INSTITUTE FOR DEFENSE ANALYSES

A New Military Retention Prediction Model: Machine Learning for High-Fidelity Forecasting

Julie Pechacek
Alan Gelder
Cullen Roberts
Joe King
James Bishop
Michael Guggisberg
Yev Kirpichevsky

June 2019
Approved for public release;
distribution is unlimited.
IDA Document NS D-10712
Log: H 19-000305

INSTITUTE FOR DEFENSE ANALYSES
4850 Mark Center Drive
Alexandria, Virginia 22311-1882

About This Publication

This work was conducted by the Institute for Defense Analyses (IDA) under contract HQ0034-14-D-0001, project BE-6-4311, "Next Generation Personnel Management Models and Modeling Environment," for the Office Under Secretary of Defense for Personnel and Readiness (OUSD (P&R)). The views, opinions, and findings should not be construed as representing the official position of either the Department of Defense, sponsoring organization, or the U.S. Government.

For More Information:

Dr. Julie A. Pechacek, Project Leader, SFRD
jpechace@ida.org, 703-578-2858
ADM John C. Harvey, Jr., USN (Ret), Director, SFRD
jharvey@ida.org, 703-575-4530

Copyright Notice

INSTITUTE FOR DEFENSE ANALYSES

IDA Document NS D-10712

A New Military Retention Prediction Model: Machine Learning for High-Fidelity Forecasting

Julie Pechacek
Alan Gelder
Cullen Roberts
Joe King
James Bishop
Michael Guggisberg
Yev Kirpichevsky

This page is intentionally blank.

Executive Summary

Improved forecasts of military personnel retention can assist Department of Defense (DOD) leaders and force managers on multiple levels. Developing a force that is lethal, efficient, and ready requires that leaders anticipate upcoming changes in the size and shape of the military workforce at a detailed level. To support the Office of the Under Secretary of Defense for Personnel and Readiness, IDA has developed the Retention Prediction Model (RPM). The RPM uses machine learning algorithms and extensive personnel records to capture rich interactions in service characteristics and predict when individual servicemembers will separate from the military. The RPM's person-level predictions can be aggregated by any desired population subset, including career field, cohort, unit, or demographics.

The RPM currently incorporates monthly records for active duty personnel between 2000 and 2018. This population encompasses approximately 4.5 million unique individuals and more than 600 administrative fields, covering career history, family, pay, and deployments. To facilitate and speed model training, a 5% sample of all individuals serving between 2000 and 2018 was split into two sub-samples. The first sub-sample, comprising 75% of the sample (approximately 169,000 individuals), was used to train the RPM. The remaining 25% of the sample was used for testing. Based on information about a service member's career and characteristics observable up to at a given point, the RPM estimates the probability that a person will continue to serve for any number of future periods. The RPM uses a survival loss function developed specifically for analytic applications where the end state in a chain of events is not observable or has not yet occurred. Categorical variables were encoded using an embedding layer to determine a mapping structure that is most useful to the predictive model.

The RPM produces individual-level predictions that closely mirror actual attrition patterns. Testing on out-of-sample data, given two randomly selected servicemembers, one of whom separates from the military within one year, the RPM identifies the correct individual 88% of the time. Extending the time horizon to four years, the model is correct 80% of the time; for any number of years up to 18, the model is correct more than 78% of the time.

Applying machine learning techniques to identify patterns in personnel data enables new insights into issues affecting military personnel retention and force planning. The DOD can leverage the RPM to anticipate shortfalls in specific occupational fields, plan for expected career lengths among a heterogeneous population, and tailor policies to retain highly sought-after personnel.

This page is intentionally blank.

