value (SIV) are recognized as ways of assessing and validating K-Means models (Wang, Wang, & Peng, 2009). The SIV is reflective of the cluster tightness and the cluster separation. Higher SIV results indicate higher quality clustering results (Wang et al., 2009).

The process of partitioning a dataset by creating a training set, a validation set, or a test set is not unique to the ML analysis sector. It is applied in a myriad of analysis tools. In this case, the ML model is built on analyzing the training data, the portion that the ML model sees and learns from (Shah, 2017). The model is then validated or tested by applying the model to the validation or test dataset, and then the outcomes or results of the validation or test check are compared to the training model results to provide an evaluation of the model and value of prediction accuracy (Shah, 2017). Further cross validation can be carried out following the training, validation, and test data partition set for an unbiased evaluation of a final model by employing the test partition data (Shah, 2017).

The process of determining partition parameters or data split ratio is largely driven by the type of data and the type of models in which it is being applied. Certain model types require large amounts of data for training and learning; other models with few hyper parameters can be validated easily such that the data split ratio can be reduced, accommodating a smaller validation set affording more training data (Shah, 2017). A common method is random partitioning whereby the model randomly selects observations and allocates them to the training or validation set at a nominal ratio of 70:30. The partitioning can be set to the desired ratio to support improved model prediction accuracy.

The partitioning approach for this analysis was to use the default ratio set at 50:50; that is, 50% data allocated to the training set and the remaining data allocated to the validation set. The use of the 50:50 ratio allowed for a standard across all models tested as part of the down-selection process. The final model selected was run utilizing both the 50:50 partition ratio and the 70:30 partition ratio to compare prediction accuracies.

V. DATA SOURCES

The following chapter outlines the source, type, and make-up of the data used in this thesis, describing the variables and subsequently their application in the analysis. The chapter also lays out the foundation for the process of cleaning, refining, and validating the data in preparation for application in the ML models. Finally, the discussion provides an understanding of the ML methodology and the down-selection process for analyzing and determining a preferred model.

A. DATA SOURCE

To analyze behavioral differences between technical and non-technical sailors it is appropriate to look at both standard manpower data from the ADF human resources (HR) system alongside survey data, which could introduce some more broad and creative data. The HR data was used for comparison to the survey data to evaluate the representative nature of the survey question responses. The dataset was previously used in an NPS thesis titled, *Length of Service/Survival Profiles Methodology for the Royal Australian Navy* (Dodds, 2018).

The HR dataset was provided by Director Navy Workforce Requirements and contains transactional movement data, including enlistment and separation, between July 2002 and May 2018. The data was cleared of all personal information, and the separation portion was used to validate the representative nature of the survey data.

The survey data utilized in this analysis is from the ADF Exit Survey (A. Ryan, email to author, August 25, 2019), which has been offered to ADF personnel voluntarily separating from the service since 1999. The survey is delivered to all personnel separating from the service sometime after they apply to separate and before the point of actual separation. The survey is delivered electronically, is anonymous and non-compulsory.

The dataset consists of 4,864 observations and covers a question set equivalent to 255 variables. The data contains variables corresponding to officers and sailors who voluntarily left the RAN between 1999 and 2018. The data contains the following traditional demographic variables: branch category, worn rank, state, YOS, years in current rank, age group, gender, and family arrangements. The remaining variables consist of responses to various questions seeking to understand the nature of the reasons for separation. Examples of questions and responses are contained in Table 5.

Table 5. Example of ADF Exit Survey Questions and Responses. Adapted from Anthony Ryan (email to author, August 25, 2019).

Demographics	
Worn rank	SMN/AB/LS/PO/CPO
Age Group	24 years and under/25 to 44 years/45 to 54 years/55 years and over
Reasons for Leaving	
Issues with day-to-day unit management of personnel matters	N/A/ Not Important/Slightly Important/ Moderately Important/ Very Important/Extremely Important
Lack of confidence in Senior Defence leadership	N/A/ Not Important/Slightly Important/ Moderately Important/ Very Important/Extremely Important
Career Questions	
How many years have you spent at your current rank? (If less than one year write 0)	integer
Did you enjoy your career in the ADF?	Yes/No
How long were you deployed on your most recent deployment?	Less than 2 weeks 2–4 weeks/ 1–3 months/ 4–6 months/ 7–12 months/ More than 12 months

The dataset contains data covering the responses for SMN to WO. As previously explained, the WO responses will not be utilized for this analysis. Initially the data and observations associated with the responses of SMN to CPO were used for comparison of ML models and the down-selection process. Following the down selection of ML models, the main focus of the analysis concentrates on the AB and LS rates and between four and eight years of service. Variables were created to identify and differentiate between participants' completed YOS and category by technical or non-technical, and further by

AB and LS. By creating the YOS variable as a dichotomous variable indicating a value equal to 1 if the observation has reached a certain YOS milestone, is a technical or non-technical sailor and an AB or LS, we can identify behaviors that affect separation behaviors in specific periods of service.

The two- to seven-year period in a sailor's career generally coincides with changes in career milestones and conforms to general economic theory of job matching. That is, in the first two years of service likely impacts to separation behaviors can be related to individual-to-organization matching. Beyond the two years-of-service point more performance, individual preference, career milestones such as qualification and promotion, and pay and recognition factors help to better understand an attitudinal shift towards separation. Key milestones that helped shape the final selection of greater than or equal to four and less than six years of service and greater than or equal to six and less than eight years of service were the periods post matching, achievement of qualifications, IMPS, and promotion. Furthermore, the final selection of an interaction term including AB and LS was based upon the highest distribution of AB and LS within the dataset between four and eight years of service.

The creation of the dummy variable, which is a dichotomous variable identifying whether a respondent is a technical sailor or a non-technical sailor, was used to delineate the two categories for side-by-side comparison of results. These comparisons are the main driver for any assertions of differences in behaviors between technical and non-technical sailors.

Interaction variables, which are dichotomous variables identifying whether a respondent is a technical sailor or a non-technical sailor, interacted with YOS variables at four, six, and eight years and combinations of periods in between were created. These variables support the analysis of technical sailors and non-technical sailors by the YOS variables to allow effective analysis and support model down selection.

B. SURVEY DATA VALIDITY

The exit survey data is reported on annually by the ADF Directorate of Strategic Personnel Planning and Research (DSPPR) and used to provide advice on the survey outcomes to the ADF Senior Leadership Group (RAN, 1999; ADF 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2010, 2011, 2012, 2014, 2016, 2018). For the purposes of assessing the validity of the survey data, reports from 1999 to 2011, 2013, and 2017 were used to understand the level of research that had been applied to the data and any claims about the credibility of the data as a representative source. The response rate traditionally has been considered low, with the average response rate of 29.46% from RAN personnel voluntarily separating between 1999 and 2017. The highest response rate of 39% was in 2000 and the lowest rate, 20%, in 2013. Although the average response rate has been 29.46% (RAN, 1999; ADF, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2010, 2011, 2012, 2014, 2016, 2018), traditionally a response rate of about 30% for an anonymous and non-compulsory survey would be considered successful or high and likely representative of the sample population.

As far as guidance on the representative nature of the data within the reports, the advice to Senior Leadership Group (SLG) would generally be that the data analysis results are indicative. The exit survey continues to exhibit a lower than desirable response rate and results can be thought of as somewhat indicative rather than representative of all discharging personnel (ADF, 2005). This caveat about the exit survey data being more indicative than predictive was largely considered in relation to the representative nature of the data until 2010. In the 2010 report the data was considered representative. Continued analysis resulted in consideration that the data, despite the relatively low response rate to the exit survey, would be a good representation of discharging personnel (ADF, 2010). The exit survey report of 2010 suggested that although previous low responses gave rise to caution when interpreting the results, the 2010 average of 39% would indicate that it captures a fairly representative sample of the ADF community (ADF, 2010).

Given the average response rate of 29.46% for RAN personnel, it was necessary to employ other ways to validate the representative nature of the survey data. To that end the HR data provided by DNWR was utilized to make some overall comparisons of major demographics. The data was used to compare gender and age group as a proportion of the overall data. The result was a direct comparison between the overall demographic of the survey data and actual separation data. The results were very close and are contained in the

Table 6. As can be seen, the gender and age group data contained in both sets are very similar. The result being a comparison of 20.8% to 19.4% female, 79.2% to 80.6% male with both average ages within the HR data falling into to the age group 25 to 44, which in turn makes up 75.46% of the survey data. The overall examination of research and reports resulting from the survey data and the comparisons between the survey and HR data allows me to make a statement that the survey data is likely to be representative of the true population.

Table 6. HR and ADF Exit Survey Demographic Comparison

	Female/Male % Ratio	**Average Age Female to Age Group**	**Average Age Male to Age Group**
HR Data	20.8% to 79.92%	38.3 years	42.34 years
ADF Exit Survey Data	19.4% to 80.6%	*25–44 years	*25–44 years

*75.46% of ADF Exit Survey Respondents fall into the age category 25–44 years.

C. CLEANING AND REFINING

The survey data was cleaned and refined to remove those observations relating to officers and to remove variables with missing values at a proportion of greater than 80%, or less than 3,000 within the 3,721of total observations. The large number of missing values is mostly due to the development of the question set over the period of its delivery. Although the nature or intent of the questions remained the same, questions were added or removed between 1999 and 2018. As such a question that was added in recent years or removed early on in the cycle would have very few responses. Another reason for lack of values may be the participant's willingness to answer the question or to provide only nil responses.

Certain variables that pertain to future employment could not be used in the analysis. Unfortunately, the variables that provided this sort of information had greater than 3,000 missing values within the total so were considered too low to be representative and thus were eliminated. Examples of the variables are: "Is this area in your preferred area of employment," "Which option best describes your future occupation," "What industry is your new job in," "What Sector is your future job in"? The lack of responses to these types of questions could be due to the sailor not having secured employment when the survey was completed, or the questions themselves may have been introduced or dropped over time thus reducing responses.

Other variables were dropped because they may have little or no bearing on the decision to separate. Examples were "Last Transition Seminar Attended" and "Was the Seminar useful?" A member does not usually attend a transition seminar until he or she has already decided to leave, as such it will have no effect on the decision-making process. Other variables dropped were: "Would you consider rejoining?" and "Other reasons not covered?" It is likely both may add no value to the decision-making process to separate from the service.

Due to a lack of responses in the data at well over an average of 3,000 missing values, an early decision was made to not include reserve feedback and concentrate on the permanent navy force. As such all Reserve Service variables relating to joining or separation from reserves were omitted.

The final result for data to be used for the analysis included 3,721 observations and covered 106 variables, including created dummy and interaction variables. From the overall data two further sets of data were created. First, a dataset was created that contained all variables with the number of missing values of less than 3,000 in the 3,721 total observations. The variables with missing values greater than 3,000 were eliminated, resulting in a remaining 102 variables. Second, a dataset was created that eliminated all variables with missing values, resulting in a total of 814 observations with 93 variables. A complete list of variables can be found at Appendix A.

D. METHODOLOGY

This section will describe the methodology used to analyze models and down select to ascertain the most effective models in this analysis. Furthermore, it will describe the model accuracies and effectiveness ultimately leading to the identification of the preferred model.

1. Initial Model Selection

Initial model selection was carried out by applying the two datasets to a range of ML models, including 11 supervised and one unsupervised model to ascertain which models would (1) provide an output and (2) provide an output that can be interpreted effectively. Each model was applied to the data utilizing the all sailor with years of service greater than four as the target variable for the supervised ML models. The unsupervised ML utilized the data filtered to include those observations that contained data utilizing the all sailor with years of service greater than four.

All models were partitioned utilizing a 50:50 data split ratio for training and validation. Table 7 lists the various models tested against both datasets and indicates their success based on whether the model provided an output that was interpretable, as represented by a tick.

Table 7. Initial ML Model Selection Success Table

Test Models to See What Works (Target YOS>4)		
Model Type	Data (<3000 missing values in the variable)	Data (no missing values)
Random Tree	√	√
K-Means (unsupervised)	√	√
CHAID	√	√
Neural Net	**X**	√
Random Forest	√	√
PCA/Factor	**X**	**X**
SVM	**X**	√
Linear SVM	**X**	√
One Class SVM	**X**	√
KNN	**X**	**X**
Bayes Net	**X**	**X**
Logistic Regression	√	√

This resulted in the selection of Random Tree, K-Means, CHAID, Linear SVM, and Logistic Regression for further down-selection modeling. It should be noted that the Logistic Regression was limited to two target variables, those being YOS greater than or equal to four and greater than or equal to six for technical sailors. The Logistic Regression models analyzed serve to highlight that in this case traditional econometric modeling techniques would not provide a valid or useful output on this type of survey data.

