NPS-HR-26-297

ACQUISITION RESEARCH PROGRAM
SPONSORED REPORT SERIES

A Machine Learning Approach to Predict Voluntary Separation from the Australian Army

March 2026

Maj Samuel S. Lam, Australian Army

Thesis Advisors: Dr. Sae Young Ahn, Assistant Professor
Dr. Ruriko Yoshida, Professor

Department of Acquisition, Finance and Manpower

Naval Postgraduate School

Prepared for the Naval Postgraduate School, Monterey, CA 93943.

ACQUISITION RESEARCH PROGRAM
DEPARTMENT OF ACQUISITION, FINANCE AND MANPPOWER
NAVAL POSTGRADUATE SCHOOL

The research presented in this report was supported by the Acquisition Research Program of the Department of Acquisition, Finance and Manpower Naval Postgraduate School.

To request defense acquisition research, to become a research sponsor, or to print additional copies of reports, please contact the Acquisition Research Program (ARP) via email, arp@nps.edu or at 831-656-3793.

ACQUISITION RESEARCH PROGRAM
DEPARTMENT OF ACQUISITION, FINANCE AND MANPPOWER
NAVAL POSTGRADUATE SCHOOL

ABSTRACT

The Australian Army faces significant workforce growth challenges over the next decade. Our study uses machine learning (ML) with human resource data to predict individual voluntary separation decisions to support retention targets. We use a two-stage approach. Stage 1 predicts "who separates" through supervised classification. Stage 2 predicts "when separation occurs" through predictive survival analysis. We evaluate the best performing models, and which features were most important to predictive performance. In Stage 1 we find that tree ensemble and gradient-boosted model perform exceptionally well. The best performing model, Light Gradient-Boosting Machine, achieves an Average Precision of 0.99. In Stage 2, Random Survival Forest accurately predicts separation timing, achieving an Uno's Concordance Index of 0.97. Unemployment rates and job-search duration consistently dominated feature importance analyses. However, we observe non-linear and non-monotonic relationships that contradict standard labor economic theory, suggesting broader structural factors or events may influence predicted separation risk. These findings show that ML is effective in processing large and complex datasets for predicting separation risk and identifying key drivers for further study and policy intervention.

THIS PAGE INTENTIONALLY LEFT BLANK

Acknowledgments

I wish to thank my co-advisors, Dr. Ruriko Yoshida and Dr. Sae Young Ahn, whose support and advice was invaluable to completing this thesis. Your instruction throughout the program has sparked interest in applying machine learning and statistical analysis towards manpower, workforce, and labor economics.

I am grateful to my foxhole buddy, Squadron Leader Darby Nelson. You motivated me throughout my time at the Naval Postgraduate School and encouraged me to expand my knowledge into machine learning. Your generous time doing multiple reviews, infectious enthusiasm, and witty banter have greatly assisted me in completing this thesis and maintaining sanity.

I would also like to thank Major Peter Stanton, whose early advice prepared me for the move to Monterey. You were instrumental to this thesis in providing all of the data I could ask for.

Finally, I am most thankful for the support of my wife, Yuki, and my children, Sakura and Kent. You uprooted your lives once again to accompany me even further on this adventure. I am eternally grateful for your love, sacrifice, and encouragement, and I will continue to be, for whatever our future holds.

THIS PAGE INTENTIONALLY LEFT BLANK

NPS-HR-26-297

ACQUISITION RESEARCH PROGRAM
SPONSORED REPORT SERIES

A Machine Learning Approach to Predict Voluntary Separation from the Australian Army

March 2026

Maj Samuel S. Lam, Australian Army

Thesis Advisors: Dr. Sae Young Ahn, Assistant Professor
Dr. Ruriko Yoshida, Professor

Department of Acquisition, Finance and Manpower

Naval Postgraduate School

Prepared for the Naval Postgraduate School, Monterey, CA 93943.

ACQUISITION RESEARCH PROGRAM
DEPARTMENT OF ACQUISITION, FINANCE AND MANPPOWER
NAVAL POSTGRADUATE SCHOOL

THIS PAGE INTENTIONALLY LEFT BLANK

Contents

List of Figures

List of Tables

THIS PAGE INTENTIONALLY LEFT BLANK

List of Acronyms and Abbreviations

ABS	Australian Bureau of Statistics
ADF	Australian Defence Force
ALE	accumulated local effect
AP	average precision
AUC	area under the curve
CatBoost	Categorical Boosting
CHF	cumulative hazard function
C-Index	concordance index
CoxRF	Cox random forest
CPH	Cox proportional hazard
CRA	compulsory retirement age
DoD	Department of Defence
DPG	Defence People Group
FN	false negative
FP	false positive
FPR	false positive rate
FY	financial year
IMPS	initial minimum period of service
KM	Kaplan–Meier
KNN	k–nearest neighbors
LASSO	least absolute shrinkage and selection operator
LightGBM	light gradient-boosting machine
MARS	Management and Reporting Solution

ML	machine learning
NB	Naïve Bayes
OHE	One Hot Encoding
PI	permutation importance
PII	personally identifiable information
PRC	precision-recall curve
RAAF	Royal Australian Air Force
RAN	Royal Australian Navy
RF	random forest
RFRSF	Random Forest, Random Survival Forest
ROC	receiver operating characteristic
ROSO	return of service obligation
RQ	research question
RSF	random survival forest
SERCAT	service category
SERVOP	service option
SHAP	SHapley Additive exPlanations
SMOTE	synthetic minority oversampling technique
SRQ	supporting research question
SVM	support vector machine
TP	true positive
TPR	true positive rate
USMC	U.S. Marine Corps
VIF	variance inflation factor
XGBoost	eXtreme Gradient Boosting
YOS	years of service

CHAPTER 1:
Introduction and Background

1.1 Strategic Context

This section establishes the strategic context and motivation for this study. We review the workforce challenge and strategic imperative facing the Australian Army, and how service in the Australian Defence Force (ADF) is rendered and ceased.

1.1.1 Defence Workforce Growth

In 2023, the Government of Australia released the Defence Strategic Review (Department of Defence, 2023b). The Review emphasizes a return to major power strategic competition, decrease in strategic warning time and a shift in alliances and reliance within the Indo-Pacific. This underscores strategic risk that requires a newer approach to force structure, force posture and increase in capabilities. The Department of Defence (DoD) addressed this through a National Defence Strategy and a Defence Workforce Plan; both specified requirements for the ADF to increase overall workforce strength, and to transition from a balanced workforce to an integrated, focused force (Department of Defence, 2024a, 2024c).

The retention of ADF members is a pressing issue to current force structure, preparedness, and future force generation. The Defence Workforce Plan intends to increase the budgeted workforce requirement for permanent ADF members from 58,850 in financial year (FY) 2024–2025 (already at a deficit) to 69,000 by the early 2030s (Department of Defence, 2024c). In addition to increasing recruitment from 5,500 to 9,000 per annum over the next decade, the median length of ADF service must increase from approximately 7 years to 12 years (Department of Defence, 2024a).

1.1.2 Australian Army Composition and Turnover

From the Australian Defence Annual Report FY 2024–2025 (Department of Defence, 2025b), Australian Army permanent force headcount—service categories (SERCATs) 6 and 7—made up 47.0% (27,701 of 58,909) of the total ADF headcount. However, Army

voluntary separations made up 54.5% (1,191 of 2,182) of total ADF separations, and 45.4% (1,191 of 2,621) of all Army separations, excluding trainee separations (Department of Defence, 2025b). This large percentage is a significant loss of human capital and experience, particularly when considering that over the course of FY 2024–2025 there were only 3,033 Army permanent workforce accessions (Department of Defence, 2025b). This drives a need to understand and improve retention, as well as improve accessions.

1.1.3 Service in the Australian Defence Force

Entry into the ADF is through enlistment (for soldiers) or appointment (for officers). Upon entry, an ADF member is subjected to a minimum service obligation (Defence Regulation 2016, 2016, Sec. 25). This is the reparation of service that pays back the investment in training and education invested by the ADF, termed an initial minimum period of service (IMPS). A further return of service obligation (ROSO) may also be enforced if further training or education is invested, and is accepted voluntarily by the member. These service obligations may also be imposed and served out concurrently or consecutively, at Service discretion.

Service in the ADF currently operates under the Total Workforce Model (Department of Defence, n.d.), in which SERCATs (1 to 7) are used to describe a spectrum of part-time through to full-time service. There are also various service option (SERVOP) (C, D and G) which are fixed contracts for continuous full-time service, deployment on operations or gap-year programs respectively. Figure 1.1 shows the Total Workforce Model. SERCATs 6 and 7 are full time service and considered the permanent force, whilst SERCATs 2 to 5 represent reserve force categories. SERCAT 1 represents Defence Civilians, and recognize the Australian Public Service for their Defence service, although not part of the uniformed force (Department of Defence, n.d.). Specific details on differences, entitlements and employment between the SERCAT and SERVOP are not significant to the completion of this study. Only SERCAT 6 and 7 members are considered in this study.

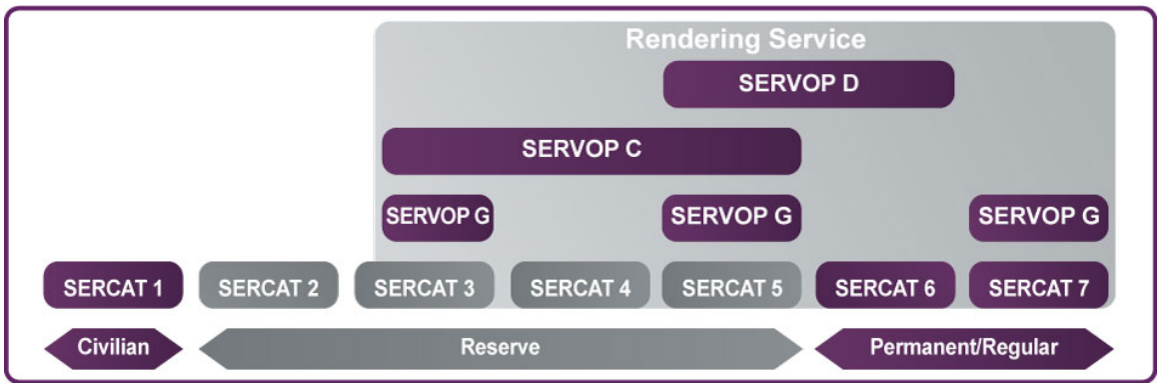

Figure 1.1. The Australian Defence Force Total Workforce Model depicting the spectrum of full-time, part-time, and fixed periods of service. Our study scope focuses on the permanent force (SERCAT 6 and 7). Adapted from (Department of Defence, n.d.)

Beyond the service obligations, an ADF member is not subject to a service contract or finite term limit (Defence Regulation 2016, 2016), and continues to render service until they are subjected to one of three modes of exiting service, generalized as:

1. Serving until a defined compulsory retirement age (CRA), which at the time of writing is 60 years old for permanent force and 65 years old for reserves;
2. A voluntary decision to transfer from the permanent force to the reserve force for at least five years; and
3. The ADF decides to involuntarily separate the member and terminate service due to medical, disciplinary or other eligibility reasons.

For the purposes of this study, the term "voluntary separation" encompasses a member's personal and voluntary decision to transfer from the permanent force (SERCAT 6 or 7) to the reserve force (SERCATs 2 to 5). The term does not include those rendering fixed periods of service under SERVOPs C, D or G.

1.2 Problem Framing

Due to the nature of ADF service, the voluntary intention or decision to separate is not clustered around specific re-enlistment milestones. Instead, it is a decision to leave the service rather than a decision to continue serving. This makes it more difficult at an organizational level to predict when or under what circumstances an individual is most at risk of voluntary

separation, thereby confounding or diffusing policies and initiatives to improve retention. It is also not feasible for an organization to continually survey every Australian Army member to gauge their intention to separate, or to reactively incentivize individuals who have already decided to separate.

Machine learning (ML), using extensive, pre-existing data to predict individual separation decisions is one potential avenue to identify who and when separation may occur. ML is highly appropriate for capturing complex patterns across a large number of features, of unknown correlation or influence on the outcome, to forecast separation and retention outcomes (Lockwood et al., 2020). An organization that can predict this decision and key influencing factors can proactively reduce the risk of an intention to separate, thereby prolonging a member's service and improving retention.

1.3 Research Questions

To address the problem statement, the primary research question (RQ) is:

RQ: Can machine learning be used to predict voluntary separation from the Australian Army?

To comprehensively address this research question, this study seeks to answer the following supporting research questions (SRQs):

SRQ1: What individual service history, demographic, and performance features provide the most significance in predicting voluntary separation?

SRQ2: What machine learning algorithms and parameters are most effective at predicting voluntary separation?

SRQ3: How can the predictive results be used to support managerial decisions and inform policy interventions?

1.4 Scope

This is an exploratory study that focuses on prediction, not causal inference. Our goal is to identify features and patterns useful for forecasting voluntary separation, not to estimate the

causal effect of specific policies or characteristics. As a result, this study is grounded more in turnover theory and organizational behavior rather than labor economics. Although labor economic indicators and inherent conditions of service are utilized within this study, we do not specifically examine or address wages, employment conditions or economic shocks. Simultaneously, we do not consider psychology or seek to identify individual factors such as values and interests in a decision to leave.

Our study explores and tests multiple ML methods. We identify preparatory measures, parameters and tuning that optimizes performance, and evaluate their predictive efficacy. We then undertake analysis to identify which features proved most important and influential in the prediction outcome. It is intended to determine suitability and configurations for future deployment, and direct future studies and organizational activities. Our study is not intended to provide an avenue for identifying specific individuals or sub-groups for targeted intervention.

The observation window encompasses FYs 2015–2016 to 2024–2025 (01 July 2015 to 30 Jun 2025).[1] During this period, the ADF saw the cessation of significant operational deployments to the Middle East, support to national emergencies presented by bushfires and the COVID pandemic , and continual high tempo deployments across Australia and the South-west Pacific to render humanitarian assistance and disaster relief (Department of Defence, 2023b). More recently, Australia's geopolitical situation has shifted to rising strategic competition in the Indo-Pacific and a return to foundational warfighting capability (Department of Defence, 2024a). This period also saw a significant decrease in Australian Army end-strength, resulting in capability challenges (Department of Defence, 2023a, 2024b). Whilst this study acknowledges these events, it can not identify or attribute workforce outcomes to specific Defence activities, policies, or events.

Our study predicts the separation of permanent force (SERCATs 6 or 7) Australian Army members only. It excludes those that render part-time service, or a defined period of full-time service through SERVOPs C, D or G, which represent fixed contracts that then force a decision to continue serving or separate. Although it is not uncommon to transfer from permanent to reserve, to permanent again later, we consider only an individual's first day of service, and only their first separation.

[1] Australian FY begins 01 July of a year, and ends 30 June of the following year.

Lastly, we seek to identify measures that support ADF objectives to increase median length of service from 7 to 12 years (Department of Defence, 2024a). Therefore this study excludes Australian Army members that have less than five years of service. First-term attrition is well studied, with various attribution towards the education, aptitude, health, and fitness of recruits, the recruiting process, and job-match fit (Hoglin, 2012; Hoglin & Barton, 2015; Hosek et al., 1989; Marrone, 2020). Inclusion of those that attrite in the first term is likely to confound identifying the separation characteristics of those in mid- and late-career stages. The majority of IMPS or ROSO are four year engagements although longer exceptions exist for technical and specialist trades. We select five years as the threshold as individuals may have a lag in deciding to separate, including approval and administrating the separation.

1.5 Objective

The objective of this thesis is to explore the use of ML methods to predict voluntary separation decisions and develop insights that may inform policy and managerial outcomes. This was conducted in two primary stages:

1. **Stage 1** is supervised classification, which aims to find insights into "who separates" and the suitability of supervised classification models. This stage trains and tests multiple models on transactional and observable personnel data to predict if an individual will voluntarily separate during an observation window. We then evaluate and choose the model with the best predictive performance. Model outputs and features are examined to determine which demonstrated the most influence on predictive performance, potentially indicating patterns in an individual's demographic, service or performance profile.

2. **Stage 2** is survival analysis, which aims to find insights into "when individuals separate" and evaluate the suitability of survival analysis models. This stage models time to separation, estimating survival functions and hazard rates that describe how the risk of voluntary separation evolves over time. These are then combined with a predictive survival model that seeks to predict the timing of separation, conditional on observed features.

We perform the following broad tasks:

1. Inspect, clean, and join data
2. Conduct feature engineering and transform data for supervised classification
3. Build and execute Stage 1: Supervised Classification models, examining best performing models and parameters
4. Examine across multiple models which features were most influential or important to classification performance
5. Transform data for survival analysis
6. Build and execute Stage 2: Survival Analysis models, examining survival and hazard outcomes for different features
7. Examine importance of features to predicting survival outcomes
8. Analyze and interpret results

1.6 Organization

In this chapter we provide background and context to the Australian Army's workforce challenge. We detail our research questions and define the scope and objectives of this exploratory study.

Chapter 2 contains a review of existing literature into turnover and attrition theory, and the use of ML and survival analysis in predicting workforce turnover, in both civilian and military contexts.

Chapter 3 details the data and methodology. We explain the conceptual framework from which we form predictive expectations and features to engineer. We discuss our data, sources, and preparation process. We describe each of the ML and survival analysis methods used, including evaluation metrics, and design choices.

Chapter 4 provides an analysis of the results. We compare the performance of the supervised classification models, and which features showed notable influence on predictive results. We detail the outcomes of survival analysis models, the survival and hazard outputs, and which features showed notable influence on survival prediction results. Lastly, we examine and interpret the results and evaluate whether expectations formed in Chapter 3 were supported.

Chapter 5 evaluates the study results against the research questions and limitations in what could be achieved. We make recommendations regarding adoption of ML techniques, Australian Army workforce planning and policy, and potential future research.

CHAPTER 2:
Academic Literature Review

This chapter reviews existing literature in four sections. The first section examines turnover theory and military attrition to identify potential indicators of turnover. We then examine studies that apply ML to predict civilian workforce turnover, observing employment of models, datasets, and metrics. We selected these studies to develop methodology for this study. The third section examines ML in predicting military workforce turnover. We observe model suitability and handling of administrative data common to military forces. The last section examines the use of survival analysis methods combined with predictive ML models. These papers represent contemporary methods for improving prediction of turnover and timing of separation based on time-variate features.

2.1 Turnover

2.1.1 Turnover Theory

This section discusses turnover theory and its application in civilian and military literature. Understanding turnover theory allows us to determine drivers of an individual's decision to voluntarily separate. Early turnover theory considered satisfaction, utility, psychological, and organizational factors correlating to turnover, but displayed only modest ability to explain variance (Hom et al., 2017). Mobley et al. (1979) proposed the now widely-accepted Conceptual Model of Employee Turnover (see Appendix A), and has been tested empirically (Johns, 2003; Johnston, 1988; Michaels & Spector, 1982). This model expresses four key determinants influencing an intention to quit. These were summarized by Johns (2003):

1. Job satisfaction, in which various aspects of the job are evaluated against individual values;
2. Attraction and utility of the present job, which is based on expectations of retaining the current job and the attainment of future outcomes internal to the current employment organization;

3. Attraction and expected utility of alternatives, which considers the attraction and attainability of alternative jobs, and the desirability and ease of leaving the current organization; and
4. Non-work values and interests, such as fulfillment of contracts, family reliance, and impact to social networks, which are considered moderating variables to an intention to quit. (Johns, 2003, p. 380)

Newer concepts add to the Mobley et al. (1979) model. Lee and Mitchell (1994) proposed the "unfolding" model, which emphasizes shocks such as unsolicited opportunities or life events that cause sudden or impulsive decisions to quit, bypassing the turnover decision chain. Individuals may not even be leaving work for alternate jobs and roles, instead exiting the workforce to study or provide parenting full time (Lee & Mitchell, 1994). Mitchell et al. (2001) also proposed "embeddedness" which offers a counter by examining the forces that influence a decision to stay, such as training opportunities, job fitness, and social links.

In a military context, there is limited research beyond first term attrition. Boesel and Johnson (1984) undertook an analysis using U.S. military data. They found that pecuniary measures were the most important determinants for re-enlistments, whilst individual characteristics and dissatisfaction were most important for attrition. Quality of life factors appeared to have greater effect in second and subsequent terms of service. Location and relocation appeared not to be powerful predictors of re-enlistment (Boesel & Johnson, 1984).

Johnston (1988) evaluates Mobley et al. (1979) turnover model using U.S. Army junior officers. He found that whilst the turnover process model is present, sub-groups responded differently, particularly to contract lengths and initial obligations. Weiss et al. (2002) undertook a broad review of existing studies of turnover in the U.S. military context. The authors identify organizational commitment and the contract and re-enlistment timings to be key factors. Their review reinforces the idea that military turnover unfolds within a unique institutional and temporal context (Weiss et al., 2002).

For the Australian Army, Tuppin (2025) presents an Army Career Theory Framework, and proposes two structures. The first is an Individual-Work-Society structure that positions career decisions as a complex combination of personal identity, organizational demand and societal forces. The second is a Time-Shock-Adaptability structure that expands on the

unfolding model raised by Lee and Mitchell (1994). Military-specific events such as deployments, promotions and family relocations are continual shocks that then require adaptability as a moderating force. Tuppin (2025) also examines the impacts of operational deployments (as a shock) on military career intentions. They find that operational deployments shortened career tenure. However this was nuanced by whether trauma was experienced during the deployment, and if the deployment was viewed as positive or negative (Tuppin, 2025).

Although Boesel and Johnson (1984) and Johnston (1988) found pay to be of significance to military turnover, there is a noted lack of contemporary research into the topic. The U.S. conducts a Quadrennial Review of Military Compensation that makes comparisons with civilian pay and recommendations on military remuneration (U.S. Department of Defense, 2025). Australia's Defence Force Remuneration Tribunal publishes reports on decisions and determinations regarding pay matters or to address recruiting attraction and targets, but there are no specific studies into effects on military turnover (Defence Force Remuneration Tribunal, n.d.).

2.1.2 Gaps in Australian Military Turnover Literature

Several quantitative studies have been conducted in recent years evaluating aspects of ADF workforce retention and separations. However, there is a clear gap in using ML for predicting individual separation decisions for the Australian Army.

For the Royal Australian Navy (RAN), Dodds (2018) conducted a survival analysis of sailors and officers, utilizing administrative data between 2002–2018 to support Kaplan–Meier (KM) and Cox proportional hazard (CPH) models. Dodds (2018) found that females had higher likelihood of separation than males in the first decade of service, whilst sailors had higher likelihood of separation than officers, particularly around early career milestones. Cole (2019) conducted an exploratory ML analysis of technical and non-technical sailors, testing multiple ML methods on exit survey data from between 1999–2018. They found that technical sailors were more likely to understand their value in the civilian sector and lack of opportunities within the military, whilst non-technical sailors' sentiments related to pay and lack of recognition for effort. Additionally, Cole (2019) tested 12 different models, with various data structures and parameters, finding that Linear support vector machine (SVM) offered the best predictive results for their situation.

For the Royal Australian Air Force (RAAF), Carr (2022) examined retention in the aviation technical workforce by conducting Markov modeling using personnel data from 2005-2020 to determine accession, separation and promotion rates and ability to meet future demand. Their study utilized a scenario for the aviation technical workforce meeting hypothetical demands in 2030–2031, and found significant workforce liabilities hindered achievement. Ashen (2022) examined the effects of families on retention, separation and re-entry behaviors of aviation officers using linear probability models. Their study utilizes personnel records from 2005–2019 and found that having children and being in recognized relationships reduced separation rates and increased re-entry rates. He also found that officers with children had higher separation rates when their eldest child commenced school, or only had one child (Ashen, 2022).

For the Australian Army, Darragh (2022) conducted a classical time-series analysis of enlisted and officer separations and endstrength using data from 2010–2021, seeking to identify separation patterns and a better methodology for strategic workforce planning. Whilst this study may enable better management of accessions to meet endstrength targets, it does not address covariates of separation to improve retention or individual separation decisions.

2.2 Machine Learning in Predicting Civilian Turnover

ML as a quantitative method is now in widespread use across many industries and fields. We reviewed eight studies to inform development of our study. Ahn and Fan (2022) utilize ML to evaluate retention outcomes in the U.S. Department of Defense's acquisition workforce utilizing structured human resource and retirement eligibility data. Sajjadiani et al. (2019) use job application form data as predictors of future work outcomes, demonstrating better predictive power than traditional administrative data. Hess (2025) utilizes a range of structured human resource data, self-reported data, and performance-related features. However, the majority of reviewed studies (Alduayj & Rajpoot, 2018; Fallucchi et al., 2020; Krishna et al., 2025; Mansor et al., 2021; Raza et al., 2022) utilize the popular synthetic IBM HR dataset, enabling comparable evaluation of methods across studies.

The most common ML methods were logistic regression, k–nearest neighbors (KNN), SVM, random forest (RF), and gradient-boosted ensembles such as eXtreme Gradient Boosting

(XGBoost). Neural networks are also tested (Ahn & Fan, 2022; Mansor et al., 2021), though it is out of scope for this study. In most cases, parametric models produced moderate performance and are used as baselines. Best performance was generally observed by non-parametric tree and gradient-boosted models, which are usually more capable in handling collinearity, differing variable types, and detecting non-linear relationships. Fallucchi et al. (2020) demonstrated that simpler algorithms such as Gaussian Naïve Bayes (NB) still offers competitive recall rates as a parametric model. In most cases, tree-based ensemble methods such as RF and XGBoost emerged as the strongest classifiers (Hess, 2025; Krishna et al., 2025; Raza et al., 2022).

Feature importance was also evaluated across models, with the most prevalent predictors being tenure with the firm, pay and working hour variables, job level and performance indicators (Alduayj & Rajpoot, 2018; Raza et al., 2022). Hess (2025) compares feature importance consistency across datasets, showing that pay and tenure are stable predictors across contexts and algorithms.

2.3 Machine Learning in Predicting Military Turnover

Except for Cole (2019), studies reviewed in this section exclusively address U.S. Army (Garcia, 2022) or U.S. Marine Corps (USMC) attrition and turnover (Falk, 2023; Gallagher, 2020; Terrazas, 2020), thereby resulting in retention and attrition decisions that centered around re-enlistment decisions and milestones. Falk (2023) and Terrazas (2020) use ML to more accurately predict retention outcomes to inform USMC endstrength modeling and forecasting. Gallagher (2020) uses prediction outcomes to inform targeting of financial bonuses for re-enlistment. Garcia (2022) conducted an exploratory study to evaluate ML algorithms and methods for potential U.S. Army implementation. Cole (2019) sought to use ML to discern the difference in behaviors and sentiments rather than simple classification.

In looking at data sources, Garcia (2022) was the only U.S. Army study longitudinal administrative data. The remaining three papers utilized USMC data sourced from the Total Force Data Warehouse also in the form of longitudinal administrative data, ranging from 2006–2021. Cole (2019) utilized novel non-compulsory exit survey data from 1999 and 2018, though it should be noted that there was a response rate of 29.5%.

In terms of methodology, all studies employed and evaluated multiple supervised classification models to predict a choice to stay or leave, with the most common models being logistic regression, regularized regression, and RF methods. An assortment of other kernel, tree-based, boosted ensemble and neural network models were also tested. Evaluation metrics were similar to metrics employed in civilian studies, utilizing receiver operating characteristic (ROC) area under the curve (AUC), average precision (AP), recall and F1 statistic. Of the studies reviewed, Falk (2023) was the only study to address class imbalance (using synthetic minority oversampling technique (SMOTE), and the only study to apply threshold tuning for classification refinement. In most cases, RF and tree-based models outperformed other models.

2.4 Machine Learning and Survival Analysis

A limitation of conventional turnover ML is that it ignores the temporal structure of job histories and features that vary over time. Additionally, traditional survival analysis models are difficult to benchmark against ML models as they use time invariant covariates or metrics such as concordance index (C-Index).

Ishwaran et al. (2008) proposed survival analysis to supplement conventional supervised classification models, called random survival forest (RSF). His method uses a tree-ensemble that splits based on right-censored survival data. RSF outputs cumulative hazard function (CHF) which can then be scored against observed survival times as concordance, measured as C-Index (Ishwaran et al., 2008). Ilarda (2021) validated the use of RSF by evaluating a CPH, RSF, and two deep survival models on the IBM synthetic dataset. Their study showed that RSF could handle non-linearity well and outperformed the baseline CPH model (Ilarda, 2021).