A New Military Retention Prediction Model: Machine Learning for High-Fidelity Forecasting

Julie Pechacek, Principal Investigator

Alan Gelder
Cullen Roberts
Joe King
James Bishop
Michael Guggisberg
Yev Kirpichevsky

JSM 2019

The views expressed here are my own
and do not represent the official position of IDA or DOD

The Institute for Defense Analyses (IDA) conducts independent research in the public interest

- Non-profit and non-partisan
- Operates three Federally Funded Research and Development Centers (FFRDCs)
- Primary sponsor is the Department of Defense (DOD)
- Bridge between academia and DOD

The DOD seeks to better forecast and understand factors that predict the retention of military personnel

- Anticipate shortfalls in career fields
- Identify changes in expected career length among a heterogeneous population
- Provide appropriate policies to encourage retention

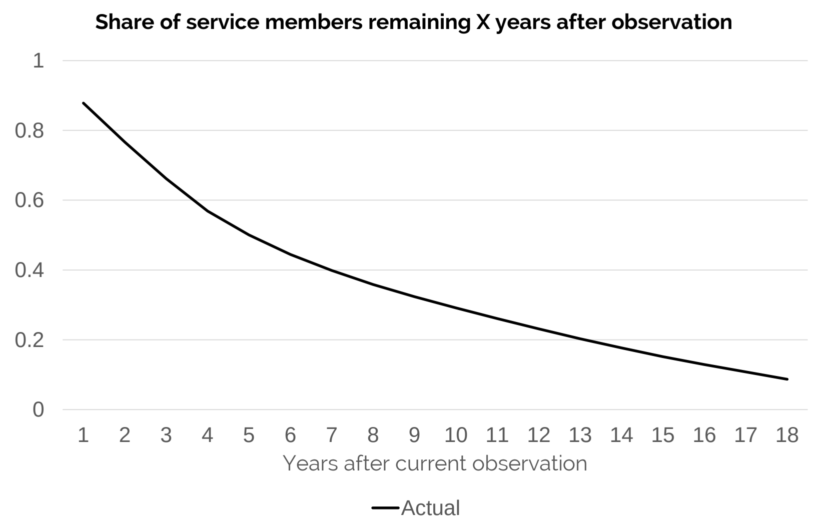
How and why do people leave?

Many leave at contract endpoints
Fixed-duration contracts often last 4 years

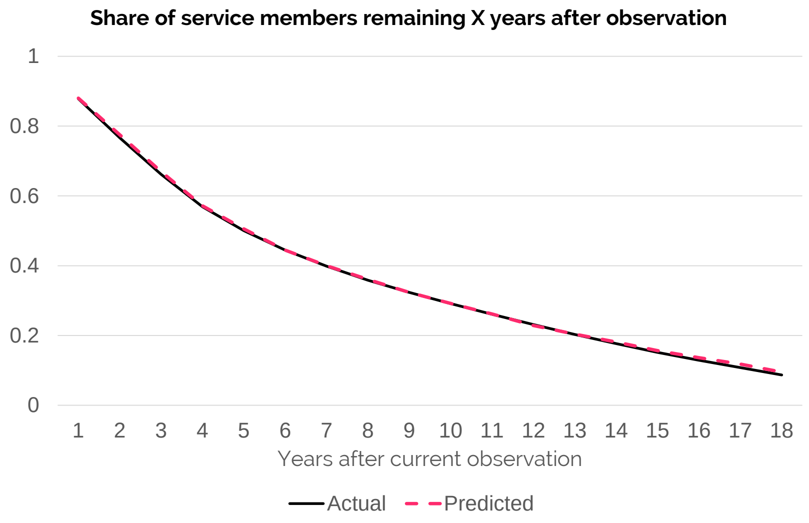
Others leave mid-contract
Poor health, discipline problems, or other reasons

We observe **entry** and **exit** by date of first and last record
Right-censoring flagged by being present at last date in data;
Entry date inferred for those present at first date in data

Observed survival function for active duty personnel



Observed survival function for active duty personnel



Out-of-sample prediction with feed forward neural network

This aggregate prediction results from millions of person-level predictions across the force

How well does the model perform for a given person?