The bulk of analysis was on the second dataset with no missing values as it is the most appropriate. This dataset provides the complete suite of responses across all variables. As such, this dataset is the most representative of the target population given personnel have likely completed all questions and answered in good faith.

The dataset with missing values of less than 3,000 was initially used for the all sailor analysis as it gave the greatest representation in general for understanding all sailor behaviors identified in the data. Also, it was useful for comparison to a more refined analysis by YOS and by technical and non-technical. The performance of this dataset, however, was poor during the all model down-selection process. This is likely due to the amount of missing data and its classification within the models.

Initially the created time period variables of YOS greater than or equal to two, greater than or equal to four, greater than or equal to six, and greater than or equal to eight plus the same variables interacted with technical and non-technical indicator variables were used for the all sailor, technical and non-technical sailor analysis, respectively. This allowed for testing of models in various target variable configurations for the down-selection process and provided a basis for some continuity across models when evaluating performance. Furthermore, this process gave early insights into the predicted attitudes and behaviors of sailor groups, along with some better understanding of model behaviors when analyzing target variables.

The modeling results can be found at Appendix B. It should be noted that for technical and non-technical sailor analysis, YOS greater than two was not utilized because all values for this variable are greater than two. All predictors tabled constitute the top ten predictors in any model.

2. Model Accuracy

The down-selected models of Random Tree, K-Means, Linear SVM, CHAID, and Logistic Regression were tested for predictive power and the results tabled for comparison. For Random Tree, Linear SVM, CHAID, and Logistic Regression the prediction accuracy is conveyed as a percentage of correct predictions within the total predictions. The accuracy of the K-Means model as an unsupervised model is given as a value percentage of the largest cluster and the SIV and description. Table 7 details the results for predictive accuracy.

As can be seen in Table 8, the results of prediction accuracy can vary in range between the models. Random tree displays high variation in accuracy between 41.5% and 97.7%. In one target variable output the random tree model only gave eight predictors, and given that in all cases I am analyzing the top ten predictors, this leads to a lack of confidence in the model. That is, most other models have greater than ten predictors. Altogether these model behaviors lead me to surmise that the model may not be behaving particularly well within this dataset. This was also confirmed with some considerable variation in predictor outputs when compared across target variables, which can be reviewed in Appendix B.

The CHAID model displays accuracy variation between 58.4% and 89.3% and in one target variable provided the lowest prediction accuracy. In one target variable case the model only provided five predictors. Like the random tree model this limit in predictors complicates model comparison and leads to a lack of confidence in the model.

By contrast, K-Means outputs are probably the simplest to analyze across other models. Although there appeared to be some consistency in predictor outputs, with this being the only unsupervised model its poor SIV results give little confidence in the model's output.

Linear SVM, although it does not have the highest prediction accuracy of all models, consistently ranked highly in accuracy across all target variables and had the least accuracy variation, between 59.53% and 91.08%. More significantly the model displayed the most consistent behavior and least variation among predicted behaviors when comparing the Linear SVM across different target variables. The predicted behaviors and the comparisons across target variables can be reviewed at Appendix B.

Logistic regression, on the other hand, scored very poorly for prediction accuracy across the two target variables tested, at 10.36%. The results and parameter estimates were largely uninterpretable and of little value when trying to analyze key predictors of the target variable. This was the expected result based on the complexity and sheer number of variables and only supports the theory that traditional statistical modeling techniques may not be always suitable for analyzing this type of data.

Table 8. Model Prediction Accuracy Table

Target Group	All Sailors	Technical	Non-technical
Data set	**<3000 Missing Values**	**No Missing Values**	**No Missing Values**
Target Variable YOS => 2 years	**Model Prediction Accuracy**	**Model Prediction Accuracy**	**Model Prediction Accuracy**
Random Tree	97.7%	N/A	N/A
CHAID	68.6%	N/A	N/A
K-Means	SIV = 0.1 Poor Largest Cluster 31.7%	N/A	N/A
Target Variable YOS=> 4 years	Model Prediction Accuracy	Model Prediction Accuracy	Model Prediction Accuracy
Linear SVM	N/A	63.13%	59.52%
Random Tree	79.6%	61.6%	59.8%
CHAID	58.4%	67.47%	63.13%
K-Means	Silhouette = 0.1 Poor Largest Cluster 32.1%	Silhouette = 0.1 Poor Largest Cluster 42.6%	Silhouette = 0.1 Poor Largest Cluster 42.4%
Target Variable YOS => 6years	Model Prediction Accuracy	Model Prediction Accuracy	Model Prediction Accuracy
Linear SVM	N/A	75.66%	72.53%
Random Tree	82%	75.8%	74.2%
CHAID	80.6%	82.41%	77.35%
Log Reg	N/A	10.36%	N/A
K-Means	Silhouette = 0.1 Poor Largest Cluster 31.1%	Silhouette = 0.1 Poor Largest Cluster 42.6%	Silhouette = 0.1 Poor Largest Cluster 42.6%
Target Variable YOS=> 8 years	Model Prediction Accuracy	Model Prediction Accuracy	Model Prediction Accuracy
Linear SVM	N/A	92.29%	91.08%
Random Tree	92.1%	91.7%	89.4%
CHAID	89.3%	91.08%	90.84%
K-Means	Silhouette = 0.1 Poor Largest Cluster 31.7%	Silhouette = 0.1 Poor Largest Cluster 29.6%	Silhouette = 0.1 Poor Largest Cluster 42.6%

Target Group	All Sailors	Technical	Non-technical
Data set	<3000 Missing Values	No Missing Values	No Missing Values
Target Variable All YOS	Model Prediction Accuracy	Model Prediction Accuracy	Model Prediction Accuracy
Linear SVM	N/A	63.61%	62.41%
Random Tree	41.5%	57.8%	60.9%
CHAID	85.4%	68.43%	68.43%
Log Reg	N/A	10.36%	N/A
K-Means	Silhouette = 0.1 Poor Largest Cluster 31.7%	Silhouette = 0.1 Poor Largest Cluster 42.6%	Silhouette = 0.1 Poor Largest Cluster 29.1%

3. Preferred Model and Data Set

Given the down-selection process and the model output comparisons, the preferred model was the Linear SVM. The Linear SVM results were the most consistent across all target variables, with consistent predictors and no excessive variation therein. The SVM provided consistently high prediction accuracies among other models when compared within the same target variable outputs. The Linear SVM from the outset displayed largely logical and intuitive predictors, which increases overall confidence and more specifically interesting behavior comparisons between technical and non-technical sailors.

The down-selection results and prediction accuracies not only allowed analysis of the models but also the two datasets. The dataset with all variables with fewer than 3,000 missing values was inconsistent in both predictors and prediction accuracies across the models that could provide an output result. As such, it was decided to utilize only the data with no missing values for the final analysis. The dataset with no missing values was used to focus the Linear SVM on analyzing the technical and non-technical sailor behaviors. The analysis uses the time period variables of greater than or equal to four and less than six YOS and greater than or equal to six and less than eight YOS. Models of all sailors were also completed to allow for comparison across the two time periods. Each final Linear SVM model was run with partitioning dataset ratios, training data to test data of 50:50 and 70:30, respectively.

THIS PAGE INTENTIONALLY LEFT BLANK

VI. FINAL ANALYSIS, FINDINGS, AND RESULTS

This chapter defines the process for the final analysis, detailing which target groups were analyzed. It explains the findings that arose from both individual analyses of the target groups and also by comparison of groups. Last, all model outputs are detailed in full.

A. FINAL ANALYSIS SET-UP

The final analysis resulted in five specific target variable models displayed in Table 9. Models were run based upon the two periods of four and five total YOS and six and seven total YOS. This is to take into consideration the IMPS or earliest point of voluntary separation of four years for the non-technical sailor community and six years for the technical sailor community. The target demographics within those periods were varied across the models for all sailors, technical only, and non-technical only.

Table 9. Target Variables Used for Final Analysis

Model	Target Variables
1	Non-Technical Sailor with four and five total YOS.
2	Non-Technical Sailor with six and seven total YOS.
3	Technical Sailor with six and seven total YOS.
4	All Sailors with four and five total YOS.
5	All Sailors with six and seven total YOS.

The sailors analyzed are of no particular exit year, age, gender, specific category, relationship profile, or family make-up. This was driven by the dataset, as refining the target variable to a more specific target quickly reduced the number of analyzed variables to a point of reduced validity. As such the highest level of refinement of the sailor profile is at the technical or non-technical level within the time periods previously explained.

The first target variable for analysis is the non-technical sailor at four and five completed YOS. That period encompasses the first available separation point and the following year. In this case, there is no comparison to the technical sailor. This is due to the minimal number of technical sailors that separate at the four- and five-year period. That is, technical sailors that separate in this period are likely doing so prior to IMPS, and therefore, their reasons for separating may be involuntary or driven by uncommon behaviors. Furthermore, because of the aforementioned reasons the technical sailor equivalent results for that time period may be skewed and unreliable for comparison. In this target period, I therefore analyzed only the top ten behavior predictors.

The second target group is the non-technical sailor group at the six and seven total YOS. This range represents the second year after IMPS and also captures that period in which many sailors promote from AB to LS. It may also be reflective of changes in the likelihood of increased family commitment due to getting married or having more children. These changes may be reflected in the attitudes displayed by the sailors in the earlier period. This target group was compared to the third target group. This allowed the direct comparison between technical and non-technical sailors.

The third target group is technical sailors, at the six and seven total YOS. This encompasses the sixth and seventh YOS and represents the first separation point at IMPS and also captures the following year and that period in which many technical sailors promote from AB to LS. This group was the first to look specifically at the technical sailor and also support the technical versus non-technical sailor comparison.

The fourth and fifth target groups analyze all sailors at four and five total YOS and six and seven total YOS, respectively. These two target groups allow for comparison of behaviors as the sailor passes through the two time periods, and seeks to identify differences during that transition. The two target groups are utilized as a comparison when looking at overall behaviors of both technical and non-technical sailors. That is, the comparison supports the identification of behaviors that are common across both technical and non-technical sailors and those that differentiate technical and non-technical sailors.

All the models covering the five target groups were run twice against two partition ratios for training and test purposes. The model prediction accuracies by partition ratio are shown in Table 10. Each target variable model with the data partition ratio providing the highest prediction accuracy as highlighted in green in Table 10, were selected for the side-by-side behavior predictor comparisons.

Table 10. Linear SVM Final Analysis Model Prediction Accuracy

Target Variable Structure	Linear SVM Prediction Accuracy with 50:50 Data Partition Ratio	Linear SVM Prediction Accuracy with 70:30 Data Partition Ratio
Non-Technical Sailor, 4 and 5 total YOS	68.67%	67.9%
Non-Technical Sailor, 6 and 7 total YOS	68.92%	70.37%
Technical Sailor, 6 and 7 total YOS	72.77%	72.84%
All Sailors, 4 and 5 total YOS	69.4%	67.9%
All Sailors, 6 and 7 total YOS	68.43%	65.84%

B. RESULTS

In the interpretation of these results some key themes arise centered around pay, lack of use of skills, an attraction to civilian employment, and a general sense of lack of recognition. There are aspects of all of these that are seen across all models affecting all sailors. There are also some similarities between technical versus non-technical sailors across each time period, which are covered in more detail later in the chapter. More interesting are the predictors that differentiate technical versus non-technical sailors, with two clear overall attitude or sentiments represented in each.

Some unexplained variables do not seem logical or intuitive based upon the period of service, or only appear once. This is attributed to some randomness in the data and is generally accepted with data of this nature, given that some respondents may have answered incorrectly by mistake, misread the question, or just misunderstood the question. One example of these is access to a pension. Ordinarily sailors at four or five YOS are not entitled to a pension; however, given the targets are not age specific there may be older sailors in that period of service who do have access to a civilian pension of some type. Another example is the question about inadequacy of rental assistance, which is a high predictor of all sailors at six or seven YOS, and only occurs in this model.

There specific sentiments are consistent across groups that all belong to a similar theme and are identified by differing predictors of a similar nature. Examples of these are predictors related to pay, recognition, and the attractiveness of civilian employment. These may give rise to respondents answering similarly related questions with varying responses as they may feel they have covered that aspect in another question. An example of this is strongly agreeing to "dissatisfaction with pay" but responding neutrally to "attractive civilian pay rates."

