NPS-HR-22-192

ACQUISITION RESEARCH PROGRAM
SPONSORED REPORT SERIES

Machine Learning in Acquisition Workforce Talent Management: New Approaches to Prediction of Workforce Retention and Promotion

June 24, 2022

Dr. Sae Young (Tom) Ahn, Assistant Professor

Dr. James Fan, Assistant Professor

Department of Defense Management

Naval Postgraduate School

The research presented in this report was supported by the Acquisition Research Program at the Naval Postgraduate School.

Executive Summary

The Department of Navy (DoN) and Department of Defense (DoD) acquisition workforce (AWF) strategic plans call for a restoration and strengthening of the civilian AWF after more than 2 decades to contraction. To reform and reshape the workforce to improve the acquisition process and delivery of world-class warfighting capability for the military, the AWF leadership must understand how attrition and retention will impact the "size, composition, and skill" needs of the workforce in "parallel with technology advances and global trends" (DoDUSD[AT&L], 2019, p. 5).

To achieve strategic talent management of the workforce, it is critical to have the ability to predict which workers are most likely to leave the AWF. Forecasting attrition will aid the leadership by identifying (1) which workers to target for retention via incentives and (2) which areas will need to increase or decrease recruitment to quickly fill personnel gaps that may arise.

This technical report is the first to evaluate whether machine learning (ML) can be a useful tool for the AWF leadership to make attrition forecasts. We first show that ordinary least squares (OLS)—which is the tool most often associated with statistical modeling of worker behavior—performs poorly, especially given the sparse administrative data set we have access to. We then test a variety of ML algorithms and find that they can predict worker attrition with a higher degree of accuracy. Our conclusion from this exploratory analysis is that, as algorithmic effectiveness increases with data set size (in terms of more worker and job/task characteristics), there may be many use cases for these algorithms in future predictive modeling for manpower and retention.

Judicious use of ML algorithms may aid the DoN to "build, develop, and sustain a balanced workforce to compete and win" (DoDUSD[AT&L], 2019).

THIS PAGE LEFT INTENTIONALLY BLANK

NPS-HR-22-192

ACQUISITION RESEARCH PROGRAM
SPONSORED REPORT SERIES

Machine Learning in Acquisition Workforce Talent Management: New Approaches to Prediction of Workforce Retention and Promotion

June 24, 2022

Dr. Sae Young (Tom) Ahn, Assistant Professor

Dr. James Fan, Assistant Professor

Department of Defense Management

Naval Postgraduate School

THIS PAGE LEFT INTENTIONALLY BLANK

Table of Contents

THIS PAGE LEFT INTENTIONALLY BLANK

Introduction

Skilled workers are in higher demand than ever before, yet they are expensive to recruit, train, and retain. Like any other private firm or government organization, the defense acquisition workforce (AWF) faces this challenge, especially for technical workers, whose skills are in high demand in the civilian sector and sensitive to industry demand. Consequently, it is critical to retain the right technical skill sets at the right levels of the organization. The *Department of Defense Acquisition Workforce Strategic Plan: FY2016–FY2021* spells out the talent management goals of the organization:

> Recruit, hire, develop, and retain a diverse, agile, highly qualified, and motivated workforce of acquisition professionals to acquire and deliver world-class warfighting capabilities for our Soldiers, Sailors, Airmen, and Marines. (Under Secretary of Defense for Acquisition, Technology, and Logistics [USD(AT&L)], 2015, p. 1)

The strategic plan further seeks to increase the overall size of the force as well as to reshape it "by deliberate and targeted growth of specific career fields" (USD[AT&L], 2015, p. 5). However, engaging in ambitious shaping of the workforce without a deep understanding of which workers are likely to leave or stay can have significant impacts on the composition of the workforce as well as impede the ability to carry out its mission.

As a recent example, Figure 1 from the AWF strategic plan shows the results of a short-sighted reduction in the workforce throughout the early 2000s. While the goal of reducing head count was achieved, the AWF lost a large number of young workers with 5 to 10 years of experience. The creation of this "bathtub" was a substantive problem for the AWF. Without active intervention, the trough of the bathtub (which indicates a large deficit in manpower with a specific amount of experience in the AWF) would push rightward with time, eventually leaving the workforce with few with the requisite experience for senior leadership. The figure also shows that despite the concerted effort by the AWF leadership to fill in the "bathtub" with additional hiring, the shape of the workforce remains unbalanced almost a decade later. With the benefit of hindsight, it is not surprising that relatively young workers who had been in the AWF long enough to acquire valuable human capital would be the most likely to voluntarily leave, because they would be in high demand in the private sector.

The strategic plan also describes future goals to change the force structure to accommodate the required skill sets, improve employee quality and professionalism, and recruit and retain a more diverse workforce to maximize effectiveness. These are ambitious goals that would be easier to achieve with a better understanding of the propensity of workers with certain sociodemographic and professional characteristics to leave the AWF.

