Retention Prediction Model - Army

WEAI 2023
Defense Sessions

Alan Gelder
Jordan Marcusse
Ed Wang
Erin Eifert
Vivian Gao
The Institute for Defense Analyses is a nonprofit corporation that operates three Federally Funded Research and Development Centers. Its mission is to answer the most challenging U.S. security and science policy questions with objective analysis, leveraging extraordinary scientific, technical, and analytic expertise.

About this Publication
The work was conducted by the Institute for Defense Analyses (IDA) under CRP C6627. The views, opinions, and findings should not be construed as representing the official position of either the Department of Defense or the sponsoring organization.

Acknowledgments
The authors thank Joe Coombs for expert editing; and Amberlee Mabe-Stanberry for proficiently producing this document.

For More Information:
Dr. Jordan R. Marcusse, Leader Author
jmarcuss@ida.org, 703-845-2147
Ms. Jessica L. Stewart, Director, SFRD
jstewart@ida.org, 703-575-4530

Copyright Notice
© 2023 Institute for Defense Analyses
730 E.Glebe Rd
Alexandria, VA 22305 • (703) 845-2000

This material may be reproduced by or for the U.S. Government pursuant to the copyright license under the clause at DFARS 252.227-7013 (Feb. 2014).
Retention Prediction Model - Army

WEAI 2023
Defense Sessions

Alan Gelder
Jordan Marcusse
Ed Wang
Erin Eifert
Vivian Gao
This page is intentionally blank.
Executive Summary

This presentation introduces the Retention Prediction Model – Army (RPM-A), a machine learning tool that forecasts individual soldier retention. The Institute for Defense Analyses (IDA) originally developed a version of the Retention Prediction Model in 2018 as part of a project for the Office of the Under Secretary of Defense for Personnel and Readiness, Military Personnel Policy. IDA is now delivering this model within the Army’s Person-Event Data Environment (PDE) along with code to facilitate updates to the data, model, and forecasts, as well as a dashboard to support the model’s use by Army personnel.

Among the sources of information that inform the model are demographics, family, career and pay, unit characteristics, casualties, deployments, the external job market, and performance data.

In this presentation, we discuss the methodology, the data inputs, and the tools we produced as part of this effort.
This page is intentionally blank.
The Retention Prediction Model - Army (RPM-A)

**Goal:** Forecast future retention and build tools for Army leaders to leverage the forecasts

**Approach:** Use machine learning toolkit to forecast retention for individual service members
- Stand up IDA’s in-house personnel data pipeline and expand it to current and historical Army personnel data within the cloud version of Army’s Person-Event Data Environment (PDE)
- Train a machine learning model on these data to produce retention forecasts for every member of the Army
- Host output forecasts in a dashboard for Army leaders

**Example use case:** Identify likely future shortfalls within populations such as occupation, rank, and performance.

**Status:** IDA has stood up the RPM-A for officers in PDE, including forecasts and a dashboard, as well as a model for the full force trained on a more limited set of data
Planned topics for today’s discussion

The RPM-A and underlying analytical methods
Hosting RPM-A within Army’s environment
Analysis of data inputs
Model performance

The RPM delivers person-level retention forecasts

OUSD(P&R) seeded the development of this flexible toolkit for retention analysis

Research applications for Services and OSD

Army: Full modeling pipeline in Army systems
ANG: Forecast and program for training slots
Navy: Identify correlates of officer exit & promotion
P&R: Examine Academy graduate ADSOs*
USU: Measure return on USU physician training
QRMC: Estimate effects of past compensation

Partnerships expand FIFE & RPM capabilities

Operationalize in shared/service environments - PDE
Link forecasts to characteristics of interest
Enhance and expand data development

*Finite Interval Forecasting Engine  **ADSO: Active Duty Service Obligation
FIFE estimates retention with machine learning, fitting a binary prediction model to each future time horizon

Input: Imbalanced panel data on individuals
Training data: \( X = \) Individual data at each time period
\( Y = \) Retention 1 period ahead, Retention 2 periods ahead, ...

FIFE’s methodology is more flexible than prevailing techniques

Traditional tools for survival analysis

Kaplan-Meier: \( H(t) \)—share surviving to time horizon \( t \)

Proportional Hazards: \( H(t) f(x) \)—now a function of feature values

Our method: \( H(t, x) \)—allows interactions with time and features
We effectively compute \( f_j(x) \) for each forecast horizon, where
\[
f_j(x) = P(\text{Remain from } t-1 \text{ to } t \mid \text{Remained in sample from } 0 \text{ to } t-1, x)
\]
\[
H(t, x) = f_1(x) f_2(x) \cdots f_j(x)
\]
Classification method - choice of $f(x)$

RPM-A uses gradient-boosted trees (via LightGBM)