1. Results for Non-technical Sailor with Four or Five Total YOS

The first model covers the behaviors of non-technical sailors at four or five total YOS. There is a definite trend in the behaviors associated with lack of recognition, with two key respondents' strong predictors being "feeling that my contribution to the ADF was not valued or appreciated" and "lack of recognition or credit for work done." The lowest predictor in the top ten is "more attractive salary package in civilian employment." This pay-related theme in this group is not especially pronounced but certainly increases in later years. This could be due to early career expectations and personal reservation wage as increases in reservation wage may coincide with someone better understanding their value in the civilian sector. This pay sentiment is related to the highest predictor of "Attractiveness of a civilian job." Although not directly stated, the transition in employment may be related to at least similar or greater pay scales in the new employment.

In this model, the predictor "General dissatisfaction with service life," can easily be related to other predictors in the set that would lead to this dissatisfaction, such as "insufficient resources to do my job" and "Lack of maintenance of trades skills due to civilianization or outsourcing." In essence, dissatisfaction with service life as represented by this data is quite generic so not particularly useful in a prediction model. The last theme is related to career management, with high predictors of "issues in information provided on my career management" and "uncertainty with career plans." Overall, we can identify three prominent themes in this group of non-technical sailors; namely, lack of recognition, pay, and career management. The top ten behaviors for non-technical sailors separating in years four or five are given in Figure 12.

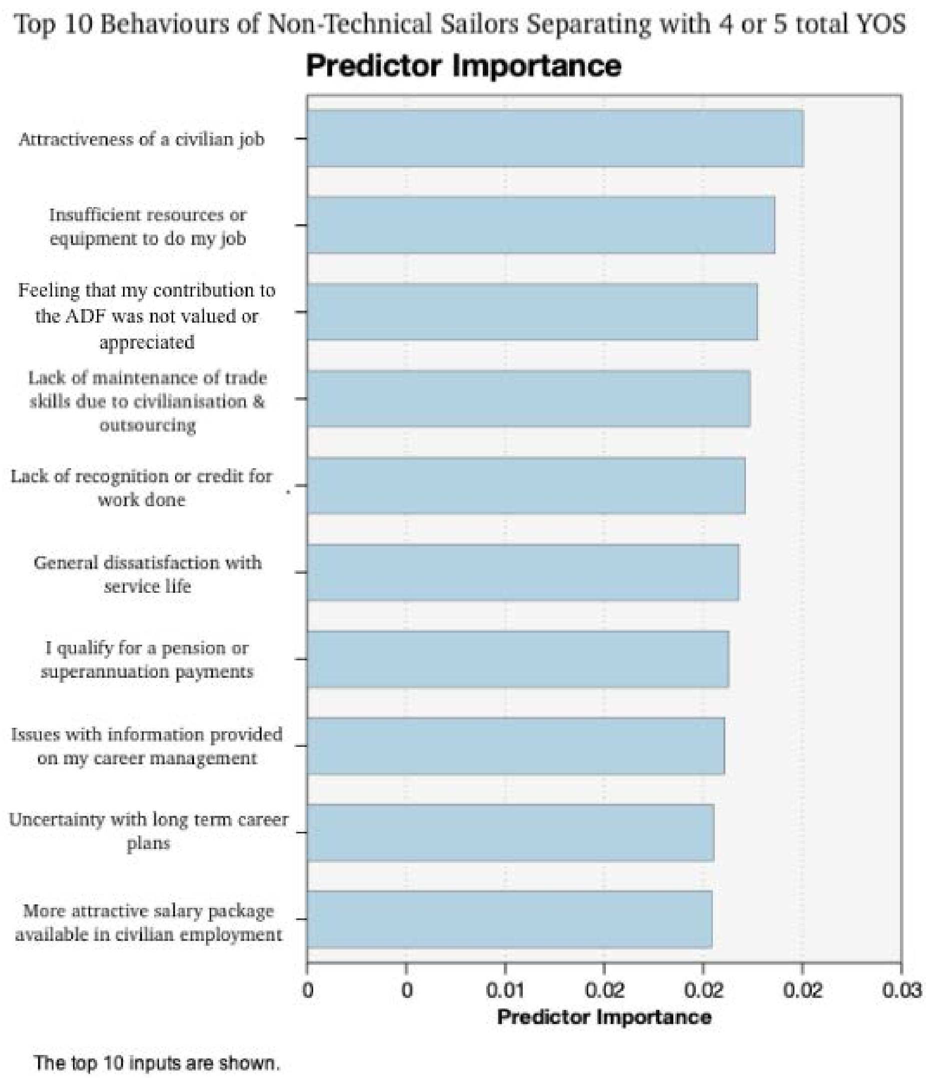

Figure 12. Top Ten Behaviors of Non-technical Sailors Separating with 4 or 5 Total YOS

2. Results for Both Technical and Non-technical Sailors at Six or Seven Total YOS

In this section each group is discussed in isolation, followed by a comparison analysis for differentiators.

a. Results for Technical Sailors at Six or Seven Total YOS

The technical sailor group separating in year six or seven had a very strong trend with respect to the understanding of their value in the civilian sector and the availability of work opportunities in that sector. These responses include: “Attractiveness of a civilian job,” “Good chance of finding a civilian job in the current economic climate,” “An offer of civilian employment,” and “Better career prospects in civilian life.” These are clear indicators of the technical sailors understanding the value of the high-quality skills achieved in service and their value in the civilian sector. This also supports the hypothesis that the technical community is sensitive to civilian technical industry growth.

The next prominent theme in the technical sailor group is the sentiment of lack of opportunity to utilize their skills due to civilianization of the military workforce. This is present in the two key predictors of “Greater integration of civilians in the workforce” and “lack of maintenance of trade and specialist skills due to civilianization or outsourcing.” The outsourcing of technical skills is often utilized to bring about efficiencies within the ADF; however, given its prominence in these results, the RAN may be underestimating its effect on retention. There are other key predictors like: “Desire to stay in one place,” “Difference between married and single entitlements,” and “lack of DHA housing” that may be related but are more likely isolated and as such difficult to interpret, without some relation to another comparison group. Overall, for technical sailors separating in years six or seven, the key themes are a general awareness of their value in the civilian sector and the lack of opportunity to employ key skills. The top ten behaviors for technical sailors separating in years six or seven appear in Figure 13.

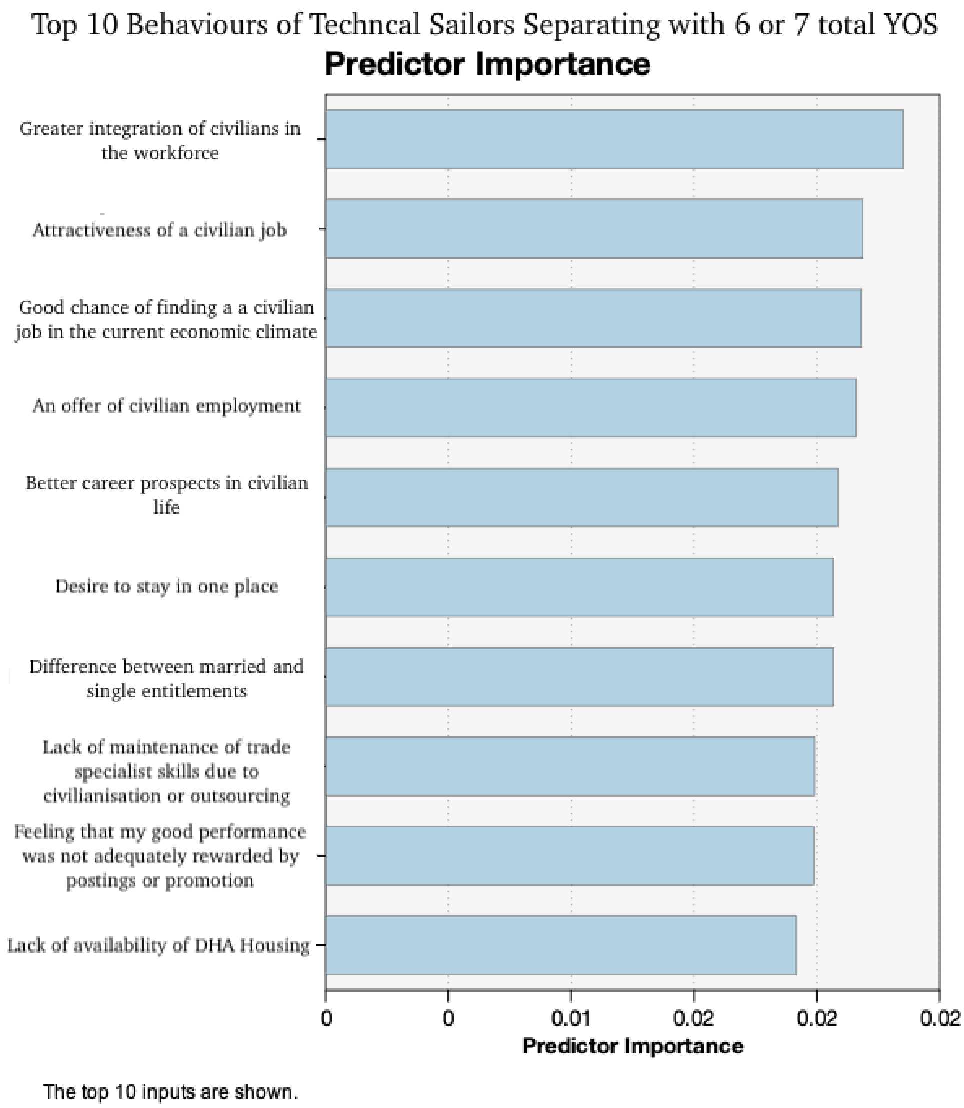

Figure 13. Top Ten Behaviors of Technical Sailors Separating with 6 or 7 Total YOS

b. Results for Non-technical Sailors at Six or Seven Total YOS

The non-technical sailor separating in years six and seven display sentiment in two main areas: pay and the amount of sea service. The top five predictors in this case are related to these two themes. Pay is the strongest sentiment among three of the top five predictors, which include: “Little financial reward for what would be considered overtime in the civilian community,” “more attractive salary package available in civilian employment,” and “Dissatisfaction with pay.” The next strongest prediction sentiment is associated with time at sea, in general, and on operations, with “Too much time spent on operational deployment” and “Sea service” being the remaining two of the top five predictors. Another predictor “Excessive work hours” may be related to the sentiment of too much time at sea, as sea service and deployments are generally linked to high workloads.

Other predictors in the non-technical group are isolated and include “Under use of or non-use of training and skills,” which is similar to predictors in the technical community. Also “My spouse or partner’s attitude to service life” is the only family-related predictor in this particular group. These types of sentiment become more prevalent in the all sailor group, as discussed later. Finally, “Personal experience of harassment or bullying” is the weakest predictor within in the ten highest and is unrelated to any other predictors. It should be noted that “Personal experience of harassment or bullying” is the only occurrence of a predictor of this nature in all of the models. The top ten behaviors for non-technical sailors separating in years six or seven appear in Figure 14.