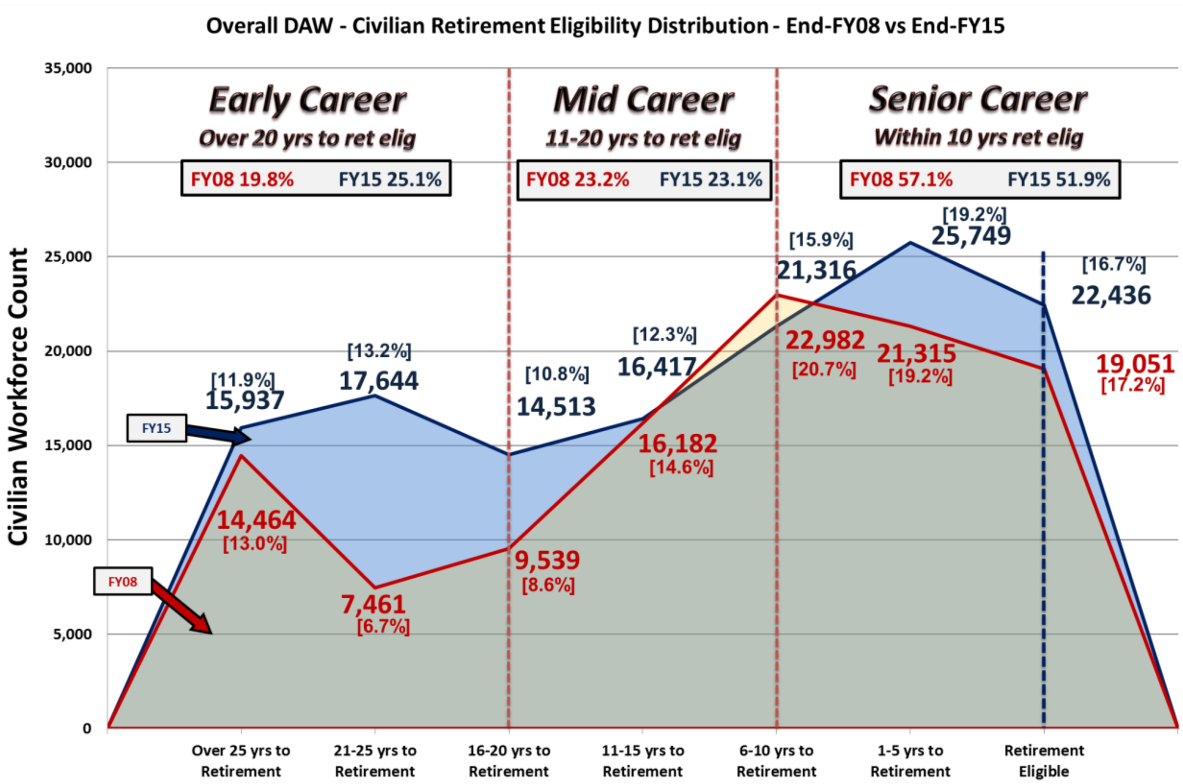

Figure 1. Civilian AWF Retirement Eligibility Distribution.
Source: USD(AT&L, 2015).

To assist the AWF leadership in these goals, this report evaluates the effectiveness of machine learning (ML) in predicting the attrition behavior of the civilian workforce. ML is a catchall term for a myriad of atheoretical methods of optimization (or algorithms) that use computer programming to analyze data to make predictions

(Alpaydin, 2020). Despite the explosion of interest and application of ML in a myriad of fields, including Department of Defense (DoD)–sponsored research, analysis of manpower policies using ML is virtually unexplored. There are at least two reasons for this relative dearth of high-quality scholarship.

First, one of the useful characteristics of ML is its applicability in taking massive amounts of real-time data and outputting forecasts immediately. As such, ML has often been used in time-critical applications. Personnel policies are, by their nature, slower in tempo. For example, the AWF strategic plan is published on a 5-year schedule. In addition, personnel data tend to be sparse and small, with data sets for analysis rarely being larger than 100 GB with fewer than 100 variables.

Second, statistical research in the acquisition field has been dominated by classical regression techniques, which focus on inferring causality. Recent examples include Brien [2019] and Pickar et al. [2019]. Due to the methodological focus of researchers funded by the Acquisition Research Program (ARP) in the past, to date, no one has attempted to evaluate whether ML could be useful in AWF personnel policy. Within the military context, there have been relatively few research papers that have married ML with active-duty personnel issues (Ahn et al., 2021; Terrazas, 2020).

We attempt in this report to provide the first evaluation of whether ML can be effectively leveraged to provide predictive insights into AWF worker retention behavior. The next section describes the data we use to analyze the AWF. We present simple summary statistics and figures showing long-run trends in retention behavior. Next, we briefly describe the ML algorithms we will be examining. We then run several models and determine the efficacy of various ML algorithms. Finally, we conclude and provide policy recommendations and potential next steps for research.

THIS PAGE LEFT INTENTIONALLY BLANK

Data

Our data consist of approximately 20 years of longitudinal data (from September 1987 to December 2018) of a subset of the AWF. We restrict our attention to DOD employees who were at any point during the 20-year span in the contracting, industrial property management, or purchasing fields (Occupation Codes 1102, 1103, and 1105). This sample includes not only workers who have had continuous employment in these AWF fields, but also those who started their careers in other DoD departments and those who moved to other DoD entities.