Zhu et al. (2019) proposed a hybrid Cox random forest (CoxRF) model. A CPH model is fed into a RF classifier, turning the problem into a binary classification task. This event-based approach enables full use of longitudinal data and evaluation by common metrics such as ROC, AUC and F1 statistic. Sumathi et al. (2021) verified the method, finding similar outcomes and performance. Jin et al. (2020) propose Random Forest, Random Survival Forest (RFRSF), similarly a two-stage process training an RSF which is then fed into a RF. This method was tested on the same data as Zhu et al. (2019) and provided superior

performance against baseline ML models. However, this CoxRF and RFRSF methods appear to be novel, with limited literature that test or use these methods.

2.5 Summary

In this literature review, we examine turnover theory to identify a conceptual framework that drives an individual's separation decision. The Mobley et al. (1979) Conceptual Model of Employee Turnover provide determinants that assist with identifying features that may indicate separation decisions. Additional concepts such as unfolding and embeddedness add shocks inherent in service life and the attractiveness of the organization to the framework. Additionally, the military context adds contractual obligations and other inherent conditions of service that may delay or moderate a decision to separate.

We examine ML used to predict turnover in both the civilian and military workforce context. From the literature reviewed, supervised classification methods using both parametric and non-parametric methods were common. Parametric models have simplifying assumptions and are used as baselines. Non-parametric methods are more elaborate, but are less restricted by these assumptions and appear to consistently outperform other models. In this study we adopt a range of common parametric and semi-parametric models to use for benchmarking (NB, logistic regression, KM, CPH). We then adopt a range of tree and gradient-boosted models with the expectation for good performance against a high-dimensional tabular dataset (RF, XGBoost, RSF).

Further predictive efficacy can be achieved by applying class imbalance handling and threshold tuning. We adopt common metrics for ranking and discrimination (ROC AUC, AP, Recall, F1, C-Index) to evaluate model performance, noting potential for bias based on data balance. Our study addresses an identified gap in Australian Army manpower research. In the next chapter we examine our data and detail the methodology for our study.

THIS PAGE INTENTIONALLY LEFT BLANK

CHAPTER 3:
Data and Methodology

This chapter describes the data and methodology employed in completing this exploratory study. We describe the conceptual framework and expectations of our study. We establish the source and structure of the data, cleaning and preparation, and how we filter the target sample from the population. The subsequent sections explain the models, metrics, and feature interpretation tools utilized within Stage 1 and 2 of our study.

All data manipulation, statistical & ML analysis, and visualization within this study was conducted using Python (Python Software Foundation, n.d.), with the following primary packages:

1. NumPy (NumPy Developers, n.d.)
2. pandas (Pandas Development Team, n.d.)
3. scikit-learn (Scikit-learn developers, n.d.)
4. Matplotlib (Matplotlib Development Team, n.d.)

The topic of military turnover is complex and multi-faceted. As this is an exploratory study, we consider a multitude of features with unknown relationships that may influence a separation decision, and employ a range of ML models, techniques, and metrics. Noting the broadness of ML and statistical techniques utilized, this chapter focuses on specific implementation and design choices, rather than explaining theoretical concepts in detail.

3.1 Conceptual Framework

Using the Mobley et al. (1979) Conceptual Model of Employee Turnover, we develop expectations for predictors toward a voluntary separation decision. We then engineer features from raw data that may contribute to predicting separation.

The first determinant is job satisfaction, which is negatively correlated to turnover. At an organizational level this is difficult to assess without surveying individuals. Other observables, such as length of service or time within an occupation, could be used as proxies to infer satisfaction. We expect increasing tenure to imply satisfaction with the Army as

an organization and the conditions of service, and a significant (negative) importance in predicting separation. We expect increasing length of time within an occupation without a trade change implies satisfaction with that role, employment and its trade structure. and will have significant (negative) importance in predicting separation. However, this timing resets after each trade change, and multiple trade changes may also indicate dissatisfaction or poor fit.

The second determinant is expected utility of internal work roles, which is future focused on the availability career opportunities internal to an organization. The hierarchical structure of the military generally creates transparent and clearly-stated requirements for promotion and higher pay. Unfortunately, these requirements may be hard to achieve or subject to competitivity in a tournament-style selection process. This could be assessed by examining the career progression or stagnation of an individual, as well as at what time or career milestones tend to trigger voluntary separation among individuals and sub-groupings. We expect increased time in rank implies a stagnation of career opportunities and a lack of upward mobility, with significant (positive) importance in predicting separation. To examine this further we engineer features that indicate proportions against their cohort, and time in rank relative to age or tenure. We also expect that at junior enlisted and junior officer ranks, less is known about navigating career and promotion systems, with less influence over their career management. Therefore we expect rank will have significant (positive) importance in predicting separation.

The third determinant is external work roles, which provides the conditions and availability of external work opportunities for a person to not only consider separating, but to go through with it. The external labor market and economic conditions (potentially up to 12 months prior to a separation decision) can indicate the general availability and ease of finding, accepting and separating from the Army for external work opportunities. When the labor market is considered tight (low unemployment rate), jobs are plentiful and harder for employers to fill; a loose market (unemployment rate is high) workers are abundant and jobs are easier to fill (Ehrenberg et al., 2021, p. 29). We expect the prevailing labor market conditions may increase or reduce external work opportunities and will have significant importance in predicting separation.

The last determinant is non-work values and interests. This speaks of two primary areas: that of whether personal values align with those of the organization, and whether individual needs/interests are being met whilst serving within the Army. The former is difficult to detect through large organizational level datasets and are usually more impacted by local level effects such as unit culture, leadership, and adverse incidents; the latter can be considered as inherent conditions of service within the military, and whether the adversity is offset by the material benefits and intrinsic value proposition of Army.

Administrative datasets are suited towards recording and detecting the multitude of changes and churn across an individual's service, personal and family life. Therefore we predict that continual and frequent change of posting locality for service requirement is disruptive. This will have significant (positive) importance in predicting separation. We also expect that as a member undergoes marital and dependent changes, the disruptive effect correlates with breakdown in marriages and families. This will have significant (positive) importance in predicting separation. Similarly to trade changes, we engineer features that indicate count, frequency and time since change.

3.2 Study Design

As we are conducting an exploratory study, we select and describe a range of models, metrics and evaluation tests. Our research questions require us to observe both model and feature performance. Our conceptual framework states what behavior we expect to see from features, and guides us in how we observe this behavior. Our overall study design is detailed in Figure 3.1. Starting from the left, we describe the form of data required to support each stage. In Stage 1, we directly compare model predictive performance, and use permutation importance (PI) to observe globally averaged feature ranking, and influence for each model. We also use SHapley Additive exPlanations (SHAP) to observe the marginal effects of individual values on the magnitude and direction of predicted separation risk.

In Stage 2, we use traditional survival analysis models to baseline a predictive survival model's performance. We then test predictive performance, feature importance, and local effects over time. Lastly, we use a technique called landmarking to test whether a model can be effective at predicting separation at various time intervals prior to separation, and which features may have influence at different time intervals.

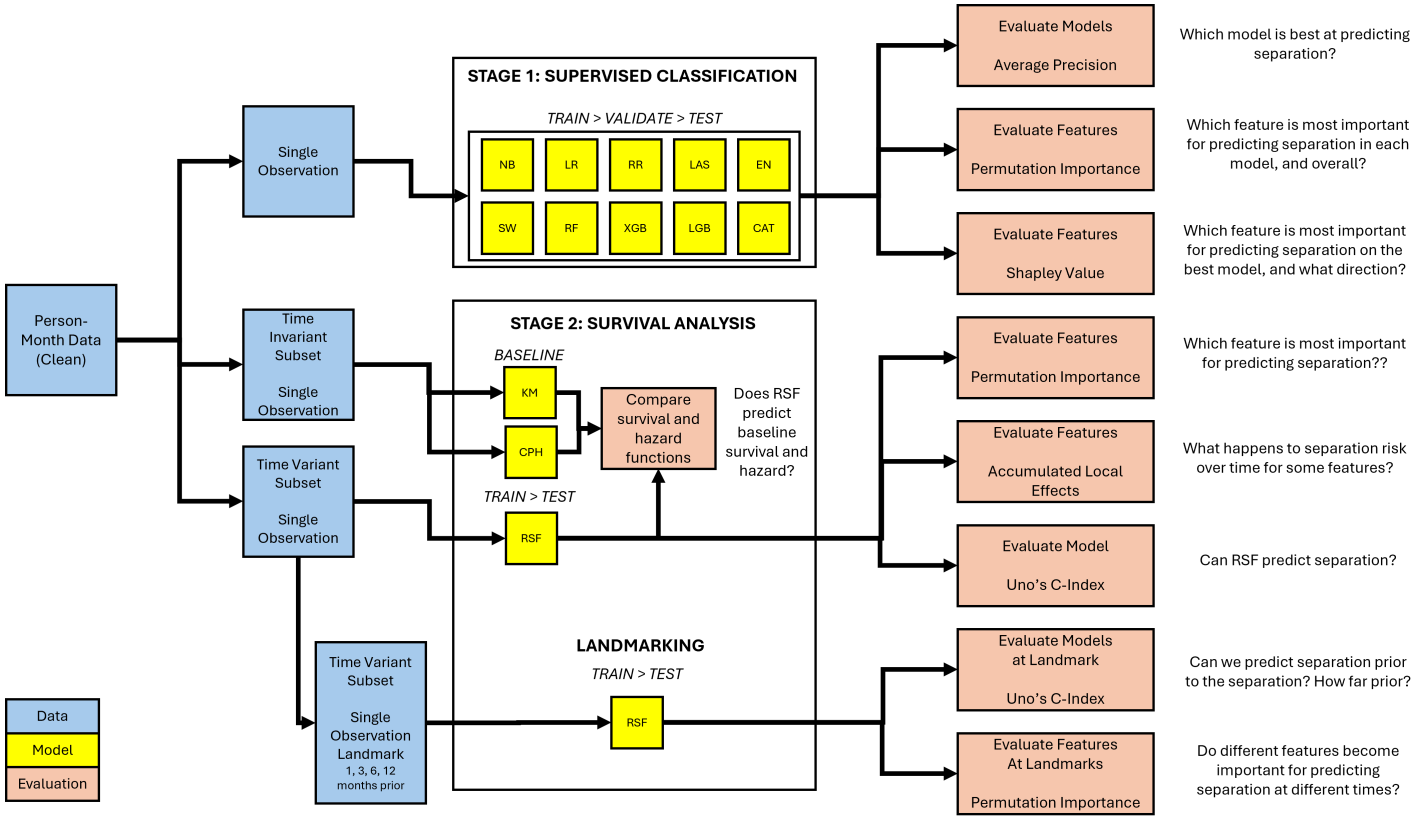

Figure 3.1. Study design showing data requirements, which models will be implemented, and what questions we seek to answer at each evaluation activity. This broad design allows us to observe and evaluate model performance, feature importance to predictive performance, and marginal effects to predicted probability of separation. This can be done at the individual model level, and globally across models.

3.3 Data

This section explains the source and scope of data for our study. This section also describes how the data was inspected, handled and prepared, including truncation and censoring to achieve our target data sample.

3.3.1 Data Sources

The primary data source is ADF human resources data stored in the Management and Reporting Solution (MARS) system and provided by the sponsoring organization, the Directorate of Workforce Analysis in the Defence People Group (DPG). This transactional data was stripped of personally identifiable information (PII) by the sponsor, with each individual assigned a unique Party ID. This data contains demographic information, personal and service-related administrative transactions, operational service records, and performance reporting data.

The data temporal range covers FY 2015–2016 to 2024–2025 (01 July 2015 to 30 June 2025), providing a maximum 120 observation months. This data included only those that rendered permanent full-time service during the observation window and excludes reservists.

The secondary data source was macro-economic data taken from the Australian Bureau of Statistics (ABS) Labour Force statistics updated to November 2025 (Australian Bureau of Statistics, 2025a). The primary records utilized from the ABS were national unemployment rate (Australian Bureau of Statistics, 2025b) and national job-search duration (Australian Bureau of Statistics, 2025c) records, observed in monthly increments.

3.3.2 Data Structure and Joining

The primary human resources data came in several report formats offering different observation points in time and required understanding so as to logically join the data. This data includes those who joined service prior to or during the observation window, those who separated within the window, and those who continued to serve through to the last observation. This section explains the different data formats and how they were joined. This section describes the data and how it was joined to create longitudinal data. An entity relationship diagram is at Appendix B.

Service and demographic data was provided as person-month rows, offering a snapshot of an individual's demographic and service status on the last day of each month. Therefore, a change such as rank, family status, or a move between units and locations must be detected as a change between two observations. A person who served partway into a month and then separated before month's end would not be observed. Separations data detailed all those who had voluntarily or involuntarily separated from the Australian Army during the observation window. This was provided as a single row record, including the specific date of separation. Therefore, when joining this data to person-month headcount data, if the separation month matched the individual's last person-month entry, then all columns from the separations data was joined to that last entry. If the separation month was later than the last person-month entry, then a new row was created by copying the last month's entry (assuming no change in non-time dependent features from the last row) and joining the separation data to the newly created row.

Performance reporting in the Australian Army is conducted annually and data was provided as an annual observation per person. When joined to person-month headcount data, promotion recommendation was joined from month of effective reporting date and carried forward until the next effective promotion recommendation change. Operational service data was provided as a single row per person, indicating the cumulative duration of operational service rendered over their career. It also contained the last observed date of operational service, as a snapshot at the end of the observation window. Macro-economic data from the Australian Bureau of Statistics was provided as monthly national averages for unemployment rate and job search duration, and joined by relevant month after feature engineering.

A longitudinal data format is required to be able to correctly join data with varying temporal ranges and to engineer features relating to changes over time. However, the ML models in this study ultimately require single record per person, and we aggregate at varying points of observation.

3.3.3 Truncation and Censoring

Our time-based snapshots are, in most cases, an incomplete record of an individual. Therefore within the data we must apply truncation and censoring in which we handle the absence of complete data about a person prior to, during and after the observation window (Klein-

baum & Klein, 2012). Although this is not expressly required for supervised classification, it is key to survival analysis, and the same underlying assumptions are used for determining eligibility within the study. We use the same dataset to ensure consistency across stages. In this section we do not provide detailed definitions of survival analysis terms or concepts—see Kleinbaum and Klein (2012)—but discuss reasoning and assumptions for inclusions or exclusions of individual observations.

In our case, exposure commences when an individual first joins military service (i.e., their Hire Date), and the event of interest is when an individual chooses to voluntarily separate from the Australian Army, which includes transfer to reserves or to another ADF service. Although it is not uncommon for an individual to separate or transfer from permanent to reserve, and re-join or become permanent again later, we consider the exposure start time as their first day of service, and ends at their first separation. The observation window starts at 01 July 2015 and ends 30 Jun 2025. Individuals who enter or join during the window but do not separate within the observation window are treated as right-censored, without considering whether they voluntarily separated after the observation window.

Truncation occurs when an observation is excluded based on time to event. As our goal is to evaluate turnover after first term attrition, we removed the confounding effect by filtering out individuals who had less than five years of service. This considers a member's four years of IMPS with an additional year for leaving service. Additionally, we removed those who were involuntarily separated.

The data also contains individuals who entered service prior to the observation window start date. Although their earlier history is unobserved, they are considered at risk of separation from the observation window start date, conditional on having survived in service up to that time. The key impact will be limitations to variables that are a cumulative count of changes or time since a change and are therefore limited to what is observed within the window.

3.3.4 Data Summary

The dataset originally started with 3,511,119 records and 58,747 unique individuals within the observation window. After involuntary separations and those with less than 5 years of service were removed, 29,022 unique individuals remained within the sample (2,452,070 records). Figure 3.2 shows the number of unique individuals within the sample at each

point in time in the sample. A healthy growth is observed up until 2020, followed by a steep decline in total individuals. Of this, 17,687 were retained at the end of the window, and 11,335 voluntarily separated during the observation window. This left a class balance ratio (retained vs. separated) of approximately 61:39.

Table 3.1 is an abbreviated table summarizing baseline demographic and service-related characteristics, after censoring and truncation is applied. A full table is at Appendix C. These statistics are descriptive and are not presented for causal analysis. The null hypothesis tested here is that there is no difference in the characteristics between retained and voluntarily separated personnel; as all p-values are <0.001 we reject the null hypothesis, and note that the difference between the retained and voluntarily separated groups is significant. However given the large sample size, statistical significance is expected and is not interpreted substantively.

Table 3.1. Baseline characteristics by separation status (analysis sample). Involuntary separations and those with less than five years of service are removed from the sample. These statistics are descriptive and are not presented for causal analysis.

Characteristic	**Retained (N=17,687)**	**Voluntary (N=11,335)**	**p-value**
Age (years)	37.14 (9.52)	34.76 (9.28)	< 0.001
Length of service (years)	14.69 (9.00)	12.92 (8.77)	< 0.001
Observed in window (months)	102.51 (23.76)	56.01 (30.03)	< 0.001
Sex			< 0.001
M	14,905 (84.3%)	9,973 (88.0%)	
Rank (collapsed)			< 0.001
Junior soldiers (E00–E04)	3,627 (20.5%)	4,069 (35.9%)	
Senior soldiers (E05–E10)	8,881 (50.2%)	5,136 (45.3%)	
Junior officers (O00–O03)	2,297 (13.0%)	969 (8.5%)	
Senior officers (O04–O10)	2,882 (16.3%)	1,161 (10.2%)	
Total sample after exclusions: N = 29,022			

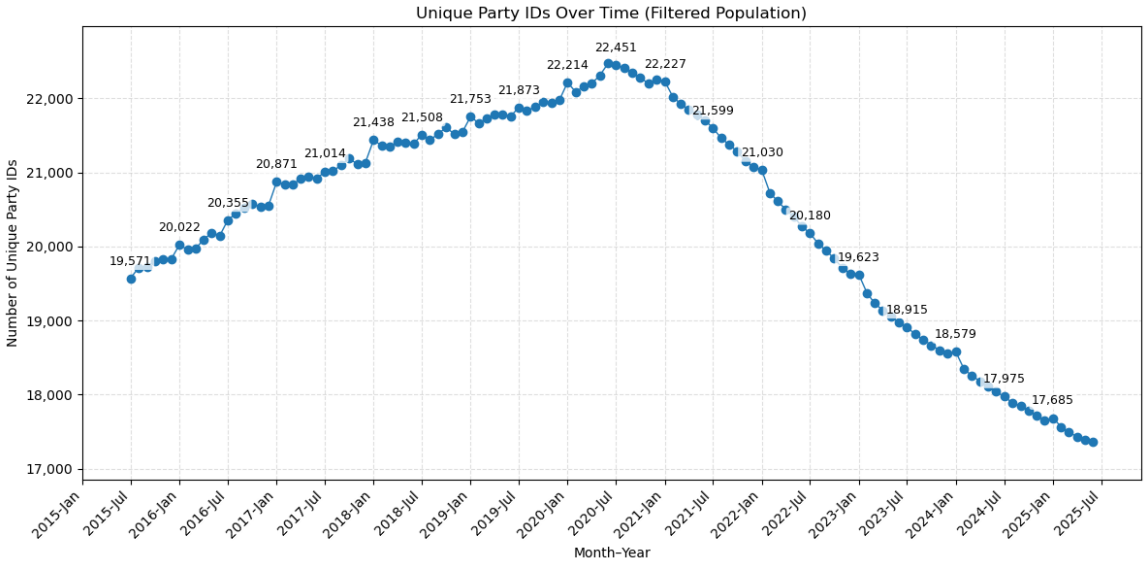

Figure 3.2. Number of unique individuals observed over time (excluding involuntary separation and less than five years of service).

3.3.5 Inspection and Cleaning

We individually inspected each feature within the data, applying a series of exclusion and inclusion rules to remove records that were incomplete, inconsistent, or outside the scope of the study. A notable exclusion was education levels, which are self-reported and were very sparsely or inconsistently filled. Another was variables relating to dependent categorization, marital status, and family composition, which were also sparsely or inconsistently filled. Removal of these variables means that family situation as a predictor of separation could not be evaluated.

We conducted internal logic checks; observations with invalid or implausible data (such as negative tenure, promotion dates preceding hire dates, individual ages exceeding 300 years old) were either removed or corrected where reliable information existed. Some duplicate records were encountered and resolved by prioritizing the most complete or most recent record where conflicts occurred.

We handled missing data conservatively by not applying imputation and preserving missingness to avoid introducing bias. Where structural missingness (due to non-availability or change in recording methodology) was observed, the feature was removed. Where necessary,

raw variables (particularly categorical variables) were standardized and lightly aggregated to improve consistency, such as combining redundant, erroneous, or synonymous categories.

Key principles governing the cleaning process were no imputation or feature construction, focusing on data validity and consistency, and adopting conservative changes. We did not apply any optimizations that would improve or optimize the prediction or classification process during this step.

3.3.6 Feature Engineering

We conducted feature engineering to transform the raw administrative records into meaningful representation of individual career histories, guided by the conceptual framework. This was required as the data did not natively depict changes between one observation to the next, nor track time since a change or cumulative changes.

We constructed cohorts and engineered temporal features that capture an individual's career progression, both individually and relative to their cohorts. Transactional records were aggregated into cumulative and time-since-event measures. This included counting the observed number of promotions, relocations or job-role changes and tracking time since a promotion, relocation, or job-role change. These features reflected the dynamic aspect of service; an individual's satisfaction or disruption may diminish or change with time, and may be compounded through frequency or cumulative count of these events.

Some additional features were engineered to identify disparity, such as cohort rank proportion, time in rank to length of service ratio, relative age in rank, time in rank compared to cohort average, and an indicator if an individual conducted a relocation off-cycle (i.e., outside the peak posting relocation period November to February).

Feature engineering was also intended to identify changes in family circumstances, such as marital status and family categorization, counting cumulative changes, and counting time since change. However, after we inspected the data, it was found to be unreliable and inconsistent. Additionally, the ADF changed family categorization systems throughout the observation window. Remediation of these features was not achievable, and these features were excised from the data. Therefore, we were unable to generate features relating to marriage, dependents or family categorization.

We also transformed macro-economic data and appended it to each record. We assume that when someone considers a voluntary separation decision, it would take approximately three months to consider the labor market, search, and apply for a job. Additionally, within the ADF, three months is considered the minimum time for processing a voluntary separation, unless there are compelling reasons (such as an active job offer), in which case separation processing may be expedited to one month (Department of Defence, 2025a). To reflect this the national unemployment rate was applied both as a three-month lag, and as a rolling three-month average. National average job search duration was similarly applied as a one-month lag, and as a rolling three-month average.

Lastly, we consolidated or aggregated some categorical features to either fewer categories, or as indicator flags. Some examples are 14 granular ethnicity categories remapped to 7 broader categories, and reducing 59 different religion categories into an indicator flag to show if the individual had recorded religion. This addresses sparsity issues and improves model stability and performance.

3.3.7 Collinearity Check

Before commencing the supervised classification, we conducted a collinearity check to assess dependence between numeric features, particularly as engineered predictors can be highly correlated. Highly correlated predictors risk inflating or destabilizing coefficient estimates by confounding true relationships between the predictor and outcome. It may also confound feature importance when influence is split across correlated features. Although some models such as tree-based models are more capable of handling correlated variables, baseline models such as logistic regression are more sensitive, and it was better to standardize which features were taken into analysis.

Collinearity of numeric features was assessed in multiple stages. Firstly we calculated pairwise correlation (using Pearson correlation coefficient) to identify where bivariate correlation may exist between features; this is shown in the correlation matrix at Figure 3.3. We observed high correlation clusters around three areas: age and years of service; macro-economic engineered features; and operational service variables. We computed variance inflation factor (VIF), which quantifies how much variance there is in the inflation of regression coefficients due to correlation (Hair et al., 1999). High VIF value features

were individually compared and redundant features were removed. Preference was given to features that had clearer temporal interpretation and alignment with the conceptual framework. We used a second round of pairwise correlation and VIF to assess the effects of the change. The final correlation matrix is at Figure 3.4. A more desirable result was achieved compared to the initial matrix, since we observe less dense clusters of correlation. Note that the separator flag (our target variable) was also introduced to observe the direction of correlation.

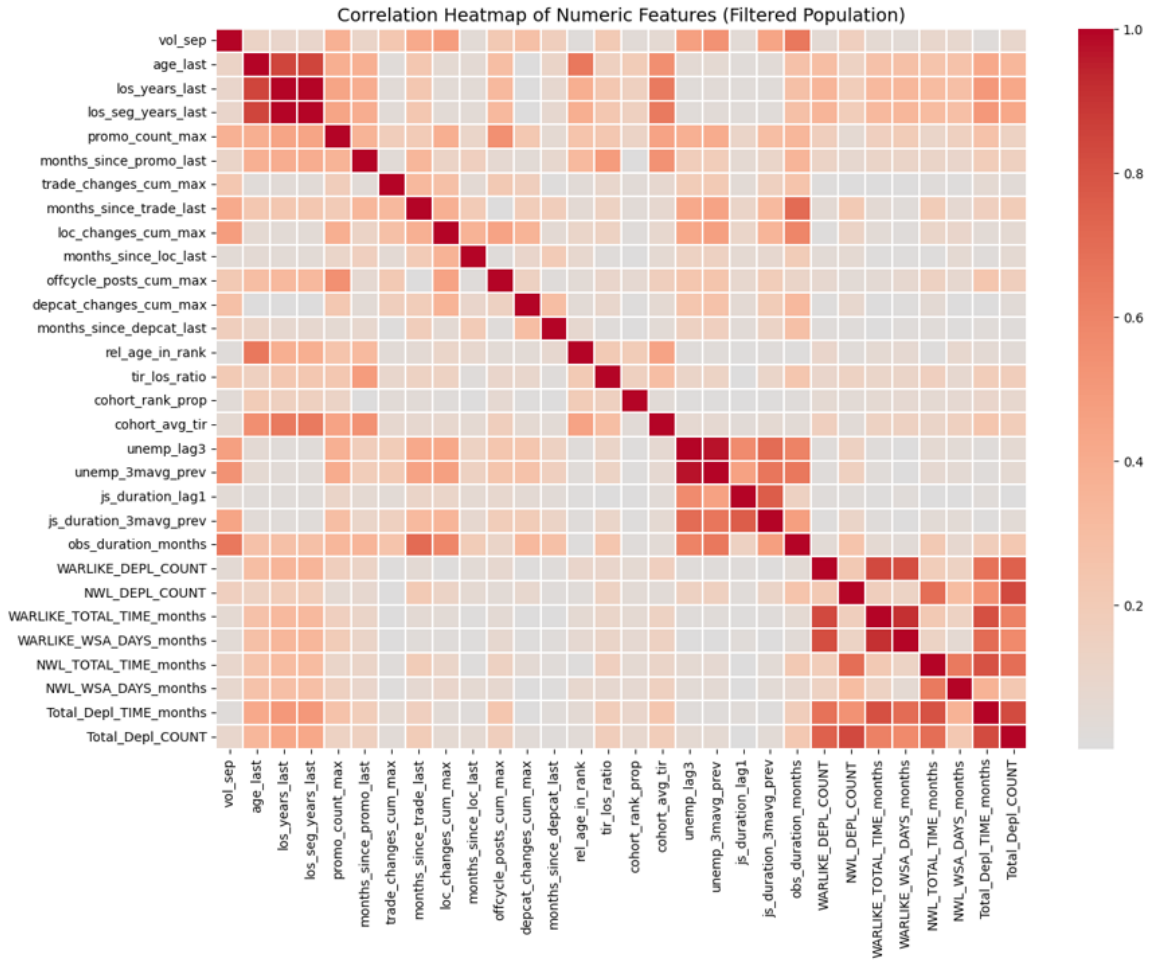

Figure 3.3. Initial correlation matrix, plotting Pearson's correlation coefficient between two variables. Darker colors indicate a higher Pearson correlation value between the two variables being compared.