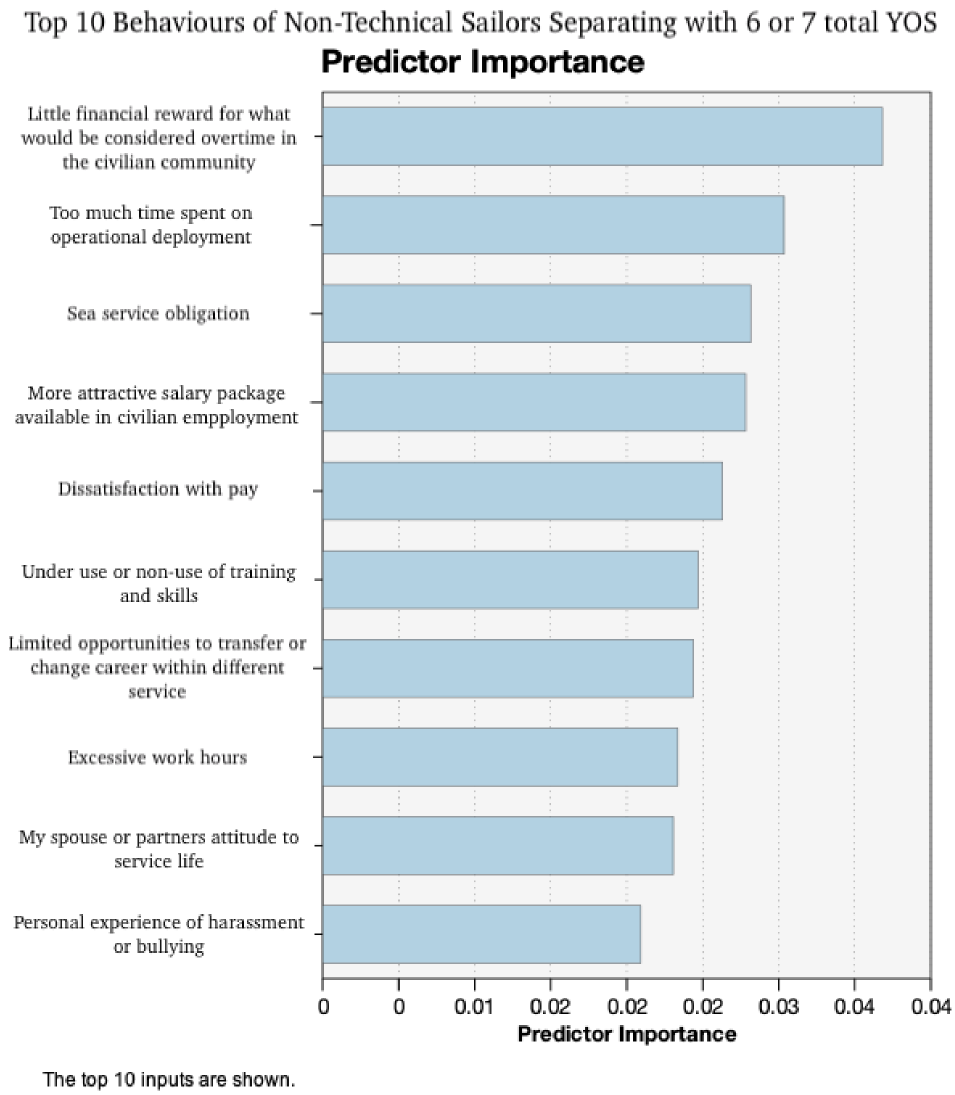

Figure 14. Top Ten Behaviors of Non-technical Sailors Separating with 6 or 7 Total YOS

3. Results from Comparing Both Technical and Non-technical Sailors with Six or Seven Total YOS

A comparison of the technical and non-technical sailor groups in the period of separation at years six and seven shows some similarities, but there are clear differentiators between both groups. One similarity is based around lack of use of skills, which is much stronger in the technical community, with three strong predictors compared to one in the non-technical group. Another similarity is a subtle link between the two in lack of reward for commitment and hard work. Both share this sentiment but the two groups differ in type of reward sought. In the non-technical sailor group the focus is financial, with three strong predictors related to pay. In contrast, the technical community sees a lack of reward in postings or promotions in the predictors, with "Feeling that my good performance was not adequately rewarded with postings or promotions" making the top ten.

There are key differentiators in overall sentiment between the two sailor groups in separating in years six and seven. These are the technical sailors' awareness of their value in the civilian sector and the impact of civilianization of the military workforce creating a lack of opportunity to employ key skills. The non-technical sailor, on the other hand, has a clearly focused sentiment in pay-related concerns and lack of recognition of effort or commitment. These differentiators are quite pronounced and not completely surprising given non-technical sailors receive lower wages and the lower wage can be construed as a lack of recognition. Conversely, the technical sailor has a higher wage, so is less likely to be linked with a lack of recognition and more likely satisfied with pay in comparison to non-technical sailors. Given this higher wage, the technical sailor will have a clear understanding of the value of his or her technical skills in the civilian sector, as this is recognized in service also. Thus, this thesis has succeeded in analyzing the key differences between these two groups. As is discussed in the final chapter, however, caution must be shown in applying these predictions to policy decisions.

4. Results from Comparing All Sailor Types between Four to Five and Six to Seven Total YOS

The final predictor comparison is analyzed to identify YOS- related predictors that vary as the sailor transitions between the periods that cover entering the earliest available separation point and possible promotion from AB to LS.

a. Results for Four to Five Total YOS

For all sailors identified in this period the key predictors are related to pay and postings, which feature highly in the top ten. These include "The nature of future postings" and "probable location of future postings." The pay-related predictors include "More attractive salary package in civilian employment," "Dissatisfaction with job related pay and allowances," and less obviously, "Attractiveness of a civilian job." Other predictors, which are similar in nature, are "Feel there is a lack of opportunities for career development" and "Limited opportunities to transfer or change career in the RAN." This leaves a couple of predictors among the top ten that are independent but have been prevalent in the technical and non-technical sailor comparisons, which are "Insufficient equipment or resources to do my job" and Greater integration of civilians into the workforce." The top ten behaviors for all sailors separating in years four or five are shown in Figure 15.

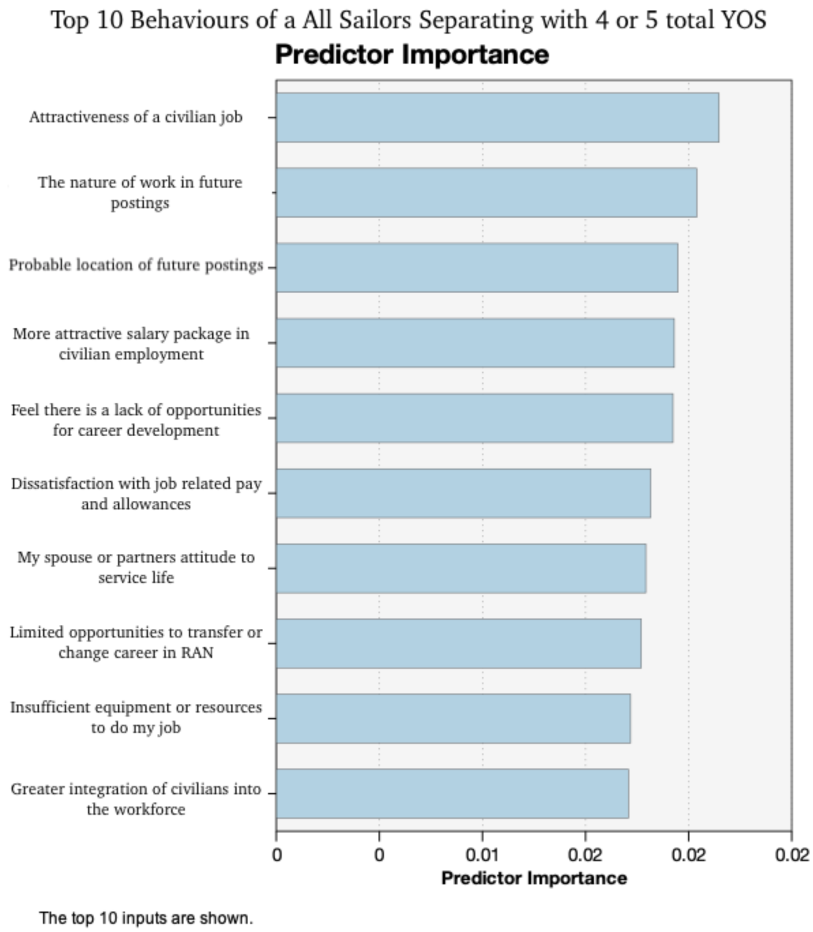

Figure 15. Top Ten Behaviors of All Sailors Separating with 4 or 5 Total YOS

a. Results for Six to Seven Total YOS

The key predictors for all sailors identified in this period are highly related to postings, family impacts, and the links between the two aspects. This is quite intuitive as sailors are likely to change family composition during this time, possibly getting married and having more children, and likely to experience the subsequent impacts of service life on the family. The predictors that enforce this sentiment about postings and family impacts are "Too much time spent away from home because of military duties," "Lack of respite postings," "Effect of postings on children's education," and "Inability to secure back to back postings during a critical stage of family and personal life."

The other sentiment seen in this group, which again has been a common theme in the technical and non-technical sailor groups, is the lure of civilian employment. In this case the predictors are “Attractiveness of a civilian job,” “An offer of civilian employment,” and “Lack of skills accreditation for civilian employment.” These predictors show that the sailors’ transitions in experience, possible promotion, and family make-up form the basis for assessing the benefits and drawbacks of service and for directly comparing them to civilian employment. The top ten behaviors for all sailors separating in years six or seven are shown in Figure 16.

From a comparison of the two periods of greater than or equal to four and less than six YOS and greater than or equal to six and less than eight YOS, it is possible to see some similarities in key predictors but also clear differences between the two periods. This is likely due to the difference in the type of sailor separating in the earlier period as compared to the later one. This is explained in that the earlier period is likely to be populated with more non-technical sailors and sailors with a lesser family commitment (e.g., not married and with fewer children). The later period is therefore likely populated with sailors—both technical and non-technical—who separate and with sailors who have possibly greater family commitments. This theory is supported by movement in predictors across the two periods. In particular, there is a greater focus on postings and the impacts on family in separations in years six and seven, and only one predictor of this nature in separations in years four and five. Conversely, pay concerns are far more prominent predictors in the separations in years four and five. Again, this is intuitive as earlier years of service are related to lower pay. Subsequently the transition to later years naturally gives rise to increased pay and a shift in sentiment in focus on other issues like the family impacts of service life.

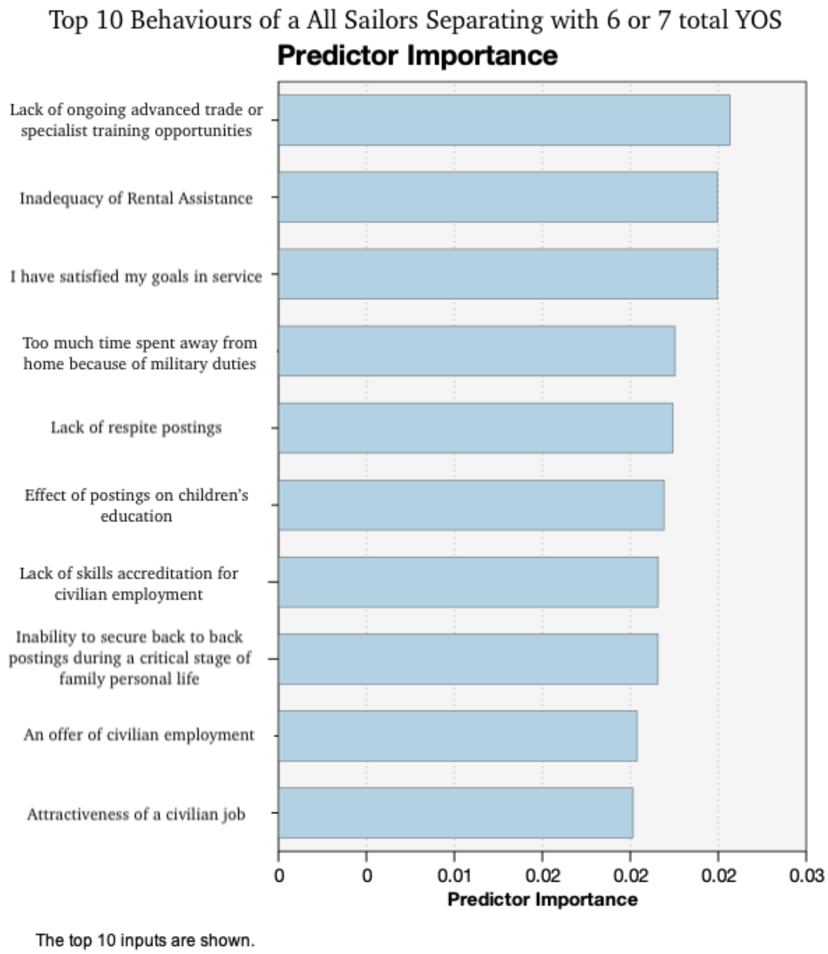

Figure 16. Top Ten Behaviors of All Sailors Separating with 6 or 7 Total YOS

As can be seen from the results and the comparisons of sentiments and attitudes of sailors across different groups and times, ML is very capable of analyzing these effectively and identifying the differences in these sentiments. In this case, the analysis has been quite broad in its approach, as it has not covered any specific individual demographics such as gender, age, and family constitution. The analysis was limited to the delineation of technical or non-technical sailors in the two periods selected. Of course, these target groups and time periods were deliberately selected to capture the best results in the most relevant periods. That said, the analysis was limited to identifying behaviors and differences in the groups discussed and as such has some limits in its application to direct policy.

VII. CONCLUSION AND RECOMMENDATIONS

This chapter summarizes the key findings and identifies potential policy recommendations and areas for future research. Finally, it offers a conclusion to the research covered by this thesis.

