IDA | Research Summary

Machine Learning Finds Patterns That Inform Decisions

Department of Defense (DoD) leaders need accurate and actionable information to make high-quality strategic and operational decisions concerning military and civilian personnel. Machine learning (ML) techniques offer an effective tool to synthesize DoD's vast and growing stores of data on personnel assignments, deployments, bonuses, aptitude, fitness, and other features. ML algorithms provide a systematic and flexible approach to identifying complex relationships and patterns within data to gain meaningful insights for effective force management and program planning.

ML techniques produce higher fidelity and more accurate characterizations and predictions than most other available techniques. The examples below illustrate how ML can apply to personnel management questions that are difficult to address with traditional statistical tools.

Personnel management example 1: *Which military career fields are likely to have staffing shortfalls in the next 5 years? Which service members are likely to renew their contracts?* Predicting when a service member will exit the force is more complex than predicting when a component on a piece of equipment will fail. IDA's Finite-Interval Forecasting Engine (FIFE) flexibly combines the mathematics of time-to-event questions with a suite of ML algorithms. IDA's Retention Prediction Model (RPM) applies the FIFE to predict the likelihood that a service member—at any career stage—will remain in service through any future point. These high-quality predictions help DoD leaders anticipate where staffing shortfalls are likely to occur.

Personnel management example 2: *Which candidates are likely to succeed in a particular training program?* The location, time, and nature of the operational environment makes answering such

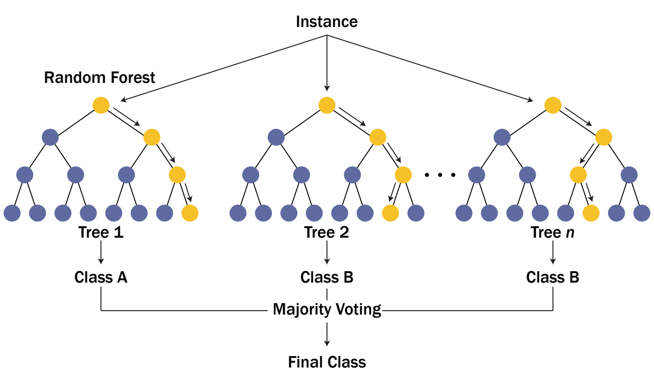
July 2021
NS D-14285

a question difficult. An ML model is more useful if it is externally valid, meaning that the model's findings generalize well when applied in environments different from that used in its development. To make best use of ML, leaders should understand that external validity can fail for at least two reasons. First, the algorithm may identify patterns in the data between inputs and outcomes that exist due to chance—a problem known as overfitting. For example, a model may identify that trainee cohorts perform worse in hot environments, when the real cause of performance decline might actually be a larger student-to-instructor ratio during summer training periods.

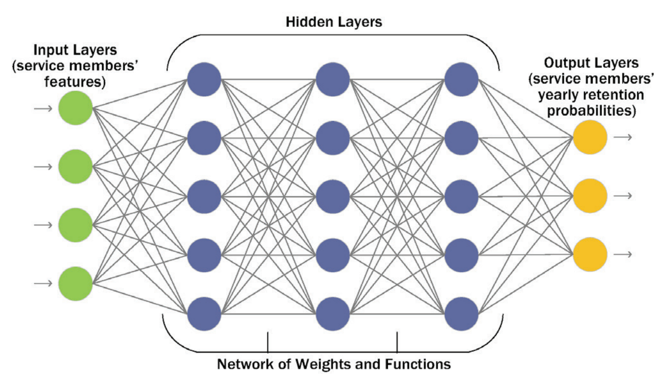
Second, a fundamental change in the environment may alter a historically valid connection between inputs and outcomes. For example, when an instruction curriculum or evaluation criteria change, the historic relationship between observed service member characteristics and training outcomes also changes. In such cases, using a "What if?" question-oriented ML analysis is more appropriate. (See IDA NS D-13236, "Using Machine Learning Effectively," for a discussion of "What if?" analyses and other analytic categories.)

Two families of ML algorithms—*tree algorithms and neural network algorithms*—provide accurate and robust predictions in complex systems. Descriptions of their mechanics follow:

- **Tree algorithms** split data into successively finer groups that share common outcomes, with individual "trees" often combined into a random forest (below left) to provide more accurate and robust predictions.
- **Neural networks** capture rich interactions within the data by using layers of neurons that are produced from preceding layers of neurons and, in turn, used as inputs to the succeeding layer (below right). The algorithm finds weights representing the relative strength of each input connection, and the final layer offers the prediction.



Simplified example of a random forest



Simplified example of a neural network

Effective use of ML to inform decisions requires: (1) an appropriate match between research questions and algorithm, (2) a careful execution of ML methodology, and (3) a proper interpretation of the results. With these requirements in place, leaders can optimize the use of ML forecasts to inform their strategic and operational decisions.



Julie A. Lockwood (formerly Pechacek), jlockwood@ida.org, leads the Human Capital Group in the Strategy, Forces and Resources Division (SFRD) of IDA's Systems and Analyses Center. For this publication, Julie led a team that included SFRD research staff members **Dave I. Cotting**, dcotting@ida.org, and **Michael R. Guggisberg**, mguggisbr@ida.org, among others.

Based on IDA P-13174, *Leveraging Machine Learning in Defense Personnel Analyses*, J. A. Lockwood, A. B. Gelder, M. S. Goldberg, J. T. Brooks, and G. M. Prugh, May 2020. Research was sponsored by the Office of the Under Secretary of Defense for Personnel and Readiness.