Pick two service members at random,
one of whom will leave within **1 year**

Our model identifies the person who leaves **88%** of the time

What if the time horizon is **4 years** instead of 1 year? **80%**
For any number of years up to 18? **78+%**

Data Overview

We harness extensive DOD administrative records on active duty personnel

Data include monthly records for all active duty personnel
- 315 million person-month records, 2000–2018
- 4.5 million unique persons
- 600 fields on career history, family, pay, deployments

To reduce training time, we
- Aggregate to yearly records
- Estimate model on a 5% subset of persons
- Train on 75%, test on 25% of this subset

DOD data from multiple administrative feeds form an unbalanced, censored panel

Censoring
- 31% of persons in service at first data month
- 30% of persons in service at last data month
- 42% of persons uncensored
- 3% in service in both first and last data month

We construct additional fields to highlight potentially salient career and life experiences, such as …

Assigned unit
- Characteristics (e.g., size, mean Armed Forces Qualification Test (AFQT) score)
- Similarity to subject person
- Proximity to subject person's home of record

Features of service
- Days until end of current contract
- Total days deployed
- Deployment features

Family Composition
- Oldest and youngest dependents
- Total dependents in various age ranges

Modeling

We seek to estimate the probability that a person continues to serve for any number of future periods

Input: information on a person's career at a given point

Output: survival probability that a person will remain in service for each successive period

Survival loss function accounts for right-censoring

Discrete survival framework

Likelihood	Censoring	Formulation
Hazard function		$h(t)$= Prob(exit at t | survival to $t-1$)
Exit at time j	Uncensored	$h(t)\prod_{t=1}^{j-1}(1-h(t))$
Survival until at least j	Censored after $j-1$	$\prod_{t=1}^{j-1}(1-h(t))$

Output is a vector of survival probabilities $(1 - h(t))$
extending out to the maximal observed survival time

Loss function implementation based on Gensheimer and Narasimhan (2019)

Machine learning models require real-valued predictors and other data transformations

Categorical fields

We have hundreds, many with hundreds of unique values

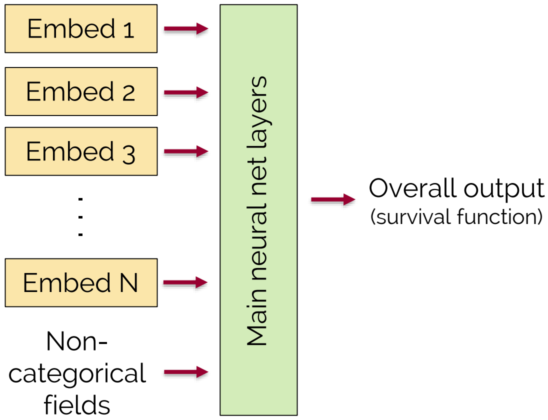
Transformation (k unique values)	Number of resulting fields
One-hot encoding	k (Boolean)
Binary encoding	$\log_2$k (Boolean)
Target mean encoding	1 (group mean outcome)
Neural network embedding	1 or more
Other (e.g., PCA, autoencoders)	1 or more

Numeric fields

Fields with a low number of unique values are treated as categorical
Others are rescaled to the unit interval [-0.5, 0.5]

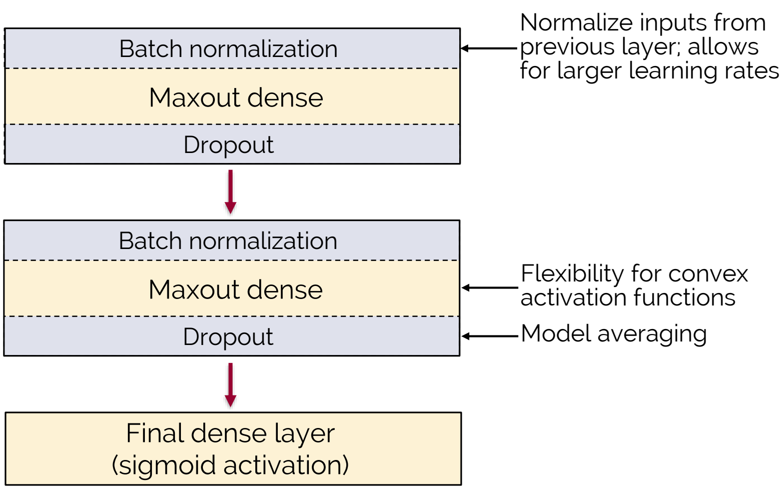
Transform each categorical field via embedding layers