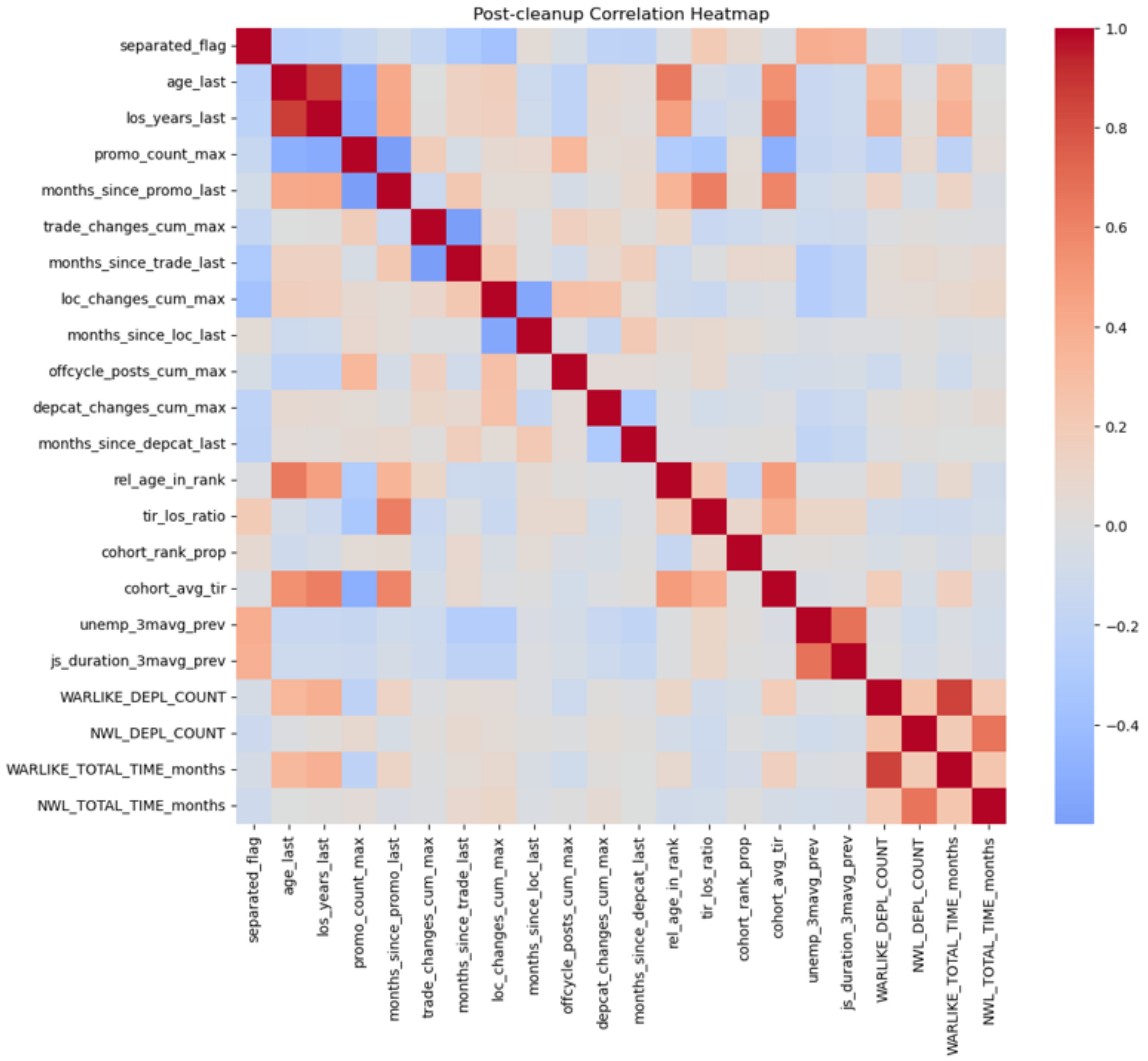

Figure 3.4. Final correlation matrix, plotting Pearson's correlation coefficient between two variables after removal of highly correlated and redundant features. Note that the separator flag is included. Darker red colors indicate higher positive correlation between the two variables being compared. Darker blue colors indicated higher negative correlation.

3.3.8 Pre-Processing and Class Imbalance

Data pre-processing was conducted by partitioning the data into 64:16:20 (training:validation:test) splits. A validation partition is required to conduct threshold tuning, which will be discussed later in this section. We held these split datasets constant across all models to ensure comparability of performance metrics. Each of these datasets were then split further by identifying

the outcome variable (separation) and predictor feature set. Data pipelines were then formed to appropriately deal with numeric and categorical feature types. Numerical features were standardized, whilst categorical features were One Hot Encoding (OHE). A data dictionary detailing the full list of variables is at Appendix D.1.

The outcome variable classes were moderately imbalanced at approximately 61:39 retained versus separated. This class imbalance was addressed in pre-processing by using SMOTE, as described by Chawla et al. (2002). This is achieved by over-sampling the minority class to balance the distribution of outcomes, thereby encouraging smoother decision boundaries lowering the likelihood of overfitting. Instead of oversampling by duplication, SMOTE interpolates between observations and their nearest neighbors (Chawla et al., 2002). This is conducted only on the training dataset to prevent leakage.

3.4 Stage 1: Supervised Classification

This section describes the preparation and execution of supervised classification models, and how model performance and feature importance is evaluated.

3.4.1 Data Preparation

For the supervised classification stage, we aggregated the longitudinal records into a single observation per individual. This was required for a binary person-level outcome and avoids a dataset with multiple highly-correlated records per person. This is key to preserving independence assumptions. Aggregation was conducted by taking the last observation for an individual; this was when they chose to voluntarily separate, or when they were censored at the end of the observation window. This snapshot captures the individual's final demographic, service and performance reporting state, the cumulative total and time since a change event (promotion, relocation, job-role change), and prevailing macroeconomic conditions.

3.4.2 Baseline Models

Naïve Bayes

NB is a probabilistic classifier based on Bayes' theorem and assumes independence between predictor features (James et al., 2023). NB provides a useful baseline because of this simplifying assumption. However, this independence assumption also limits the ability to capture interactions between covariates (James et al., 2023). In this study, NB is used as a benchmark for supervised classification performance.

Logistic Regression

Logistic regression models the probability of voluntary separation as a function of a linear combination of predictors. It is useful as a baseline for binary classification problems as it is interpretable with well-understood statistical properties and is commonly used in ML and statistical research (James et al., 2023). However, the logistic regression model requires interpretation on a log-odds scale, and can be sensitive to multi-collinearity. In this study, logistic regression is used as a benchmark for comparing other ML models, as it does not assume independence and may capture interactions and linear relationships.

Stepwise Regression

Stepwise regression extends logistic regression by iteratively adding or removing predictors based on statistical tests and other metrics. This approach automates feature selection to reduce dimensionality. However, it is relatively computationally intensive and sensitive to sampling variation, and may not select best fitting models (James et al., 2023). In this study, we use forward stepwise selection, scored by F1 (discussed later in this chapter), and is included for comparison, rather than as a preferred model.

Regularized Regression

Regularized regression techniques are a shrinkage method that applies penalty terms to the logistic regression function. It seeks to mitigate multi-collinearity and reduce insignificant or redundant variables, thereby reducing model complexity and preventing overfitting. In this study we utilized Ridge, least absolute shrinkage and selection operator (LASSO), and Elastic Net techniques. Ridge regression applies a penalty which shrinks some coefficients to near-zero (without setting them to zero), LASSO applies a penalty that sets some coefficients

to zero, whilst Elastic Net combines both types of penalties simultaneously (James et al., 2023; Tibshirani, 1996). These models balance interpretability and predictive performance and are more suited to handling correlated features. This imposed feature selection usually performs better than ordinary logistic models (James et al., 2023; Tibshirani, 1996). We have included them here for comparison, though we note that previous collinearity reduction actions will likely minimize their performance gains.

3.4.3 Tree and Boosted Models

Random Forest

RF is an ensemble bagging method that constructs multiple decision trees through bootstrap sampling and randomized selection of features. Final predictions are obtained by averaging (for regression) or take majority voting (for classification).This reduces variance and reduces overfitting (Breiman, 2001). RF are capable of capturing non-linear relationships, are usually robust to multi-collinearity, and require minimal scaling. However, RF are not as interpretable as parametric models (Breiman, 2001). In this study, RF is used as a benchmark for non-parametric models.

eXtreme Gradient Boosting

XGBoost is an optimized variant of forest models, building gradient-boosted decision trees in parallel (rather than serial like other gradient-boosted methods) and is highly scalable (Chen & Guestrin, 2016). XGBoost iteratively fits trees with the objective of correcting residual errors of the previous iteration, with the final outcome being a weighted sum of all tree predictions. XGBoost usually produces high predictive performance but requires parameter tuning to avoid overfitting (Chen & Guestrin, 2016). We have included it here as a best-in-class and widely used model for comparison.

Light Gradient-Boosting Machine

light gradient-boosting machine (LightGBM) is a gradient boosting method that is suited to very large datasets. It uses histogram-based learning by binning continuous features that then reduce the need to continually run re-sorting operations (Ke et al., 2017). LightGBM grows trees leafwise by only expanding the leaf that yields the greatest reduction in loss;

this is opposed to the XGBoost method of growing depthwise, in which nodes are fully expanded before proceeding to the next (Ke et al., 2017). LightGBM is computationally fast, however is prone to overfitting without careful parameter management (Ke et al., 2017). It is evaluated here as a high-performing alternative to XGBoost.

Categorical Boosting

Categorical Boosting (CatBoost) is another gradient-boosting method that incorporates better native handling of categorical features through encoding (Prokhorenkova et al., 2018). This reduces the need to separately pre-process categorical features and usually performs well in datasets with many mixed data types, such as the one used in this study. CatBoost is utilized here to assess whether specific treatment of categorical features improves predictive performance in comparison to other boosted models.

3.4.4 Performance Evaluation

Comparing the predictive performance of ML models is not simplistic. Each model processes and outputs differently, and different metrics capture varying aspects of predictive behavior and performance. This challenge is made more difficult where a class imbalance exists within the data and the outcome of interest is more rare. Metrics that perform well under balanced class distributions may be biased and unreliable for imbalanced data. In this study, multiple metrics are used to discuss model behavior, with particular emphasis placed on discrimination, class-specific performance, and decision-threshold sensitivity. Therefore, not all metrics are treated as equally informative for model comparison.

Classification metrics are calculated by comparing if a predicted positive or negative value was truly positive or negative, and is represented in the confusion matrix in Table 3.2. Common metrics using this confusion matrix are shown in Equations 3.1 to 3.5.

Table 3.2. Confusion matrix for binary classification.

	Predicted Negative	**Predicted Positive**
Actual Negative	True Negative (TN)	False Positive (FP)
Actual Positive	False Negative (FN)	True Positive (TP)

$$\text{True Positive Rate (TPR)} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3.1}$$

$$\text{False Positive Rate (FPR)} = \frac{\text{FP}}{\text{FP} + \text{TN}} \tag{3.2}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{3.3}$$

$$\text{Recall} = \text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3.4}$$

$$\text{F1} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{3.5}$$

Receiver Operating Characteristic and Area Under the Curve

The ROC curve plots true positive rate (TPR) against false positive rate (FPR) across all possible classification thresholds, providing a threshold-independent measure of discrimination (Hanley & McNeil, 1982; Swets, 1988). The area under ROC is widely used as it can easily be interpreted as the probability that a randomly chosen actual positive instance is ranked above an actual negative instance. Therefore, ROC AUC evaluates a model's ability to rank observations correctly rather than its performance at any specific decision threshold.

A key limitation of using ROC AUC for imbalanced datasets is that FPR are normalized by negative observations, as usually the negative class is more numerous, while the positive class is rare. This results in ROC AUC remaining high even in the presence of high false positive (FP), with little penalization (Davis & Goadrich, 2006; Saito & Rehmsmeier, 2015). In practice a model that predicts many FP at any threshold can still achieve a high ROC AUC and is misleading.

Precision-Recall Curve and Average Precision

The precision-recall curve (PRC) provides an alternative view of classification performance by plotting precision against recall across decision thresholds. Unlike ROC, PRC focuses more specifically on the positive class and is therefore more informative against imbalanced datasets (Davis & Goadrich, 2006). AP, which is calculated as the area under the PRC, reflects the trade off between correctly predicting true positive (TP) cases and avoiding FP (Davis & Goadrich, 2006). Saito and Rehmsmeier (2015) shows that PRC and AP are more sensitive to class imbalances, and therefore more useful when the real-life cost of a FP (such as medical or safety settings) can have significant consequences. In this study, we observe ROC AUC, but as our sample is moderately imbalanced, we prioritize PRC AP as the primary metric for comparing model performance.

Threshold-Dependent Metrics, Tuning, and Cross-Validation

While ROC AUC and PRC AP are threshold-independent, the practical employment of ML requires selection of a decision threshold, which is a point at which the model evaluates a probability as positive or negative. This can be arbitrary (usually 0.5) or a decision maker may accept risk or offset the practical costs of FP or false negative (FN) by adjusting the threshold. In this study we conduct automated threshold tuning in which the threshold is iteratively adjusted to find a defined balance between true and false predictions. Our threshold tuning prioritizes F1 (see Equation 3.5), which is a balanced mean between precision and recall when both FP and FN carry significance (Fernández et al., 2018). Additionally, automatic threshold tuning using F1 decouples our choice of decision threshold from the comparison of models, as each model is then performing at its "best."

To improve training of an ML model, we conduct k-fold cross-validation. Cross-validation works by randomly splitting and sampling the partitioned validation data set into k number of subsets, iteratively evaluating each as a test set whilst keeping the remainder for training, and then averaging results. This works to reduce underfitting or overfitting when the model is evaluated against the hold-out test data set (James et al., 2023). In this study, we conducted a five-fold stratified cross-validation.

3.4.5 Feature Evaluation

Feature evaluation across multiple classification models is not possible with native outputs and presents a comparability and interpretability challenge. Linear models such as the logistic and regularized regression models produce log-odds coefficients, whilst tree-based models provide feature importance using metrics such as split frequency or impurity reduction (Altmann et al., 2010). These measures are model-specific, preventing features from being directly comparable.

In this study, we use PI and SHAP. Feature evaluation was conducted on the partitioned test data set, using models already trained on the training data, preventing leakage. All data preparation, imputation, scaling and encoding was preserved. Feature evaluation results are interpreted as indicators of predictive ability only, not causal effect.

Permutation Importance

To enable consistent and model-agnostic comparison, we use PI as described by Fisher et al. (2019), which is performance-based and builds on feature importance introduced by Breiman (2001). This method is conducted by iterating through the set of features available to a model, permuting (i.e., shuffling) the values across observations within that feature, and measuring the decrease in the model's predictive performance. A feature is 'unimportant' if there is negligible change in the predictive performance (Molnar, 2019). In Stage 1 of this study, we conduct PI and measure change using PRC AP as the defining metric. Results are then collated, normalized, and ranked globally. This method however only shows whether there is a global effect by that feature on model performance.

SHapley Additive exPlanations

SHAP is another method that complements PI. This method iterates through an observation's features, and for each value comparing it to other observations, and determining the marginal contribution of that value to the predicted outcome (Molnar, 2019). Shapley importance analysis is able to attribute the marginal contribution of an individual observation to the prediction probability at an individual level, as well as direction in relation to the average outcome. Shapley values can be aggregated to discern average global contribution of a feature. SHAP is model-agnostic, and is considered as an 'explainable' framework for understanding the black box of ML models, including complex high dimensional models

and data sets. However, it's disadvantages are that it is computationally expensive, and results can be difficult to interpret (Louhichi et al., 2023; Lundberg & Lee, 2017; Molnar, 2019). In this study, SHAP is used to complement PI by providing insight into the direction and variability of feature effects. As it is computationally expensive, SHAP is conducted on only the best-performing model.

3.5 Stage 2: Survival Analysis

Stage 2 applies survival analysis methods to model time-to-voluntary separation, complementing the supervised classification conducted in Stage 1. Whilst Stage 1 addressed the question of 'who separates', survival analysis addresses 'when separation occurs'. Survival analysis has a long history in medicine, workforce attrition, and reliability analysis (Kleinbaum & Klein, 2012). Although survival analysis is framed as a traditional statistical modeling technique, we implement an ensemble-based predictive survival model as the primary objective, which aligns it with ML approaches. In this study, we create baselines using KM and CPH model. We then implement RSF, evaluate it for predictive performance, feature importance, and observe time-based effects.

3.5.1 Data Preparation

For KM and CPH models, a baseline survival dataset was constructed based on the longitudinal person-month dataset prepared prior to Stage 1. This dataset was collapsed to one observation per person, containing only time-invariant demographic and service attributes (sex, officer vs. enlisted, migrant status, ethnicity, age-at-hire, cohort decade) as recorded at their entry time (start of observation). An event indicator was added, identifying whether they separated during the observation window. An additional column was also added, recording the number of months observed prior to separation (or censoring).

For RSF, a second longitudinal dataset was also prepared, using the longitudinal person-month dataset prepared prior to Stage 1, but is aggregated at the temporal point of analysis for landmarking (discussed later in this chapter). Additional columns include time observed and an event indicator for each month. This dataset preserves the temporal ordering of events and allows RSF to model time-to-event outcomes. At each observation, feature values could only contain information available up to that point in order to prevent leakage, such as how

many re-locations a member had experienced or how long a person had spent in a rank, at that point in time. Of note, operational service data had to be removed as it was an aggregated dataset.

3.5.2 Descriptive Survival Analysis

Kaplan-Meier

KM estimation was used to determine the probability of remaining in service beyond a given time (Kleinbaum & Klein, 2012). KM survival curves were used descriptively to visualize basic separation patterns and subgroup differences (sex, officer versus enlisted, migrant, ethnicity, age-at-hire, cohort decade). This provided a contextual baseline for subsequent survival analysis. No specific covariate effects or individual-level predictions were inferred from KM estimates.

3.5.3 Semi-Parametric Survival Models

Cox Proportional Hazard

The CPH model was used as a semi-parametric baseline for modeling the relationship between covariates and the hazard of voluntary separation. The CPH model estimates hazard ratios under the assumption of proportional hazards, and a log-linear relationship between covariates and the hazard function (Kleinbaum & Klein, 2012). Although widely used for estimating effects of covariates, the CPH model is not well-suited to complex or high-dimensional interactions and non-linear relationships. The CPH model is used here as a benchmark for comparative evaluation against other models (Ilarda, 2021).

3.5.4 Predictive survival models

Random Survival Forest

We use the RSF model as proposed by Ishwaran et al. (2008), which extends the RF framework to right-censored time-to-event data. Similar to RF, RSF uses an ensemble of trees, bootstrap sampling and random feature selection. However, tree splits are decided using log-rank criteria that maximize separation in survival between nodes while accounting for censoring (Ishwaran et al., 2008). At termination, each tree estimates a CHF, with the

ensemble prediction obtained by averaging the CHF across all trees. RSF models capture non-linear covariate effects and interactions. Prior comparative studies have demonstrated that RSF may outperform classical CPH models in complex and high-dimensional environments (Ilarda, 2021). RSF does not assume proportional hazards like CPH, and is a non-parametric predictive model. In this study, we use it to make survival predictions, and examine local effects based on sub-groups through accumulated local effect (ALE) plots (Molnar, 2019).

Landmarking

To complement survival modeling, landmarking was adopted to examine how observable characteristics evolved in the period leading up to a voluntary separation. This is a common approach in time-to-event modeling (Pickett et al., 2021; Sheu et al., 2023). Rather than modeling survival over the entire observation window, landmarking aims to identify signals at 1, 3, 6 and 12 months prior to separation. By evaluating RSF models using information at these time intervals, we determine predictive performance and feature importance over time. This supports the study's objectives by examining not only who separates and when separation occurs, but what risk signals develop and change over time prior to separation.

3.5.5 Performance Evaluation

Concordance

Survival model performance was evaluated using concordance (via C-Index), which measures whether predicted risk scores align with the expected order of observed events (Schmid et al., 2016). Concordance represents the probability that, for a randomly selected pair of individuals, the individual with the higher predicted risk experiences the event earlier. Harrell's C-Index is commonly used and provides a threshold-free measure of discrimination, commonly used to compare survival models with censored data (Schmid et al., 2016).

However, Graf et al. (1999) and Uno et al. (2011) recognized that Harrell's C-Index can be biased by right-censored data since observations at the end of the window under-represent longer-surviving individuals. Uno et al. (2011) proposed an inverse probability weighting method (called Uno's C-Index) that accounts for this censoring bias. A key input to Uno's C-Index is τ, which is the observation period chosen to appropriately truncate the observa-

tion window for pair comparisons such that the inverse probability weighting remains stable in the heavily censored tail of survival distributions (Uno et al., 2011). In our study, both concordance metrics are used as the primary metric for comparing survival model predictive performance. Noting that our data is heavily right-censored, we treat Uno's C-Index as the more robust and reliable measure.

3.5.6 Feature Evaluation

Permutation Importance

Similar to Stage 1, feature importance for survival models was evaluated using PI applied to RSF. This approach assesses the predictive contribution of each feature by permuting its values in the test data and measuring the resulting decrease in Uno's C-Index, while keeping the trained model fixed (Molnar, 2019). PI was evaluated separately at separation and for each landmark (1, 3, 6, and 12 months prior to separation), allowing examination of how predictive signals emerge and change as separation approaches. PI results are interpreted as indicators of predictive performance rather than causal effects.

3.6 Summary

Our exploratory study implements a wide range of ML techniques that answer varying aspects of the problem. We compare which ML models and methods are effective given current data and conditions. We identify which features are key to predicting separation and examine predictive indicators in the lead up to separation. In this chapter, we described the models, techniques and metrics, and design choices to optimize model performance and comparison.

We commenced by establishing the source and structure of the dataset, which contained a combination of demographic, service, performance, operations and macro-economic data. We then detailed the conceptual framework for our study, based on turnover theory established by Mobley et al. (1979).

We described how our data set was inspected and cleaned, and engineered additional features. Unfortunately, we discovered that variables relating to marriage, dependents, and family categorization was inconsistent and unreliable, and could not be used to test family

dynamics as a predictor of voluntary separation. Collinearity checks are completed to remove highly-correlated and redundant features to improve model performance and stability. Data is split into training, validation and test partitions.

We use supervised classification modeling as Stage 1. We use a range of parametric models to provide benchmarks, and non-parametric tree and boosted models that represent current best in class. We utilize SMOTE to address class imbalance, and for each model we conduct threshold-tuning and cross-validation, We described the range of metrics that will be used to assess model predictive performance. We then describe the use of PI and SHAP to understand which features have the most influence on predictive performance, providing an indicator of which features matter in understanding separation decisions.

We use classical and ML survival models in Stage 2. KM and CPH are used to provide survival and hazard baselines, followed then by RSF to predict survival outcomes. We use PI to understand which features are key to predictive performance. We extend this to landmarking, in which we seek to identify predictive indicators of a separation decision at 1, 3, 6 and 12 months prior to separation. In the next chapter, we present and discuss the results of our study.

THIS PAGE INTENTIONALLY LEFT BLANK

CHAPTER 4:
Results

In this chapter we detail the results of the analysis. We evaluate the performance of models via ranking metrics, and which features are significant to predicting separation and separation timing. We interpret the results against our conceptual framework. The results in this chapter relate to predictive performance and cannot be used to infer causation.

In Stage 1 supervised classification, we find that tree and boosted models perform exceedingly well with the current data and configuration, scoring highly across multiple threshold independent and dependent metrics. The best performing model was LightGBM, achieving PRC AP of 0.99. PI and SHAP analyses found that prediction probabilities are dominated primarily by macro-economic features, followed by service-related attributes.

In Stage 2 survival analysis, we find that the RSF model also performed well under current conditions, scoring highly (Uno's C-Index of 0.97) indicating good discriminatory ability to rank individuals by predicted probability of separation. Landmarking at 1, 3, 6 and 12 months prior to separation also produced consistently high Uno's C-Index of 0.96-0.97. PI in this stage also found that macro-economic features dominated predictive influence, followed by service-related attributes.

Interpretation against predictive expectations achieved mixed results. Macro-economic features and trade changes proved to be significant as we expected. Others relating to length of service, promotion, rank, and location change proved insignificant. Expectations relating to family could not be tested due to data limitations.

4.1 Stage 1: Supervised Classification

In this section, we compare and discuss the predictive performance of classification models, the results of PI, select the best performing model and SHAP analysis.

4.1.1 Model Performance

Ten classification models were trained, validated and tested. Uniform methodology included the use of SMOTE to address imbalance, 5-fold cross validation, and threshold tuning (prioritizing F1). A Model Comparison Dashboard with ROC and PRC plots is shown at Figure 4.1. Model Performance Metrics are summarized in Table 4.1, ordered by greatest PRC AP. We chose PRC AP as it is threshold independent, and makes model comparisons more robust, particularly where class imbalance may result in ROC AUC to be overly optimistic.

It can be seen from Table 4.1 that the model with the best predictive performance was LightGBM based on the highest PRC AP of 0.99, and also performed strongly across ROC AUC and F1. A practical consideration is that LightGBM performed only marginally better than XGBoost but took longer and was more computationally expensive; XGBoost could be considered as more optimal under current tuning configuration.

Figure 4.2 shows the detailed performance dashboard for LightGBM. For each model, ROC and PRC plots were generated, as well as a confusion matrix. A test probability distribution histogram shows the distribution of predicted separation probabilities, stratified by true class (red indicates true negatives, green indicates true positives). This plot communicates the calibration of probabilities, if class predictions are separated, and the implications of threshold selection. This model has predicted near zero probability for most Actual 0 (retained) personnel, and probabilities near one for most Actual 1 (voluntary separation). The dashed line indicates the optimum decision threshold (evaluated by F1). There is little overlap between the two distributions, indicating strong class separation and discrimination performance. Appendix E contains the results dashboards for each of the other nine models.

Looking more broadly at the results, we observe three tiers of performance. In the top tier, the tree and boosted models performed nearly identically with marginal differences at the third decimal place, with excellent predictive performance exceeding the other model types. This is not unexpected given that these models are currently considered best in class for tabular classification.

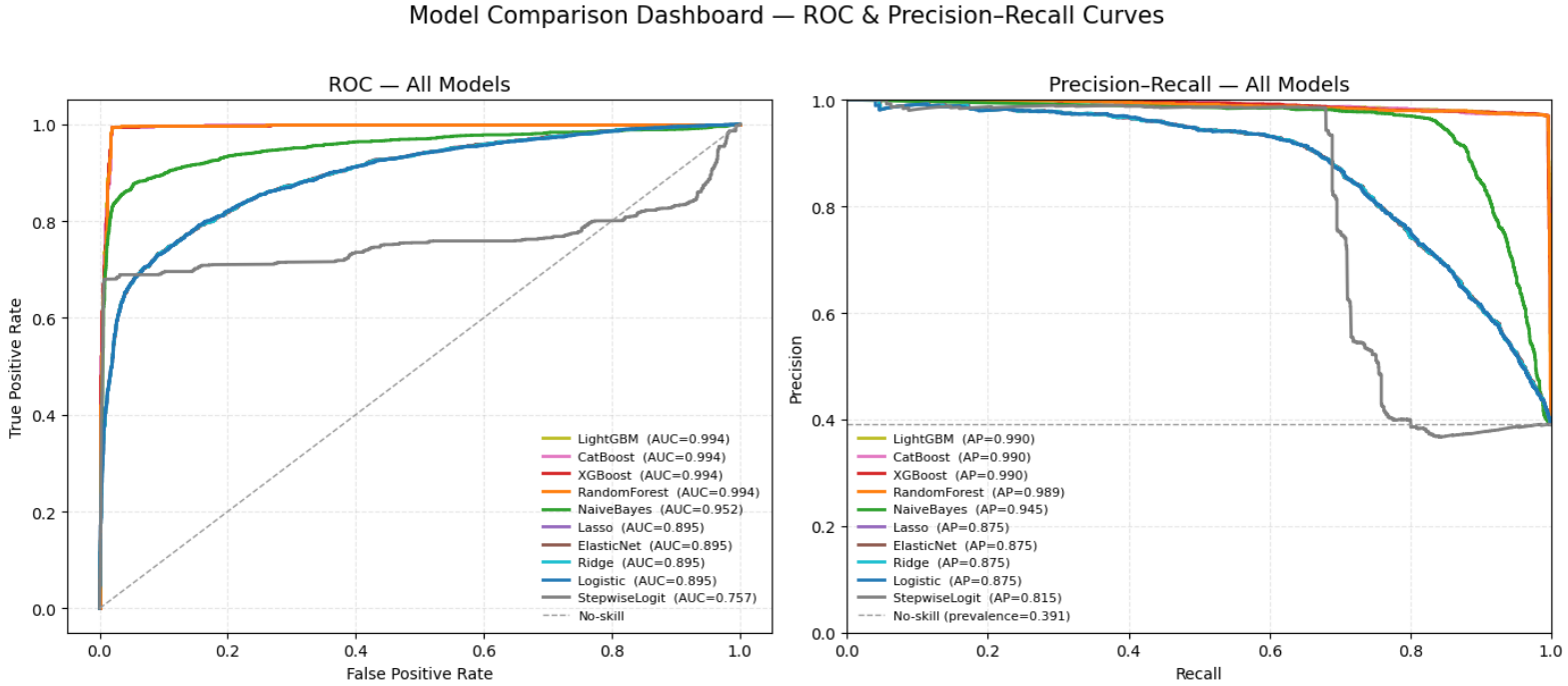

Figure 4.1. Model results dashboard, showing receiver operating characteristic and precision–recall curves.

Table 4.1. Classification model performance comparison, ordered by average precision.