A. SUMMARY OF KEY FINDINGS

The analysis was able to identify predictors of separation from the service for all sailors, technical as well as non-technical sailors. Furthermore, it was able to clearly identify differences in behaviors between the sailor groups and separation in the two periods studied. Some clear differences emerged between the technical and non-technical sailor communities and also in their likelihood of separating at four and five years of service in comparison to six and seven years of service.

The technical sailor community clearly showed strong predictors of sentiment with respect to the understanding of their value in the civilian sector and the availability of work opportunities in that sector. Moreover, this community showed strong predictors of sentiment related to lack of opportunity to utilize their skills due to civilianization of the military workforce. Conversely, the non-technical sailor community's sentiments related to separation were prominently based in pay and recognition of effort and commitment.

The analysis of all sailors in the period greater than or equal to four and less than six YOS and in the period greater than or equal to six and less than eight YOS clearly identified a change in sentiment between the two periods. The earlier period identified pay and postings as key predictors of separation behaviors. In the later period, by contrast, sentiment shifted more towards family-related predictors along with the impact of service on those families.

Overall there were similarities and similar themes that emanated across all models, such as pay and civilianization of the workforce. Nevertheless, the models were able to identify key group and time period differentiators.

The application of ML in this context and its effectiveness was proven. The final model selection of Linear SVM was able to accurately identify key behavioral predictors of both technical sailors and non-technical sailors and subsequently the differences. This has demonstrated the need for further research in this area.

B. POTENTIAL POLICY CONSIDERATIONS

Despite the limitations of the data detailed in Chapter V, it is reasonable to consider the results of this study as a success in the sense that the capability of ML and its application in this method of behavior analysis is applicable and valuable. However, the results are not the proverbial silver bullet. As such it would be remiss of me not to afford some caution in consideration of using these results as they stand to drive any real policy decision making. That said, with resources applied to further recommended research and associated activity, it is highly likely that further analysis of this nature will support future retention policy. Despite the reservations expressed in the earlier statement, the concept of the application of ML in analysis of separation behaviors is sound and may support the following potential policy initiatives.

There may be scope to utilize this type of analysis to refine divisional system monitoring, recognition, and intervention. In this case, ML may be able to support profiling of sailors more effectively during divisional interviews when displaying known separation attitudes, behaviors, and sentiment. The profile of the sailor, including those sentiments among high predictors of separation, can then shape suitable intervention measures as required, such as more specific career management, incentives, postings, and training.

The use of this type of analysis to better profile sailors at risk of separation in the workplace will directly support the recent DCN signal detailing the vision for supporting the efforts of the RAN to increase its recruitment numbers and reduce attrition (RAN, 2018). The analysis may support the recommendations for greater individual support and personal focus within the divisional system to reduce separation (RAN, 2018). This analysis can also apply to larger groups above the individual level, focusing on workgroups or categories or areas of capability by analyzing other surveys, such as culture surveys and snapshot surveys that also capture sentiment responses.

A review of the ADF Exit Survey policy is needed. First, policy must be revised to accept this type of analysis as potentially a part of normal business. Further, it should make survey completion more robust in turn to support the analysis. This will provide better survey data on which to analyze and subsequently better advise commanders in making decisions on retention policies.

C. FUTURE RESEARCH AND ASSOCIATED ACTIVITY RECOMMENDATIONS

First, free text answers within the dataset in this instance were not used to support the analysis, as the dataset contained only approximately 10% of personnel who actually took the time to write some feedback associated with question, "Why are you leaving?" As such, those responses were not used in this analysis. Continued analysis of exit survey data, possibly across the whole of ADF, to look at a larger dataset and include some method of sentiment analysis to free text answers may provide valuable results. This would allow quantification of the sentiment, negative or otherwise, to be factored into the analysis and separate those leaving who provided positive feedback.

Second, the dataset used in this analysis included only the exit surveys of sailors to meet the thesis objective. I recommend expanding the focus to a more holistic group to implement a larger dataset and better exploit the ML applications in predicting separation behaviors of a wider set of personnel, possibly the whole of RAN. This could also be applied across a longer period of time to ascertain differences through periods of major change, such as culture changes like the New Generation Navy or changes to service conditions such as the Military Superannuation & Benefit Scheme.

Third, I would consider a research topic that specifically looks at a comparison of the output of ML versus output from traditional modeling techniques. Possibly identifying a specific critical branch in service and applying both ML and traditional regression modeling might be done to see whether there is a correlation or disparity between both methods. Such an approach could also identify the value of ML beyond or in complementing traditional modeling methods.

Fourth, although the analysis was able to identify differences in attitudes and behaviors at the level of technical versus non-technical sailor, further analysis by specific technical category using the same methodology may identify specific behaviors affecting branches with acute retention issues. It could also identify findings that may support highly focused retention policy.

Last, analysis of the exit survey and its implementation is needed to examine ways of improving the response rate and quality. The premise would be first to explore how to improve the response rate, up to and including making completion of the survey compulsory although that may impact response quality. Furthermore, it would be useful to analyze the application of the survey to better capture sentiment or attitude at the point of deciding to separate. Such analysis would improve the data quality and the representative nature of the target population survey data.

D. CONCLUSION

This thesis set out to support DNWR in the ongoing critical sailor retention efforts of the RAN, by analyzing the separation behaviors of RAN sailors using ML. Furthermore, it set out to identify the validity of ML analysis in this context and its application to exit survey data such that it could identify key predictors of sentiments, attitudes, and behaviors of sailors. In particular, this thesis identifies differentiators in those sentiments, attitudes, and behaviors between technical and non-technical sailors and the whole sailor community across critical career milestones like IMPS.

Consequently, the analysis has also proven ML to be effective in this context. The outcome of this was first to confirm that ML is very capable and applicable in analyzing survey type data gleaned from responses to attitudinal questions. Second, it confirmed that ML could not only analyze the data but provide outputs clearly defining separation behaviors. Finally, the outcomes of the analysis identified that in this case traditional statistical modeling techniques, such as logit or probit, are much less effective in analyzing this type of data when compared to ML. This application of ML may also be considered for other types of applications. For example, with further research and analysis, areas such as talent identification, unit performance, and medical categorization may be considered appropriate for the application of ML technologies in analyzing these areas.

Overall, I can say that the research and results within this thesis are very positive with respect to both the application of ML and in analyzing attitudinal data, and this study has set a foundation for a future research and its application in supporting retention policy in the RAN.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX A. VARIABLES TABLE ADAPTED FROM THE ADF EXIT SURVEY QUESTION SET DATA

Table data retrieved from Anthony Ryan (email to author, August 25, 2019).

Question/Variable	Response Value
Cells in italics and bold indicate variables that were omitted once all missing values had been accounted for	
Demographic	
Technical (dummy variable created)	0/1 Indicator
Non-Technical Technical (dummy variable created)	0/1 Indicator
YOS Greater than or equal to 2, 4, 6, and 8 years Technical (dummy variable created)	0/1 Indicator
Interaction Variables Technical*YOS Greater than or equal to 2, 4, 6, and 8 years (dummy variable created)	0/1 Indicator
Interaction Variables Technical*YOS Greater than or equal to 4 and less than 6, & Greater than or equal to 6 and less than 8 (dummy variable created)	0/1 Indicator
Interaction Variables Non-Technical*YOS Greater than or equal to 2, 4, 6, and 8 years (dummy variable created)	0/1 Indicator
Interaction Variables Non-Technical*YOS Greater than or equal to 4 and less than 6, & Greater than or equal to 6 and less than 8 (dummy variable created)	0/1 Indicator
Interaction Variables AB*YOS Greater than or equal to 4 and less than 6, &, Greater than or equal to 6 and less than 8 (dummy variable created)	0/1 Indicator
Interaction Variables LS*YOS Greater than or equal to 4 and less than 6, & Greater than or equal to 6 and less than 8 (dummy variable created)	0/1 Indicator
Year Surveyed	E.g., 2018
What is your Category?	E.g., Maritime Logistics
Worn rank	SMN/AB/LS/PO/CPO
State (derived from locality)	NSW/ACT/SA/QLD/WA/NT/VIC
How many years have you spent at your current rank? (If less than one year write 0)	integer
In total, how many years of service have you completed with the RAN (If less than one year write 0)	Integer
Age Group	24 years and under/25 to 44 years/45 to 54 years/55 years and over
Do you identify as?	Male/Female/Prefer not to say
What are your family/living arrangements? (Note, if a dependent child/child lives with you some of the time please select living with you)	Couple - living together, no dependent child / children Couple - living apart, dependent child / children living away from you / Single, no dependent child / children/ Single, dependent child / children living away from you / other
Economic	
How will the salary of your new job compare to your Defence job? -	***Less than Defence/ More than Defence/ About the same as Defence***
What sector is your new job in?	Public Sector/Private Sector/Unsure
Service Related	
How long were you deployed on your most recent deployment?	Less than 2 weeks/2-4 weeks/1-3 months/4-6 months/7-12 months/ More than 12 months

Question/Variable	Response Value
Where did you serve on your most recent overseas operational deployment or United Nations mission? - Note MEAO encompasses Afghanistan	I have not served on an operation/Solomons/East Timor/ Christmas Island/Afghanistan/MEAO/Other middle east/ Other/ Bouganville/Gulf
During the last 12 months, how many months were you away due to operational time at sea?	***Less than 1 month/1-3 months/ 3-5 months/5-7 months/ 7-10 months/10-12 months***
During the last 12 months, how many months were you away due to domestic disasters or civil emergencies?	
During the last 12 months, how many months were you away due to unit training or field exercises?	
During the last 12 months, how many months were you away due to foreign humanitarian missions?	
During the last 12 months, how many months were you away due to military education?	
During the last 12 months, how many months were you away due to non-deployed time at sea?	
During the last 12 months, how many months were you away due to other work-related travel?	
During the last 12 months, how many months were you away due to Peacekeeping ops?	
During the last 12 months, how many months were you away due to warlike ops?	
How long is it since you returned from your most recent deployment?	***Currently deployed/Less than 3 months/3 months or more but less than 6/6 months or more but less than 1 year/I year or more but less than 2 years/2 years or more but less than 3 years/3 years or more ago***
Attitudinal or Sentiment Based (What was important in your decision to leave)	
Is there anything the ADF could have done that would have encouraged you to alter your decision to leave?	No/Yes
Did you enjoy your career in the ADF?	Yes/No
Responses to following questions = N/A/Not Important/Slightly Important/Moderately Important/Very Important/ Extremely Important	
Issues with day-to-day unit management of personnel matters -	Lack of maintenance of trade / specialist skills due to civilianization / outsourcing of functions -
Lack of confidence in senior Defence leadership	Lack of ongoing / advanced trade or specialist training opportunities -
Inability to access Long Service Leave or Leave Without Pay -	Inability to utilise training opportunities due to other work commitments / demanding work schedule -
Dissatisfaction with pay -	General dissatisfaction with service life -
Little financial reward for what would be considered overtime in the civilian community -	Excessive workload -
More attractive salary package available in civilian employment -	Excessive work hours -
Need for spouse / partner to get stable employment to supplement family income -	Insufficient equipment or resources to do my job -
Ineligibility for a retention bonus or allowance -	Sea service obligation -
Dissatisfaction with job-related allowances and benefits -	Desire for less separation from family -
Difference between single and married entitlements -	Impact of job demands on family / personal life -
Attractiveness of a civilian job supplemented by a pension	My spouse's / partner's attitude to service life -
Lack of provision for a Defence Force pension under MSBS	Too much time spent away from home because of military duties -
I qualify for a pension or superannuation payments -	The nature of the work in future postings -
Better career prospects in civilian life	A desire for more challenging work -
To make a career change while still young enough -	Desire to stay in one place -

Question/Variable	Response Value
Good chance of finding a civilian job in the current economic climate	Desire to return to my home location -
An offer of civilian employment -	Probable location of future postings -
Limited opportunities in my present Category	Posting conflicts with spouse's / partner's career -
Limited opportunities to transfer / change career within SAME Service -	Effect of postings on children's education -
Limited opportunities to transfer / change career into a DIFFERENT Service -	Effect of postings on family life -
Issues with promotion prospects -	Lack of recognition or credit for work done -
Issues with information provided on my career management -	Feeling that my contribution to the ADF was not valued or appreciated -
Feel there is a lack of opportunities for career development -	Little appreciation of the personal sacrifices made during my time in the ADF -
Lack of skills accreditation for civilian employment -	Lack of availability of DHA housing -
Desire to pursue further education that is not available through or relevant to Defence -	Adequacy of rental assistance -
Favoritism in the allocation of postings -	Uncertainty with long term career plans -
Selections or promotions not based entirely on merit -	Ongoing difficulties with spouse employment -
Posting preferences appear not to be considered -	To look after children -
Feeling that my good performance was not adequately rewarded by postings or promotion -	Lack of adequate child care -
Underuse or non-use of training and skills -	Difficulty managing work and family commitments as a single parent
Insufficient personnel in units to do the work -	Greater integration of women in the Service
Too much time spent on operational deployment -	Traumatic incident/s related to work -
Not enough opportunities for overseas deployments -	Personal experience of harassment or bullying -
Inability to secure back to back postings during a critical stage of family / personal life -	I have satisfied my goals in the Service -
Lack of 'respite' posting opportunities -	Lack of job satisfaction -
Greater integration of civilians in the work environment -	Insufficient personnel in units to do the work -