<table>
<thead>
<tr>
<th>Benefits</th>
<th>Drawbacks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Best performance among classifiers</td>
<td>Cannot perform mathematical operations across features</td>
</tr>
<tr>
<td>Relatively stable performance across hyperparameter specifications and data inputs</td>
<td>$(x - y &gt; a)$, time trends, ...</td>
</tr>
<tr>
<td>Fast computation</td>
<td>Feature engineering helps to resolve some of these issues</td>
</tr>
</tbody>
</table>

Alternative classifiers available in FIFE

- Feed-forward neural network (via Keras)
- Proportional hazards (via a constrained neural network)
- Group rates (via pandas groupby; tantamount to fixed effects)

Hosting RPM-A within Army’s environment
The RPM-A in Army’s Person-Event Data Environment

### Data inputs to RPM-A

- Demographics
- Dependents
- Career and pay
- Unit traits
- Casualty
- Deployments
- External job market
- Loss categories
- Performance
- Rater characteristics
- Fitness
- Drug testing

**Bolded** data inputs are Army-specific.

### Operationalizing the RPM-A puts the full pipeline in Army’s hands

Data preparation, modeling, and dashboard reside in the Person-Event Data Environment (PDE)

Army can use the retention forecasts within a business intelligence system

Army can directly control access to and applications of RPM-A outputs

---

#### We streamlined the data-modeling-analysis pipeline

1. Query raw data
   - Access data from Army, DoD, and other sources

2. Clean data
   - Standardize missing values
   - Map recoded values
   - Merge across reformat
   - Other cleaning processes

3. Engineer features
   - Merge tables
   - Store in quickly read, column-based format (.feather)

4. Merge datasets

**Examples**

- Dates → Time since/until event
  - Officer has 748 days until her ADSO date

- Past events → Event histories
  - Officer’s most recent senior rater evaluation was “Most Qualified”
  - Officer was deployed for 271 days in the past 12 months

- Descriptions of dependents → Family structure
  - Officer has 2 children under the age of 6

- Status → Change in status
  - Officer got divorced in the past year

- Aggregate data from other units of observation
  - Unit/Rater/Military Occupational Specialties (MOS) characteristics
Model training

With this efficient data pipeline, we can train versions of the model during scheduled high-compute days.

We assess model performance and feature importance across a variety of specifications, such as:

- Feature inputs
- Training period
- Forecast date

To train a new model: specify a population, feature inputs, time period, and hyperparameters.

We train a set of predefined models and store their forecasts as quarterly data updates arrive:

- Administrative RPM-A
- Research RPM-A (includes data with restrictions on use or limited dates)

The RPM-A dashboard supports analysis and planning.
Dashboard uses: Aggregate forecasts over user-defined populations

Dashboard uses: Identify data elements coincident with retention/exit
Dashboard uses: Evaluate performance of forecasts from prior periods against actual outcomes

Considerations for applications of the RPM-A

The RPM-A forecasts assume that patterns in the past will continue

The RPM-A cannot identify the causal impact of data elements

Using forecasts for individual-level career management is risky
- Accuracy may be insufficient for person-level actions
- Even with perfect accuracy, there are concerns of ethics, fairness, creation of unintended incentives, and other matters

The RPM-A’s retention probabilities can supplement other analyses
Model Performance: All officers, 3 years

Predictions of retention from September 2019 to September 2022:

**DMDC and economic data, or all data**

With a prediction threshold of 50% for assignment to remain/stay\(^1\)

RetentionPolicy: 75.4%

<table>
<thead>
<tr>
<th>Forecast</th>
<th>Outcome</th>
<th>Exit</th>
<th>Stay</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exit</td>
<td>7,217</td>
<td>2,300</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7,495</td>
<td>2,125</td>
<td></td>
</tr>
<tr>
<td>Stay</td>
<td>12,111</td>
<td>56,797</td>
<td></td>
</tr>
<tr>
<td></td>
<td>11,833</td>
<td>56,972</td>
<td></td>
</tr>
</tbody>
</table>

**Statistics\(^2\)**

<table>
<thead>
<tr>
<th>Metric</th>
<th>Current</th>
<th>Baseline</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>81.6%</td>
<td>82.2%</td>
</tr>
<tr>
<td>Precision</td>
<td>75.8%</td>
<td>77.9%</td>
</tr>
<tr>
<td>Recall</td>
<td>37.3%</td>
<td>38.8%</td>
</tr>
<tr>
<td>AUROC</td>
<td>0.812</td>
<td>0.824</td>
</tr>
</tbody>
</table>

\(^1\) Other thresholds trade off false positives and false negatives, improving precision or recall, and may be more appropriate for specific use cases

\(^2\) Accuracy: rate of actual exits & stays being identified; precision: rate of forecasted exits being correct; recall: rate of actual exits being identified

AUROC: Area Under the Receiver Operating Characteristic
Select informative features by information gain

**1-4 quarters from forecast date**
- Days until projected end of service
- Assigned UIC (Unit Identification Code)
- Active Service Loss Incentive status
- Duty UIC
- Strength accounting code
- Years of active federal military service
- Days deployed in career
- Date
- Pay status
- Days spent in pay grade