Model	Threshold	Val F1	ROC AUC	PRC AP	F1	Precision	Recall	Fit (s)	Tune (s)	Eval (s)	Total (s)
LightGBM	0.398	0.979	0.994	0.990	0.983	0.970	0.996	6.8	0.2	4.6	11.6
CatBoost	0.755	0.979	0.994	0.990	0.982	0.971	0.994	29.8	0.2	1.8	31.7
XGBoost	0.595	0.979	0.994	0.990	0.983	0.971	0.996	2.3	0.1	3.0	5.4
RandomForest	0.502	0.979	0.994	0.989	0.983	0.971	0.996	8.1	0.7	2.9	11.6
NaiveBayes	0.003	0.902	0.952	0.945	0.895	0.936	0.857	0.2	0.0	0.9	1.2
LASSO	0.592	0.785	0.895	0.875	0.778	0.843	0.722	140.8	0.0	1.0	141.8
ElasticNet	0.600	0.784	0.895	0.875	0.777	0.848	0.717	523.2	0.0	3.5	526.8
Ridge	0.593	0.784	0.895	0.875	0.778	0.845	0.720	3.5	0.1	3.8	7.4
Logistic	0.610	0.782	0.895	0.875	0.776	0.853	0.712	4.1	0.0	1.3	5.4
StepwiseLogit	0.498	0.806	0.757	0.815	0.804	0.985	0.679	177.5	0.0	0.9	178.5

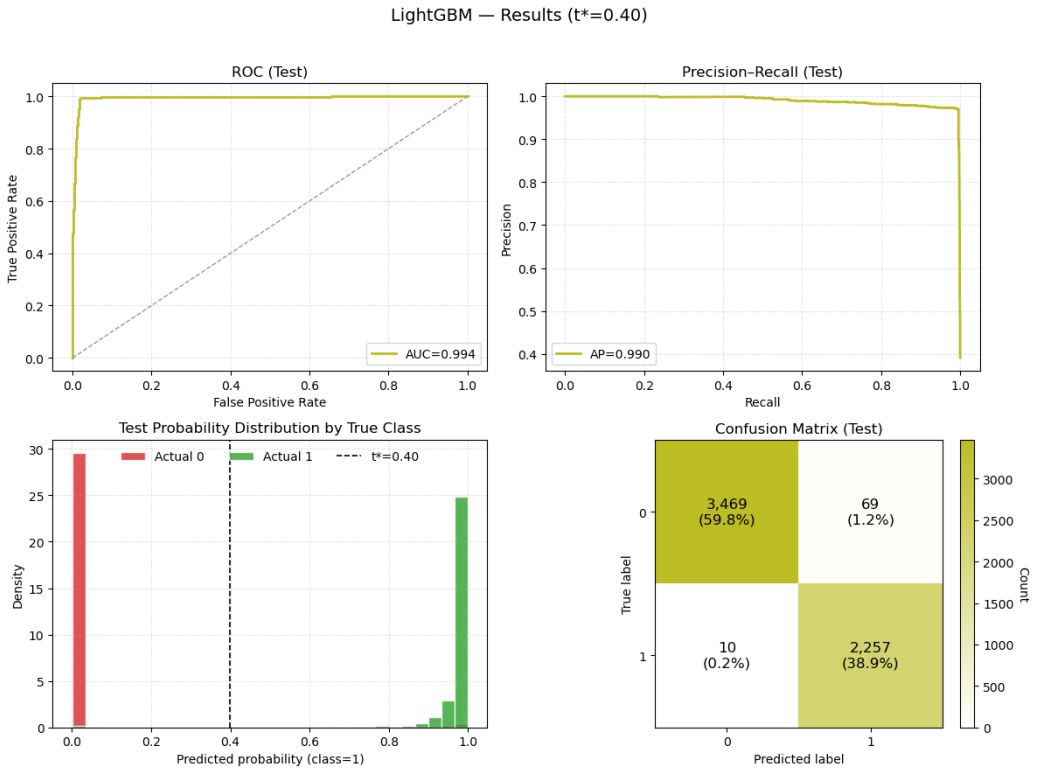

Figure 4.2. Results dashboard for performance of Light Gradient Boosting-Machine model, which scored the highest PRC AP (0.99) of all models. t* indicates optimum decision threshold (0.40), tuned prioritizing F1.

In the second tier, NB performed notably given its simplicity, outperforming logistic, stepwise, and regularized models. However, we note the decision threshold was optimized to 0.003, which is extremely low compared to others in the 0.4–0.7 range. This implies many predicted probabilities are extremely compressed to near zero, and is likely due to severe probability miscalibration, or multiplicative likelihood shrinkage. This is demonstrated in the calibration plot at Figure 4.3, comparing Naive Bayes to LightGBM. An ideal plot would follow the diagonal; the NB curve being above the diagonal indicates likely conservative underestimation of probabilities despite having good ranking outcomes, and the skew to the left confirms tight clustering of prediction probabilities near zero. Improvement of NB for employment would require further calibration. However, this highlights shortcomings at looking at metrics in isolation, and the need to conduct calibration. Noting that LightGBM follows the diagonal, we consider it to be empirically calibrated and utilize it for further analysis.

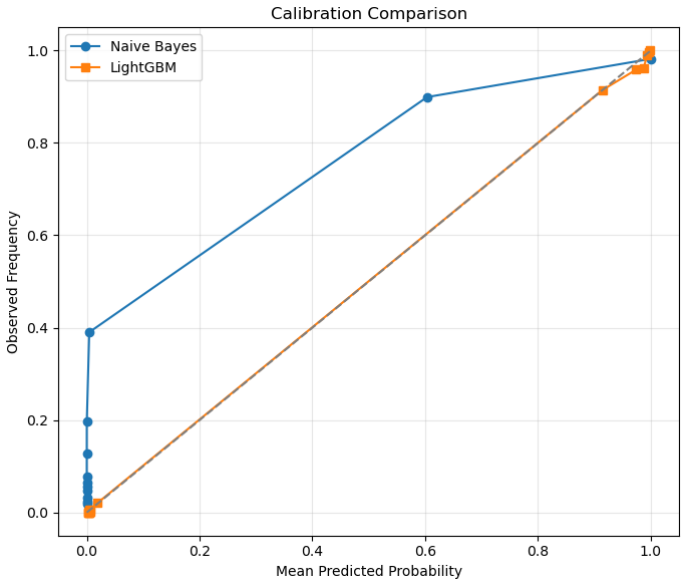

Figure 4.3. Comparison of calibration plots for NB and LightGBM models.

In the last tier, the linear regularized models did not perform as well compared to other models, with results clustered tightly together. This is likely due to the inability to capture non-linear relationships detected by the boosted non-parametric models. Within group, the regularized models (LASSO, Elastic Net, Ridge) all have similar PRC AP and F1 to the logistic model. This indicates that regularization has not significantly improved predictive performance and is likely due to the collinearity checks conducted early in the data preparation process.

Of interest, stepwise regression had the lowest PRC AP, but highest precision of all models. As precision measures TP out of all positive predictions (TP + FP), this indicates that when the model predicts separation, it is very conservative with minimal FP. However, many actual separators are missing (i.e., higher FN) resulting in lower recall. It is likely that the stepwise model selected a subset of core features that gave the strongest prediction signals to produce a narrower decision boundary. However, the model also dropped weaker predictors and misclassified more cases. We also note that the stepwise and regularized models were computationally expensive with significantly longer processing times.

From the results, we learn several lessons. Firstly, ensemble tree and gradient-boosted models had excellent predictive performance against our data and parameters as expected. This was likely due to the nature of our high-dimensional tabular data, and being able to capture

non-linear relationships. These types of models are primary candidates for employment in future studies or deployment. We also learn that the measures we undertook has significantly optimized model performance: collinearity checks, over-sampling, threshold-tuning and cross-validation are all measures that should be undertaken when possible for future deployment. However, if we address collinearity in pre-processing then the use of regularized models are redundant, and ordinary logistic regression is likely sufficient for benchmarking. Lastly, our choice of threshold-dependent and independent metrics appear to be appropriate, allowing us to check for consistency and confidently compare performance against known biases. Additionally, these metrics allowed us to diagnose and understand model behavior, and determine the need for additional measures such as calibration.

4.1.2 Permutation Importance

We ran PI across all models after training and testing. A heatmap showing normalized importance for each feature, by model, is at Figure 4.4. It depicts PI scores for individual models, sum-normalized within model. From this heatmap, the clear dominant feature is is the rolling 3-month average of the unemployment rate, and is near the maximum importance for stepwise, RF, XGBoost, LightGBM, and CatBoost models.

The second ranked feature for those models is the rolling 3-month average of job search duration. This indicates that for the stepwise, tree and gradient-boosted models that short-term macro-economic conditions were the most consistent driver of predicted voluntary separation, whilst most other variables tended towards 0. However, for the linear models, PI is spread across age and rank at separation, promotion count, and months since trade change, indicating no single dominant driver. We also observe that macro-economic features tended toward 0 here, and this is likely due to being unable to capture potential interaction and non-linear effects between macro-economic variables and voluntary separation.

An interesting pattern is also observed with CatBoost, in which job-search duration shows greater importance than unemployment rate, and may be indicative of other patterns or relationship with voluntary separation not detected by other models. The heatmap also indicates that after macro-economic and service-related features, nearly all demographic features contributed negligibly to model predictions, with the exception of age-related features.

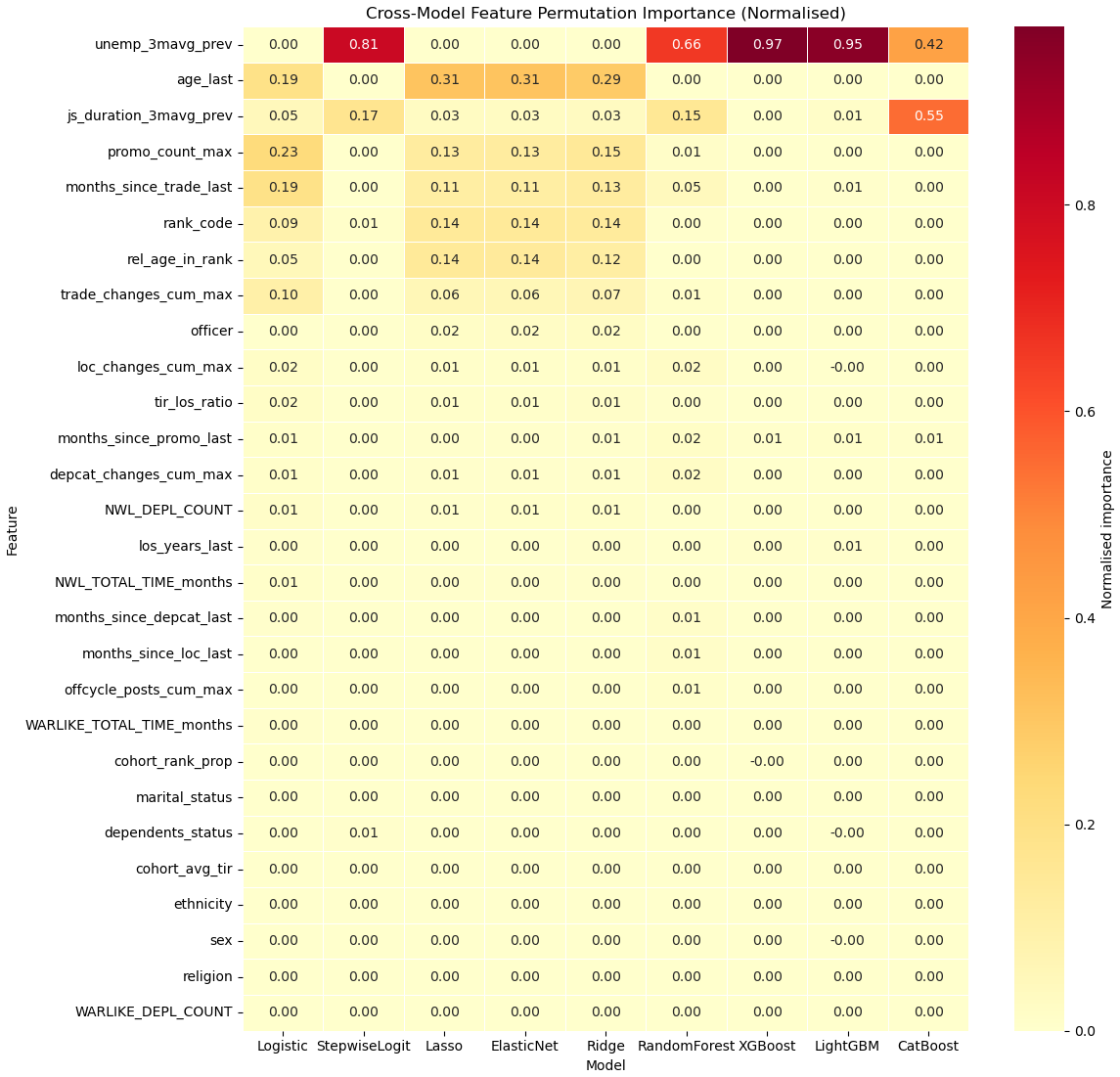

Figure 4.4. Heatmap indicating each feature's permutation importance, by model, sum-normalized within model. Darker red color indicates higher normalized importance.

A combined overall feature importance is shown at Figure 4.5, and depicts which features were consistently important across models. The plot shows the magnitude of decrease in PRC AP from running PI. This plot also uses the sign from log-odds ratio (generated by the logistic regression model) to assist with interpretation. A positive sign indicates higher values of this feature have a higher probability of voluntary separation, whilst a negative sign indicates higher values have lower probability of voluntary separation. However, we treat this cautiously as this may not hold true in non-linear relationships, and directional signs may not be uniform across all models. As PI indicates impact to predictive power only, this plot shows which features are most important, whilst the signs from logistic regression indicate whether those features are predicted to have a more protective or higher risk towards voluntary separation (assuming linear log-odds assumption). The logistic regression coefficients and log-odds are included in Appendix F for reference.

From this plot, the PI of unemployment rate is larger than any other feature. This is consistent with the heatmap and does not appear to be model-specific noise. Age is a secondary driver, and likely represents maturity effects. Noting the negative sign, this indicates that older members have lower predicted probability risk. A middle cluster of features (relative age in rank, promotions and trade changes) indicate career progression predictors are also important towards prediction. This may hint at the impact of internal opportunities.

One unexpected observation is that the directional signs for unemployment rate and job-search duration are positive, interpreted as higher values of these features increase the log-odds of voluntary separation. This is counter-intuitive to traditional economic labor market logic, since we expect lower unemployment rates and lower job-search duration to indicate a tight labor market, increasing external opportunities and the prediction probability of separation. This is discussed in greater detail later.

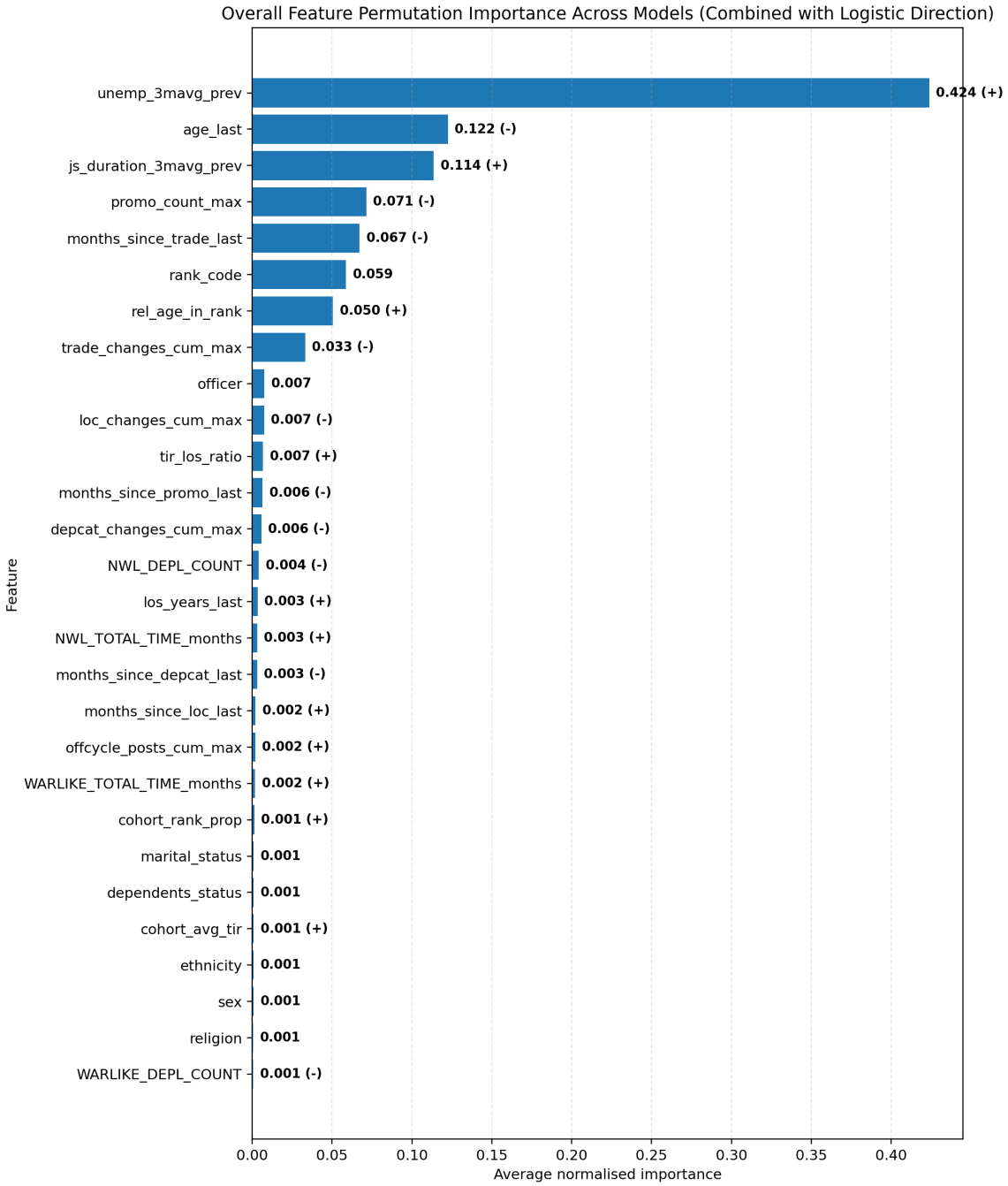

Figure 4.5. Overall feature influence across models, with all features, utilizing average normalized permutation importance. Logistic odds-ratio direction is included as a gross indicator of directional effect, and should not be interpreted as causal.

In examining both plots together, we summarize that macro-economic conditions dominate, followed closely by age and career progression factors, whilst demographic features had minimal influence on prediction. This may indicate that voluntary separation decisions are more sensitive to external labor market conditions and service factors, rather than individual demographic characteristics. A plausible reason for this is that individuals already have intention to separate, but choose to time their separation with prevailing economic conditions. Additionally, non-linear relationships with voluntary separation risk are likely to be present, and that tree/boosted models concentrate these importances sharply, whilst linear models spread these more evenly over other features.

4.1.3 SHapley Additive exPlanation

We ran SHAP on LightGBM as the best performing model. The top 20 global SHAP importance values are plotted in Figure 4.6, showing the absolute average contribution of each feature to the model output. In this plot, we observe that unemployment rate again dominates feature importance. Noting its ranking here and in PI, we can conclude that unemployment rate is not just permutation noise and directly drives model outputs. This assessment is further reinforced with job-search duration as second-ranked driver, followed by a cluster of career progression features. Of significance, these career progression variables relate to time since a significant change event (at point of separation/censoring), length of service and time in rank. It is also notable that since we are examining LightGBM only, age ranks much lower and appears to have less influence on predictive outcomes. Again, demographic features contribute little to the predictive outcomes. Interestingly, we also note no categorical features made it into the top 20.

A SHAP beeswarm plot is at Figure 4.7, and indicates direction and magnitude of influence on predicting separation risk. Each dot indicates one individual's feature values (by color) and that observation's influence on the direction and magnitude of influence on predicting separation (Lundberg, 2018). Examining the plot, unemployment rate and job-search duration again dominate predictive outcomes. These outcomes are unexpected in that higher values of unemployment rate and job-search duration appear to be drivers of increasing predicted probability of separation, which is counter-intuitive to traditional economic labor market logic. However, the direction of these effects is reinforced by the log-odds signs previously discussed during PI, ruling out issues of inverted signs. Additionally, since this

is consistent in both linear (with stricter assumptions) and non-linear models, this outcome is unlikely to be some type of modeling error or artifact, but instead reflects a structural pattern in the data.

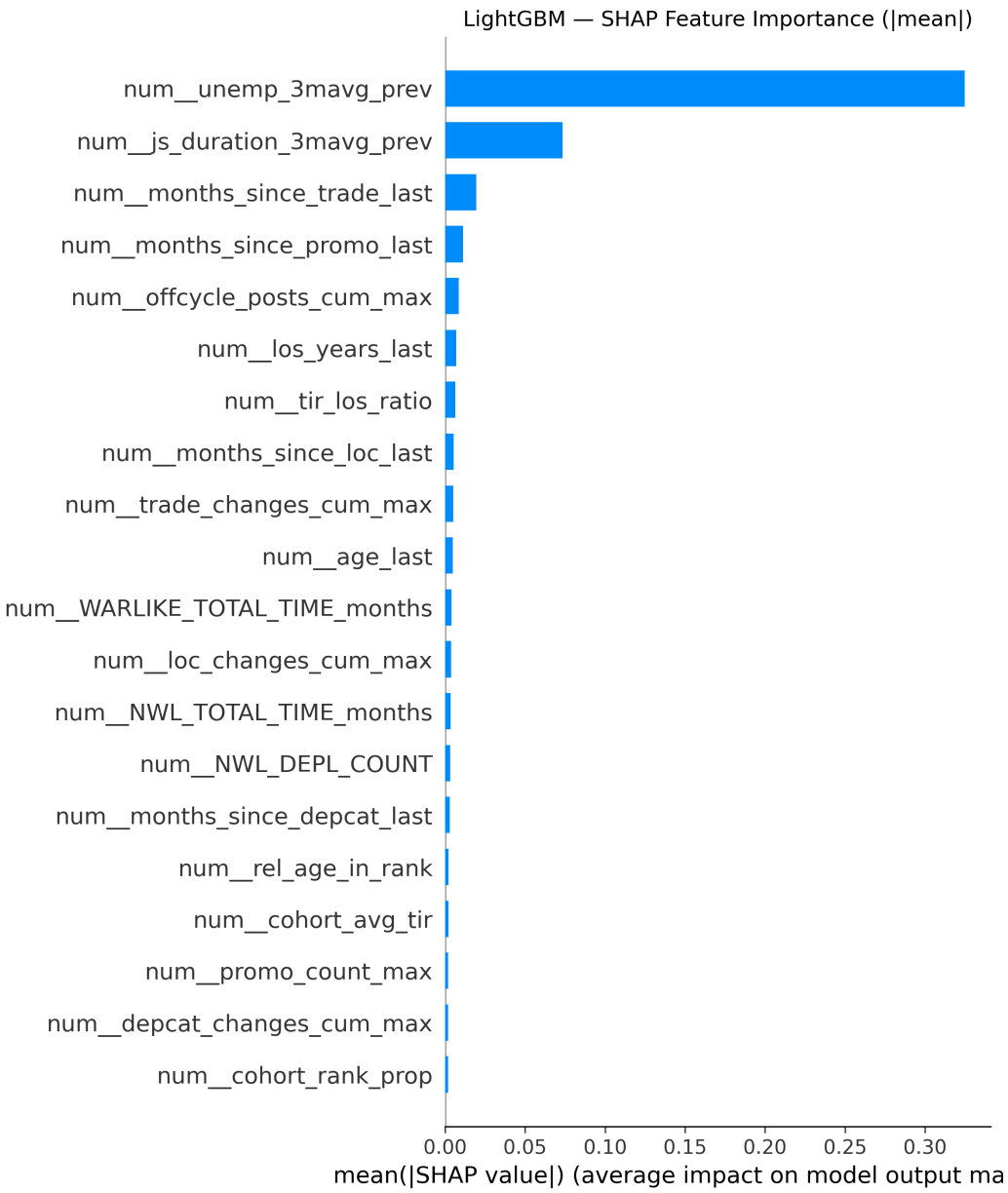

Figure 4.6. Global Shapley feature importance plot, showing top 20 scoring features.

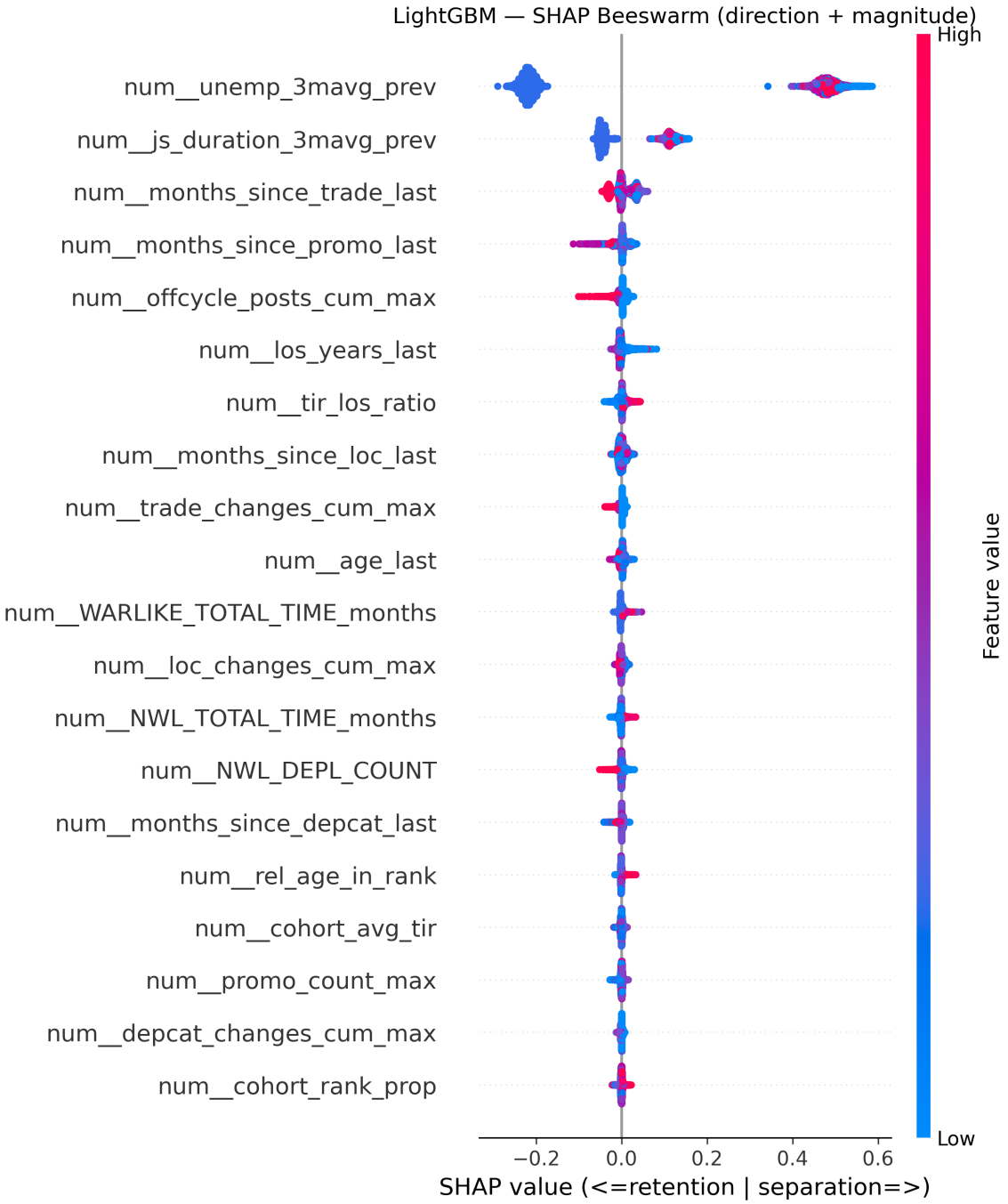

Figure 4.7. Shapley beeswarm plot, indicating direction and magnitude of feature importance. Each dot indicates one individual, with color indicating the original observed feature value, and location on the x-axis indicating separation risk magnitude and direction. Darker red color indicates a relatively high value for that feature, whilst darker blue indicates a relatively low value.

4.1.4 Stage 1 Summary

Stage 1 classification results show a clear hierarchy between different models for predictive performance against the dataset, utilizing SMOTE, five-fold cross validation, and threshold tuning (prioritizing F1). Tree and gradient-boosted models performed extremely well and near identically with PRC AP of 0.990, substantially out-performing linear models. LightGBM achieved the highest PRC AP (0.99), although XGBoost also provided comparable performance with lower computational cost.