Store for use in subsequent models



Model rerun after embeddings are calculated (“freezing the embeddings”)
L2 regularization applied to output of each embedding layer
Adam optimizer (AMSGrad variant, see Reddi et al. 2018)

Main neural network layers consider all possible interactions while reducing overfitting

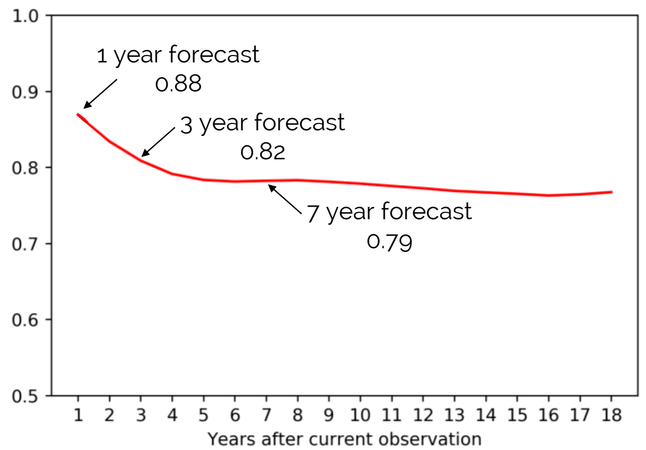


Batch normalization (Loffe and Szegedy 2015); Maxout dense (Goodfellow et al. 2013)

