Proceedings of the Annual General Donald R. Keith Memorial Conference
West Point, New York, USA
April 28, 2022
A Regional Conference of the Society for Industrial and Systems Engineering

Analyzing Army Readiness Through Unsupervised Machine Learning

Brennan Coulson, Andrew Harvey, Zacharia Schalliol, William Eckland, and Jeffrey Demarest

Department of Systems Engineering
United States Military Academy
West Point, NY 10996

Corresponding author's Email: Brennan.Coulson@westpoint.edu

Author Note: This document is the culmination of several months of work by the Cadet Capstone Team, Rockbound Readiness. The group consists of Cadets Coulson, Eckland, Harvey, and Schalliol working for the Deputy Assistant Secretary of Defense for Force Readiness (DASD(FR)) as part of the Capstone program within the Department of Systems Engineering. We would like to thank the Project Owner, Mr. Jeffrey Demarest, and our key clients at DASD(FR), Ms. Kimberly Jackson, Mr. George Landis, and Lt Col Veronica Reyes, for their sponsorship and guidance throughout the project. We would also like to thank MAJ JD Cadell for his advice and assistance during the data preparation phase of the project.

Abstract: Defense readiness is hard to measure due to current reporting structures, and unit, operational, and environmental variables. This work, conducted in support of the Deputy Assistant Secretary of Defense for Force Readiness (DASD (FR)), from August 2021 to May 2022 proposes that Army readiness metrics can be improved with unsupervised machine learning. The objective is to identify a comprehensive list of potential variables that can supplement the existing Army readiness model and use these to create an unsupervised machine learning model that improves readiness metrics. Utilizing data sources in Vantage, the team exported a set of ten predictor variables with 1,056 observations into Python and created a K-Means Cluster algorithm with four clusters to classify different groupings of active component Army battalion sized units. Initial results show that cluster 3 represents the readiest units with cluster 2 units lacking in expert weapons qualifications. Future work should increase the number of predictor variables and include external operational demands.

Keywords: Unsupervised Machine Leaning, Python, Army Readiness, K-Means Cluster, Army Vantage

1. Background

AR220-1 is the Army doctrine for how commanders report readiness which defines readiness as "...the ability of the U.S. military forces to fight and meet the demands of assigned missions" (Headquarters, 2010). Each of the military services collects and analyzes readiness information on its forces under general readiness reporting guidelines laid out in Joint instruction. This instruction requires Joint and service unit commands to evaluate, in near real-time, the readiness of forces to accomplish assigned and potential tasks through the Defense Readiness Reporting System (DRRS)— Department of Defense's (DOD) system of record for readiness data. Specifically, the instruction requires that the commanders of each unit assess and report on the readiness of units, at least monthly (Headquarters, 2010).

The Army currently decomposes readiness into four main categories: personnel, maintenance, equipment, and training readiness (Headquarters, 2010). Each of these categories measure how proficient units accomplish tasks, keep facilities up to code, and other metrics relevant to each category. Categories are measured on a scale called the Readiness Assessment (RA) scale where they are assigned a number between 1-4 (1 being the highest possible score or level of readiness). Once each of the four readiness categories have their scores, the overall unit readiness score, known as the C-level, is assigned on a similar 1-5 scale, where the C-level is determined by the lowest score from any category (Headquarters, 2010). Infrequently, a C-5 is given to units going through a major personnel changeover, modernization, or location change. The C-level is the main reporting score used in assessing whether a unit can perform its core functions (Headquarters, 2010). The Army collates these reports once a month.

This method of measuring readiness is simple to report and visualize. Unfortunately, it fails to fully encompass the complexity of military readiness. In his seminal book, *Military Readiness*, Betts (2010) argues that simplistic measurement systems fail to provide enough detail to decision makers to allow them to make decisions about resource allocation and employment. An example of a nuanced decision with tradeoffs is the conflict between increasing training and decreasing equipment readiness. The military uses large scale training exercises to increase immediate readiness; however, this can break equipment decreasing longer term readiness (Betts, 1995). Readiness measurement methodologies must incorporate more than a simple C-rating.