These civilian DoD employee data were extracted and cleaned by the Defense Manpower Data Center (DMDC). The raw data were from the 1,000-byte Appropriated Funds (APF) Civilian Personnel Master File. The data files were transferred to Naval Postgraduate School (NPS) researchers' secured workstations after encryption and anonymization by DMDC. The data were in a flat ASCII format and are 0.98 GB in size. We converted these data into Stata data file format (.dta) and SPSS file format (.sav) for analyses in these two software packages.

Restricting the AWF sample substantively decreased the size of the population to be analyzed. In addition, we dropped workers born before January 1, 1950, and those born after December 31, 1980. Workers born prior to 1950 would have retired or be nearing retirement, which may result in sample selection bias. Additionally, it may be difficult to model decisions based on unobservable factors such as health or family circumstances. Further, these workers' primary labor market experiences, in the 1970s and 1980s, may be less relevant for predicting the behavior of current or future workers. Employees born in 1981 or later may be too young to provide relevant information on long-term career decisions. After restricting the sample, we obtained more than 2 million worker-month variables. Approximately 5,700 AWF employees are tracked in the sample.

A partial list of variables received from DMDC is in Table 1. Previous research exploring the same topic using dynamic programming methods (Ahn & Menichini, 2021) used variables that are in bold. To exploit the power of ML algorithms, we used all variables that are bolded.

Table 1. List of Variables. Source: Defense Manpower Data Center (DMDC, 2022).

Variables
Unique ID
Gender
Race Code
Education Level
Tenure
Disability
Prior Military Experience
Basic Pay
Locality Adjustment
Adjusted Basic Pay
Total Salary
Nature of Action (1)
Nature of Action (2)
File As of Date

Table 2 presents some summary statistics for the sample.

Table 2. Summary Statistics for the Acquisition Workforce. Source: DMDC (2022).

Variables	Mean (Std. Dev)
Female	0.654
White	0.797
African American	0.216
Hispanic	0.042
Asian	0.031
Native American / Native Alaskan	0.010
Has Identified Disability	0.145
Prior Military Service	0.361
Has Bachelor’s Degree	0.374
Has Post-Graduate Degree	0.228
Gained Additional Education in AWF	0.441
Career Length in AWF (in years)	9.09 (8.04)
Age at Entry (in years)	33.0 (8.2)
Age at Exit (in years)	48.2 (10.55)
Position Type: Professional	0.497
(Ever Held) Technical	0.164
Blue-Collar	0.009
White-Collar	0.174
Ever Ranked Not Fully Satisfactory	0.723
Highest Salary (in dollars)	56,015.22 (29,088.28)
Observations	5,692

The sample of workers is majority female and 80% white. African Americans make up a little over 20%, and there are smaller fractions of Hispanic, Asian, and Native American/Alaskan workers. The workers are highly educated, with over 60% of the sample having a bachelor's degree or above. Among these workers, 44% acquired additional education or training while in the AWF. More than one-third of the sample is prior-enlisted, indicating that this is one of the main pathways by which workers are recruited.

Average starting age for the workforce sample is 33, implying that this is not their first job. Tenure in the job is approximately 12 years. Once a worker enters the DoD labor force, the job-match is remarkably stable. That said, transfers into and out of the AWF fields, but within the DoD, is not treated as a separation event in this study. In combination with starting age, advanced education, and the fact that most of these workers are in professional or technical positions, the average of the maximum annual salary of the sample is over $56,000, with the top observation at over $180,000. One curious aspect of the data is that over 70% of workers have at some point been evaluated as "not fully satisfactory." Whether this is due to the standards in the AWF being particularly high or some fundamental, widespread problem in the workforce remains unclear.

Figures 1 to 3 show empirical Kaplan–Meier "survival" estimates of AWF attrition. Looking at Figure 1, we can see that a representative (i.e., average) AWF worker has about a 60% chance of staying in the acquisition field for at least 5 years. At about 20 years, roughly 75% of the workforce has left the acquisition field. Figure 2 shows that male and female attrition behavior is similar. Figure 3 shows that there is a large difference in attrition behavior between workers with no military experience, compared to those who joined the AWF with prior-enlisted experience. Workers with prior-enlisted experience are likely to stay much longer in the AWF. These figures demonstrate that workers with different characteristics may behave in vastly different or surprisingly similar ways when it comes to long-run career decisions.