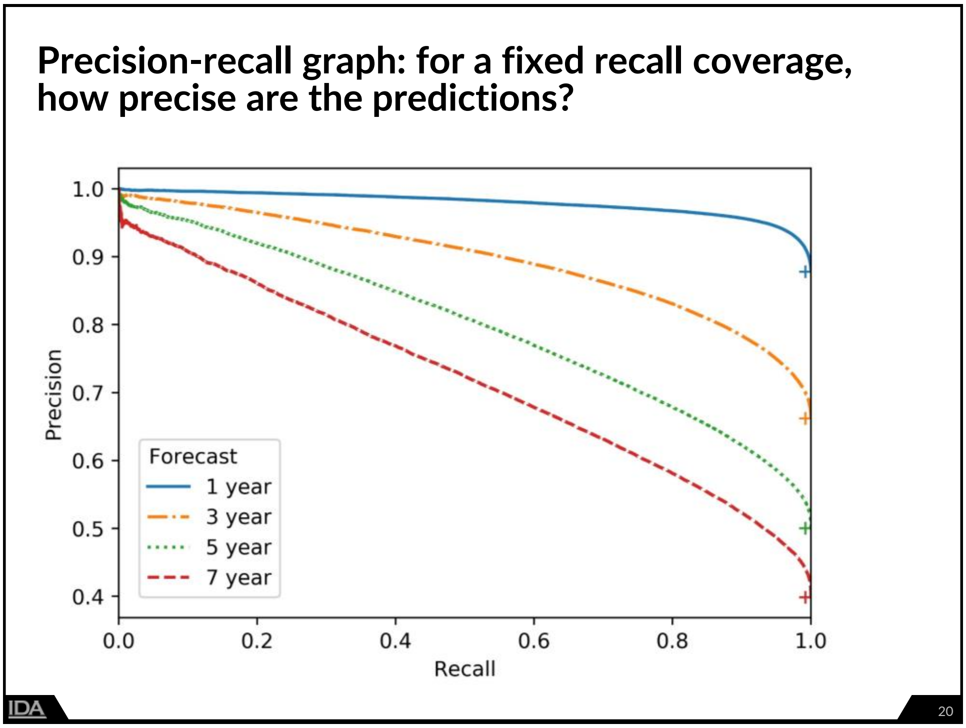
Results

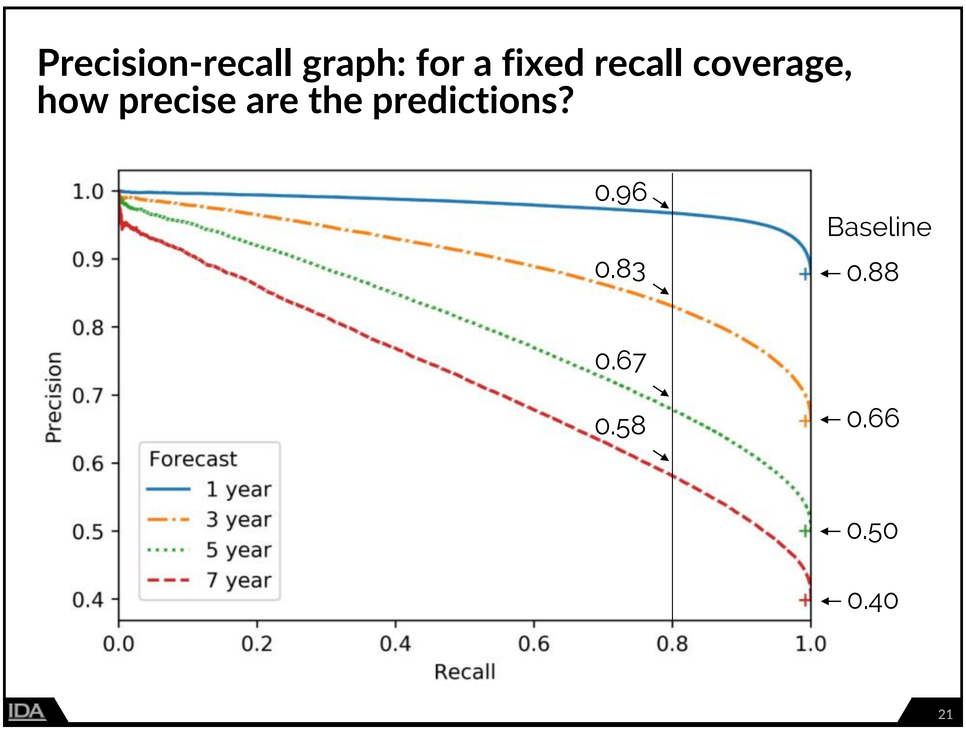
Area Under Receiver Operator Characteristic (AUROC)

Probability that a random draw from the set of actual positives will **rank** higher than a random draw from the set of actual negatives



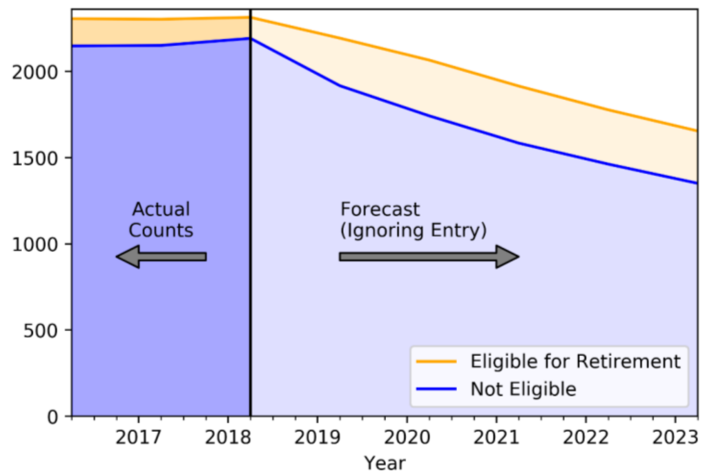
Measured on out-of-sample observations


Precision-recall graph: for a fixed recall coverage, how precise are the predictions?
Precision
Recall
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.0
0.2
0.4
0.6
0.8
1.0
Forecast
1 year
3 year
5 year
7 year
IDA
20