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX B. COMPARISON OF RANDOM TREE, K-MEANS, LINEAR SVM, CHAID, AND LOGISTIC REGRESSION RESULTS FOR MODEL DOWN SELECTION

Model Test and selection for final model Results Comparison table					
YOS => 2 years					
Target Group Data Target Variable/ Model	All Sailors		Technical		Non-technical
	<3000 Missing Values		No Missing Values		No Missing Values
YOS => 2 years	Top 10 Behavior Predictors		Interacted Tech*YOS Target Variable		Interacted Non-Tech*YOS Target Variable
Random Tree	1. Worn rank 2. Family living arrangements 3. Last 12 months at sea operational domestic disasters civil emergencies 4. Age group 5. State 6. Where deployment 7. Last 12 months deployed operational at sea 8. How long deployed 9. How salary compares to Defence job 10. Do you identify as	Linear SVM	Not Valid all values >2	Linear SVM	Not Valid all values >2
Model Accuracy	97.7% Prediction Accuracy	N/A	N/A	N/A	N/A
	Top 10 Behavior Predictors				
K-Means (unsupervised)	1. Difficulty managing work and family commitments as a single parent 2. Dissatisfaction with job related allowances and benefits 3. Feeling that my contribution to the ADF was not valued or appreciated 4. Feeling that my good performance was not adequately rewarded by postings or promotion 5. Inability to utilize training opportunities due to other work commitments and demanding work schedule 6. Lack of confidence in SLG 7. Lack of ongoing advanced trade or specialist training opportunities 8. Little appreciation of the personal sacrifices made during my time in the ADF 9. More attractive salary package available in civilian employment 10. Uncertainty with long term career plans	Random Tree	Not Valid all values >2	Random Tree	Not Valid all values >2
Model Accuracy	Silhouette = 0.1 Poor Largest Cluster 31.7%	N/A	N/A	N/A	N/A
	Top Behavior Predictors	K-Means	Not Valid all values >2	K-Means	Not Valid all values >2
CHAID (Chi-square automatic interaction detector)	1. Years at current rank 2. Uncertainty with long term career plans 3. Lack of ongoing advanced trade or specialist training opportunities				
Model Accuracy	68.6% prediction accuracy				

YOS => 4 years					
YOS => 4 years	Top 10 Behavior Predictors	YOS > 4 years	Top 10 Behavior Predictors	YOS > 4 years	Top 10 Behavior Predictors
Linear SVM	N/A model does not work with missing values	Linear SVM	1. Limited Opportunities in my current category 2. The nature of work in future postings 3. General dissatisfaction with service life 4. Desire to pursue further education not available in Defence 5. Worn rank 6. Too Much time spent on operational deployment 7. Greater integration of civilians in the work environment 8. Age group 9. Desire for less separation from family 10. Too much time spent away from home because of military duties	Linear SVM	1. The nature of the work in future postings 2. Limited opportunities in my present category 3. Age group 4. Lack of skills accreditation 5. I have satisfied my goals in service 6. My spouse's or partner's attitude to service life 7. I qualify for pension or superannuation payments 8. Not enough opportunities for overseas deployments 9. Feel there is a lack of opportunities for career development 10. Dissatisfaction with job related allowances and benefits
Model Accuracy			63.13% Prediction Accuracy		59.52% Prediction Accuracy
Random Tree	1. Worn rank 2. Family living arrangements 3. Where deployment 4. State 5. Last 12 months at sea operational domestic disasters civil emergencies 6. Age group 7. Last 12 months away unit training field exercises 8. Hd 9. Do you identify as 10. How salary compares to Defence job	Random Tree	**1.** Lack of confidence in SLG **2.** How long deployed **3.** Worn rank **4.** Inability to access LSL or LWOP **5.** Family living arrangements **6.** Little financial reward for what would be considered overtime in civilian sector **7.** Issues with day to day unit management of personnel issues **8.** Is there anything else the ADF could have done after the decision to leave? **9.** Dissatisfaction with pay **10.** Age group	Random Tree	1. Worn rank 2. Where deployment 3. Family living arrangements 4. Issues with day to day unit management of personnel issues 5. Dissatisfaction with pay 6. Lack of confidence in SLG 7. How long deployed 8. Inability to access LSL or LWOP 9. Do you identify as 10. Is there anything else the ADF could have done after the decision to leave?
Model Accuracy	79.6% Prediction Accuracy		61.6% Prediction Accuracy		59.8% Prediction Accuracy
K-Means (unsupervised)	1. Difficulty managing work and family commitments as a single parent 2. Dissatisfaction with job related allowances and benefits 3. Feeling that my contribution to the ADF was not valued or appreciated 4. Feeling that my good performance was not adequately rewarded	K-Means	1. Feeling that my contribution to the ADF was not valued or appreciated 2. Favoritism in the allocation of postings 3. Posting preferences appear to not be considered 4. Feeling that my good performance is not adequately rewarded by postings or promotion 5. Inability to utilize training opportunities due to other	K-Means	1. Feeling that my contribution to the ADF was not valued or appreciated 2. Posting preferences appear to not be considered 3. Favoritism in the allocation of postings 4. Inability to utilize training opportunities due to other work commitments and demanding work schedule 5. Feeling that my good performance is not

YOS => 4 years					
YOS => 4 years	Top 10 Behavior Predictors	YOS > 4 years	Top 10 Behavior Predictors	YOS > 4 years	Top 10 Behavior Predictors
	by postings or promotion 5. Inability to utilize training opportunities due to other work commitments and demanding work schedule 6. Lack of confidence in SLG 7. Lack of ongoing advanced trade or specialist training opportunities 8. Little appreciation of the personal sacrifices made during my time in ADF 9. More attractive salary package available in civilian employment 10. Uncertainty with long term career plans		work commitments and demanding work schedule 6. Little appreciation of the personal sacrifices made during my time in the ADF 7. Lack of respite posting opportunities 8. Lack of ongoing advanced trade or specialist training opportunities 9. Lack of recognition or credit for work done 10. Issues with day to day unit management of personnel issues		adequately rewarded by postings or promotion 6. Little appreciation of the personal sacrifices made during my time in the ADF 7. Lack of ongoing advanced trade or specialist training opportunities 8. Lack of respite posting opportunities 9. Lack of recognition or credit for work done 10. Lack of maintenance of trade specialist skills due to civilian outsourcing
Model Accuracy	Silhouette = 0.1 Poor Largest Cluster 32.1%		Silhouette = 0.1 Poor Largest Cluster 42.6%		Silhouette = 0.1 Poor Largest Cluster 42.4%
		Log Reg	Most significant betas (In word doc)		Most significant betas (In word doc)
Model Accuracy			10.36% Prediction Accuracy		
CHAID (Chi-square automatic interaction detector)	1. Last 12 months at sea operational domestic disasters civil emergencies 2. Last 12 months at sea operational domestic disasters civil emergencies 3. Issues with promotion prospects 4. Effect of postings on children's education 5. Desire to return to my home location 6. Desire for more challenging work 7. Age group 8. How long since returned from most recent deployment 9. Worn rank 10. Year surveyed	CHAID (Chi-square automatic interaction detector)	1. Worn rank 2. Need for spouse or partner to get stable employment to supplement income 3. Lack of maintenance of specialist skills due to civilian outsourcing 4. Posting preferences appear to not be considered 5. Age group 6. General dissatisfaction with service life 7. Do you identify as 8. Too much time spent on operational deployment 9. Little financial reward for what would be considered 10. Personal experience of harassment or bullying	CHAID (Chi-square automatic interaction detector)	1. Lack of maintenance of specialist skills due to civilian outsourcing 2. An offer of civilian employment 3. Underuse or non-use of training and skills 4. Limited opportunity to transfer change career in different service 5. Posting conflicts with spouse's or partner's career 6. Year surveyed 7. Where deployment 8. Issues with promotion prospects 9. Not enough opportunity for overseas deployment 10. How long deployed
Model Accuracy	58.4% Prediction Accuracy		67.47% Prediction Accuracy		63.13% Prediction Accuracy

YOS => 6 years					
YOS => 6 years	Top 10 Behavior Predictors	YOS > 6 years		YOS > 6 years	
Linear SVM	N/A	Linear SVM	1. Desire to pursue further education not available in Defence 2. Desire for more challenging work 3. General dissatisfaction with service 4. Greater integration of civilians in the workforce 5. Insufficient equipment or resources to do job 6. Limited opportunities to transfer within same service 7. The nature of the work in future postings 8. Feeling that my good performance was not adequately rewarded by postings or promotion 9. Attractiveness of a civilian job 10. Desire for less separation from family	Linear SVM	1. The nature of the work in future postings 2. Desire to pursue further education not available in Defence 3. Where deployment 4. Lack of recognition or credit for work done 5. Desire for less separation from family 6. Greater of integration of women into service 7. More attractive salary package 8. Limited opportunities to transfer within same service 9. Insufficient equipment or resources to do job 10. Little financial reward for what would be considered overtime in civilian sector
Model Accuracy	N/A		75.66% Prediction Accuracy		72.53% Prediction Accuracy
Random Tree	1. Worn rank 2. Where deployment 3. State 4. Age group 5. Family living arrangements 6. How salary compares to Defence job 7. How long deployed 8. Last 12 months at sea operational domestic disasters civil emergencies 9. Last 12 months away unit training field exercises 10. Do you identify as	Random Tree	1. Family arrangements 2. Worn rank 3. Lack of confidence in SLG 4. Is there anything else the ADF could have done after the decision to leave? 5. Age group 6. Issues with day to day unit management of personnel matters 7. How long deployed 8. Inability to access LSL or LWOP 9. More attractive salary pack in civilian sector 10. Little financial reward for what would be considered overtime in the civilian sector	Random Tree	**1.** Where deployment **2.** Family arrangements **3.** Lack of confidence in SLG **4.** Worn rank **5.** Dissatisfaction with pay **6.** Age group **7.** Is there anything else the ADF could have done after the decision to leave? **8.** Do you identify as **9.** Inability to access LSL or LWOP **10.** Issues with day to day unit management of personnel matters
Model Accuracy	82% Prediction Accuracy		75.8% Prediction Accuracy		74.2% Prediction Accuracy
K-Means (unsupervised)	1. Difficulty managing work and family commitments as a single parent 2. Dissatisfaction with job related allowances and benefits 3. Feeling that my contribution to the ADF was not valued or appreciated	K-Means	1. Posting preferences appear to not be considered 2. Feeling that my contribution to the ADF was not valued or appreciated 3. Inability to utilize training opportunities due to other work commitments and demanding work schedule	K-Means	1. Feeling that my contribution to the ADF was not valued or appreciated 2. Posting preferences appear to not be considered 3. Favoritism in the allocation of postings 4. Inability to utilize training opportunities due to other