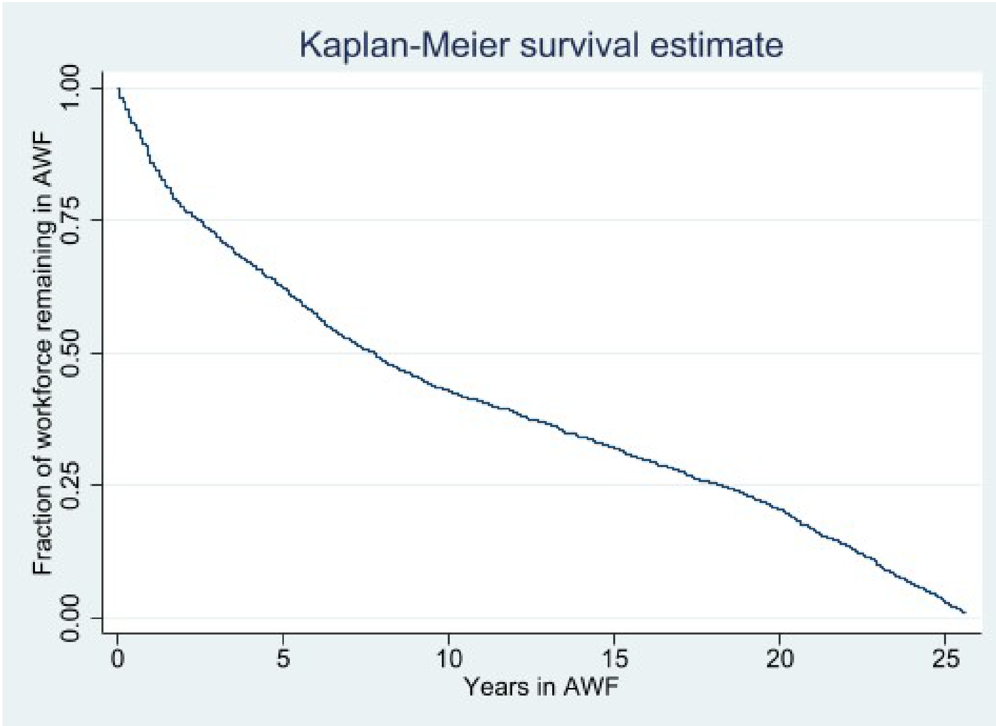

Figure 1: Career Trajectory of Representative AWF Worker

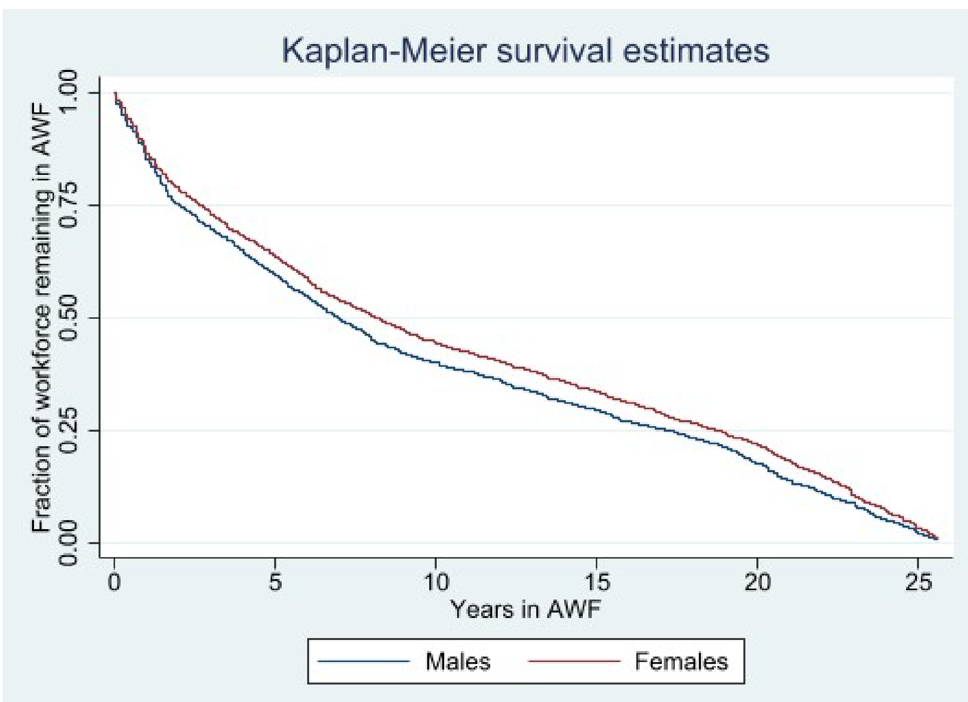

Figure 2: Career Trajectory of Workers by Gender

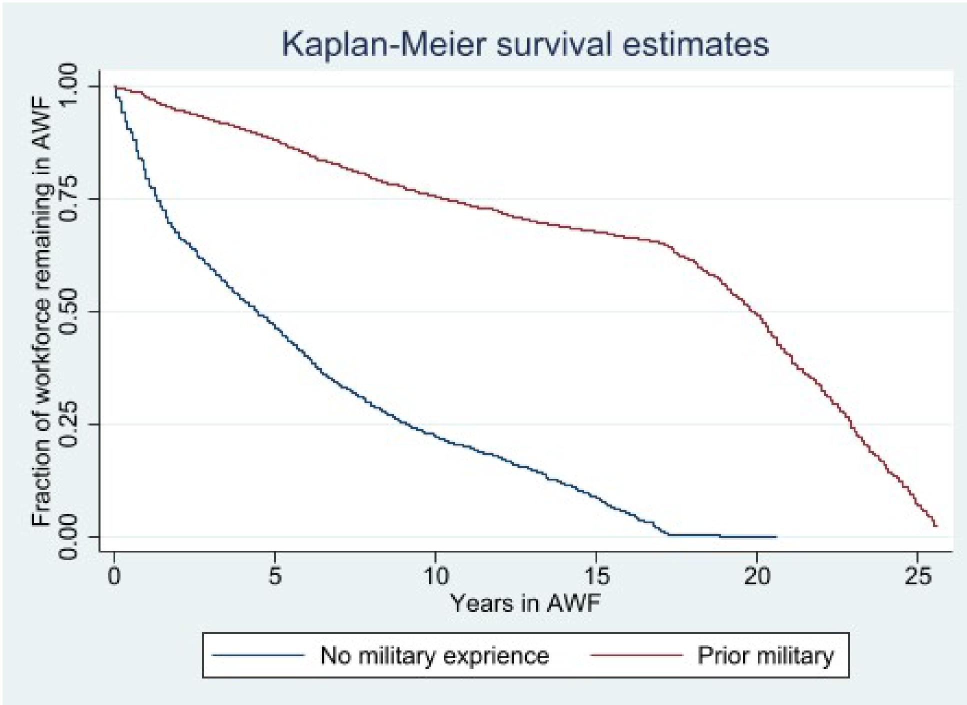

Figure 3: Career Trajectory of Workers by Prior Military Experience

THIS PAGE LEFT INTENTIONALLY BLANK

Machine Learning Algorithms

ML focuses on the use of data and different algorithms to imitate the way that humans learn and improve future predictions based on historical data. ML algorithms in (relatively) common use include Nearest Neighbor, Naive Bayes, Decision Trees, Linear Regression, SVM, Extreme Gradient Boosting, and Neural Networks. Below, we describe some common ML algorithms in the literature.

SVM models are used for classification or regression via a linear combination of features (Delen, 2010). Decision trees is a classifier method that is built from many decision trees, where each tree is a simple function that will provide an output when a set of predictor values is entered (Delen, 2010).

Neural networks, also known as artificial neural networks, mimic biology—in particular, the human brain. The networks are comprised of a node layers, containing an input layer, one or more hidden layers, and an output layer. Each node, or artificial neuron, connects to another and has an associated weight and threshold. If the output of any individual node is above the specified threshold value, that node is activated, sending data to the next layer of the network. This model is capable of analyzing very complex nonlinear functions (Delen, 2010).

Random forest algorithms are quick and able to handle variables with varying measurement units (Mather & Mahesh, 2003). The structure of a random tree model consists of a collection of small trees or networks. In this model, input variable values in each tree are compared to the target variable and classified against the target. This comparison process is carried out across all trees, and the output class is the one that is identified most often; consequently, the response is the average of all the responses (Kalmegh, 2015).