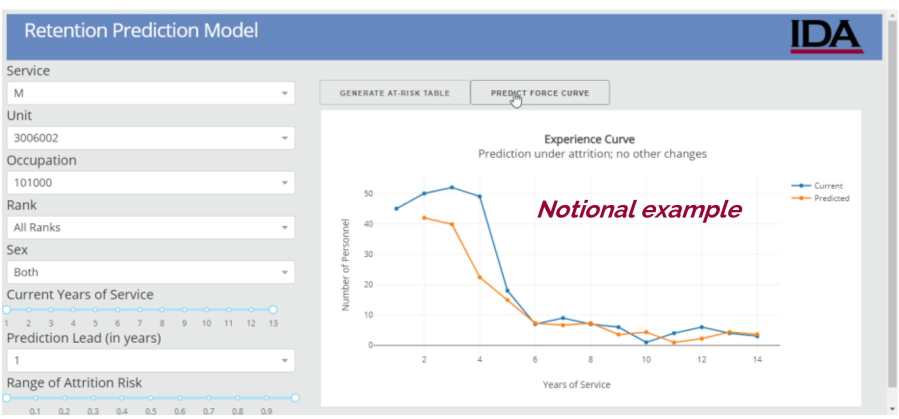

Precision-recall graph: for a fixed recall coverage, how precise are the predictions?
Precision
Recall
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.0
0.2
0.4
0.6
0.8
1.0
0.96
0.83
0.67
0.58
Baseline
0.88
0.66
0.50
0.40
Forecast
1 year
3 year
5 year
7 year
IDA
21

Person-level predictions can be aggregated by career field, cohort, unit, demographics, and so forth

Estimates for Air Force fighter pilots (11F)



A graphical interface enables DOD leaders to visualize these predictions for custom groups



For a given set of drill-down characteristics, plot the current and predicted force curve by years of experience

This capability can predict events for personnel, units, equipment, and operations

Examples: **accurately predicting …**

whether a potential recruit will succeed in basic training

which applicants will likely pass special forces evaluation

when an equipment item will fail

The inputs are flexible, only requiring panel data with

time and person (or object) identifiers

regularly spaced time intervals

Questions or Comments?

Feel free to contact Julie Pechacek
jpechace@ida.org

Thank You

Backup

We examined four classes of models

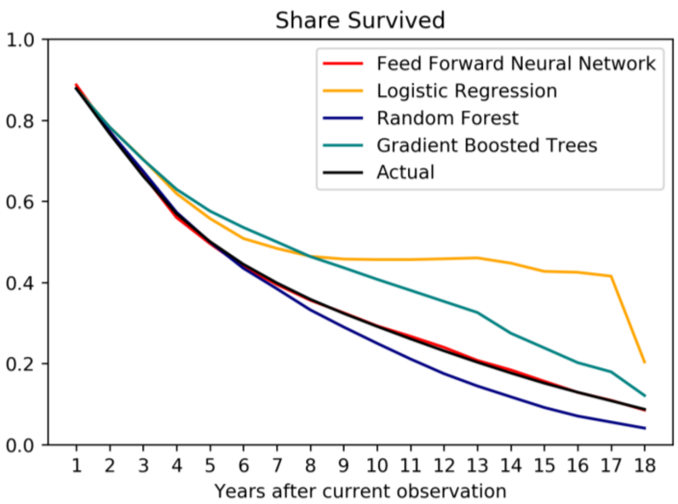
Feed-forward Neural Network
    Two 256-node Maxout layers, each with 25% dropout