YOS => 6 years					
YOS => 6 years	Top 10 Behavior Predictors	YOS > 6 years		YOS > 6 years	
	4. Feeling that my good performance was not adequately rewarded by postings or promotion 5. Inability to utilize training opportunities due to other work commitments and demanding work schedule 6. Lack of confidence in SLG 7. Lack of ongoing advanced trade or specialist training opportunities 8. Little appreciation of the personal sacrifices made during my time in the ADF 9. More attractive salary package available in civilian employment 10. Uncertainty with long term career plans		4. Favoritism in the allocation of postings 5. Under use or non-use of training and skills 6. Lack of respite postings 7. Dissatisfaction with job related benefits and allowances 8. Feeling that my good performance is not adequately rewarded by postings or promotion 9. Little appreciation of the personal sacrifices made during my time in the ADF 10. Lack of ongoing advanced trade or specialist training opportunities		work commitments and demanding work schedule 5. Little appreciation of the personal sacrifices made during my time in the ADF 6. Feeling that my good performance is not adequately rewarded by postings or promotion 7. Lack of ongoing advanced trade or specialist training opportunities 8. Lack of respite posting opportunities 9. Lack of recognition or credit for work done 10. Dissatisfaction with job related benefits and allowances
Model Accuracy	Silhouette = 0.1 Poor Largest Cluster 31.1%		Silhouette = 0.1 Poor Largest Cluster 42.6%		Silhouette = 0.1 Poor Largest Cluster 42.6%
CHAID (Chi-square automatic interaction detector)	1. Worn rank 2. Years at current rank 3. I qualify for a pension or superannuation 4. Lack of provision of a Defence force pension 5. A desire for more challenging work 6. Feel there is a lack of opportunities for career development 7. Year surveyed 8. Age group 9. Issues with promotion prospects 10. Lack of recognition or credit for work done	CHAID (Chi-square automatic interaction detector)	1. Worn rank 2. Year surveyed 3. Dissatisfaction with pay 4. I qualify for a pension or superannuation payments 5. How long deployed 6. Greater integration of women into service 7. Ineligibility for retention bonus or allowance 8. Limited opportunity to transfer change career within different service 9. Lack of provision for Defence force pension under MSBS 10. Posting preferences appear not to be considered	CHAID (Chi-square automatic interaction detector)	1. Worn rank 2. Year surveyed 3. Traumatic incidents related to work 4. Greater integration of women into service 5. Inability to utilize training opportunities due to other work commitments 6. More attractive salary package available in civilian sector 7. Lack of maintenance of specialist skills due to civilian outsourcing 8. Where deployment 9. Posting conflicts with partner's or spouse's career 10. To make a career while still young enough
Model Accuracy	80.6% Prediction Accuracy		82.41% Prediction Accuracy		77.35% Prediction Accuracy
		Log Reg	10.36% Accuracy		

YOS => 8 years					
YOS => 8 years	Top 10 Behavior Predictors	YOS > 8 years	Top 10 Behavior Predictors	YOS > 8 years	Top 10 Behavior Predictors
Linear SVM	N/A	Linear SVM	1. Too much time spent on operational deployment 2. Desire to pursue further education not available in Defence 3. General dissatisfaction with service life 4. Adequacy of RA 5. The nature of work in future postings 6. Too much time spent away from home because of military duties 7. Favoritism in the allocation of postings 8. Ongoing difficulties with spouse's employment 9. Lack of provision for a Defence force pension under MSBS 10. Posting preferences appear to be not considered	Linear SVM	1. Effect of postings on children's education 2. Too much time spent on operational deployment 3. The nature of work in future postings 4. Desire to pursue further education not available in Defence 5. Need for spouse or partner to get stable employment to supplement income 6. Dissatisfaction with job related benefits and allowances 7. Personal experience of harassment. 8. Lack of skills accreditation 9. Where deployment 10. Lack of recognition or credit for work done
Model Accuracy	N/A		92.29% Prediction Accuracy		91.08% Prediction Accuracy
Random Tree	1. Worn rank 2. Family living arrangements 3. Where deployment 4. Age group 5. How long deployed 6. State 7. Last 12 months at sea operational 8. Do you identify as 9. How salary compares to Defence job 10. Last 12 months at sea operational domestic disasters civil emergencies	Random Tree	1. Family arrangements 2. How long deployed 3. Issues with day to day unit management of personnel matters 4. Inability to access LSL or LWOP 5. Worn rank 6. More attractive salary pack in civilian sector 7. Lack of confidence in SLG 8. Little financial reward for what would be considered overtime in the civilian sector 9. Dissatisfaction with pay 10. Age group	Random Tree	1. Dissatisfaction with pay 2. Do you identify as 3. Family arrangements 4. Dissatisfaction with pay 5. How long deployed 6. Worn rank 7. More attractive salary pack in civilian sector 8. Lack of confidence in SLG 9. Where deployment 10. Inability to access LSL or LWOP
Model Accuracy	92.1% Prediction Accuracy		91.7% Prediction Accuracy		89.4% Prediction Accuracy
K-Means (unsupervised)	1. Difficulty managing work and family commitments as a single parent 2. Dissatisfaction with job related allowances and benefits 3. Feeling that my contribution to the ADF was not valued or appreciated 4. Feeling that my good performance was not adequately rewarded	K-Means	1. Posting preferences appear to not be considered 2. Inability to utilize training opportunities due to other work commitments and demanding work schedule 3. Favoritism in the allocation of postings 4. Feeling that my contribution to the ADF was not valued or appreciated 5. Under use or non-use of training and skills	K-Means	1. Feeling that my contribution to the ADF was not valued or appreciated 2. Posting preferences appear to not be considered 3. Favoritism in the allocation of postings 4. Inability to utilize training opportunities due to other work commitments and demanding work schedule 5. Feeling that my good performance is not

YOS => 8 years					
YOS => 8 years	Top 10 Behavior Predictors	YOS > 8 years	Top 10 Behavior Predictors	YOS > 8 years	Top 10 Behavior Predictors
	by postings or promotion 5. Inability to utilize training opportunities due to other work commitments and demanding work schedule 6. Lack of confidence in SLG 7. Lack of ongoing advanced trade or specialist training opportunities 8. Little appreciation of the personal sacrifices made during my time in the ADF 9. More attractive salary package available in civilian employment 10. Uncertainty with long term career plans		6. Lack of respite postings 7. Dissatisfaction with job related benefits and allowances 8. Feeling that my good performance is not adequately rewarded by postings or promotion 9. Little appreciation of the personal sacrifices made during my time in the ADF 10. Lack of ongoing advanced trade or specialist training opportunities		adequately rewarded by postings or promotion 6. Little appreciation of the personal sacrifices made during my time in the ADF 7. Lack of respite posting opportunities 8. Lack of ongoing advanced trade or specialist training opportunities 9. Lack of recognition or credit for work done 10. Dissatisfaction with job related benefits and allowances
Model Accuracy	Silhouette = 0.1 Poor Largest Cluster 31.7%		Silhouette = 0.1 Poor Largest Cluster 29.6%		Silhouette = 0.1 Poor Largest Cluster 42.6%
CHAID (Chi-square automatic interaction detector)	1. Worn rank 2. Years at current rank 3. Impact of job demands on family personal life 4. Feeling that my contribution to the ADF is not valued or appreciated 5. Feeling there is a lack of opportunities for career development	CHAID (Chi-square automatic interaction detector)	1. Too much time spent on operational deployment 2. Selections or promotions not based entirely on merit 3. Difficulty managing work and family commitments as a single parent 4. Attractiveness of a civilian job supplemented by a pension 5. Lack of provision of a Defence force pension under MSBS 6. Difference between married and single entitlements	CHAID (Chi-square automatic interaction detector)	1. Year surveyed 2. Lack of skills accreditation for civilian employment 3. Worn rank 4. Did you enjoy ADF career 5. Good chance of finding a civilian job in the current economic climate
Model Accuracy	89.3% Prediction Accuracy		91.08% Prediction Accuracy		90.84% Prediction Accuracy

ALL YOS					
All YOS	Top 10 Behavior Predictors	Technical (Technical Target Variable)	Top 10 Behavior Predictors	Technical (Technical Target Variable)	Top 10 Behavior Predictors
Linear SVM	N/A	Linear SVM	1. Lack of maintenance of trade specialist skills due to civilian outsourcing 2. The nature of work in future postings 3. Feeling that my good performance is not adequately rewarded by postings or promotion 4. Age group 5. Limited opportunities in my present category 6. Dissatisfaction with pay 7. Attractiveness of a civilian job supplemented by a pension 8. General dissatisfaction with service 9. Lack of skills accreditation 10. Desire for less separation from family	Linear SVM	1. The nature of work in future postings 2. Lack of maintenance of trade specialist skills due to civilian outsourcing 3. Age group 4. Feeling that my good performance is not adequately rewarded by postings or promotion 5. Limited opportunities in my present category 6. General dissatisfaction with service 7. Lack of skills accreditation 8. Dissatisfaction with pay 9. Attractiveness of a civilian job supplemented by a pension 10. Desire for less separation from family
Model Accuracy	N/A		63.61% Prediction Accuracy		62.41% Prediction Accuracy
Random Tree	1. Age group 2. Worn rank 3. ias 4. How salary compares to Defence job 5. State 6. Family living arrangements 7. How long deployed 8. Last 12 months at sea operational	Random Tree	1. How long deployed 2. Do you identify as 3. Family arrangements 4. Issues with day to day unit management of personnel matters 5. Dissatisfaction with pay 6. Worn rank 7. Little financial reward for what would be considered overtime in the civilian sector 8. Is there anything else the ADF could have done after the decision to leave 9. Inability to access LSL or LWOP 10. Lack of confidence in SLG	Random Tree	1. Where deployment 2. Issues with day to day unit management of personnel matters 3. Family arrangements 4. Do you identify as 5. Dissatisfaction with pay 6. Worn rank 7. Lack of confidence in SLG 8. Inability to access LSL or LWOP 9. Is there anything else the ADF could have done after the decision to leave 10. Age group
Model Accuracy	41.5% Prediction Accuracy		57.8% Prediction Accuracy		60.9% Prediction Accuracy
K-Means (unsupervised)	1. Difficulty managing work and family commitments as a single parent 2. Dissatisfaction with job related allowances and benefits 3. Feeling that my contribution to the ADF was not valued or appreciated 4. Feeling that my good performance was not adequately rewarded	K-Means	1. Feeling that my contribution to the ADF was not valued or appreciated 2. Favoritism in the allocation of postings 3. Posting preferences appear to not be considered 4. Feeling that my good performance is not adequately rewarded by postings or promotion 5. Inability to utilize training opportunities due to other	K-Means	1. Posting preferences appear to not be considered 2. Feeling that my contribution to the ADF was not valued or appreciated 3. Inability to utilize training opportunities due to other work commitments and demanding work schedule 4. Favoritism in the allocation of postings 5. Under use or non-use of training and skills

ALL YOS					
All YOS	Top 10 Behavior Predictors	Technical (Technical Target Variable)	Top 10 Behavior Predictors	Technical (Technical Target Variable)	Top 10 Behavior Predictors
	by postings or promotion 5. Inability to utilize training opportunities due to other work commitments and demanding work schedule 6. Lack of confidence in SLG 7. Lack of ongoing advanced trade or specialist training opportunities 8. Little appreciation of the personal sacrifices made during my time in the ADF 9. More attractive salary package available in civilian employment 10. Uncertainty with long term career plans		work commitments and demanding work schedule 6. Lack of respite postings 7. Little appreciation of the personal sacrifices made during my time in the ADF 8. Lack of ongoing advanced trade or specialist training opportunities 9. Lack of recognition or credit for work done 10. Dissatisfaction with job related benefits and allowances		6. Lack of respite postings 7. Dissatisfaction with job related benefits and allowances 8. Feeling that my good performance is not adequately rewarded by postings or promotion 9. Lack of ongoing advanced trade or specialist training opportunities 10. Little appreciation of the personal sacrifices made during my time in the ADF
Model Accuracy	Silhouette = 0.1 Poor Largest Cluster 31.7%		Silhouette = 0.1 Poor Largest Cluster 42.6%		Silhouette = 0.1 Poor Largest Cluster 29.1%
	N/A	Log Reg	10.36% Accuracy		N/A
CHAID (Chi-square automatic interaction detector)	1. Years at current rank 2. Worn rank 3. Age group 4. Feeling there is a lack of opportunities for career development 5. Lack of provision for a Defence force pension under MSBS 6. Did you enjoy career ADF 7. Desire to pursue further education that is not available in Defence 8. A desire for more challenging work 9. Attractiveness of a civilian job supplemented by a pension 10. More attractive salary package available in civilian employment	CHAID (Chi-square automatic interaction detector)	1. Lack of maintenance of specialist skills due to civilianization of workforce 2. Do you identify as 3. Traumatic incidents related to work 4. Impact of job demands on family life 5. Greater integration of women in then service 6. Need for spouse or partner to get stable employment to supplement family income 7. Feeling that my good performance is not adequately rewarded by postings or promotion 8. Adequacy of rental assistance 9. Lack of job satisfaction 10. To make a career change while still young enough	CHAID (Chi-square automatic interaction detector)	1. Lack of maintenance of specialist skills due to civilianization of workforce 2. Do you identify as 3. Traumatic incidents related to work 4. Impact of job demands on family life 5. Greater integration of women in then service 6. Need for spouse or partner to get stable employment to supplement family income 7. Feeling that my good performance is not adequately rewarded by postings or promotion 8. Adequacy of rental assistance 9. Lack of job satisfaction 10. To make a career change while still young enough
Model Accuracy	85.4% Prediction Accuracy		68.43% Prediction Accuracy		68.43% Prediction Accuracy

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF REFERENCES

Alpaydin, A. (2010). *Introduction to machine learning* (2nd ed.). Cambridge, MA: MIT Press.