2. Literature Review

2.1 Machine Learning

Machine learning, from a simple perspective, is the idea that a computer-based model can learn and train itself over time. The subsets of machine learning are Supervised and Unsupervised machine learning. Since there is no known data set with associated response variables for the readiness problem set, this study focuses on unsupervised methods.

2.1.1 Unsupervised Machine Learning

Unsupervised machine learning has been used and implemented in a wide range of industries. Suominen, Toivanen, and Seppanen, conducted an analysis of patent data and explained the benefits and constraints of machine learning in this space. The group states that the "recent proliferation of machine learning text analysis methods is changing the traditional patent data analysis methods and approaches" (Suominen, Toivanen, & Seppanen, 2017). They are on the leading edge, attempting to bring machine learning algorithms into the process. The group leveraged research from Blei and Lafferty (2007) to understand how to apply Latent Dirichlet Allocation (LDA) as a topic model. Similar research into LDA's have produced much of the unsupervised machine learning algorithms currently used. The group used the mobile telecommunication industry as a case study for their analysis. Using a repository of full-text patent descriptions the group was able to create tokenized data in Python. They ran the data through an online variational Bayes algorithm for LDA (Suominen, Toivanen, & Seppanen, 2017). The program chunks the data into pieces until every piece has been analyzed. Using a trial-and-error approach they created K-values to create 75 topics. The algorithm created two matrices: document probabilities and word probabilities (Suominen, Toivanen, & Seppanen, 2017). After inputting the matrices into visualization software, they were able to create word clouds with defined edges. From this they were able to use a grouped time series model proposed by Hyndman and Athanasopoulos (2014) to forecast knowledge trajectories (Suominen, Toivanen, & Seppanen, 2017). Their application of unsupervised machine learning allowed them to see a managerial view of various developments in patents.

A 2019 survey by various authors developed a comprehensive report on recent advancements in unsupervised machine learning and applications to various learning tasks. The group states that there are increasingly voluminous amounts of unstructured unlabeled network data that is ripe for analysis using unsupervised machine learning (Usama, et al., 2019). There is a necessity for this type of machine learning in the modern era. This group's research develops a comprehensive survey of unsupervised machine learning techniques. Furthermore, they conclude that unsupervised learning can be broken into six major categories: hierarchal learning, data clustering, latent variable models, dimensionality reduction, and outlier detection (Usama, et al., 2019). Additionally, their work provided a clear and understandable diagram of all unsupervised machine learning techniques as well as an easy to grasp list of acronyms. Both provide great insight into the future use and application of unsupervised machine learning. After breaking down each technique and explaining its uses and function they apply the technique to the real world. Their work provides a very professionally written catalog of unsupervised learning techniques and creates an easy-to-understand map to implement this type of artificial intelligence.

One field where machine learning is regularly used is in medicine. A case study in this field can be found in the research article written by Wenbo Zheng et. Al. which discusses the use of machine learning to diagnose cases of COVID-19 at the beginning of the pandemic. The initial problem consisted of three challenges: The lack of training samples and small datasets due to COVID being a new disease, limited images and public data concerning COVID-suspected patients, and the level of urgency to diagnose early symptoms of COVID as well as the virus itself (Zheng, et al., 2021). After aggregating data and researching previously used methods of diagnosis, the team constructed three datasets: a training, support, and testing set. These datasets were developed by collecting X-ray and CT images from radiology medical reports and through a process of data augmentation to create a meta-learning based model (Zheng, et al., 2021). Meta-learning is the practice of training a new machine learning algorithm by using some combination of established algorithms to train the new one (Zheng, et al., 2021). The team trained their model using a two-step process. First, they develop a model called a feature extraction model to use meta-learning. Their second model was a relation model that took the feature map created by the feature extraction model and input them into an 8-layer network that could classify the images as being healthy, having pneumonia, or having COVID (Zheng, et al., 2021). Their resulting models had an overall accuracy rating of almost 97 percent, almost 7 percent higher than previously used methods in the field (Zheng, et al., 2021).