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

K-Means sorts the data into clusters of related data points to create latent classes. The K-Means clustering method partitions the data in a number of groups (k) and then continues to create clusters and refine the models. This process of creating and refining clusters continues until there is no change in centers and the model has converged to give the most powerful cluster output (Wagstaff et al., 2001).

Gradient boosting is a standard supervised learning technique that attempts to build strong classifiers by aggregating weak classifiers. Extreme gradient boosting assigns weights to independent variables in sequentially formed trees. Weights are reevaluated based on correct and incorrect predictions on each iteration, and weights on wrongly predicted variables are "boosted" (increased) in the sequentially formed tree (Friedman, 2001).

While there are hundreds of distinct ML algorithms, in this study, we will focus on those that have gained wide acceptance as being useful in the private sector.

Estimation Technique

We leverage the suite of ML algorithms listed above as well as traditional linear regression, with years of service (YOS) as the dependent variable. Independent variables are listed in Table 1. We split the data in half; one datum is used in the training set and one half is used to test prediction accuracy. Empirical studies predicting military retention and attrition has been conducted often using traditional econometric techniques, which is why we use traditional linear regression as a baseline for comparison. Our goal is to assess the potential of ML algorithms in the context of manpower. Do these new predictive models offer insights that standard econometric models cannot? We present results in the following section.

Results

Ordinary Linear Regression

We start by reporting linear regression estimations with YOS as the dependent variable. For independent variables, we use all variables in Table 1 as well as second level interaction effects between indicator variables. Figure 4 ranks the predictor importance of the independent variables in our linear estimation.