Gradient Boosted Trees
    64 trees, each at most 8 branches deep

Random Forest
    128 trees, each leaf with at least 256 observations

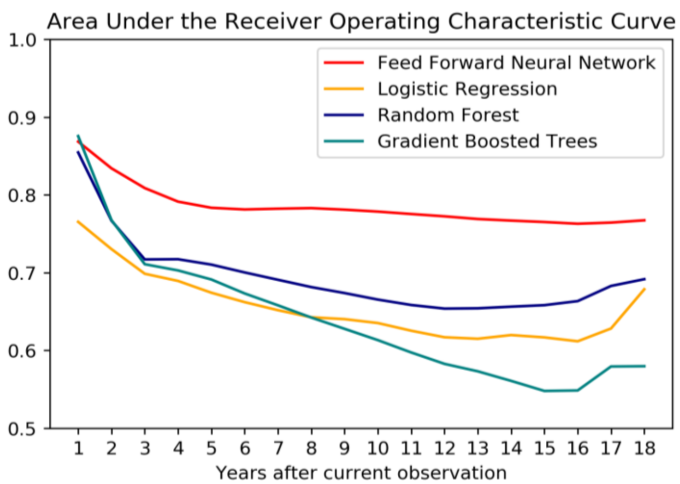
Logistic Regression
    L2 regularization of 0.125

Neural network predictions best match actual aggregate attrition patterns



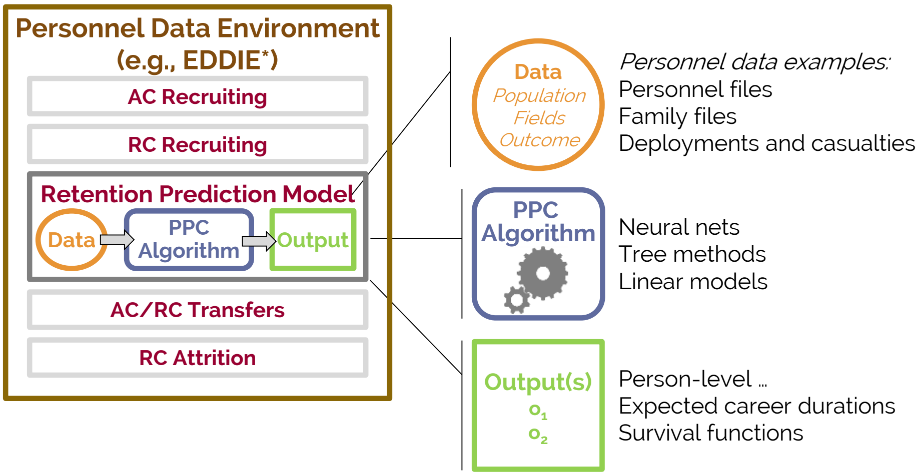
Measured on out-of-sample observations

AUROC: Comparing model classes



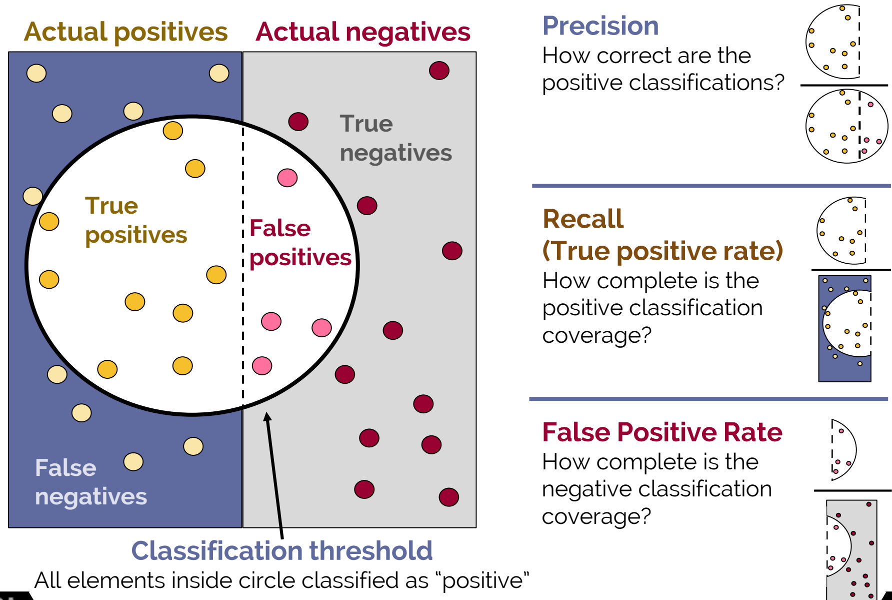
Measured on out-of-sample observations

The Persistence Prediction Capability and supporting data have been designed for efficient reuse



