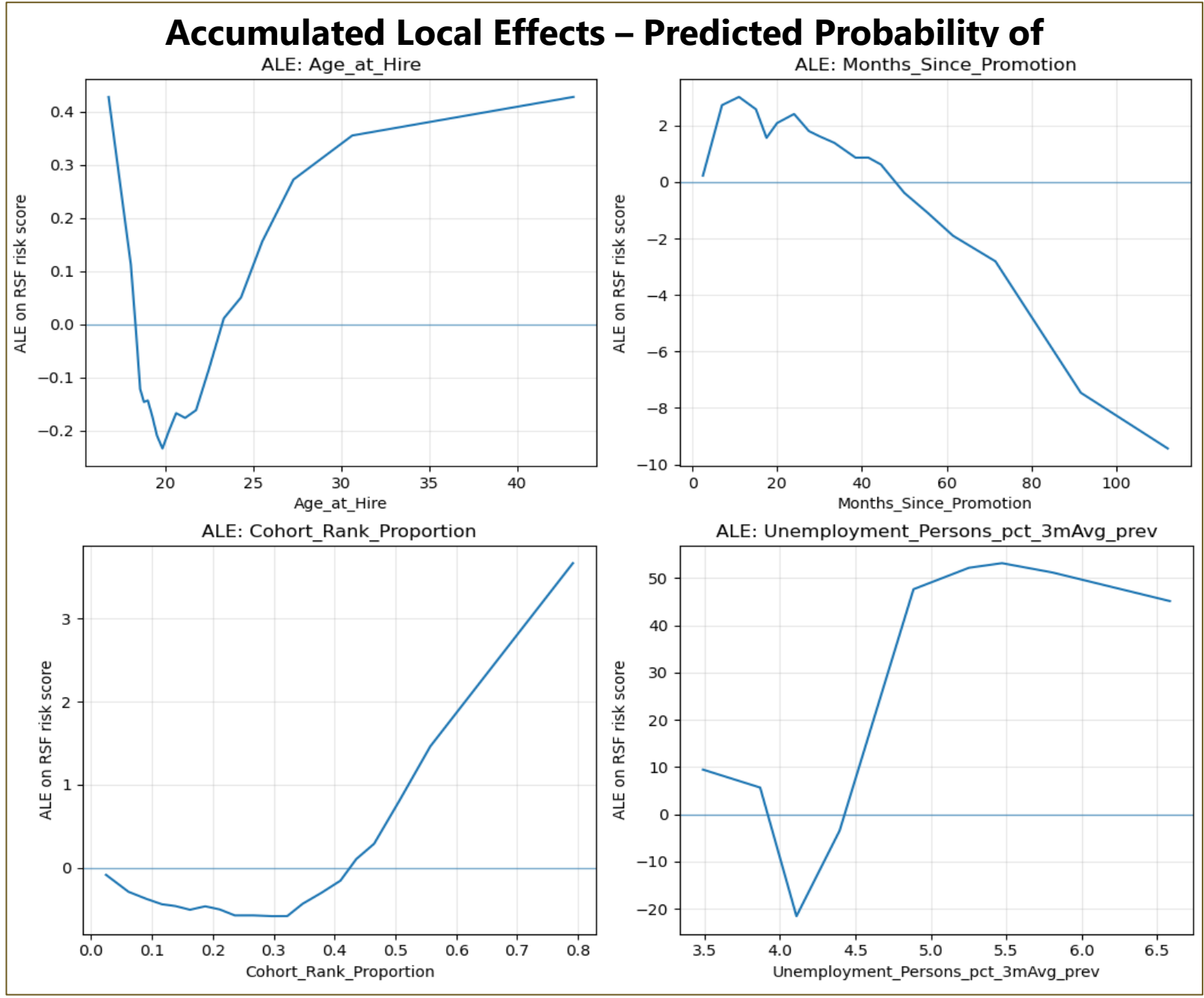
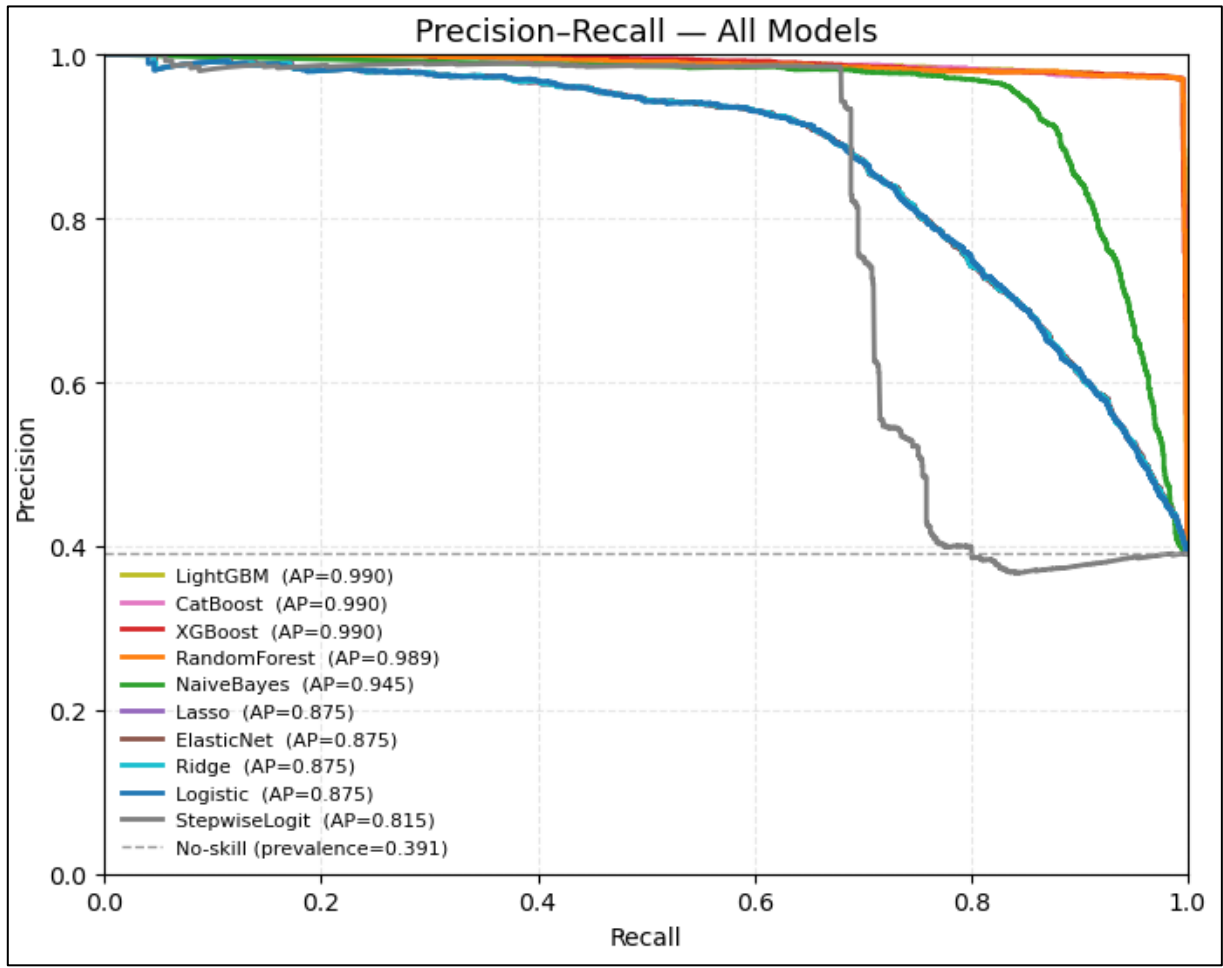
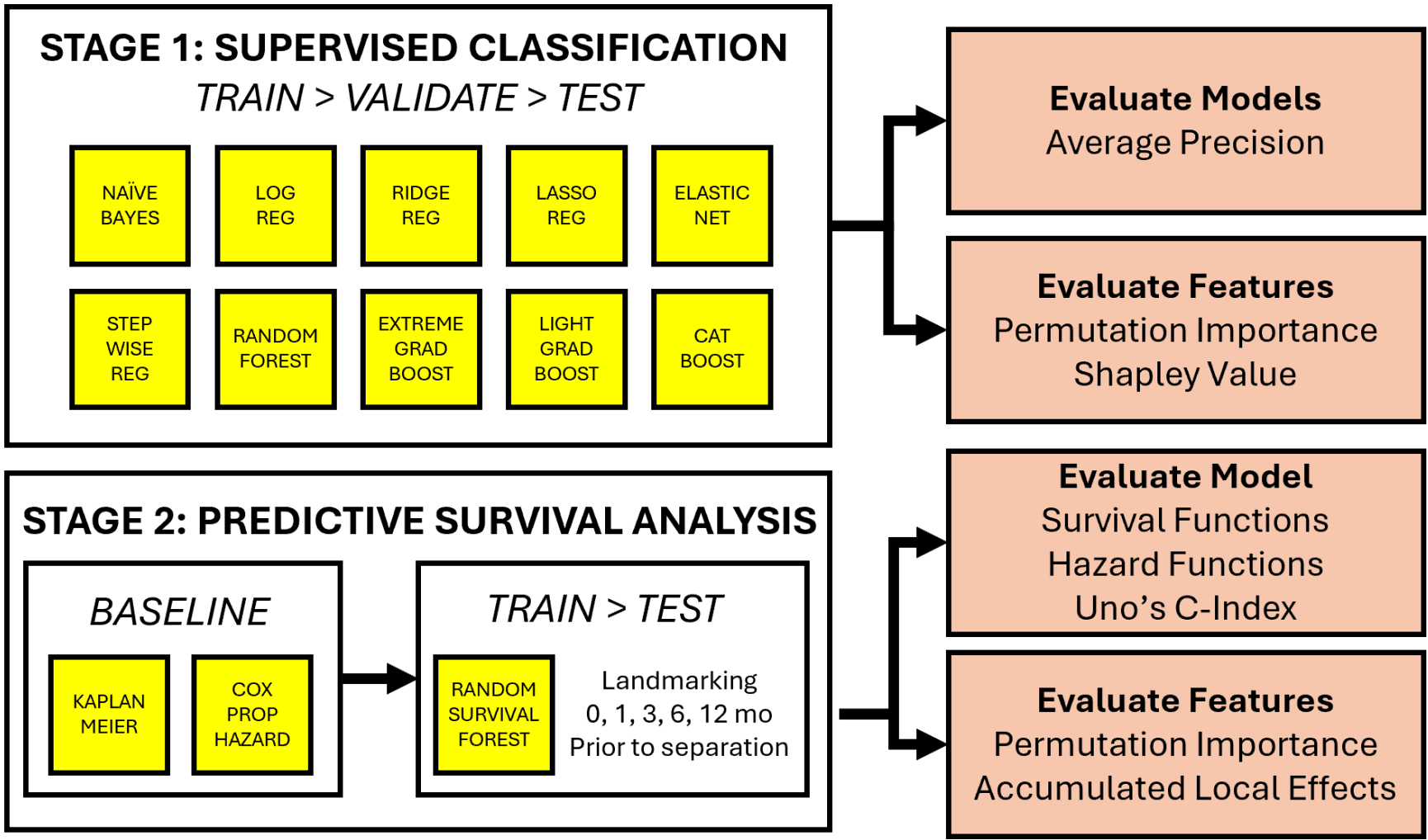


Abstract

- 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.
- Stage 1 found 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.

Methods

- Person-month administrative dataset joined with performance, operations, separation and economic data (FY2015-16 to FY2024-25)
- Use workforce turnover theory to engineer features
- Two stage ML process
 - Stage 1: Supervised Classification
 - Stage 2 Predictive Survival Analysis
- Compare model predictive performance
- Assess and interpret feature importance using Permutation Importance, Shapley Additive Explanation, and Accumulated Local Effects
- Landmark at 1, 3, 6, and 12 months prior to separation and test model predictive performance



Results & Impact

- Tree ensemble and gradient-boosted models performed strongly in both stages.
- Strong predictive ability at 1, 3, 6, and 12 month landmarks prior to separation
- Macro-economic and career dynamic features were most important to predicting separation. Demographic features displayed little influence.

Future Research

- Evaluate neural networks and deep survival models for predictive performance using longitudinal data
- Investigate macro-economic effects on separation and whether they proxy structural or real-world events
- Utilize ML findings to direct causative studies and retention policies
- Utilize ML to predict potential outcomes of interventions