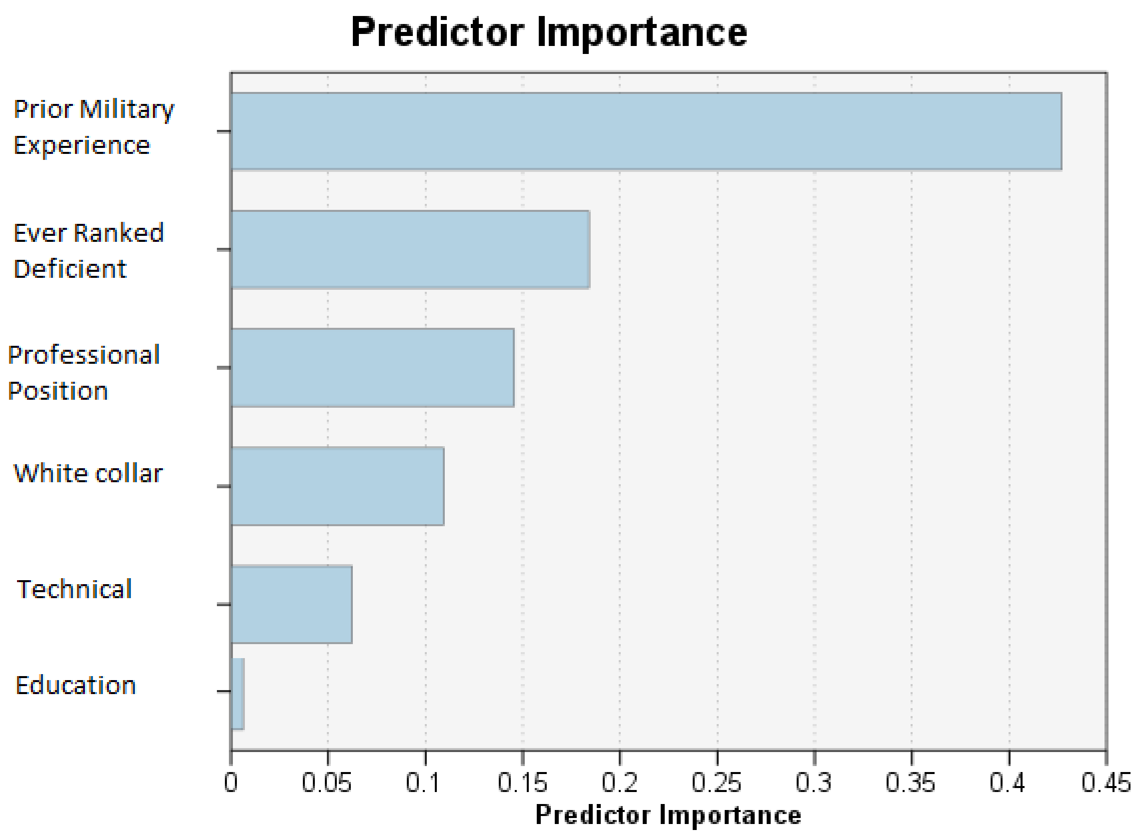

Figure 4: Predictor Importance in Linear Regression

Table 3 shows the estimation results of interaction effects, with redundant and/or insignificant interactions suppressed for brevity.

Table 3: OLS Regression of Interaction Effects on YOS

Variable	Coefficient	Std. Error	t statistic	p-value
Intercept	25.392	1.694	14.991	>0.001
Never rank deficient * Prior military experience	-13.971	0.58	-24.07	>0.001
Rank deficient * No prior military experience	-3.701	1.758	-2.106	0.035
Professional * Not white collar	3.843	1.268	3.031	0.002
Professional * No prior military experience	-2.456	0.536	-4.579	>0.001
Not professional * Never rank deficient	3.51	0.356	9.85	>0.001
Professional * Not technical	3.024	1.017	2.974	0.003
Male * Prior military experience	-1.238	0.402	-3.079	0.002
Not technical * No prior military experience	-1.938	0.517	-3.749	>0.001
Bachelor * Not white	1.451	0.672	2.16	0.031
Never rank deficient * No disability	0.586	0.312	1.877	0.061
Not white * No prior military	-1.324	0.477	-2.772	0.006
Not blue collar * No prior military experience	-4.006	1.354	-2.958	0.003

Note. Adjusted R-squared: 0.73

Figure 5 shows the difference between the model's predicted YOS and actual YOS from the data. As expected, we see significant errors between the model predictions and the actual data. We will use this as a baseline for comparison with the suite of ML algorithms.