NB performed surprisingly strongly with PRC AP of 0.945, but required an extremely low decision threshold of 0.003, reflecting a requirement to conduct calibration despite good ranking ability. Linear and regularized models clustered around PRC AP of approximately 0.875, with little gain from the regularization process. This outcome met our expectations, with tree and gradient-boosted models excelling with high-dimensional tabular data. This group of models, with current pre-processing and configuration steps, are most suitable for future deployment.

PI and SHAP analysis consistently identifies that short-term macro-economic variables were dominant predictors across models, followed by a range of career progression features. Notably, demographic features (with the exception of age) contributed very minimally to predictive performance. It was also noted that the macro-economic variables contributed significantly to stepwise, tree and boosted models, whilst linear and regularized models had predictive performance spread across a range of variables with little contribution from macro-economic variables, and suggests that tree-based models may be capturing interactions and effects from macro-economic variables that are not captured by linear and regularized models.

Lastly, a significant outcome was noted in the direction and magnitude of unemployment rate and job-search duration variables, in which it seems higher values resulted in increased predicted probability of voluntary separation, which is counter-intuitive to traditional labor economic theory. This was confirmed to be consistent across both linear and non-linear models, indicating a likely structural data pattern rather than an issue with the modeling.

4.2 Stage 2: Survival Analysis

In this section, we discuss the outcomes of traditional survival analysis models, the predictive performance of the RSF model, and observe landmark features at 0, 1, 3, 6 and 12 months prior to voluntary separation.

4.2.1 Kaplan-Meier

KM provides a descriptive baseline to Stage 2. Survival curves were modeled for the sample as a whole, and then subgroups based on sex, officer versus enlisted, migration status, age at hire, and cohort by decade. Figure 4.8 shows the survival curve for the all individuals within the sample, showing a steady decline to approximately 0.56 over after 120 months, indicating 44% cumulative voluntary separation after 10 years. The curve also shows a steeper slope between 60–100 months indicating a slightly higher attrition rate between the end of the first-term until mid- career milestones (between 5 and 9 years of service (YOS)) before flattening out beyond 100 months. No extreme early cliff is observed, reflecting our removal of first term attritors (less than 5 YOS).

Sub-group plots are at Figure 4.9. In the sex sub groups, it can be seen that females have a consistently higher survival rate than males during the observation period. In officers versus enlisted, officers are observed to have consistently better survival rates than enlisted, with a notable widening of the gap between groups and a steeper slop for enlisted from 60 months onward. In migration status, there is a slightly better retention among migrants, though the curves are essentially the same shape with an almost negligible separation between curves.

In examining ethnic groups, the Not Provided group shows much lower survival, whilst the Indigenous group has a notably higher survival rate. However, we note the sample size is much smaller (689 individuals) resulting larger confidence intervals to reflect uncertainty. The survival of Not Provided group may not be easily attributed to due to the lack of information. In examining age at hire, the less than or equal to 18 group has the lowest survival rates, whilst greater than 30 group has the highest. This likely reflects mature entry and mid-career opportunity costs. Cohort by decade shows some interesting curves, specifically in that 1980s cohort had lowest survival whilst the 1990s cohorts had the highest, and the 2010s cohort had initially very high retention, with steep decline after 60 months. These differences are potentially linked to generational differences, structural workforce

reforms and increase in operational tempo, though censoring must also be acknowledged specifically for the 2010s cohort.

Although KM survival curves provide unadjusted comparisons as a baseline, they do not account for simultaneous covariate effects or formal hazard estimates. We estimate this with CPH.

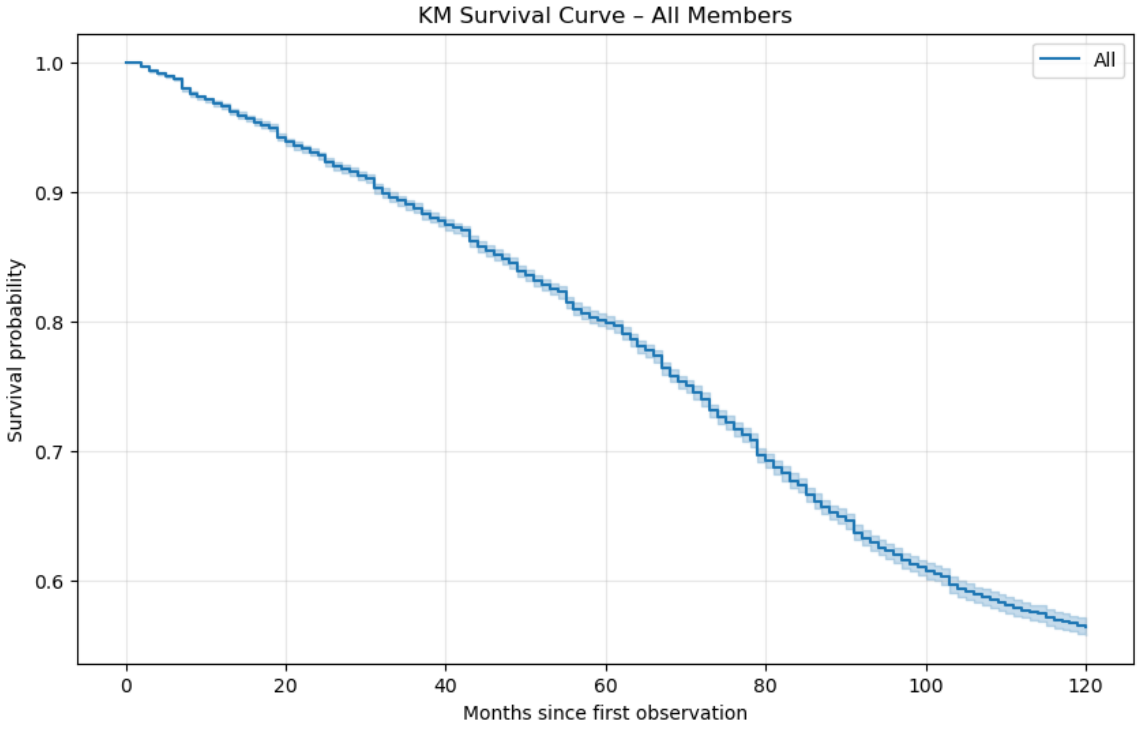

Figure 4.8. Kaplan-Meier survival curve for the full filtered data sample. Shaded region represents 95% confidence interval.

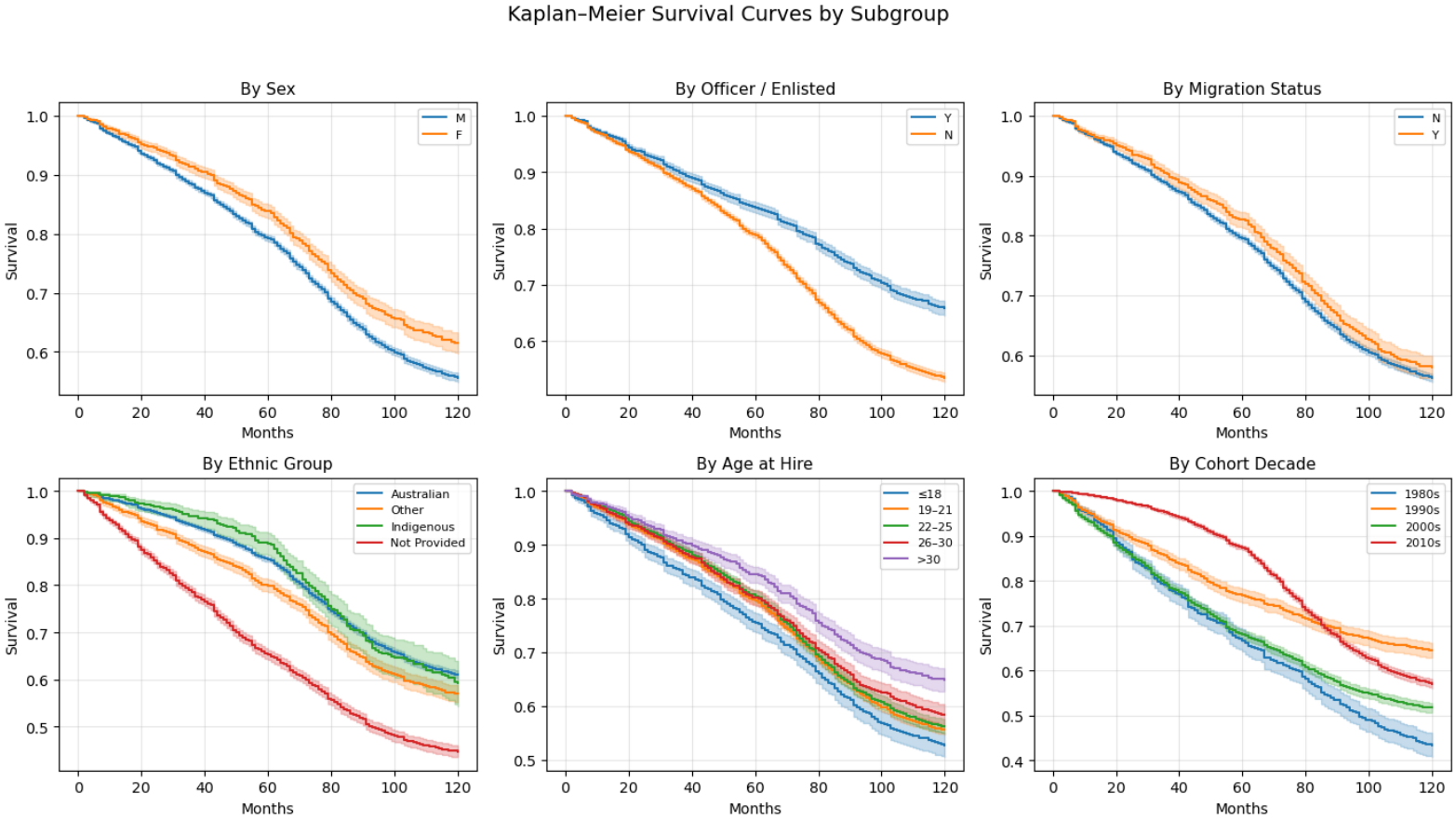

Figure 4.9. Kaplan-Meier survival curves by subgroup, across months since first observation. Shaded region represents 95% confidence interval.

4.2.2 Cox Proportional Hazard

CPH provides an interpretable hazard structure to Stage 2. An initial CPH model was fitted with age at hire, officer versus enlisted, migration, cohort bands, and sex variables and is at Table 4.2. From this table we see that all of the cohort sub-groups have extreme and statistically significant hazard ratios (0.007-0.164) relative to the omitted cohorts (1970s). This likely indicated a violation of the proportional hazard assumption, and we confirmed through scikitlearn assumption. This suggests that hazard functions are significantly different for each of the cohort bands rather than sharing a common baseline, and may indicate that each cohort needs to be treated differently. This is also reinforced by observing the significant variance between KM survival curves in Figure 4.9. This was achieved by running a second CPH model, stratified by cohort (shown at Table 4.3), thereby ensuring that covariates reflect within-cohort effects. From this we observe more reasonable and consistent hazard ratios, and no significant assumption violations when running tested in scikitlearn.

In examining Table 4.3 results, we make several observations. For age at hire, a hazard ratio of 0.987 indicates for each additional year older at hire results in 1.3% reduction in separation hazard. This implies maturity at entry translates to longer term retention, even in the mid-career population noting that first-term attritors have been removed. For sex, males have a HR of 1.117, indicating that there is a 11.7% higher hazard of separation than females. This is notable when comparing against the findings of Dodds (2018), in which RAN females were 10.7% higher hazard of separation than men. However, we note that this was a different service, and included first term attritors. Our results also show officers had 36% lower hazard of separation compared to soldiers. Dodds (2018) calculated officer versus sailor hazards differently and are not directly comparable, but broadly aligns with their findings that sailors had higher rates of separation.

We also note that migrants had 13.7% lower hazard of separation than Australian-born members. Ethnicity sub-groups are referenced against Australian ethnicity and require cautious interpretation. Indigenous members are noted to have a lower separation hazard than the Australian group, though the result is not statistically significant and likely due to the small sample size. The Not Provided sub-group had significantly higher separation, however we cannot interpret it to be a demographic effect, and likely more data incompleteness or missing administratively. The Other ethnic sub-group has a 31.9% higher separation hazard, and potentially warrants further interrogation, though outside this study's scope..

It is likely that Cohort sub-groups deviate from the baseline hazard function significantly enough that it violates proportionality assumptions and could not be estimated. This is notable in that RSF utilize cumulative hazard, and does not require restrictive proportional hazard assumptions (Ishwaran et al., 2008). RSF is therefore likely better suited to capturing time-varying and cohort-specific hazard in our predictive study.

Table 4.2. Cox proportional hazards model (unstratified) coefficients and hazard ratios.

Covariate	coef	SE	HR	95% CI	p
Age_at_Hire	-0.012	0.002	0.988	[0.984, 0.993]	<0.001***
Sex Code_M	0.108	0.029	1.114	[1.052, 1.178]	<0.001***
Cohort_Band_1980s	-1.810	0.091	0.164	[0.137, 0.197]	<0.001***
Cohort_Band_1990s	-2.408	0.089	0.090	[0.075, 0.107]	<0.001***
Cohort_Band_2000s	-2.011	0.086	0.134	[0.113, 0.159]	<0.001***
Cohort_Band_2010s	-2.292	0.085	0.101	[0.085, 0.119]	<0.001***
Cohort_Band_2020s	-5.007	0.263	0.007	[0.004, 0.011]	<0.001***
Migrated_From_Foreign_Country_Indicator_Y	-0.152	0.035	0.859	[0.805, 0.919]	<0.001***
Ethnic_Group_Collapsed_Indigenous	-0.031	0.071	0.970	[0.848, 1.108]	0.670
Ethnic_Group_Collapsed_Not_Provided	0.573	0.021	1.774	[1.703, 1.850]	<0.001***
Ethnic_Group_Collapsed_Other	0.275	0.033	1.316	[1.238, 1.398]	<0.001***
Rank_Officer_Indicator_Y	-0.449	0.025	0.638	[0.614, 0.670]	<0.001***

Notes: HR = exp(coef). 95% confidence intervals are shown for hazard ratios.
Significance: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

Table 4.3. Cox proportional hazards model stratified by cohort decade. Coefficients and hazard ratios now indicate within-cohort effects.

Covariate	coef	SE	HR	95% CI	p
Age_at_Hire	-0.013	0.002	0.987	[0.983, 0.992]	<0.001***
Sex Code_M	0.111	0.029	1.117	[1.056, 1.181]	<0.001***
Migrated_From_Foreign_Country_Indicator_Y	-0.147	0.035	0.863	[0.809, 0.918]	<0.001***
Ethnic_Group_Collapsed_Indigenous	-0.022	0.071	0.978	[0.860, 1.116]	0.770
Ethnic_Group_Collapsed_Not_Provided	0.573	0.021	1.774	[1.703, 1.850]	<0.001***
Ethnic_Group_Collapsed_Other	0.277	0.033	1.319	[1.239, 1.404]	<0.001***
Rank_Officer_Indicator_Y	-0.446	0.025	0.640	[0.608, 0.674]	<0.001***

Notes: Model stratified by Cohort_Band, allowing separate baseline hazards for each cohort decade while estimating common covariate effects. HR = exp(coef). 95% confidence intervals are shown for hazard ratios. Significance: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

4.2.3 Random Survival Forest

RSF provides predictive survival modeling to Stage 2. The first iteration collapses the data to a single observation at either separation or censoring. RSF was fitted on the training set using 500 trees and a minimum node size of 10. We evaluate model performance using Harrell's C-Index on both the training and test data, and Uno's C-Index on the test data and is shown in Table 4.4. We observe that Harrell's C-Index between training and test sets in near identical (0.971 vs 0.970) and indicates minimal overfitting. Due to the potential for bias due to right censoring, Uno's C-Index was also evaluated on the partitioned test data truncated at $\tau = 119$ months, scoring 0.972, and is consistent with Harrell's C-Index.

Table 4.4. Random Survival Forest discrimination on train and test sets, evaluated using Harrell's C-Index and Uno's C-Index.

Metric	Train	Test
Harrell's C-index	0.971	0.970
Uno's C-index	–	0.972

Notes: Uno's C-index was evaluated on the test set truncated at $\tau = 119$ months.

Figure 4.10 compares KM survival curves with mean predicted survival curves from RSF. Across the full sample, it can be seen from visual inspection that the RSF curve closely tracks with the KM curve. This indicates strong calibration, preserving the empirical survival pattern at an aggregate level whilst deriving the individual survival probabilities.

Comparison of the sex sub-groups and officer versus enlisted sub-groups also show consistency between the RSF and KM baseline. This indicates strong calibration at the sub-group level without any significant deviation. Comparison of cohort survival curves are somewhat aligned except for the two extreme ends of the groups (1970s and 2020s). This is likely due to the very small number of individuals within both cohorts, whilst the 2020s cohort would also have nearly all individuals being censored. Overall, we assess that the RSF model is well calibrated against the baseline hazard and provides no visual evidence of overfitting. The RSF model is also stable at the sub-group level.

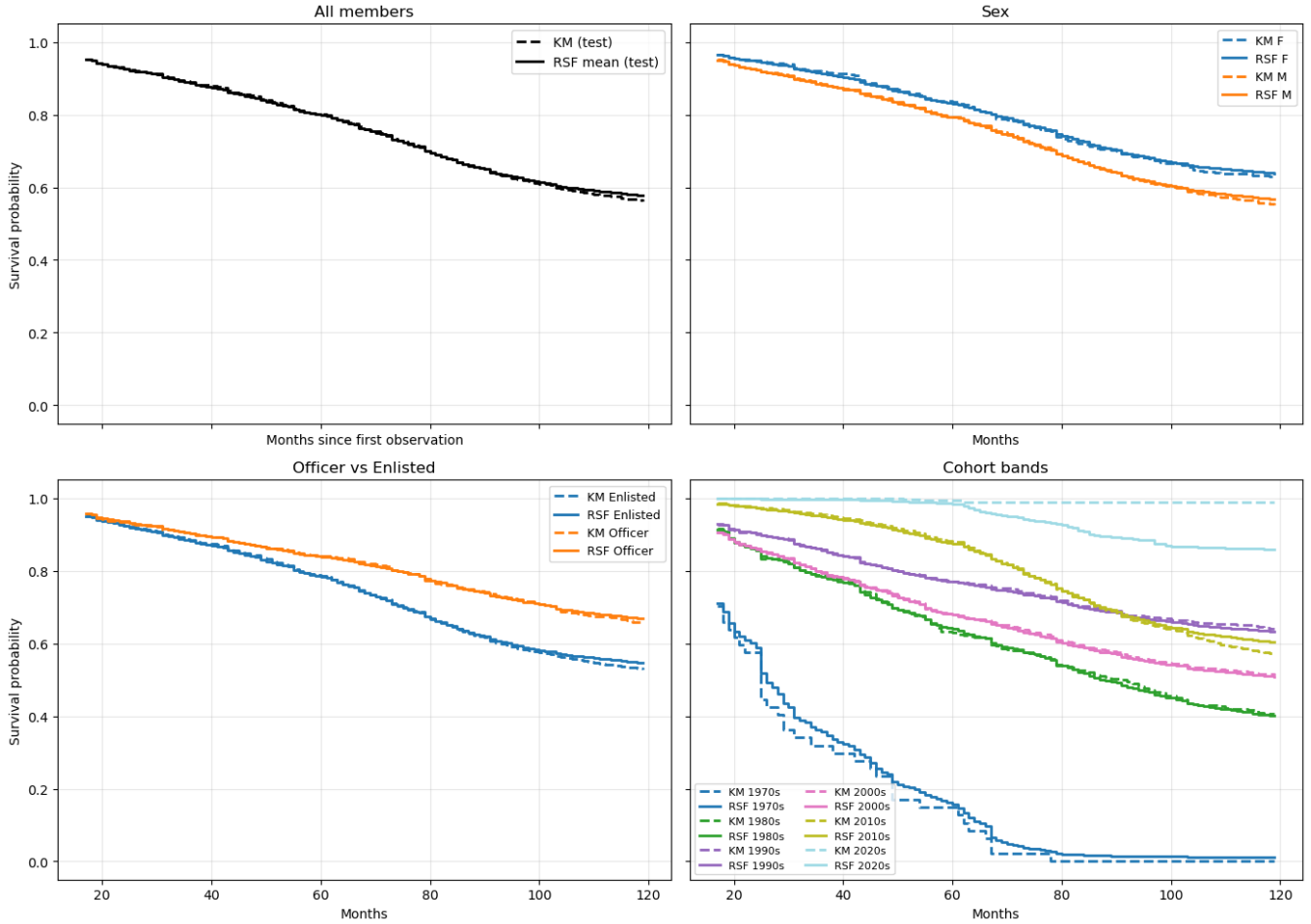

Figure 4.10. Random Survival Forest survival curves plotted and compared against Kaplan-Meier survival curves. We observe whether curves are visually consistent and for indicators of calibration and overfitting.

4.2.4 Permutation Importance

PI was conducted on RSF using the test data set, scored by change in Uno's C-Index. It can be seen in Figure 4.11 that macro-economic variables dominate survival prediction with unemployment rate making up 60.8% of the model's normalized total permutation importance, whilst job-search duration is ranked second at 22.3%. This is significant, since macro-economic variables make up 83% of total importance when summed. This likely indicates predictive performance that is more dominated by external rather than internal or individual drivers. A second tier of career- and service-related variables occupy the remainder of importance, showing that career mobility (around trade and promotion) is associated with separation risk. The remainder of features are near 0, with little contribution to discrimination.

Of note, cohort bands within this PI are near zero, despite our expectations of higher importance after CPH required stratification. This does not imply that cohort bands are not important, but that their importance is likely captured by other features during RSF. In this case, macro-economic or career dynamic features likely provide a proxy for structural differences that were previously associated with cohort bands.

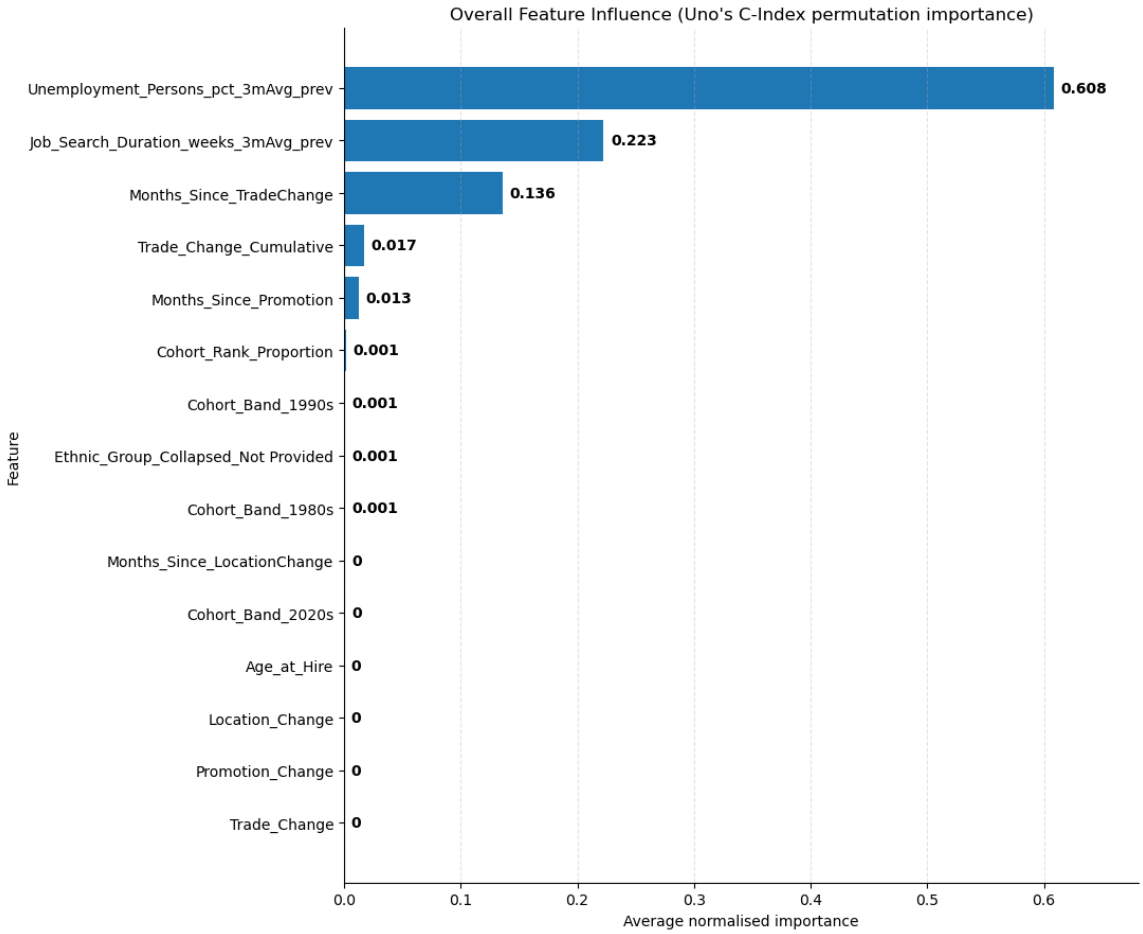

Figure 4.11. Overall feature permutation importance for random survival forest using Uno's C-Index. Scores are sum-normalized to 1.

4.2.5 Accumulated Local Effects

To examine the relationship between key predictors and voluntary separation risk, we generated ALE plots for select features at Figure 4.12. ALE estimates the marginal effect of a feature on the model's predicted risk score (Molnar, 2019). The vertical axis indicates change to predicted risk score (relative to the average prediction), with positive values indicating increasing risk, and negative values indicated reduced risk. The horizontal axis indicates increasing feature value.

From the plot for age at hire, we observe that young entrants (17-18 years old) have very high separation risk. This risk declines for entrants between 19-23 years old, but then increasing from age 24 onward. As we noted in CPH this likely reflects maturity effects. However, we see the distribution of the risk over different ages as opposed to a single hazard ratio.

In months since promotion, we observe high separation risk shortly after promotion, followed by a gradual decline in risk. This may be due to an adjustment period towards increased responsibility, or promotion aligning with service obligation milestones. In both cases there is likely a reassessment of career trajectory after promotion attainment. This separation risk appears to lessen after approximately 40 months.

In months since location change, we observe a similar profile to promotion, with an initial high separation risk that declines with increasing time in location. This may be indicative of personal or family disturbance and an adjustment period. This is also representative of the Time-Shock-Adaptability structure described by Tuppin (2025). Similarly, this risk appears to lessen after approximately 40 months. This is also notable in that Army posting orders are generally of 2-3 years tenure; some individuals and their families may continually by cycling through shock and adaptation with each move, with no prolonged settling period. This warrants further exploration, though current features around marriage, dependents and family categorization are unusable.

For cohort rank proportion, there appears to be lesser risk when an individual is promoted ahead of their cohort. However, separation risk increases at higher proportions of a cohort attaining that rank, indicating stagnation within their career progression or promotion bottlenecks. This likely results in reassessment of career prospects and comparison of own performance against peers.

We note some interesting patterns with unemployment rates and job search duration ALE plots. In both plots we observe higher separation risk to the left of the plots, when unemployment rates and job search duration are low. This aligns with expected behavior during tight labor markets. We observe that separation risk drops steeply at approximately 4% unemployment and around 12 weeks job search duration. However separation risk then rises sharply with higher unemployment rate and job-search duration to the right of the plots, peaking at 5.5% and 17 weeks respectively. This portion aligns with Stage 1 observations that are counter-intuitive to labor economic logic. As this separation risk follows a non-linear relationship, this suggests that these features are acting as proxies; there may be workforce structure or real-world events that coincide with changes in the economic environment, and worth further investigation.

Overall, the ALE plots have been insightful. Career mobility plots have shown time decay effects following an event, and demographic features have weaker and smaller magnitude effects. These are consistent with Stage 1 analysis. However, an interesting pattern has been observed with macro-economic variables that warrant investigation and discussion.