*Enterprise Data to Decisions Information Environment

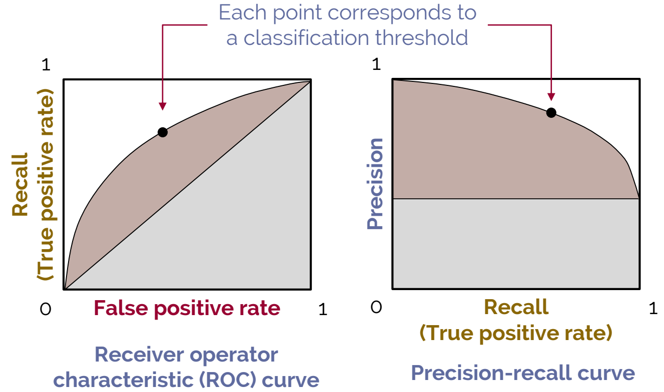
Metrics primer: precision, recall, false positive rate



Metrics primer: changing classification thresholds

Our output is a set of survival probabilities
Binary stay-exit predictions require a classification threshold
Example: stay if survival probability > 0.5; exit otherwise

How does the model perform as we vary the threshold?

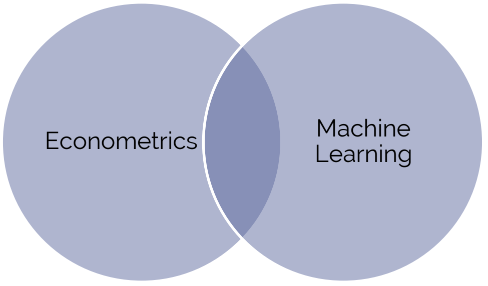


Summary of missing data

Missing data values
We remove the 30% of fields that are more than 99.9% missing
54% of the remaining fields are at least 50% missing
Missing values filled with person's previous non-missing value

Future work will allow for more accurate predictions, measures of uncertainty, and counter-factual studies

Bayesian optimization of hyper-parameters
Bootstrapping to estimate sampling uncertainty
Address prediction interactions with legally protected fields
Move toward counter-factual studies



REPORT DOCUMENTATION PAGE	Form Approved OMB No. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From – To)
XX-06-2019	Final	

4. TITLE AND SUBTITLE

A New Military Retention Prediction Model: Machine Learning for High-Fidelity Forecasting

5a. CONTRACT NO.
HQ0034-14-D-0001

5b. GRANT NO.

5c. PROGRAM ELEMENT NO(S).

6. AUTHOR(S)

Julie Pechacek, Alan Gelder, Cullen Roberts, Joe King, James Bishop, Michael Guggisberg, Yev Kirpichevsky

5d. PROJECT NO.
BE-6-4311

5e. TASK NO.

5f. WORK UNIT NO.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Institute for Defense Analyses
Strategy, Forces and Resources Division
4850 Mark Center Drive
Alexandria, VA 22311-1882

8. PERFORMING ORGANIZATION REPORT NO.
IDA Document NS D-10712
Log: H 19-000305

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
OUSD (P&R)
Pentagon

10. SPONSOR'S / MONITOR'S ACRONYM(S)
OUSD (P&R)

11. SPONSOR'S / MONITOR'S REPORT NO(S).

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT
Using machine learning algorithms and 18 years of data, we predict individual-level attrition among active duty personnel in all military Services, with hold-out sample prediction accuracies typically exceeding 70%. Importantly, our methodology accommodates both right and left-censoring of observed career paths, and significantly outperforms traditional survival analysis. Using these individual-level predictions, we generate aggregate predicted force profiles which closely align with historical actuals. This and other features offer a rich slate of observations for further empirical analysis, and suggest new policy levers for managing attrition.

15. SUBJECT TERMS
Retention, attrition, machine learning, forecast, military personnel

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NO. OF PAGES	19a. NAME OF RESPONSIBLE PERSON David M. Percich
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER (Include Area Code)
U	U	U	UU	26	(703) 693-2238

This page is intentionally blank.