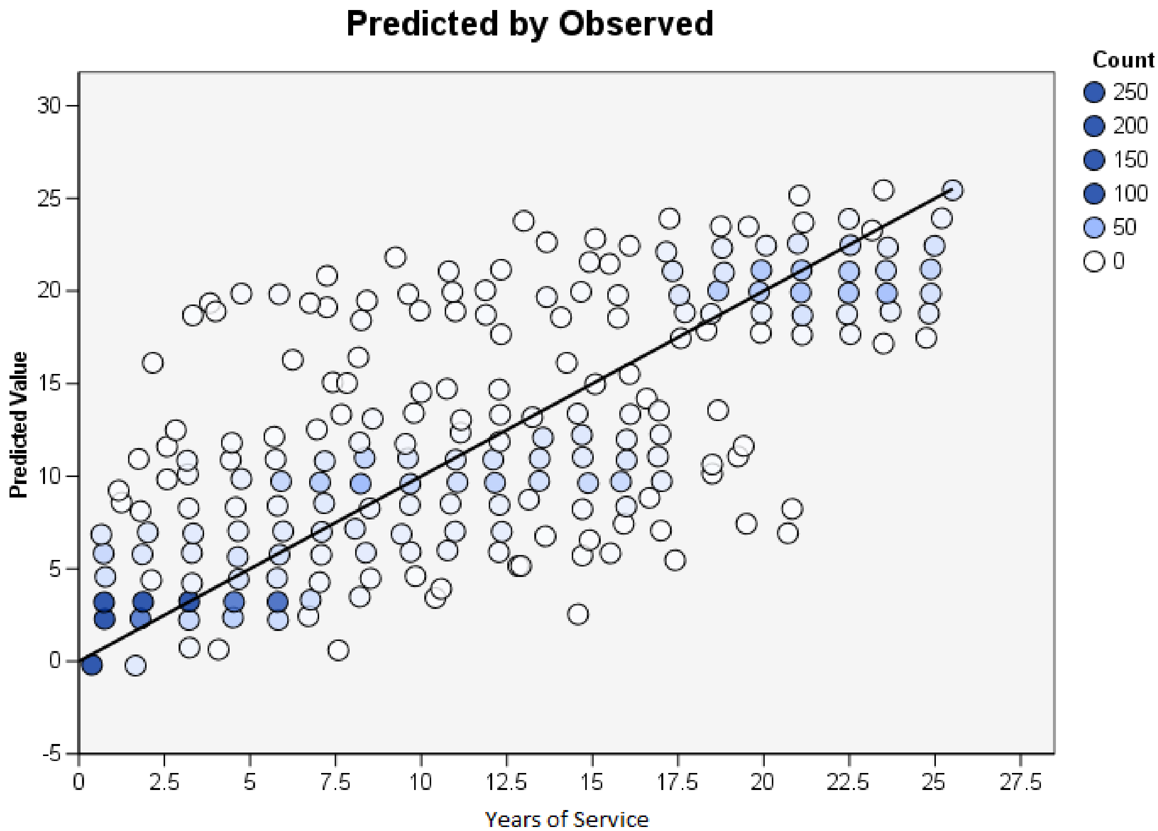

Figure 5: Linear Regression Difference Between Predicted Versus Observed Values

Machine Learning Algorithms

Next, we analyze the suite of ML algorithms and their predictive power for total YOS. Table 4 lists the five best fitting models, their correlation coefficient, and the relative error; individual model output is suppressed for brevity. ML models that typically perform well with structured data sets and a clearly defined dependent variable, in our case total YOS, unsurprisingly perform well in our study. Standard CART and random forest algorithms, as well as the sophisticated extreme gradient boosted tree algorithm, are capable of classification and prediction of YOS with a high degree of accuracy. CHAID and neural net algorithms also provide high correlation coefficients and low relative error. All five of the best fitting models provide high predictive power compared to traditional linear regression models.

Figure 6 ranks the predictor importance of the independent variables, aggregated across all ML algorithms. We note that, just as before, prior military experience is the highest predictor of future service. In fact, the top predictors of total YOS between the suite of ML algorithms and the simple linear regression estimation are the same. We see that professional experience, having been ranked deficient, white-collar positions, and technical positions strongly predict total YOS for an acquisition worker.

Perhaps what is most surprising is the degree of accuracy with which ML algorithms are able to predict YOS given the relatively small data set. All top five models have correlation coefficients >0.9 in the prediction set. This represents the biggest contrast between traditional linear regression and ML algorithms—the strength of correlation and the predictive power of ML algorithms unsurprisingly surpasses linear models using the same data set. With a relatively small set of predictive variables, and over a relatively short time horizon, ML algorithms are able to predict attrition to a high degree of accuracy for acquisition workers. In comparison, the traditional econometric model using linear regression has difficulty capturing the nonlinearities in our data (see Figure 5). That is not to say that ML algorithms are always superior—no argument for causality can be made when utilizing these models. They are another instrument, a potentially powerful one, to improve analysis of past trends in manpower. With sufficient data, decision-makers can glean insights from past policies' successes and failures to improve future policy guidance.

Table 4: Best-Fitting ML Models by Correlation Coefficient

Model	Correlation Coefficient	Relative Error
Extreme Gradient Boosted Tree	0.917	0.16
CHAID	0.915	0.163
CART	0.914	0.164
Random Forest	0.913	0.167
Neural Net	0.911	0.171

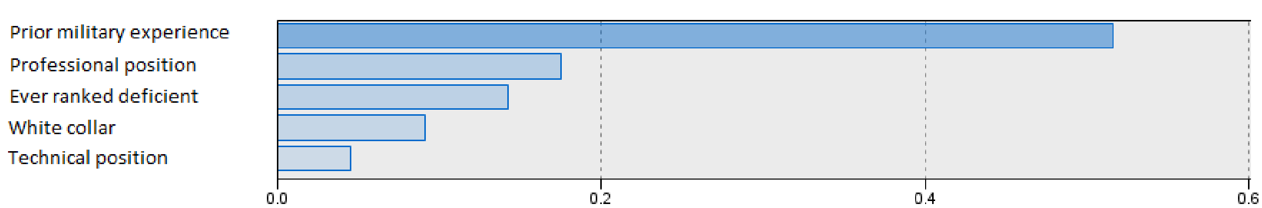

Figure 6: Predictor Importance in ML Algorithms

Conclusions

Identifying workers who are likely to attrit is an integral part of talent management. Without being able to predict those at higher risk of attriting, effective strategies to make the firm or organization a more attractive place to work cannot be explored. In an age of increased competition across firms and government organizations for human capital, retaining the best and the brightest has become a top priority.