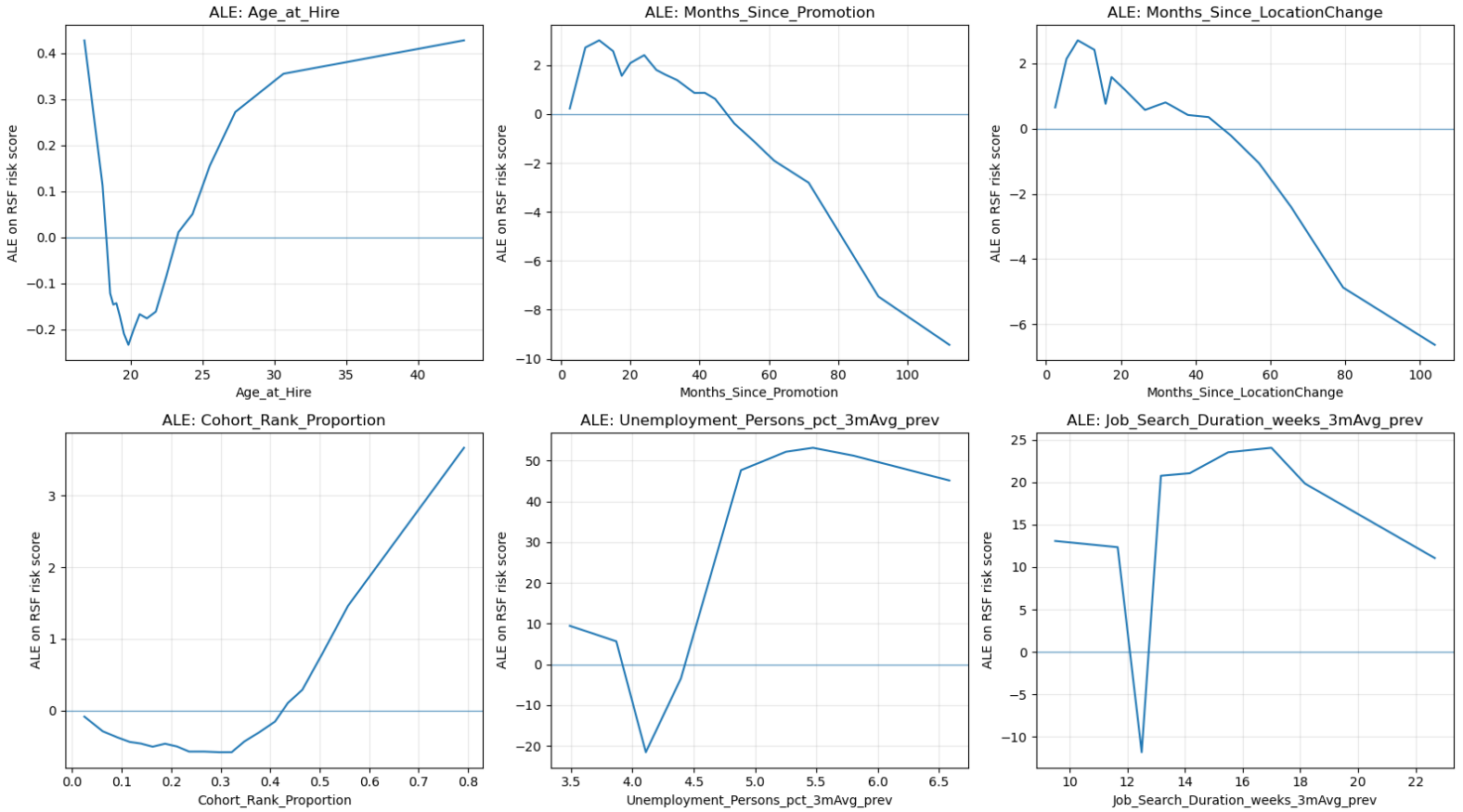

Figure 4.12. Random survival forest accumulated local effects plots for various sub-groups.

4.2.6 Random Survival Forest Across Landmarks

In order to determine early indicators of voluntary separation, we use landmarking. This is where RSF is fitted using the observations of an individual at 1, 3, 6 and 12 months prior to separation. As the number of months increases, we are predicting further into the future with less observations. We expect that predictive performance decreases with less immediate information and short term signals. We also expect that different features may be more important at different landmarks prior to separation.

Model Performance Across Landmarks

Table 4.5 shows C-Index across landmarks. Using the point of separation as the anchor, it can be seen that Uno's C-Index declines initially over the 1 and 3 month marks, followed by an increase at the 6 and 12 month marks. We have previously observed that our RSF model is heavily driven by macro-economic features and career/service dynamics. The separation risk from these features are slower to change and are more persistent over time; therefore it is reasonable to believe that these are relatively stable across our landmark range. From this we assess that voluntary separation is not dominated by last-minute or acute decision making, and it is possible to predict this up to at least 12 months. It is possible that individuals develop intention to separate very early; a person in their mid-career likely understands their career prospects, organizational direction, internal opportunities, and ranking relative to peers. As there is no timing imperative pushing a decision to leave, it is possible that individuals await favorable economic conditions and external opportunities to separate.

Table 4.5. Random survival forest performance metrics across landmarks, which is the number of months prior to separation.

Landmark (months)	Harrell's C (Train)	Harrell's C (Test)	Uno's C (Test)	τ_{safe}
0	0.971	0.971	0.972	119
1	0.969	0.969	0.970	118
3	0.967	0.965	0.967	116
6	0.970	0.968	0.969	113
12	0.971	0.970	0.971	107

C-index was computed using inverse probability of censoring weights estimated from the training data and truncated at τ_{safe} months.

Permutation Importance across Landmarks

PI was also conducted across landmarks and feature compared as time increased. Figure 4.13 shows a heatmap with PI normalized within landmark. Broadly, we observe some minor variance between landmarks. Macro-economic features continue to dominate across landmarks, although relative contribution varies. Simultaneously, we note that Months Since Trade Change is ranked highly and is more important than macro-economic features at 3 and 12 month landmarks. Unfortunately from just this we can not observe direction, and whether effects are related to high or low values of jobs since trade change. Further SHAP or ALE would need to be iterated and compared to observe changes at each landmark; this is not done here due to computational expense. However, if a recent trade change (low value) increases separation risk this may be due to an adjustment period, where dissatisfaction emerges after exposure to the new trade. If a long time since trade change (high value) increases risk then this may be due to a lack of career mobility or stagnation. Lastly, we note that demographic and cohort variables contribute minimally to each landmark. This is consistent with previous observations that generational differences are related or proxied by macro-economic and career factors rather than intrinsic cohort effect.

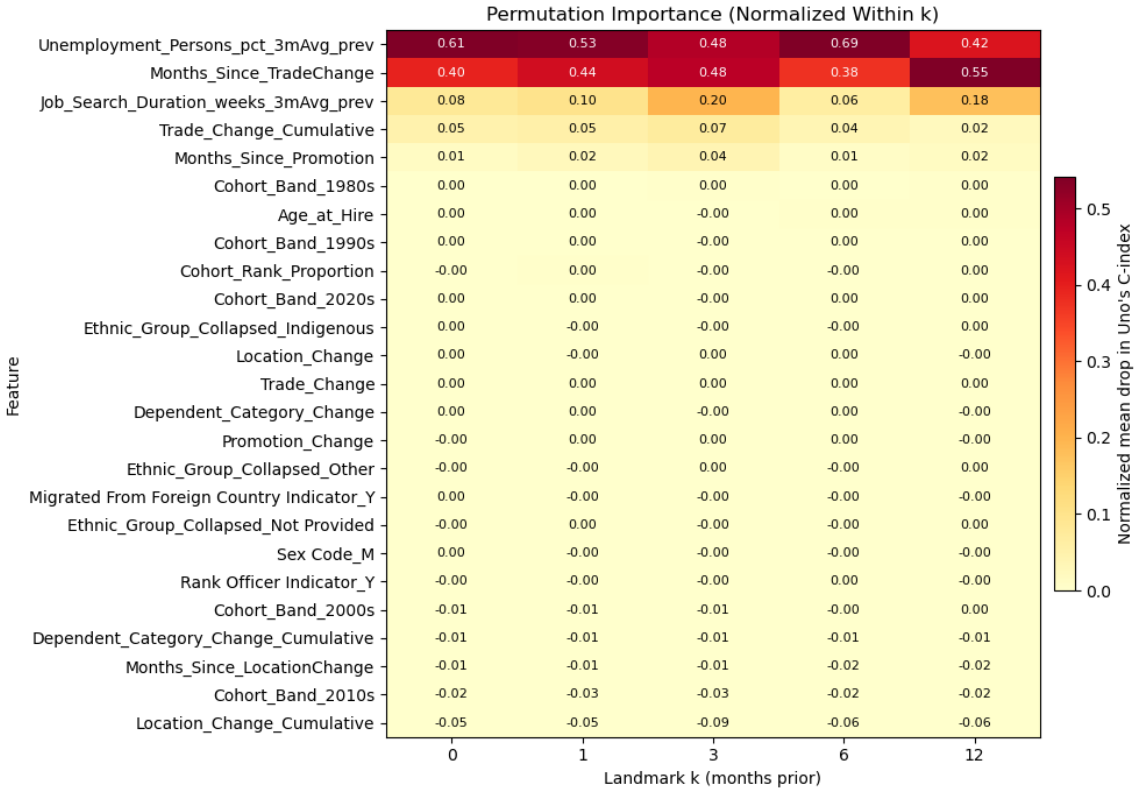

Figure 4.13. Feature permutation importance for Random Survival Forest across landmarks using Uno's C-Index. Darker red colors indicate higher drops in Uno's Concordance Index when values are permuted. Normalization is applied within each landmark period.

4.2.7 Stage 2 Summary

Stage 2 applied survival analysis to model the timing of voluntary separation. KM curves provided a baseline, indicating approximately 44% cumulative voluntary separation over 120 months. Females and officers are observed to have higher survival rates over time, compared to males and soldiers, respectively. Cohorts also showed differences in survival curves, with those in the 1990s showing higher survival rate over the observation window.

CPH models were used to estimate interpretable hazard baseline. An initial model produced extreme hazard ratios. This was likely due to cohort band hazard function being significantly different to baseline, failing proportional hazard assumptions. This may be attributed to workforce structure, economic or operational tempo differences between cohorts. A second CPH model was estimated, stratified by cohort, producing more reasonable hazard ratios that were consistent with KM curves. This highlights some of the limitations of proportional hazards for predictive modeling.

RSF was then fitted to provide a non-parametric predictive survival model without proportional hazards assumptions. The model produced extremely high and stable Harrell's and Uno's C-Index (0.97). Mean predicted survival curves closely tracked the baseline survival curves produced under KM providing evidence of good calibration and ability to model survival and hazard at the individual and sub-group level. PI using Uno's C-Index was dominated by short-term unemployment and job-search duration features, followed then by career mobility variables.

ALE plots provided significant insights and further reinforced understanding of these variables. This was particularly interesting when observing direction changes to predicted probability of voluntary separation over time. Notably, macro-economic features appeared consistent with Stage 1 observations of being counter-intuitive to labor market norms, whilst promotion changes and location changes showed decaying risk profiles.

Lastly, RSF at 0, 1, 3, 6 and 12 month landmarks were fitted, with predictive performance (evaluated using Uno's C-Index) remaining high and stable. This indicates that voluntary separation risk is not driven by short-term signals but can be observed up to 12 months prior to separation. Macro-economic features remained dominant, followed closely by career mobility features

4.3 Results Summary and Discussion

This chapter presented the outcomes of a two-stage predictive analysis for voluntary separation. In this section we summarize outcomes across modeling and analyze feature importance.

4.3.1 Supervised Classification Modeling

Stage 1 evaluated ten supervised classification models to predict who separated in the observation window. Tree-based and boosted models achieved extremely good predictive performance (PRC AP = 0.99), significantly outperforming linear and regularized models. LightGBM produced the highest overall performance in the current tuning configuration, XGBoost provided comparable performance at lower computational cost. We assess that this outcome met our expectations, with tree and gradient-boosted models excelling with high-dimensional tabular data. Combined with pre-processing and tuning steps (collinearity checks, SMOTE, threshold tuning, cross-validation, calibration), this group of models are capable and suitable for future deployment against this type of task and data.

4.3.2 Survival Analysis Modeling

Stage 2 established baselines by utilizing standard survival analysis techniques, followed by RSF for predictive modeling. Our results found that RSF modeling provided much more flexibility and performed very well in predicted survival probabilities (Uno's C-Index = 0.97). This was also followed by RSF conducted across landmarks (1, 3, 6, 12 months prior to separation), also producing Uno's C-Index = 0.96 to 0.97 across landmarks. This disproved our expectation that prediction performance at longer periods prior to separation would decline. Under current configuration, RSF is suitable for future deployment against this type of task and data.

4.3.3 Feature Importance

Throughout this study we assessed the importance of features towards predictive performance through PI and SHAP analyses. Further insights were gained by using ALE to examine separation risk over time for certain features. Across all stages, the overwhelming result was that macro-economic features (unemployment rate and job search duration) dominated measures of predictive performance. This was consistent across Stage 1 PI (Figure 4.5),

Stage 1 SHAP absolute feature importance (Figure 4.6) and beeswarm (Figure 4.7), and Stage 2 PI across landmarks (Figure 4.11, Figure 4.13). The surprising part of this result was that the direction of predicted probability did not align with common labor market logic and intuition. We observed higher unemployment rates and job-search duration result in higher predicted separation risk. This was consistent across the range of PI and SHAP analyses in different stages, and suggests that the results are robust, and not an outcome of a modeling or permutation approach.

Examining ALE plots for both unemployment rate and job-search duration provides valuable insight (Figure 4.12). In both ALE plots we observe higher local separation risk to the left of the plots when unemployment rate and job-search duration are low. This aligns with expected labor market behavior. However, separation risk then declines steeply to a local minimum, followed then by a steep rise to right of the plots, counter to expected behavior. This indicates that the relationship between these features is non-monotonic and non-linear, and contrast our expectations of monotonic behavior. This observation indicates that unemployment rate or job-search duration may not directly reflect individual separation decisions; instead these features act as proxies for broader structural or policy shifts that coincide with economic conditions over time. Some possibly-linked events may be the COVID19 pandemic, operational tempo and conflicts in the Middle-East and Europe, or ADF force and capability restructures. Although macro-economic features contribute significantly to predictive performance, their importance likely reflects confounding temporal effects rather than causal labor market effects.

After macro-economic features, various career progression and change features were most prominent in predictive performance. Across the study, frequent highly-ranked features were: months since a trade change, cumulative trade changes, months since promotion, and relative age in rank. However, none of these were consistent in rank or magnitude across modeling approaches. This suggests that importance of these features are more model-specific. These measures of internal work opportunities are still significant to predictive performance, and should be retained for future deployment, albeit tested across multiple models. Recalling our conceptual framework, only trade change seemed to meet our expectations in terms of significance and direction. All others proved to be insignificant and may not have been in the expected direction.

Lastly, demographic features appeared to have little influence in predictive performance once macro-economic and career progression features are included. In this study, demographic features are more useful for comparing survival and hazard rates, or regressing to understand marginal effects. Usage in future predictive studies should be balanced on cardinality, dimensionality, and practicality of their inclusion.

4.4 Summary

In both stages we observed that prediction probabilities were dominated primarily by macro-economic features. This was followed by service-related attributes, whilst demographic variables had negligible importance. However, examination of ALE plots indicated non-linear and non-monotonic relationships between macro-economic features and risk of separation, and indicates that these features proxy broader structural effects or timing that require further investigation.

In comparison of results against our conceptual framework, only macro-economic features and trade changes met our expectations. The framework itself served as a useful theoretical foundation for building features from Australian Army data. However, the results challenge some of the determinants of turnover, and may require further analysis to confirm if this was specific to our study design or data. Lastly, our results indicate predictive performance and feature importance on predicting separation risk only, and do not support causal inferences.

THIS PAGE INTENTIONALLY LEFT BLANK

CHAPTER 5:
Conclusion and Recommendations

5.1 Conclusion

In this section we conclude our study by summarizing whether our exploration of methods and results answered our research questions. We also discuss limitations to our study, and potential avenues for future work.

5.1.1 Research Questions

RQ: Can machine learning be used to predict voluntary separation from the Australian Army?

Our exploratory study demonstrates that ML can be used to predict voluntary separations from the Australian Army, and could also be utilized across the ADF workforce. We demonstrate that ML models are highly capable and suitable for processing large tabular datasets with high dimensionality and a range of numerical and categorical data. The best performing supervised classification models achieved very good predictive performance, detecting non-linear and non-monotonic relationships. Similarly, predictive survival analysis in current configuration also demonstrated good predictive performance. We tested this at various landmark times prior to separation, and demonstrated that this predictive performance was retained to at 12 months prior to separation.

ML outputs can be difficult to directly interpret. However, we adopted a range of interpretability measures that analyzed importance of features to predictive performance. From this we can determine the magnitude and direction of a feature's influence on the predicted outcomes. Greater importance to prediction can be used as an indicator to determine features to be investigated more thoroughly for causative investigation. This permits more efficient and effective resource allocation towards policy research efforts.

SRQ1: What individual service history, demographic, and performance features provide the most significance in predicting voluntary separation?

In this study there were several significant findings. The first is that macro-economic variables had the most significance to predictive performance. Using PI and SHAP in both stages, we found that macro-economic variables consistently ranked highly across both classification and survival modeling, including during landmarking. It was expected that this represented the overwhelming effect of external work opportunities as a turnover determinant. However, through ALE we found that macro-economic features did not have a monotonic effect as we might expect from conventional labor economic theory. Instead there was a complex non-monotonic and non-linear relationship. This indicated that macro-economic features were proxies that represented some other broader structural changes or events that aligned with labor economic outcomes and warrants further investigation. However, this does not change the observation that higher separation risk still prevailed at times of high unemployment rate or job search duration. This indicates that individuals could still choose and time their separation even when labor economic conditions were not conventionally favorable.

The second finding was that career progression and career mobility features had the next greatest importance after macro-economic features. Time since a trade change had moderate importance, and as time increases predicted separation risk decreases. This is likely an indicator of internal opportunities, and a desire to stay in service even if there were job fit or a change in circumstances requiring a trade change. Time since promotion had the next greatest importance, although results were surprising. From our SHAP analysis, increasing time since promotion increases predicted separation risk. However the ALE shows that this effect is highest and localized to the first 24 months; after this time, predicted risk diminishes and after 40 months actually had an increasingly protective effect. Other service-related features were compared and did not meet our expectations. These other features were observed to have very little importance to predictive performance.

Notably, demographic, operational, and performance features all had negligible importance to predictive performance. These were absent from top importance rankings across the study despite expectations that there would be some significance.

SRQ2: What machine learning algorithms and parameters are most effective at predicting voluntary separation?

In our study, we found that tree ensemble and gradient-boosted ML models had the best predictive performance for this type of data and task. For classification, LightGBM achieved the best PRC AP of 0.99, although this was marginally better than the other models at the third decimal place. XGBoost, CatBoost, and RF all performed comparably, and exceeded in other metrics such as computational cost. For survival, RSF in the current configuration was more than suitable, achieving good calibration with baselines, and good discrimination with Uno's C-Index of 0.96-0.97 even across landmarks up to 12 months prior to separation. These models are highly suitable for future employment against similar tasks with similar data.

To improve ML performance we undertook a range of preparatory activities. We cleaned the data and aggregated select categories to reduce cardinality and dimensionality, improving model stability. We conducted collinearity checks and removed redundant features. We sought to improve training outcomes by using SMOTE to address class imbalance, and cross-validation. Threshold tuning was also conducted (prioritizing F1) in order to more readily compare models on best ability. These measures are commonly implemented by modelers to improve predictive outcomes and are suitable with little penalty here.

Lastly we chose metrics that were robust to class imbalance or bias issues. We prioritized PRC AP and F1 as classification metrics that balance between precision and recall. Other metrics, when evaluated concurrently also assist in diagnosing issues of calibration or under/over-estimation. Uno's C-Index was prioritized as the more robust measure of concordance in predictive survival modeling noting the data is right-censored. Lastly, interpretability was achieved through the use of PI and SHAP. These tools allowed us to understand the magnitude and direction of each feature's importance and influence on predictive performance. We observe that ALE was a powerful tool for examining and visualizing the local relationship between each feature and the risk of separation over time.

SRQ3: How can the predictive results be used to support managerial decisions and inform policy interventions?

We observe that Australian Army members remain at high risk of separating when traditional labor economic logic would expect a decrease. Therefore, policies such as retention bonuses or non-monetary benefits should persist even in economic downturns. Additionally, trade changes are observed to have a negative effect on predicted separation risk. The offer of

trade changes present an alternate intervention tool for those intending to or have decided to separate.

From ALE, we have observed that predicted separation risks are highest in the first 24 months after promotion, and change in location. Policy that respectively focuses on mentoring newly promoted members or providing support to families in the first 24 months could potentially reduce predicted separation risk.

We also consider the absence of importance for specific demographic and service sub-groups such as age, job, or rank categories. Therefore, when interventions are implemented to improve retention, these should be capability focused rather than targeting specific demographic sub-groups.

Lastly this exploratory study demonstrates that predictive ML can be used to identify patterns, groups and outcomes. Therefore, ML employment should be presented as a method available to decision and policy makers for workforce and policy analysis, both pre- and post-policy implementation. This may then inform policy decision, or direct further studies, particularly to make causal inference.

5.1.2 Limitations

The primary limitation of this study is its predictive nature, and that results could not be used for causal inference. We are able to identify a potential pattern or relationship, without the ability to understand why. For example the relationship between macro-economic features and predicted separation probability will require dedicated analytical studies to prove if an actual relationship exists. As such ML is best employed to understand patterns in large datasets, with outcomes then being used to direct allocation of limited resources or effort to causal studies.

An inherent limitation with our data is that we are unable to capture when a person actually makes the decision to separate. Our dataset contains the date that someone's separation comes into effect. However, it is possible that the decision was made well in advance, and the chosen separation date aligns with administrative reasons or convenience. An inspection of our data indicates separation dates are consistently highest in January each year; a likely reason is alignment with the Australian Army posting cycle. An individual may complete

their duties or hand over their responsibilities to an incoming replacement, maintaining goodwill with the workplace. The Australian school year and summer holidays also align with December/January and may be convenient for timing family milestones and activities. Additionally, the ADF permits a person to undertake varying forms of transition training or paid and unpaid leave prior to separation taking effect. It is common for individuals to have a 'last day of work' 6 or 12 months prior to the recorded separation date. A potential solution to this is to identify the date of submission of the voluntary separation application form (AC853 ADF application to transfer within or separate from the ADF). This is not currently collected at an organizational level or stored in MARS, but is the closest measure for an individual's separation decision.

Another limitation encountered was in data consistency. Notable issues were found within marriage, family categorization, dependent and education data noticeable at our person-month granularity. This included missing dates or categorizations, or illogical dates. As some of this data is self-reported and entered manually, this highlights consistency and data quality issues. Information assurance activities such as quality audits and checks of input fields in digital forms may address some of these issues.

Lastly, our observation of each individual was limited to a defined period. An alternative way to frame the study population is to analyze the full service history of an individual regardless of the window start date. This gives more reliability to cumulative features, noting the extensive career of some individuals.

5.2 Future Work

Several areas of future work are identified in this section. We employed a range of linear, regularized, tree ensemble and gradient-boosted models in our study and achieved high predictive performance scores. An emerging area of ML we did not test are neural networks, which are more sophisticated in design but may produce better predictive outcomes, and identify more complex patterns and insights. Neural networks may be designed to conduct classification, regression, or survival neural network architecture design options are complex and endless, and implementation can have high computational cost. Future work could be directed toward exploring neural networks, architectural design choices, and implementation on high performance computing for this type of task and data. Potential candidates are feed-

forward neural networks (such a multi-layer perceptrons), recurrent neural networks with attention, temporal convolution networks, and deep survival models. These may also be capable of taking longitudinal data without significant transformation. Models and results from our exploratory study may provide a comparison point.

Similarly, we identified novel predictive survival modeling such as RFRSF and CoxRF that were not tested in this study. These methods may provide a point of comparison to survival modeling conducted here.

In designing our study, we chose several preparatory and optimization measures such as collinearity reduction, threshold tuning, cross-validation and under-/over-sampling techniques. We chose to implement all of these 'best-practice' measures to optimize model performance. Future work could be directed towards understanding which combinations and configurations are suited toward our data and task.

A potential application for our classification and survival framework is to predict end-strength using individual predictors. This differs to classic time series analysis such as that undertaken by Darragh (2022), which relies on empirical patterns, seasonality and trends from aggregated headcount over time. Our ML framework may be applied to future workforce records to identify attrition, compared to time series analysis, and potentially guide future short term recruiting or retention targets.

Our literature review also identified pay and pecuniary measures as being linked to turnover and attrition, but it was not tested within our study and is an identified shortfall. Potential future research could build on this study by considering a member's pay over time, and comparing trades or employment sectors in the civilian market. This may also provide further insight into our results regarding macro-economic unemployment data.

Lastly, future research could be directed toward understanding more about specific communities and sub-groups. Forecasting and managing Australian Army trades are currently using workforce pyramids and Markov modeling. ML in both classification and survival modeling may complement these methods.

Our exploratory study demonstrates that ML can be applied to tabular individual administrative data to predict voluntary separation decisions. By conducting a two-stage analysis,

we have identified characteristics and timing that influence the predicted risk of separation. We have compared models, and implemented optimization and interpretation techniques. These have proven successful and insightful, and indicates suitability of our approach (in current configuration) for this type of data and task. Our study contributes to Australian Army workforce understanding, and establishes a starting point from which future work may result in policy and decisions that improve retention outcomes and achieve ADF workforce objectives.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX A:
Conceptual Model of Employee Turnover

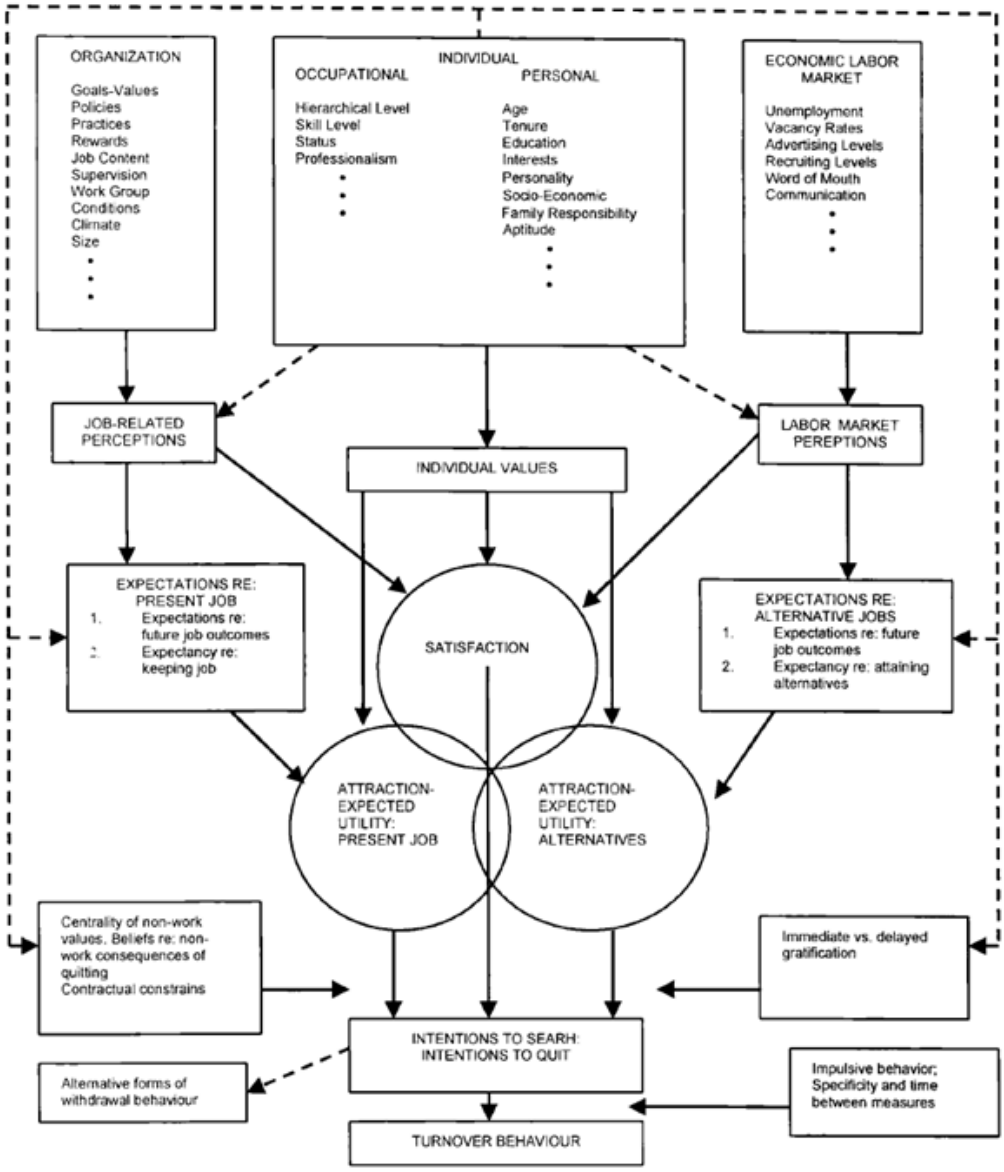

Figure A.1. The three top boxes indicate organizational individual and economic labor market factors. These are considered together by an individual to determine internal and external work opportunities, job satisfaction and alignment with individual values to make a decision to separate. Adapted from Mobley et al. (1979).