**9-12 quarters from forecast date**
- Days until projected end of service
- Duty UIC
- Years of active federal military service
- Pay grade
- Occupation code (MOS)
- ZIP code of home address
- Quarter
- Days spent in pay grade

**Army-specific data elements in top 10-100 features:**
- Most recent Officer Evaluation Report (OER) was referred
- Ever had a referred OER
- Reason not rated on most recent OER
- Box check on most recent OER
- Percent of box checks “Most qualified” on OER
- Ever had a profile on OER
- Fitness score percentile
- 2 mile run percentile

Model performance, input importance vary over time

Deployment activity was an important predictor from 2008-2016, but is much less of a predictor now

Model performance of RPM-A trained on DMDC (Defense Manpower Data Center) + fitness data for all officers, trained on rolling 4-year windows of data preceding the forecast date
FIFE is available in the open source

https://pypi.org/project/fife/

https://fife.readthedocs.io/

https://github.com/IDA-HumanCapital/fife

What’s in the FIFE package?

Panel Data Processor
Computes survival durations, identifies censorship, drops degenerate and duplicate features, and identifies training/validation sets

Survival Modelers
Can select from gradient-boosted trees (via LightGBM), feed-forward neural network (via Keras), proportional hazards, or group rates

State Modelers
Computes the future value of a feature conditional on survival

Exit Modelers
Computes competing risk of exit under various conditions

Feature Importance Attribution
Identifies the change in predictive power using SHAP analysis
Model Performance: All officers, 1 year

Predictions of retention from September 2019 to September 2020:

DMDC and economic data, or all data

With a prediction threshold of 50% for assignment to remain/stay\(^1\)

Retention rate: 92.5%

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Exit</th>
<th>Stay</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forecast</td>
<td>3,455</td>
<td>336</td>
</tr>
<tr>
<td>Exit</td>
<td>3,533</td>
<td>400</td>
</tr>
<tr>
<td>Stay</td>
<td>2,545</td>
<td>72,059</td>
</tr>
<tr>
<td></td>
<td>2,467</td>
<td>72,025</td>
</tr>
</tbody>
</table>

**Statistics\(^2\)**

<table>
<thead>
<tr>
<th></th>
<th>Accuracy</th>
<th>Precision</th>
<th>Recall</th>
<th>AUROC</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\text{forecast} = \text{outcome})</td>
<td>96.2%</td>
<td>90.4%</td>
<td>57.6%</td>
<td>0.925</td>
</tr>
<tr>
<td>(\text{all observations})</td>
<td>96.3%</td>
<td>89.8%</td>
<td>58.9%</td>
<td>0.933</td>
</tr>
</tbody>
</table>

\(^1\) Other thresholds trade off false positives and false negatives, improving precision or recall, and may be more appropriate for specific use cases

\(^2\) Accuracy: rate of actual exits & stays being identified; precision: rate of forecasted exits being correct; recall: rate of actual exits being identified

AUROC: Area Under the Receiver Operating Characteristic

**Frequency of exit types**

Exit counts by exit type

Within 2 years of Q2 of the listed year

- Other
- Retirement
- Performance/misconduct
- Involuntary release
- Early release
- Disability/Death
- Pregnancy/parenthood
- End of ADSO
Mean retention forecast by types of subsequent exit

2 year-ahead retention forecasts for officers who exited within 2 years of the forecast date

Mean retention forecast by exit type

- Disability/Death
- Performance/misconduct
- Early release
- End of ADSO
- Retirement

Low mean retention probabilities mean that the model does relatively well at predicting that exit type.

A core set of features capture a large amount of the predictive power over a 1-year time horizon

Population: All officers
Forecast date: December 2019
Retention to: December 2020
Training data: March 2000-September 2019

<table>
<thead>
<tr>
<th>Accuracy</th>
<th>AUROC</th>
</tr>
</thead>
<tbody>
<tr>
<td>DMDC and econ data</td>
<td>96.6%</td>
</tr>
<tr>
<td>28 selected fields</td>
<td>94.8%</td>
</tr>
</tbody>
</table>

Exit/Remain Classification Threshold = 0.5
Complete list of features included in Single Node RPM-A

28 Data Input Fields

Direct Indicators of Exit
- Active Duty Service Projected End Date
- Strength Accounting Code
- Personnel Strength Status Code
- Active Service Loss Incentive

Unit Information
- TOE/TDA type of Assigned Unit
- TOE/TDA type of Assigned Unit
- Assigned Unit Major Command Group
- Assigned Base
- Duty Base

Military Career
- Primary AOC
- Secondary AOC
- Duty AOC
- Years in Paygrade
- Months of Military Experience
- Current/previous command
- Component

Education/Quality
- Education Level
- Joint Professional Military Education
- Professional Military Education
- Source of Accession

Demographics
- Age
- Gender
- Race
- Ethnicity
- Source of U.S. citizenship

Return
This page is intentionally blank.