This study is a first step toward evaluating whether ML algorithms can be leveraged to help AWF decision-makers predict with higher accuracy the behavior of their workforce and more precisely pinpoint workers who are most likely to leave the AWF.

We initially showed that using ordinary least squares (OLS) leads to poor forecasting. While OLS has been the workhorse of statistical inference for all manners of data analysis, clearly it is not up to the task of predicting which workers may leave the AWF. We then explored whether using different ML algorithms yielded better results. We found that, despite the parsimonious set of independent variables, several ML algorithms can predict attrition among acquisition workers with a high degree of accuracy. Given that ML algorithmic effectiveness scales with data set size, our results suggest many use cases for these algorithms in future predictive modeling for manpower and retention.

While ML algorithms have demonstrated their potential, we must call attention to a serious trade-off. While OLS usually makes explicit what factors impact the probability of attrition by how much, ML algorithms are very much a black box. The algorithms work to maximize the correlation between the outcome (attrition) and the inputs. However, the statistical relationship they estimate is entirely a-theoretic. As such, no causal interpretation can be assigned. This means that while the algorithm may be helpful for identifying individuals at risk for attrition, it holds no insight in how to dissuade workers from attriting or how to design or change policy to encourage a longer career. For example, even if an algorithm tells us that prior-enlisted experience is a salient predictor of longevity in the AWF, we should not alter recruiting and hiring strategies to favor workers with prior-enlisted experience in the hopes of getting the average employee to stay longer.

Decision-makers should consider ML algorithms as a useful tool to predict worker behavior yet be aware of its inherent limitations.

References

Ahn, T. & Menichini, A. (2021). Optimal Long-Run Talent Management of the Department of Defense Acquisition Workforce in Response to COVID-19: A Dynamic Programming Approach. *Defense Acquisition Research Journal*, 29(1), 50–77.

Ahn, T., Cole S., Fan, J., & Griffin, C. (2021). Predicting worker retention: a machine-learning approach with the Royal Australian Navy. *Military Operations Research Journal* (forthcoming).

Alpaydin, E. (2020). Introduction to machine learning. MIT press. Brien, S. (2019). Attrition among the DoD civilian workforce. (Report No. NPS-HR-20-004) Acquisition Research Program

Defense Manpower Data Center. (2022). [Unpublished raw data on civilian DoD employees]. Department of Defense.

Delen, D. (2010). A comparative analysis of machine learning techniques for student retention management. *Decision Support Systems*, *49*(4), 798–506. https://doi.org/10.1016/j.dss.2010.06.003

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, *29*(5), 1189–232. https://doi.org/10.1214/aos/1013203451

Mather, P., & Mahesh, P. (2003). An assessment of the effectiveness of decision tree methods for land cover classification. *Remote Sensing of Environment*, *86*(4), 554–565. https://doi.org/10.1016/S0034-4257(03)00132-9

Pickar, C., Franck, R., Hildebrandt, G. & Udis B. (2019). In the fullness of time: Towards realistic acquisition schedule estimates. (Report No. NPS-AM-19-016) Acquisition Research Program

Kalmegh, S. (2015). Analysis of WEKA data mining algorithm REP tree, simple cart and random tree for classification of Indian news. *International Journal of Innovative Science, Engineering, & Technology*, *2*(2), 438–446. https://ijiset.com/vol2/v2s2/IJISET_V2_I2_63.pdf

Díaz-Pérez, F. & Bethencourt, C. (2016). CHAID algorithm as an appropriate analytical method for tourism market segmentation. *Journal of Destination Marketing & Management*, *5*(3), 275–282. https://doi.org/10.1016/j.jdmm.2016.01.006

Terrazas, G. (2020) Evaluation of machine learning applicability for USMC reenlistment. [Master's thesis, Naval Postgraduate School]. NPS Archive: Calhoun. https://calhoun.nps.edu/handle/10945/64888.

Under Secretary of Defense for Acquisition, Technology, and Logistics. (2015). *Department of Defense acquisition workforce strategic plan: FY2016–FY2021.* *https://www.hci.mil/docs/DoD_Acq_Workforce_Strat_Plan_FY16_FY21.pdf*

Under Secretary of Defense for Acquisition, Technology, and Logistics. (2019). Department of Defense acquisition workforce strategic plan: FY2019–FY2024. https://www.secnav.navy.mil/rda/workforce/Documents/StrategyPolicy/AWF_Strategic_Plan_-_24_April_2019.pdf

Wagstaff, K., Cardie, C., Rogers, S., & Schroedl, S. (2001). Constrained K-means clustering with background knowledge. *Conference: Proceedings of the Eighteenth International Conference on Machine Learning*. https://www.cs.cmu.edu/~./dgovinda/pdf/icml-2001.pdf

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

www.acquisitionresearch.net