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX B:
Data Entity Relationship Diagram

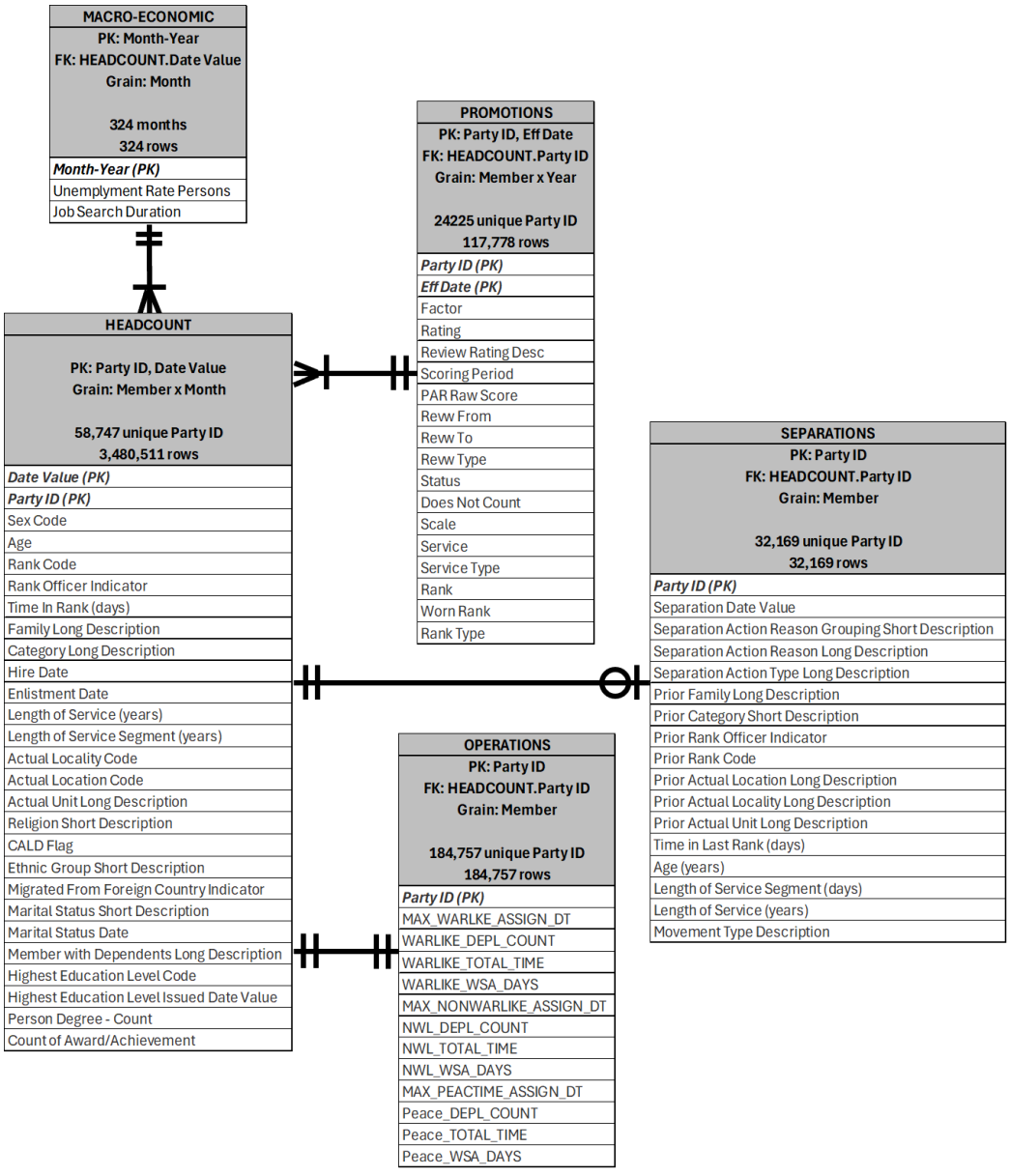

Figure B.1. Modified entity relationship diagram, indicating keys and cardinality. Headcount, Separations, Operations and Promotions data was sourced from the ADF data warehouse through the sponsor. Macro-economic data was sourced from the Australian Bureau of Statistics

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX C:
Data Sample Balance Table

Table C.1 details baseline demographic and service-related characteristics, after censoring and truncation is applied. These statistics are descriptive and are not presented for causal analysis. The null hypothesis tested here is that there is no difference in the characteristics between retained and voluntarily separated personnel. As all p-values are <0.001 we reject the null hypothesis, and note that the difference between the retained and voluntarily separated groups is significant. However given the large sample size, statistical significance is expected for many variables and is not interpreted substantively.

Table C.1. Baseline characteristics by separation status (analysis sample).

Characteristic	**Retained (N=17,687)**	**Voluntary (N=11,335)**	**p-value**
Continuous characteristics			
Age (years)	37.14 (9.52)	34.76 (9.28)	< 0.001
Length of service (years)	14.69 (9.00)	12.92 (8.77)	< 0.001
Observed in window (months)	102.51 (23.76)	56.01 (30.03)	< 0.001
Sex			
F	2,782 (15.7%)	1,362 (12.0%)	
M	14,905 (84.3%)	9,973 (88.0%)	
			< 0.001
Officer			
N	12,508 (70.7%)	9,205 (81.2%)	
Y	5,179 (29.3%)	2,130 (18.8%)	
			< 0.001
Rank			
E00	12 (0.1%)	34 (0.3%)	

Continued on next page

Characteristic	Retained (N=17,687)	Voluntary (N=11,335)	p-value
E01	19 (0.1%)	7 (0.1%)	
E02	21 (0.1%)	4 (0.0%)	
E03	2,489 (14.1%)	3,005 (26.5%)	
E04	1,086 (6.1%)	1,019 (9.0%)	
E05	4,140 (23.4%)	3,205 (28.3%)	
E06	2,200 (12.4%)	982 (8.7%)	
E08	1,784 (10.1%)	613 (5.4%)	
E09	756 (4.3%)	336 (3.0%)	
E10	1 (0.0%)	0 (0.0%)	
O00	33 (0.2%)	26 (0.2%)	
O01	0 (0.0%)	1 (0.0%)	
O02	504 (2.8%)	73 (0.6%)	
O03	1,760 (10.0%)	869 (7.7%)	
O04	1,759 (9.9%)	825 (7.3%)	
O05	809 (4.6%)	286 (2.5%)	
O06	218 (1.2%)	36 (0.3%)	
O07	68 (0.4%)	5 (0.0%)	
O08	24 (0.1%)	3 (0.0%)	
O09	4 (0.0%)	5 (0.0%)	
O10	0 (0.0%)	1 (0.0%)	
			< 0.001
Marital status			
MS Svc Rqt	0 (0.0%)	52 (0.5%)	
Married	12,587 (71.2%)	7,158 (63.1%)	
Single	5,100 (28.8%)	4,125 (36.4%)	
			< 0.001
Dependents			
Accompanied Resident Family	11,959 (67.6%)	5,569 (59.8%)	
Member	4,448 (25.2%)	3,221 (34.6%)	

Continued on next page

Characteristic	Retained (N=17,687)	Voluntary (N=11,335)	p-value
Unaccompanied Resident Family	1,275 (7.2%)	524 (5.6%)	
			< 0.001
Religion			
N	9,757 (55.2%)	5,724 (50.5%)	
Y	7,930 (44.8%)	5,611 (49.5%)	
			< 0.001
Ethnicity			
Aboriginal & Torres SI	468 (2.6%)	221 (1.9%)	
African & ME	126 (0.7%)	70 (0.6%)	
Americas	47 (0.3%)	36 (0.3%)	
Asian Pacific	1,030 (5.8%)	693 (6.1%)	
Australian	11,581 (65.5%)	5,994 (52.9%)	
European	990 (5.6%)	620 (5.5%)	
Not Provided	3,445 (19.5%)	3,701 (32.7%)	
			< 0.001
Total sample after exclusions: N = 29,022			

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX D:
Data Dictionary

Table D.1. Comprehensive data dictionary.

Variable	**Type**	**Source / Engineered**	**Description**
Keys and Indexing			
Party ID	Numerical	Source	Unique person identifier
Date Value	Datetime	Source	Observation date (person-month record)
Year	Numerical	Engineered	Calendar year extracted from Date Value
Month	Numerical	Engineered	Calendar month extracted from Date Value
Month Year	Categorical	Engineered	Month-year period label
Month Year Ordinal	Numerical	Engineered	Ordinal encoding of Month Year
Outcome and Separation Variables			
Separated Flag	Binary-indicator	Source	Indicator of separation event
Separation Date Value	Datetime	Source	Date of separation
Separation Action Reason Grouping Short Description	Categorical	Source	Grouped separation reason

Continued on next page

Variable	Type	Source / Engineered	Description
Separation Action Reason Long Description	Categorical	Source	Detailed separation reason
Separation Action Type Long Description	Categorical	Source	Separation action type
Voluntary Separation	Binary-indicator	Engineered	Indicator for voluntary separation
Involuntary Separation	Binary-indicator	Engineered	Indicator for involuntary separation
Vol Sep Reason	Categorical	Engineered	Collapsed voluntary separation reason
Core Demographics			
Sex Code	Categorical	Source	Sex / gender code
Age	Numerical	Source	Age at observation (years)
Religion Short Description	Categorical	Source	Religion (short description)
Religion Indicator	Categorical	Engineered	Collapsed religion grouping (recoded)
CALD Flag	Categorical	Source	Culturally and Linguistically Diverse flag
Ethnic Group Short Description	Categorical	Source	Ethnicity (short description)
Ethnic Group	Categorical	Engineered	Collapsed ethnicity grouping (recoded)
Migrated From Foreign Country Indicator	Categorical	Source	Migration indicator

Continued on next page

Variable	Type	Source / Engineered	Description
Marital Status Short Description	Categorical	Source	Marital status (short label)
Marital Status	Categorical	Engineered	Collapsed marital status grouping (recoded)
Marital Status Date	Datetime	Source	Date marital status recorded
Member with Dependents Long Description	Categorical	Source	Dependent category
Highest Education Level Code	Categorical	Source	Highest education level
Highest Education Level Issued Date Value	Datetime	Source	Education record issue date
Person Degree - Count	Numerical	Source	Number of degrees held
Count of Award/Achievement	Numerical	Source	Count of awards/achievements
Service and Career Structure			
Hire Date	Datetime	Source	Date of first hire
Enlistment Date	Datetime	Source	Date of enlistment (if re-entry)
Rank Code	Categorical	Source	Rank code at observation
Rank Officer Indicator	Binary-indicator	Source	Officer vs non-officer indicator
Family Long Description	Categorical	Source	Job Family (long label)
Category Long Description	Categorical	Source	Job Category (long label)
Actual Locality Code	Categorical	Source	Locality code

Continued on next page

Variable	Type	Source / Engineered	Description
Actual Location Code	Categorical	Source	Defence Establishment Location code
Actual Unit Long Description	Categorical	Source	Unit name (long description)
Movement Type Description	Categorical	Source	Separation event description
Tenure and Time in Role			
Length of Service (years)	Numerical	Source	Length of service in years
Length of Service Segment (years)	Numerical	Source	LOS segment (years) (if re-entry)
Time In Rank (years)	Numerical	Engineered	Time in rank in years
Age at Hire	Numerical	Engineered	Age at hire
PreWindow Service	Numerical	Engineered	Service accrued before observation window start
TIR to LOS Ratio	Numerical	Engineered	Time-in-rank to LOS ratio
Time in Last Rank (days)	Numerical	Engineered	Time spent in previous rank
Performance			
AO Review Rating Desc	Categorical	Source	Performance review rating (Assessing Officer)
SAO Review Rating Desc	Categorical	Source	Performance review rating (Senior Assessing Officer)
Career and Mobility Events			

Continued on next page

Variable	Type	Source / Engineered	Description
Promotion Change	Binary-indicator	Engineered	Promotion event indicator
Months Since Promotion	Numerical	Engineered	Months since last promotion
Promotion Count Cumulative	Numerical	Engineered	Cumulative promotion count
Trade Change	Binary-indicator	Engineered	Trade change indicator
Trade Change Cumulative	Numerical	Engineered	Cumulative trade changes
Months Since TradeChange	Numerical	Engineered	Months since last trade change
Location Change	Binary-indicator	Engineered	Location change indicator
Location Change Cumulative	Numerical	Engineered	Cumulative location changes
Months Since Location-Change	Numerical	Engineered	Months since last location change
OffCycle Posting	Binary-indicator	Engineered	Off-cycle posting indicator
OffCycle Posting Cumulative	Numerical	Engineered	Cumulative off-cycle postings
Dependent Category Change	Binary-indicator	Engineered	Dependent category change indicator
Months Since DependentChange	Numerical	Engineered	Months since dependent category change
Dependent Category Change Cumulative	Numerical	Engineered	Cumulative dependent category changes

Continued on next page

Variable	Type	Source / Engineered	Description
Prior Family Long Description	Categorical	Engineered	Prior period family category
Prior Category Short Description	Categorical	Engineered	Prior period service category
Prior Rank Officer Indicator	Binary-indicator	Engineered	Prior officer indicator
Prior Rank Code	Categorical	Engineered	Prior rank code
Prior Actual Location Long Description	Categorical	Engineered	Prior location (long label)
Prior Actual Locality Long Description	Categorical	Engineered	Prior locality (long label)
Prior Actual Unit Long Description	Categorical	Engineered	Prior unit (long label)
Cohort and Peer-Comparison Aggregates			
Hire Month Year	Datetime	Engineered	Hire month-year period
Cohort Month Year	Datetime	Engineered	Cohort entry month-year
Cohort Rank Proportion	Numerical	Engineered	Proportion of cohort in same rank
Rank Age Mean	Numerical	Engineered	Mean age within rank
Rank Age SD	Numerical	Engineered	SD of age within rank
Relative Age in Rank	Numerical	Engineered	Age relative to rank mean
Cohort Avg TIR	Numerical	Engineered	Cohort average time-in-rank
Cohort SD TIR	Numerical	Engineered	Cohort SD time-in-rank

Continued on next page

Variable	Type	Source / Engineered	Description
TIR vs Cohort Avg	Numerical	Engineered	Difference from cohort average TIR
TIR vs Cohort Z	Numerical	Engineered	Z-score of TIR vs cohort
Cohort SD TIR Missing	Binary-indicator	Engineered	Missingness indicator for cohort SD
Macro-Economic Variables			
Unemployment Persons pct Lag 3	Numerical	Engineered	Unemployment rate lagged 3 months
Unemployment Persons pct 3mAvg prev	Numerical	Engineered	Unemployment 3-month trailing average
Job Search Duration weeks Lag 1	Numerical	Engineered	Job search duration lagged 1 month
Job Search Duration weeks 3mAvg prev	Numerical	Engineered	Job search duration 3-month average
Deployment and Operational Exposure			
MAX WARLKE ASSIGN DT	Datetime	Source	Most recent warlike assignment date
WARLIKE DEPL COUNT	Numerical	Source	Count of warlike deployments
WARLIKE TOTAL TIME	Numerical	Source	Total time in warlike deployments
WARLIKE WSA DAYS	Numerical	Source	Days in warlike service area
MAX NONWARLIKE ASSIGN DT	Datetime	Source	Most recent non-warlike assignment date

Continued on next page

Variable	Type	Source / Engineered	Description
NWL DEPL COUNT	Numerical	Source	Count of non-warlike deployments
NWL TOTAL TIME	Numerical	Source	Total time in non-warlike deployments
NWL WSA DAYS	Numerical	Source	Days in non-warlike service area
MAX PEACTIME ASSIGN DT	Datetime	Source	Most recent peacetime assignment date
Peace DEPL COUNT	Numerical	Source	Count of peacetime deployments
Peace TOTAL TIME	Numerical	Source	Total time in peacetime deployments
Peace WSA DAYS	Numerical	Source	Days in peacetime service area

APPENDIX E:
Individual Classification Model Results

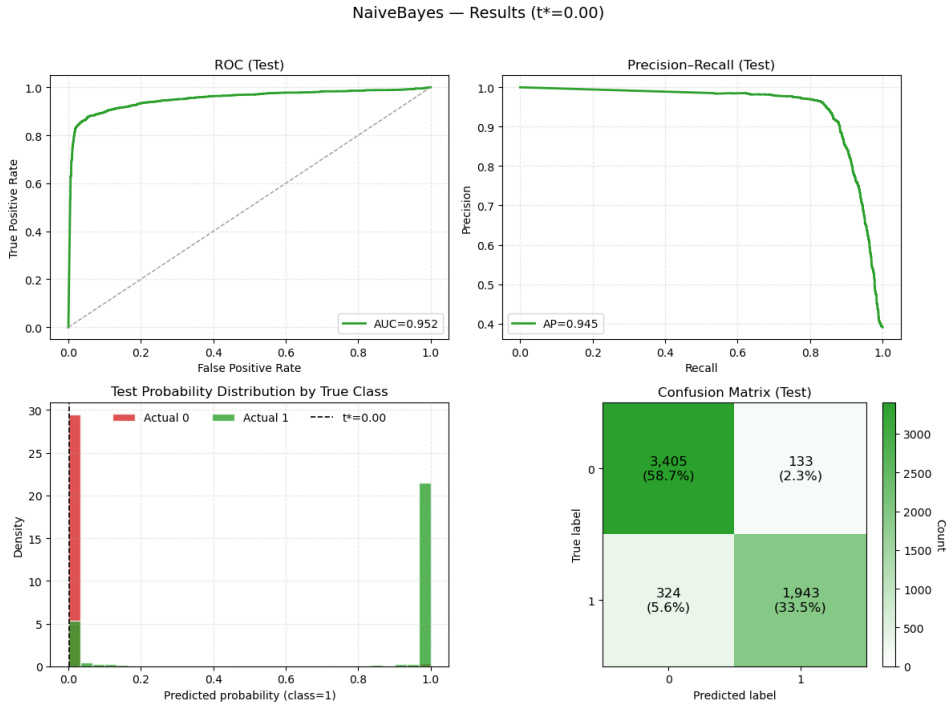

Figure E.1. Results dashboard for performance of Naive Bayes.

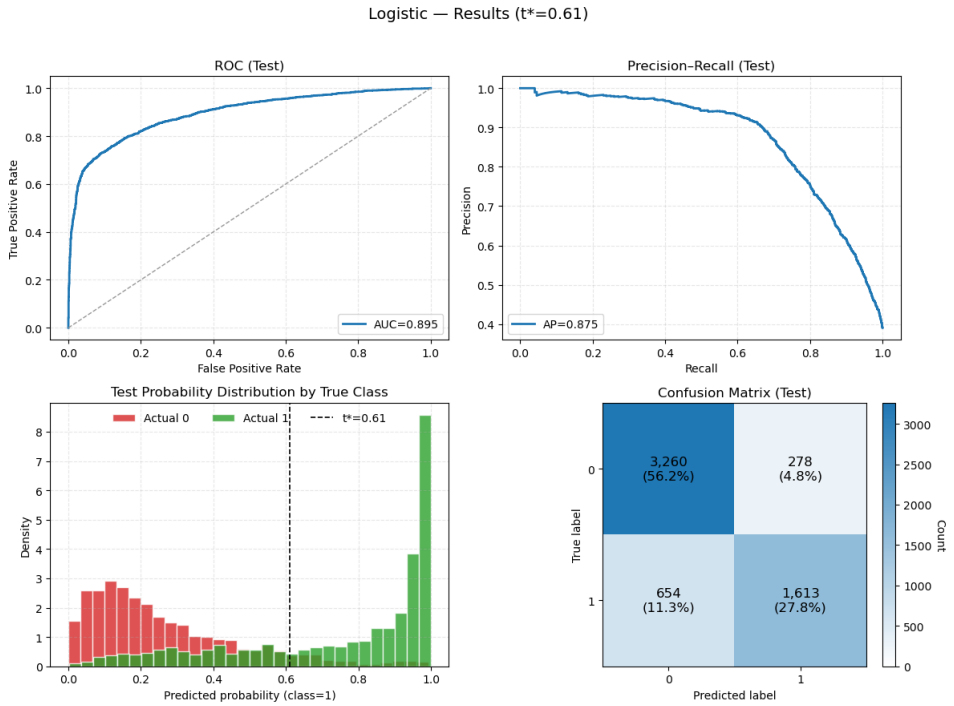

Figure E.2. Results dashboard for performance of logistic regression.

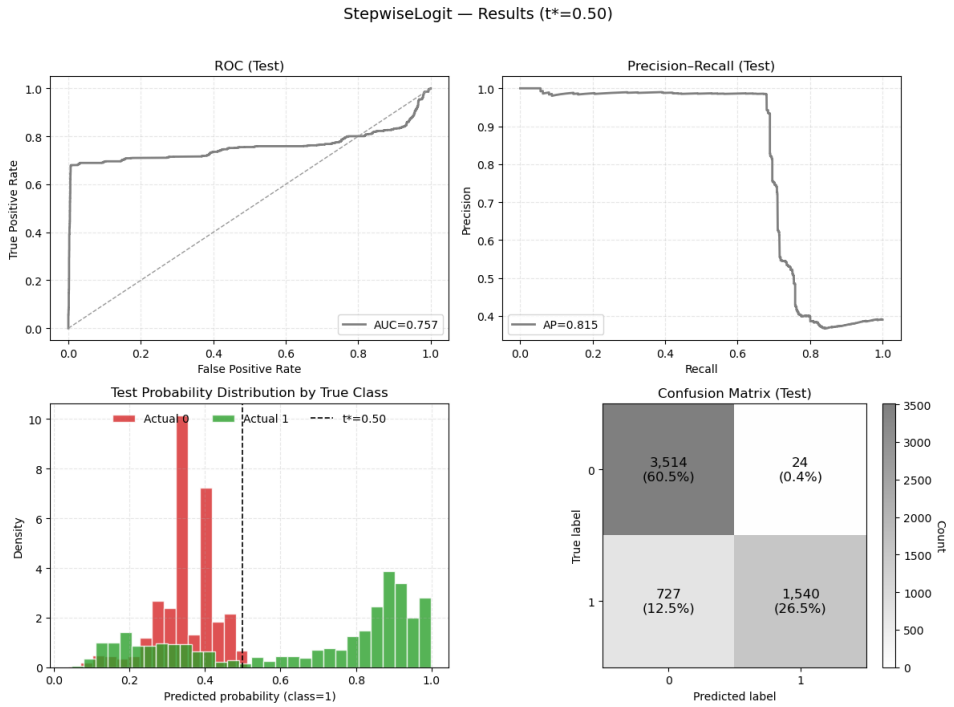

Figure E.3. Results dashboard for performance of stepwise logistic regression.

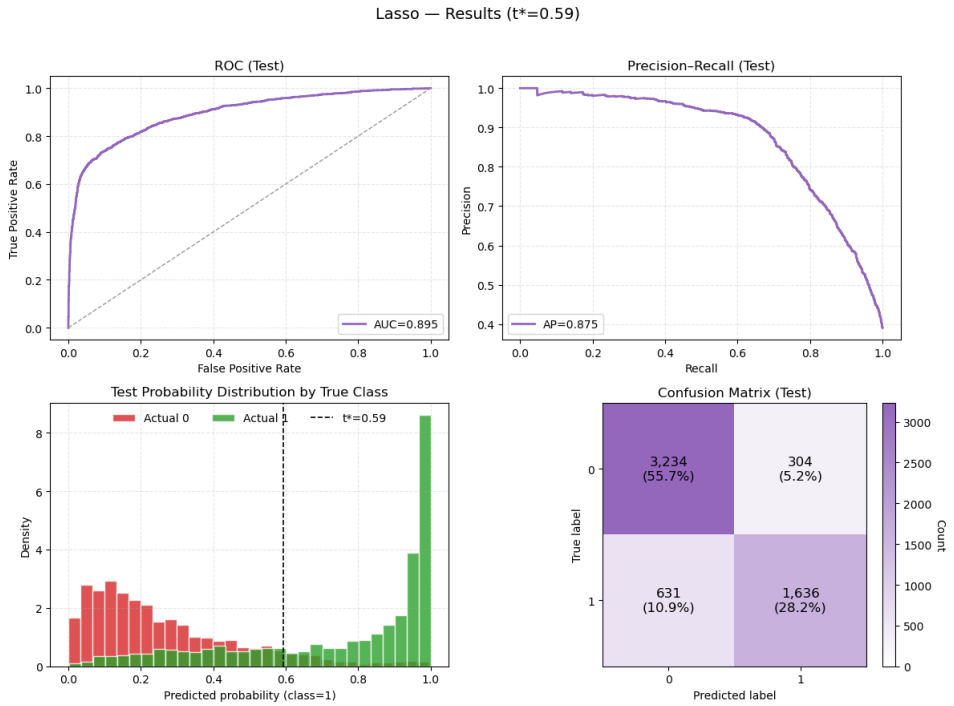

Figure E.4. Results dashboard for performance of LASSO logistic regression.

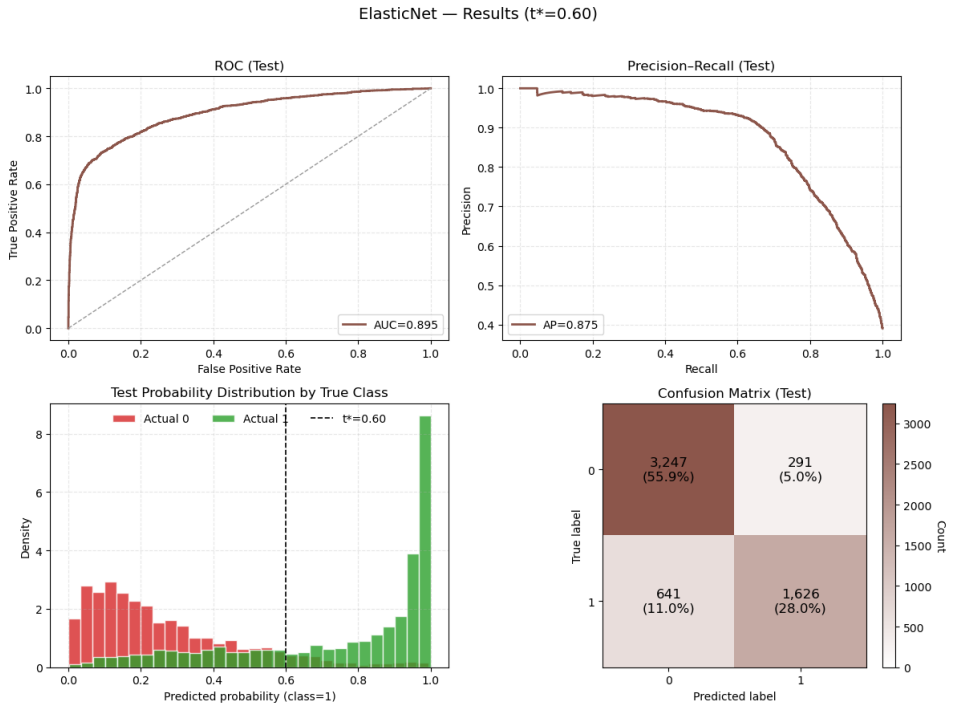

Figure E.5. Results dashboard for performance of elastic net logistic regression.

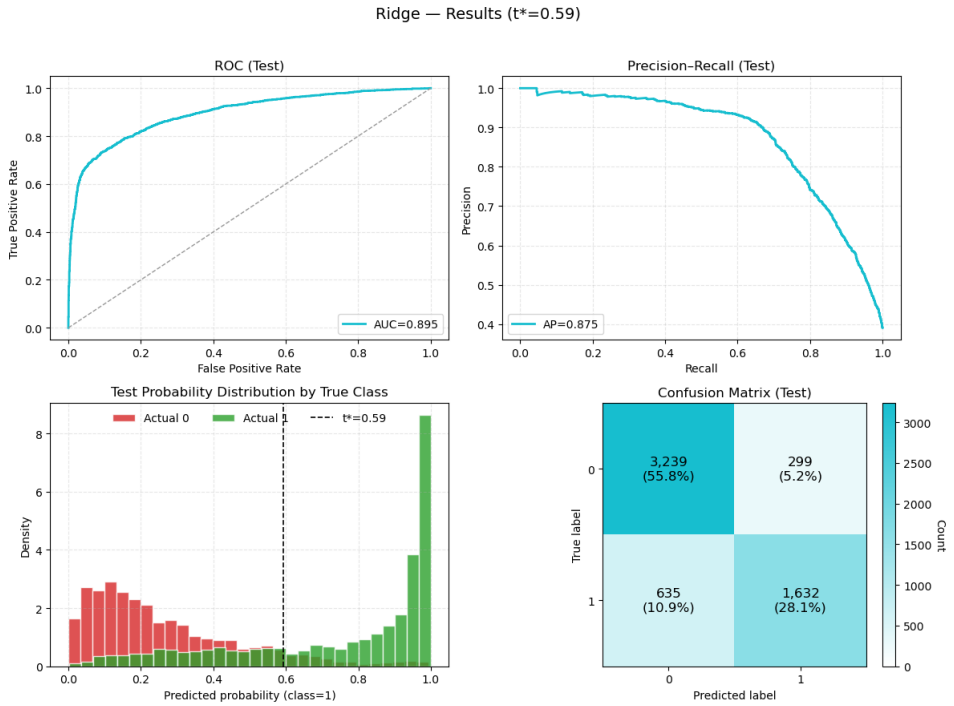

Figure E.6. Results dashboard for performance of ridge logistic regression.