Arkes, J. (2019). *Regression analysis: A practical introduction* (1st ed.). Abington, United Kingdom: Routledge.

Asch, B. J., Hosek, J. R., & Warner, J. T. (2007). New economics of manpower in the Post–Cold War era. In T. Sandler & K. Hartley (Eds.), *Handbook of defense economics* (Vol. 2, pp. 1075–1138). Amsterdam, The Netherlands: Elsevier. https://doi.org/10.1016/S1574-0013(06)02032-1

Australian Defence Force (ADF). (2000). *ADF Report of the Australian Defence Force exit survey - Reasons for leaving*. (DSPPR Research Note 2/2000). Canberra, Australia: Directorate of Strategic Personnel Planning.

Australian Defence Force (ADF). (2002). *ADF Exit Survey report reasons for leaving 2001*. (DSPPR Research Report 4/2002). Canberra, Australia: Directorate of Strategic Personnel Planning.

Australian Defence Force (ADF). (2003). *ADF Exit Survey report reasons for leaving 2002*. (DSPPR Research Report 6/2003). Canberra, Australia: Directorate of Strategic Personnel Planning.

Australian Defence Force (ADF). (2004). *ADF Exit Survey report reasons for leaving 2003*. (DSPPR Research Report 4/2004 April 2004). Canberra, Australia: Directorate of Strategic Personnel Planning.

Australian Defence Force (ADF). (2005). *ADF Exit Survey report reasons for leaving 2004*. (DSPPR Research Report 26/2005). Canberra, Australia: Directorate of Strategic Personnel Planning.

Australian Defence Force (ADF). (2006). *ADF Exit Survey report reasons for leaving 2005*. (DSPPR Research Report 23/2006). Canberra, Australia: Directorate of Strategic Personnel Planning.

Australian Defence Force (ADF). (2007). *ADF Exit Survey report reasons for leaving 2006*. (DSPPR Research Report 38/2007). Canberra, Australia: Directorate of Strategic Personnel Planning.

Australian Defence Force (ADF). (2008). *ADF Exit Survey report reasons for leaving 2007*. (DSPPR Research Report 38/2008). Canberra, Australia: Directorate of Strategic Personnel Planning.

Australian Defence Force (ADF). (2010). *ADF Exit Survey report reasons for leaving 2009*. (DSPPR Research Report 59/2010). Canberra, Australia: Directorate of Strategic Personnel Planning.

Australian Defence Force (ADF). (2011). *ADF Exit Survey report reasons for leaving 2010*. (DSPPR Research Report 14/2011). Canberra, Australia: Directorate of Strategic Personnel Planning.

Australian Defence Force (ASDF). (2012). *ADF Exit Survey preliminary report 2011*. (DSPPR, 28 November 2012). Canberra, Australia: Directorate of Strategic Personnel Planning.

Australian Defence Force (ADF). (2014a). *ADF your say leaving defence report 2013*. (DSPPR Research Report 6/2014). Canberra, Australia: Directorate of Strategic Personnel Planning.

Australian Defence Force (ADF). (2014b). *ADF sour say reasons for leaving and retention - ADF 2014 data*. (DSPPR Research Report 108/2015). Canberra, Australia: Directorate of Strategic Personnel Planning.

Australian Defence Force (ADF). (2016). *ADF your say service intentions and reasons for Leaving 2015 ADF Results*. (DSPPR Research Report 38/2016). Canberra, Australia: Directorate of Strategic Personnel Planning.

Australian Defence Force (ADF). (2018). *ADF your say starting and leaving defence Survey Results—Navy 2017*. (DSPPR Research Report 34/2018). Canberra, Australia: Directorate of Strategic Personnel Planning.

Australian defence Force (ADF). (n.d.-a). *ADF Careers.* Retrieved August 3, 2019, from https://www.defencejobs.gov.au

Australian defence Force (ADF). (n.d.-b). *Graduated Other Ranks Pay Structure.* Retrieved August 3, 2019, from https://www.defence.gov.au/Publications/NewsPapers/Navy/editions/5206/5206.pdf

Australian Defence Force (ADF). (n.d.-c). *Military superannuation and benefits scheme retention benefit.* Retrieved October 11, 2019, from https://www.defence.gov.au/payandconditions/ADF/Members-Guide/Salaries-and-bonuses/MSBSRetentionBenefit.asp

Australian Defence Force (ADF). (n.d.-d). *National naval shipbuilding enterprise*. Retrieved October 11, 2019, from https://www.defence.gov.au/Naval Shipbuilding

Australian defence Force (ADF). (n.d.-e). *Pay and Conditions*. Retrieved October 5, 2019, from https://www.defence.gov.au/PayAndConditions/ADF/Resources/WRA.pdf

Australian Defence Force (ADF). (n.d.-f). *Total Workforce Model*. Retrieved October 5, 2019, from http://drnet.defence.gov.au/navy/eport/pages/ ADF%20Total%20Workforce%20Model.aspx

Australian National Audit Office (ANO). (2016). *Recruitment and retention of specialist skills for Navy* (Auditor General Report No.17 2014–2015). Canberra, Australia: Department of Defence.

Australian Government National Service Act of 1964. (National Service Act, 1964). https://www.awm.gov.au/articles/encyclopedia/viet, accessed 10 Oct 2019.

Australian Qualifications Framework. (n.d.). Home page. Retrieved October 5, 2019, from https://www.aqf.edu.au

Australian Trade Commission. (n.d.). Economic analysis. Retrieved August 3, 2019, from https://www.austrade.gov.au/news/economic-analysis/2017-18-gdp-growth-rate-of-2-9-per-cent-confirms-the-resilience-of-our-economy

Delen, D. (2010). A comparative analysis of machine learning techniques for student retention management. *Decision Support Systems*, *49*(4), 798–506. Retrieved from https://doi.org/10.106/j.dss.2010.06.003

Díaz-Pérez, F. & Bethencourt, C. (2016). CHAID algorithm as an appropriate analytical method for tourism market segmentation. *Journal of Destination Marketing & Management Volume 5 (2016) 275–382.* University of La Laguna, Tenerife – Canary Islands, Spain.

Dodds, D. (2018). *Length-of-service/survival profiles methodology for the Royal Australian Navy* (Master's thesis). Retrieved from https://calhoun.nps.edu/handle/ 10945/61350

Ehrenberg, R., & Smith, R. (2016). *Modern labor economics* (12th ed.). Ithaca, NY: Theory and Public Policy/School of Industrial and Labor Relations, Cornell University.

Kalmegh, S. (2015). *Analysis of WEKA data mining algorithm REP tree, simple cart and random tree for classification of Indian news*. Associate Professor, Department of Computer Science, Sant Gadge Baba Amravati University Amravati, Maharashtra.

Mather, P., & Mahesh, P. (2003, May 3). An assessment of the effectiveness of decision tree methods for land cover classification. *Remote Sensing of Environment 86*(4), pp. 554–565.

Parnell, S. (2012). Army Recruits drop out to join mining moom. *The Weekend Australian.* Retrieved from https://www.theaustralian.com.au/national-affairs/army-recruits-drop-out-to-join-mining-boom/news-story/7ba599905f8126513a4e55a9a96976dd

Robertson, A. (2019, 23 August). Tech jobs are waiting for Australians but not enough people have the skills, companies say. ABC News. Retrieved from https://www.abc.net.au/news/2019-08-24/jobs-are-waiting-to-be-created-but-skilled-workers-lacking/11442270

Rodney, D. (2017). *Navy manpower planning*. Arlington, VA: CNA. https://www.cna.org/CNA_files/PDF/CPP-2017-U-015038-Final.pdf

Rosen, S. (1986). The theory of equalizing differences. In O.C. Ashenfelter & R. Layard (Eds.), *Handbook of labor economics* (Vol. 1, pp. 641–692). Amsterdam, The Netherlands: Elsevier. https://doi.org/10.1016/S1573-4463(86)01015-5

Royal Australian Navy (RAN). (1999). *RAN exit survey report Jan 1999*. (DSPPR Research Note 1/99). Canberra, Australia: Directorate of Strategic Personnel Planning.

Royal Australian Navy (RAN). (2018, July 7). *Retention of Navy people*. Retrieved from https://www.forcenet.gov.au/community/viewgroup?Id=09a8177c-8f45-4bca-84d1-c37e6c739430

Royal Australian Navy (RAN). (n.d.). *Navy Retention Initiative*. Retrieved October 5, 2019, from http://drnet.defence.gov.au/navy/eport/documents/NRI%20Payment%20Signal%2020090632Z%20May%202019.pdf

Shah. (2017, December 6). Retrieved from https://towardsdatascience.com/train-validation-and-test-sets-72cb40cba9e7

Strategic Workforce Planning & Analysis (Navy) (SWPA(N)). (2019a, May 1). *Sailor pocket brief*. Retrieved from http://drnet.defence.gov.au/People/WP/Strategic-Workforce-Planning-and-Analysis/Pages/SWPA-Published-Products.aspx

Strategic Workforce Planning & Analysis (Navy) (SWPA(N)). (2019b, May 1). *Sailor snapshot*. Retrieved from http://drnet.defence.gov.au/People/WP/Strategic-Workforce-Planning-and-Analysis/Pages/SWPA-Published-Products.aspx

Strategic Workforce Planning & Analysis (Navy) (SWPA(N)). (2019c, May 1). *Marine technician family pocket brief*. Retrieved from http://drnet.defence.gov.au/People/WP/Strategic-Workforce-Planning-and-Analysis/Pages/SWPA-Published-Products.aspx

Strategic Workforce Planning & Analysis (Navy) (SWPA(N)). (2019d, May 1). *Electronics technician family pocket brief*. Retrieved from http://drnet.defence.gov.au/People/WP/Strategic-Workforce-Planning-and-Analysis/Pages/SWPA-Published-Products.aspx

Strategic Workforce Planning & Analysis (Navy) (SWPA(N)). (2019e, May 1). *Aviation technician aircraft pocket brief*. Retrieved from http://drnet.defence.gov.au/People/WP/Strategic-Workforce-Planning-and-Analysis/Pages/SWPA-Published-Products.aspx

Strategic Workforce Planning & Analysis (Navy) (SWPA(N)). (2019f, May 1). *Aviation technician avionic pocket brief*. Retrieved from http://drnet.defence.gov.au/People/WP/Strategic-Workforce-Planning-and-Analysis/Pages/SWPA-Published-Products.aspx

Strategic Workforce Planning & Analysis (Navy) (SWPA(N)). (2019g, May 1). *Maritime logistics - Supply chain family pocket brief*. Retrieved from http://drnet.defence.gov.au/People/WP/Strategic-Workforce-Planning-and-Analysis/Pages/SWPA-Published-Products.aspx

Strategic Workforce Planning & Analysis (Navy) (SWPA(N)). (2019h, May 1). *Combat systems operator pocket brief*. Retrieved from http://drnet.defence.gov.au/People/WP/Strategic-Workforce-Planning-and-Analysis/Pages/SWPA-Published-Products.aspx

Strategic Workforce Planning & Analysis (Navy) (SWPA(N)). (2019i, May 1). *Navy monthly workforce status report*. Retrieved from http://drnet.defence.gov.au/People/WP/Strategic-Workforce-Planning-and-Analysis/Pages/SWPA-Published-Products.aspx

Wagstaff, K., Cardie, C., Rogers, S., & Schroedl, S. (2001). Constrained K-means clustering with background knowledge. Kiri Wagstaff & Claire Cardie Department of Computer Science, Cornell University, Ithaca, NY 14853 USA. Seth Rogers Stefan Schroedl, DaimlerChrysler Research and Technology Center, 1510 Page Mill Road, Palo Alto, CA 94304 USA

Wang, K., Wang, B., & Peng, L. (2009, 24 April). CVAP: Validation for cluster analysis. School of Mathematics and Computer Science, Fujian Normal University, Fuxhou 35007, P.R. China. School of Computer Science & Technology, Xidian University, Xian710071, P.R. China.

Warner, J. T., & Asch, B. J. (1995). The economics of military manpower. In T. Sandler & K. Hartley (Eds.), *Handbook of defense economics* (Vol. 1, pp. 347–398). Amsterdam, The Netherlands: Elsevier. https://doi.org/10.1016/S1574-0013(05)80015-8

THIS PAGE INTENTIONALLY LEFT BLANK

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California