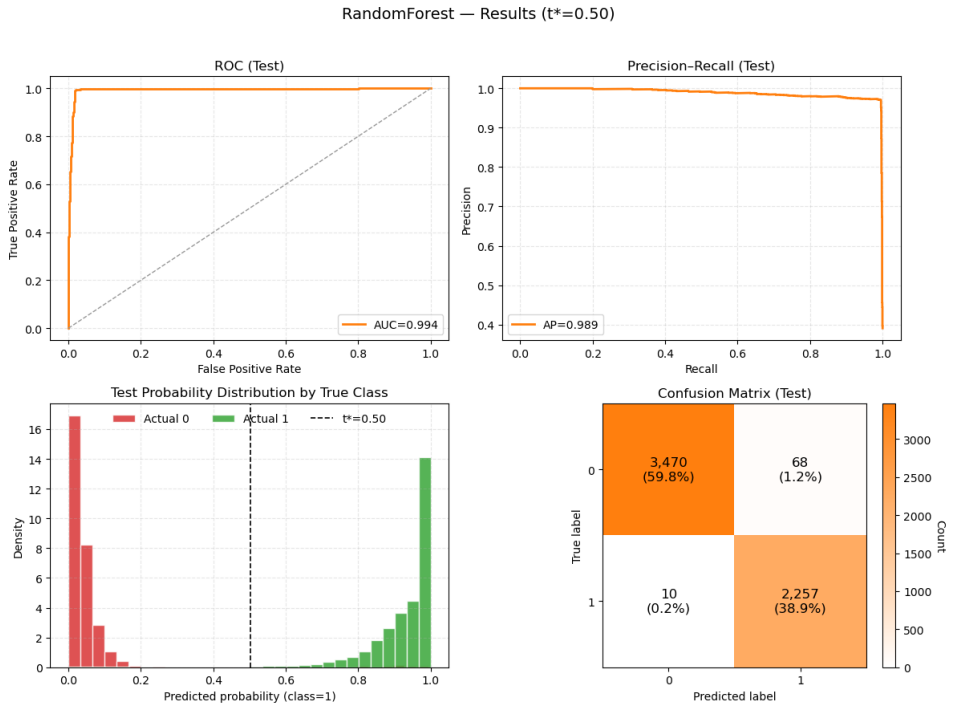

Figure E.7. Results dashboard for performance of random forest.

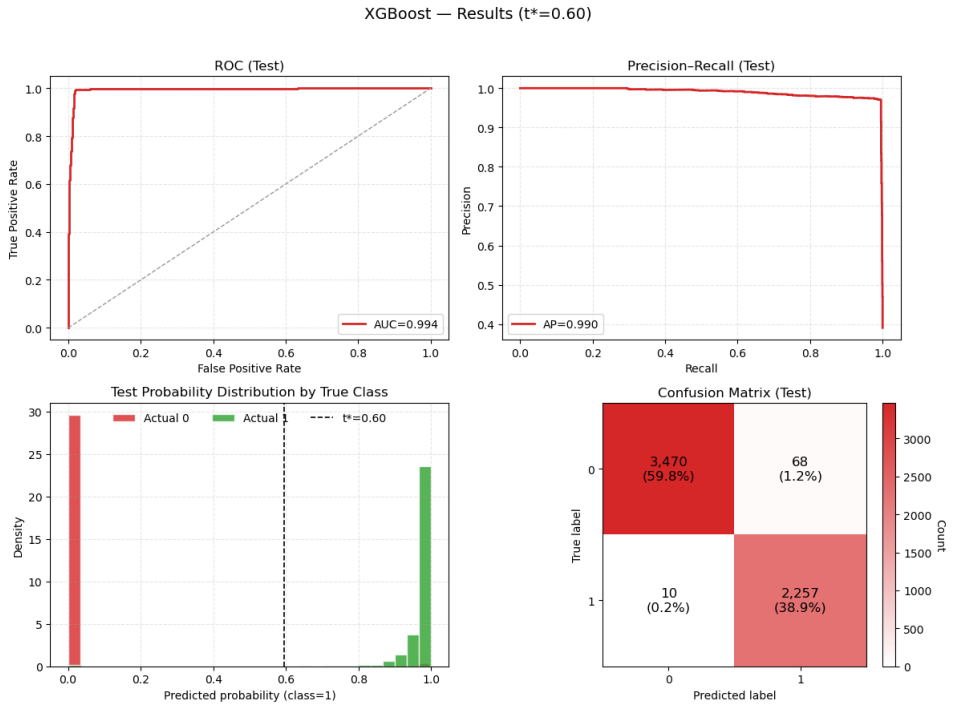

Figure E.8. Results dashboard for performance of eXtreme Gradient Boosting.

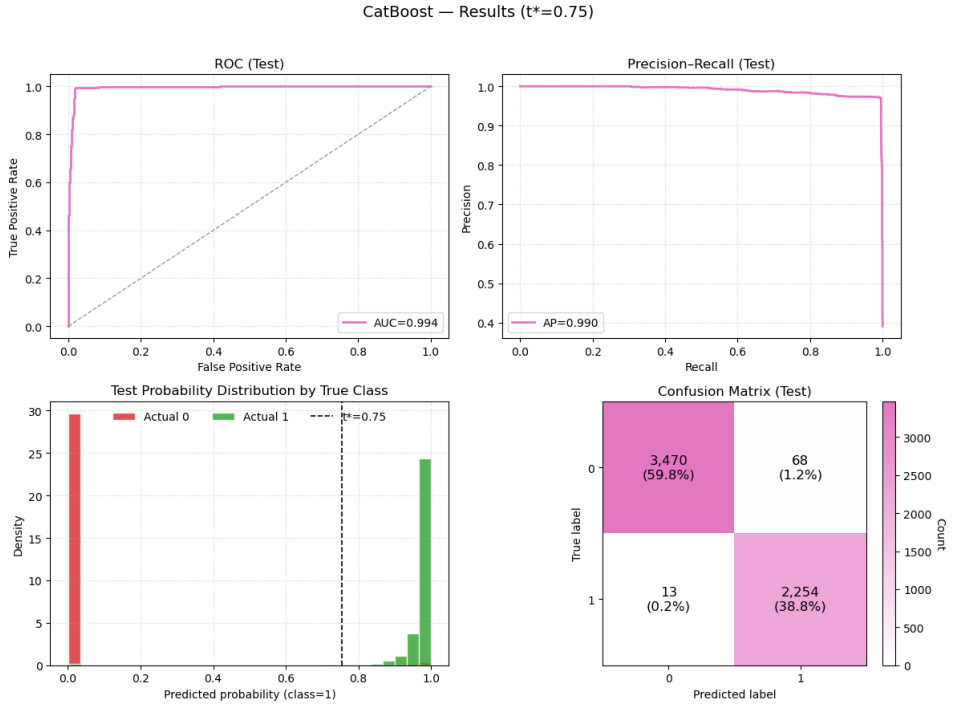

Figure E.9. Results dashboard for performance of categorical boosting.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX F:
Logistic Regression Log-Odds

Table F.1. Logistic Regression Coefficients and Odds Ratios (Grouped by Sign and Ordered by $|Coeff|$)

Feature	Coefficient	Odds Ratio
Variables Increasing Separation Risk ($Coeff > 0$)		
cat__rank_code_E09	1.529	4.612
cat__rank_code_E05	1.424	4.153
cat__rank_code_E04	1.066	2.903
cat__marital_status_MS Svc Rqt	1.040	2.828
cat__rank_code_O06	1.001	2.722
cat__rank_code_O09	0.968	2.633
cat__rank_code_E06	0.948	2.580
cat__rank_code_O05	0.916	2.499
cat__rank_code_O04	0.860	2.363
cat__rank_code_E08	0.800	2.225
num__rel_age_in_rank	0.484	1.623
num__js_duration_3mavg_prev	0.333	1.395
num__tir_los_ratio	0.250	1.284
cat__rank_code_O01	0.188	1.207
cat__rank_code_E03	0.183	1.200
num__NWL_TOTAL_TIME_months	0.163	1.178

Continued on next page

Feature	Coefficient	Odds Ratio
num__cohort_rank_prop	0.151	1.163
num__WARLIKE_TOTAL_TIME_months	0.132	1.141
cat__sex_M	0.125	1.133
num__los_years_last	0.119	1.127
cat__rank_code_O03	0.117	1.124
cat__ethnicity_African & ME	0.071	1.073
cat__ethnicity_Not Provided	0.064	1.066
num__cohort_avg_tir	0.060	1.062
cat__rank_code_O08	0.054	1.056
cat__dependents_status_Member	0.041	1.042
cat__ethnicity_European	0.041	1.042
cat__dependents_status_Accompanied Resident	0.041	1.042
cat__officer_Y	0.040	1.041
num__months_since_loc_last	0.031	1.031
num__unemp_3mavg_prev	0.031	1.031
cat__religion_Y	0.029	1.029
num__offcycle_posts_cum_max	0.018	1.018
Variables Decreasing Separation Risk ($Coeff < 0$)		
cat__rank_code_O02	-2.806	0.060
cat__rank_code_E01	-2.523	0.080
cat__rank_code_E02	-2.444	0.087
cat__rank_code_E00	-1.124	0.325
num__age_last	-1.086	0.338

Continued on next page

Feature	Coefficient	Odds Ratio
cat__rank_code_000	-1.009	0.365
num__months_since_trade_last	-0.921	0.398
num__promo_count_max	-0.918	0.400
num__trade_changes_cum_max	-0.715	0.489
cat__marital_status_Single	-0.626	0.535
cat__marital_status_Married	-0.516	0.597
num__NWL_DEPL_COUNT	-0.294	0.746
num__loc_changes_cum_max	-0.286	0.752
cat__rank_code_007	-0.250	0.779
num__months_since_promo_last	-0.237	0.789
cat__sex_F	-0.227	0.797
cat__ethnicity_Aboriginal & Torres SI	-0.210	0.811
num__depcat_changes_cum_max	-0.208	0.812
cat__dependents_status_Unaccompanied Resident	-0.185	0.831
num__months_since_depcat_last	-0.143	0.867
cat__officer_N	-0.142	0.868
cat__religion_N	-0.131	0.878
num__WARLIKE_DEPL_COUNT	-0.105	0.900
cat__ethnicity_Australian	-0.032	0.969
cat__ethnicity_Americas	-0.021	0.979
cat__ethnicity_Asian Pacific	-0.016	0.985

THIS PAGE INTENTIONALLY LEFT BLANK

List of References

Ahn, T., & Fan, J. (2022). *Machine learning in acquisition workforce talent management: New approaches to prediction of workforce retention and promotion* (Technical report No. NPS-HR-22-192). Naval Postgraduate School. Retrieved February 24, 2026, from https://dair.nps.edu/handle/123456789/4713

Alduayj, S. S., & Rajpoot, K. (2018). Predicting employee attrition using machine learning. *2018 International Conference on Innovations in Information Technology (IIT)*, 93–98. https://doi.org/10.1109/INNOVATIONS.2018.8605976

Altmann, A., Toloşi, L., Sander, O., & Lengauer, T. (2010). Permutation importance: A corrected feature importance measure. *Bioinformatics*, *26*(10), 1340–1347. https://doi.org/10.1093/bioinformatics/btq134

Ashen, R. (2022). *Family ties: The relationship between family and workforce behaviors (retention, separation, and re-entry) in the royal australian air force's officer aviation workforce* [Master's thesis]. Naval Postgraduate School. Retrieved February 24, 2026, from https://hdl.handle.net/10945/69611

Australian Bureau of Statistics. (2025a, November). *Labour force, Australia, detailed* (No. 6291.0.55.001). Australian Bureau of Statistics. Retrieved January 13, 2026, from https://www.abs.gov.au/statistics/labour/employment-and-unemployment/labour-force-australia-detailed/nov-2025

Australian Bureau of Statistics. (2025b, November). *Labour force, Australia, detailed, November 2025: Table 6291002*. Retrieved January 13, 2026, from https://www.abs.gov.au/statistics/labour/employment-and-unemployment/labour-force-australia-detailed/nov-2025/6291002.xlsx

Australian Bureau of Statistics. (2025c, November). *Labour force, Australia, detailed, November 2025: Table 6291014c*. Retrieved January 13, 2026, from https://www.abs.gov.au/statistics/labour/employment-and-unemployment/labour-force-australia-detailed/nov-2025/6291014c.xlsx

Boesel, D., & Johnson, K. (1984). *Why service members leave the military: Review of the literature and analysis* (Technical Report No. DMDC/MTC/TR-84-3). Defense Manpower Data Center. Arlington, VA.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Carr, G. (2022). *Retention in the royal australian air force aviation technical workforce: Is it changing and how feasible is the future demand?* [Master's thesis]. Naval Postgraduate School. Retrieved February 24, 2026, from https://hdl.handle.net/10945/69621

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). Smote: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, *16*, 321–357. https://doi.org/10.1613/jair.953

Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672.2939785

Cole, S. (2019). *Using machine learning to predict early service separation of technical and non-technical sailors* [Master's thesis]. Naval Postgraduate School. Retrieved February 24, 2026, from https://hdl.handle.net/10945/64123

Darragh, T. (2022). *A time series analysis of australian regular army enlisted and officer separations* [Master's thesis]. Naval Postgraduate School. Retrieved February 24, 2026, from https://hdl.handle.net/10945/69627

Davis, J., & Goadrich, M. (2006). The relationship between precision-recall and roc curves. *Proceedings of the 23rd International Conference on Machine Learning (ICML '06)*, 233–240. https://doi.org/10.1145/1143844.1143874

Defence Force Remuneration Tribunal. (n.d.). *Defence force remuneration tribunal*. Retrieved February 26, 2026, from https://www.dfrt.gov.au/

Defence Regulation 2016 (2016). Retrieved February 19, 2026, from https://classic.austlii.edu.au/au/legis/cth/consol_reg/dr2016147/

Department of Defence. (2023a). *Defence annual report 2022–23*. Australian Government. Retrieved February 24, 2026, from https://www.defence.gov.au/sites/default/files/2023-10/Defence-Annual-Report-2022-23.pdf

Department of Defence. (2023b). *Defence strategic review 2023*. Australian Government. Retrieved February 24, 2026, from https://www.defence.gov.au/about/reviews-inquiries/defence-strategic-review

Department of Defence. (2024a). *2024 national defence strategy and 2024 integrated investment program*. Australian Government. Retrieved February 24, 2026, from https://www.defence.gov.au/about/strategic-planning/2024-national-defence-strategy-2024-integrated-investment-program

Department of Defence. (2024b). *Defence annual report 2023–24*. Australian Government. Retrieved February 24, 2026, from https://www.defence.gov.au/sites/default/files/2024-10/Defence-Annual-Report-2023-24_FA.pdf

Department of Defence. (2024c). *Defence workforce plan 2024*. Australian Government. Retrieved February 24, 2026, from https://www.defence.gov.au/sites/default/files/2024-11/2024DefenceWorkforcePlan.pdf

Department of Defence. (2025a). *Australian defence force transition manual*. Australian Government. Retrieved February 24, 2026, from https://www.defence.gov.au/sites/default/files/2026-02/ADFMemberandFamilyTransitionGuide.pdf

Department of Defence. (2025b). *Defence annual report 2024–25*. Australian Government. Retrieved February 24, 2026, from https://www.defence.gov.au/sites/default/files/2025-10/Defence-Annual-Report-2024-25.pdf

Department of Defence. (n.d.). Adf total workforce system. Retrieved February 2, 2026, from https://pay-conditions.defence.gov.au/career-and-service/adf-total-workforce-system

Dodds, D. (2018). *Length-of-service / survival profiles methodology for the royal australian navy* [Master's thesis]. Naval Postgraduate School. Retrieved February 24, 2026, from https://hdl.handle.net/10945/61350

Ehrenberg, R. G., Smith, R. S., & Hallock, K. F. (2021). *Modern labor economics: Theory and public policy* (14th). Routledge. https://doi.org/10.4324/9780429327209

Falk, A. (2023). *Who leaves? individual-based predictive modeling of non-end of active service attrition for enlisted marines* [Master's thesis]. Naval Postgraduate School. Retrieved February 24, 2026, from https://hdl.handle.net/10945/72000

Fallucchi, F., Coladangelo, M., Giuliano, R., & De Luca, E. W. (2020). Predicting employee attrition using machine learning techniques. *Computers*, *9*(4), 86. https://doi.org/10.3390/computers9040086

Fernández, A., García, S., Galar, M., Prati, R. C., Krawczyk, B., & Herrera, F. (2018). *Learning from imbalanced data sets*. Springer International Publishing. https://doi.org/10.1007/978-3-319-98074-4

Fisher, A., Rudin, C., & Dominici, F. (2019). All models are wrong, but many are useful: Learning a variable's importance by studying an entire class of prediction models simultaneously. *Journal of Machine Learning Research*, *20*(177), 1–81.

Gallagher, P. J. (2020). *Predicting marine corps retention behavior with machine learning* [Master's thesis]. Naval Postgraduate School. Retrieved February 24, 2026, from https://hdl.handle.net/10945/66074

Garcia, M. F. (2022). *Insights on the use of machine learning to predict retention of career soldiers in the united states army* [Master's thesis]. The Pennsylvania State University. Retrieved February 24, 2026, from https://etda.libraries.psu.edu/catalog/23011mfg5614

Graf, E., Schmoor, C., Sauerbrei, W., & Schumacher, M. (1999). Assessment and comparison of prognostic classification schemes for survival data. *Statistics in Medicine*, *18*(17-18), 2529–2545. https://doi.org/10.1002/(SICI)1097-0258(19990915/30)18:17/18<2529::AID-SIM274>3.0.CO;2-5

Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1999). *Multivariate data analysis* (5th). Prentice Hall.

Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (roc) curve. *Radiology*, *143*(1), 29–36. https://doi.org/10.1148/radiology.143.1.7063747

Hess, R. A. (2025). *A comparison of modern machine learning methods for applied attrition modeling* [PhD dissertation]. University of Georgia. Retrieved February 24, 2026, from https://openscholar.uga.edu/record/26905?v=pdf

Hoglin, P. (2012). *Analysis of first term attrition in the australian defence force* [PhD dissertation]. UNSW Sydney. https://doi.org/10.26190/UNSWORKS/15673

Hoglin, P., & Barton, N. (2015). First-term attrition of military personnel in the australian defence force. *Armed Forces & Society*, *41*(1), 43–68. https://doi.org/10.1177/0095327X13494743

Hom, P. W., Lee, T. W., Shaw, J. D., & Hausknecht, J. P. (2017). One hundred years of employee turnover theory and research. *Journal of Applied Psychology*, *102*(3), 530–545. https://doi.org/10.1037/apl0000103

Hosek, J., Antel, J. J., & Peterson, C. E. (1989). *Who stays, who leaves? attrition among first-term enlistees* (N-2967-FMP). RAND Corporation. Retrieved February 24, 2026, from https://www.rand.org/pubs/notes/N2967.html

Ilarda, L. (2021). *Predicting deliberate employee attrition using survival analysis* [Master's thesis]. Tilburg University. Retrieved February 24, 2026, from http://arno.uvt.nl/show.cgi?fid=157352

Ishwaran, H., Kogalur, U. B., Blackstone, E. H., & Lauer, M. S. (2008). Random survival forests. *The Annals of Applied Statistics*, *2*(3), 841–860. https://doi.org/10.1214/08-AOAS169

James, G., Witten, D., Hastie, T., Tibshirani, R., & Taylor, J. (2023). *An introduction to statistical learning: With applications in python* (2nd). Springer. https://doi.org/10.1007/978-3-031-38747-0

Jin, Z., Shang, J., Zhu, Q., Ling, C., Xie, W., & Qiang, B. (2020). Rfrsf: Employee turnover prediction based on random forests and survival analysis. In Z. Huang, W. Beek, H. Wang, R. Zhou, & Y. Zhang (Eds.), *Web information systems engineering – wise 2020* (pp. 503–515, Vol. 12343). Springer International Publishing. https://doi.org/10.1007/978-3-030-62008-0_35

Johns, R. (2003). Why do employees quit: A case study of the complexities of labour turnover. In G. White, S. Corby, & C. Stanworth (Eds.), *Regulation, de-regulation and re-regulation: The scope of employment relations in the 21st century*. International Employment Relations Association.

Johnston, I. (1988). *Turnover of junior army officers: A test of the mobley, griffeth, hand and meglino model of personnel turnover using structural equation techniques* [Master's thesis]. Naval Postgraduate School.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems*, *30*, 3146–3154.

Kleinbaum, D. G., & Klein, M. (2012). *Survival analysis: A self-learning text* (3rd). Springer. https://doi.org/10.1007/978-1-4419-6646-9

Krishna, G. S., Aravind, D., Anusha, M., Chandra, K. U., Kalangi, P. K., & Swamy, I. M. K. (2025). Employee retention prediction by machine learning models. *Proceedings of the 2025 5th International Conference on Intelligent Technologies (CONIT)*, 1–5. https://doi.org/10.1109/CONIT65521.2025.11166715

Lee, T. W., & Mitchell, T. R. (1994). An alternative approach: The unfolding model of voluntary employee turnover. *Academy of Management Review*, *19*(1), 51–89. https://doi.org/10.2307/258835

Lockwood, J., Gelder, A., Goldberg, M., Brooks, J., & Prugh, G. (2020). *Leveraging machine learning in defense personnel analyses* (Report No. P-13174). Institute for Defense Analyses.

Louhichi, M., Nesmaoui, R., Mbarek, M., & Lazaar, M. (2023). Shapley values for explaining the black box nature of machine learning model clustering. *Procedia Computer Science*, *220*, 806–811. https://doi.org/10.1016/j.procs.2023.03.107

Lundberg, S. M. (2018). *Beeswarm plot — shap documentation*. Retrieved February 22, 2026, from https://shap.readthedocs.io/en/latest/example_notebooks/api_examples/plots/beeswarm.html

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 4768–4777.

Mansor, N., Sani, N. S., & Aliff, M. (2021). Machine learning for predicting employee attrition. *International Journal of Advanced Computer Science and Applications*, *12*(11). https://doi.org/10.14569/IJACSA.2021.0121149

Marrone, J. (2020). *Predicting 36-month attrition in the u.s. military: A comparison across service branches* (Research report). RAND Corporation. https://doi.org/10.7249/RR4258

Matplotlib Development Team. (n.d.). *Matplotlib* (Version 3.8.4). Retrieved February 28, 2026, from https://matplotlib.org/

Michaels, C. E., & Spector, P. E. (1982). Causes of employee turnover: A test of the mobley, griffeth, hand, and meglino model. *Journal of Applied Psychology*, *67*(1), 53–59. https://doi.org/10.1037/0021-9010.67.1.53

Mitchell, T. R., Holtom, B. C., Lee, T. W., Sablynski, C. J., & Erez, M. (2001). Why people stay: Using job embeddedness to predict voluntary turnover. *Academy of Management Journal*, *44*(6), 1102–1121. https://doi.org/10.2307/3069391

Mobley, W. H., Griffeth, R. W., Hand, H. H., & Meglino, B. M. (1979). Review and conceptual analysis of the employee turnover process. *Psychological Bulletin*, *86*(3), 493–522. https://doi.org/10.1037/0033-2909.86.3.493

Molnar, C. (2019). *Interpretable machine learning: A guide for making black box models explainable*. Leanpub. https://leanpub.com/interpretable-machine-learning

NumPy Developers. (n.d.). *Numpy* (Version 1.26.4). Retrieved February 28, 2026, from https://numpy.org/

Pandas Development Team. (n.d.). *Pandas* (Version 2.2.2). Retrieved February 28, 2026, from https://pandas.pydata.org/

Pickett, K. L., Suresh, K., Campbell, K. R., Davis, S., & Juarez-Colunga, E. (2021). Random survival forests for dynamic predictions of a time-to-event outcome using a longitudinal biomarker. *BMC Medical Research Methodology*, *21*, 216. https://doi.org/10.1186/s12874-021-01375-x

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). Catboost: Unbiased boosting with categorical features. *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, 6639–6649.

Python Software Foundation. (n.d.). *Python* (Version 3.12.12). Retrieved February 28, 2026, from https://docs.python.org/3/

Raza, A., Munir, K., Almutairi, M., Younas, F., & Fareed, M. M. S. (2022). Predicting employee attrition using machine learning approaches. *Applied Sciences*, *12*(13), 6424. https://doi.org/10.3390/app12136424

Saito, T., & Rehmsmeier, M. (2015). The precision-recall plot is more informative than the roc plot when evaluating binary classifiers on imbalanced datasets. *PLOS ONE*, *10*(3), e0118432. https://doi.org/10.1371/journal.pone.0118432

Sajjadiani, S., Sojourner, A. J., Kammeyer-Mueller, J. D., & Mykerezi, E. (2019). Using machine learning to translate applicant work history into predictors of performance and turnover. *Journal of Applied Psychology*, *104*(10), 1207–1225. https://doi.org/10.1037/apl0000405

Schmid, M., Wright, M. N., & Ziegler, A. (2016). On the use of harrell's c for clinical risk prediction via random survival forests. *Expert Systems with Applications*, *63*, 450–459. https://doi.org/10.1016/j.eswa.2016.07.018

Scikit-learn developers. (n.d.). *Scikit-learn* (Version 1.7.2). Retrieved February 28, 2026, from https://scikit-learn.org/stable/

Sheu, Y.-h., Sun, J., Lee, H., Castro, V. M., Barak-Corren, Y., Song, E., Madsen, E. M., Gordon, W. J., Kohane, I. S., Churchill, S. E., Reis, B. Y., Cai, T., & Smoller, J. W. (2023). An efficient landmark model for prediction of suicide attempts in multiple clinical settings. *Psychiatry Research*, *323*, 115175. https://doi.org/10.1016/j.psychres.2023.115175

Sumathi, K., Balakrishnan, D., Naveen, V., Hariharan, P., & Iniyan, M. R. (2021). Talent flow employee analysis based turnover prediction on survival analysis. *Annals of the Romanian Society for Cell Biology*, *25*(4), 3844–3857.

Swets, J. A. (1988). Measuring the accuracy of diagnostic systems. *Science*, *240*(4857), 1285–1293. https://doi.org/10.1126/science.3287615

Terrazas, G. A. (2020). *Evaluation of machine learning applicability for usmc reenlistment* [Master's thesis]. Naval Postgraduate School. Retrieved January 30, 2026, from https://hdl.handle.net/10945/64888

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *58*(1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Tuppin, K. A. (2025). *The career trajectories of australian army officers: How the army impacts individual career outcomes* [Doctoral dissertation]. Macquarie University. https://doi.org/10.25949/30141667.v1

Uno, H., Cai, T., Pencina, M. J., D'Agostino, R. B., & Wei, L. J. (2011). On the c-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. *Statistics in Medicine*, *30*(10), 1105–1117. https://doi.org/10.1002/sim.4154

U.S. Department of Defense. (2025). *Report of the fourteenth quadrennial review of military compensation: Volume i* (Report). U.S. Department of Defense. Retrieved February 26, 2026, from https://militarypay.defense.gov/Portals/3/Documents/QRMC_14_Vol1_final_web.pdf

Weiss, H. M., MacDermid, S. M., Strauss, R., Kurek, K. E., Le, B., & Robbins, D. (2002). *Retention in the armed forces: Past approaches and new research directions* (Research report). Military Family Research Institute at Purdue University.

Zhu, Q., Shang, J., Cai, X., Jiang, L., Liu, F., & Qiang, B. (2019). Coxrf: Employee turnover prediction based on survival analysis. *Proceedings of the 2019 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing &*

Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation, 1123–1130. https://doi.org/10.1109/SmartWorld-UIC-ATC-SCALCOM-IOP-SCI.2019.00212

Acquisition Research Program
Naval Postgraduate School
555 Dyer Road, Ingersoll Hall
Monterey, CA 93943

www.acquisitionresearch.net