3. Methodology

3.1 Research Question

The fundamental research question is whether an unsupervised machine learning model can be applied to create a viable and improved way to measure Army Readiness. Derived from the research question, this study hypothesizes that it is possible to use data collected from identified Army data sources to develop an unsupervised machine learning algorithm (using K means clustering) to predict and classify units into various readiness clusters. The primary goal is to develop an objective, unsupervised, machine learning algorithm to improve Army readiness metrics for senior decision makers.

3.2 CRISP-DM Framework

This project utilized the Cross Industry Standard Process for Data Mining (CRISP-DM) methodology as the engineering framework. Using an iterative approach, developers first refine their business and data understanding within the context of the problem domain. This is done primarily through communication between the clients and the project team as well as background research done by developers. The next steps are data preparation and model creation, two steps that are alternated multiple times until an acceptable model is fully developed. The client then evaluates the model to determine whether it answers the business problem. The core of this methodology is having sufficient data to create a reliable model. Chapmans "Step-By-Step Model Guide" states that there are six data mining steps: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation and Deployment. (Chapman, et al., 2000).

3.2.1 Business Understanding

Through stakeholder analysis and background research, the study team understands that there are shortcomings with the current method of measuring Army readiness. The current method described above, that collates everything into a single C rating, is convenient, but does not provide enough detail to make specific resourcing or employment decisions. It is focused primarily on the supply-side of readiness and does not address any specific operational or environmental demands. There are the different types of readiness that Betts describes, and it is not always clear which one the current system measures. There is also a subjective component to the current readiness ratings in which commanders can adjust readiness measures in P, R, S or T. This paper will not argue for or against commander's ratings but will focus on creating an objective-based model.

3.2.2 Data Understanding

The data associated with Army readiness is not only hard to find and access, but also extremely difficult to concatenate into one database. Army readiness data is spread out across multiple platforms to include DRRS-A, GCCS-A, MEDPRO, DTMS, and many more. The initial search for potential variables that describe attributes of readiness was very broad to be as comprehensive as possible. The search used a systems thinking approach to categorize potential variables into a comprehensive catalog (Parnell, Driscoll, & Dale, 2011). For ease of tracking, the "variable catalog" was subdivided into four categories: Personnel (P), Equipment (E), Training (T), and External factors (X). This list included 53 variables. Each variable was further classified to identify its characteristics to include *currently measured, current data source, type, and measurement method.* Some variables of note that are not currently measured by DRRS-A include: *Number of personnel not making height/weight (P), Mean time between equipment failures (E), Days since most recent deployment (T),* and *Officer red cycle taskings (X).*

An analysis of what is currently measured, accessible, and applicable reduced the list to about twenty feasible variables. At the current moment the most accessible, comprehensive, data platform in the Army is Vantage. In Vantage, custom contours were created to create paths that linked, filtered, joined, and sorted the necessary data. Not all 53 variables were available or accessible in Vantage. The contours allowed for a repeatable subset of concatenated data to be created providing a final data set at the battalion level (Caddell, 2022).

At this point, it also became clear that there was not a response variable in the data. The C-rating could have been used, but this would have been at odds with the goal of the study due to the inherent issues with the current rating system. Additionally, unit assessments are rare. An assessment is done of Brigade Combat Teams (BCT) at the National Training Center, however every BCT receives the same assessment as "certified." Since the outcome is unknown, an unsupervised machine learning method to train and predict whether a unit is ready or not became the best option.

3.2.3 Data Preparation

The contour function of Vantage enabled extensive filtering of the data and a focus on a specific force element. Vantage also allowed the creation of variables by using functions of other variables. Specifically, *percent weapons tests passed, percent deployable over assigned, percent failed height weight over assigned, pacing FMC GCSS actual percent,* and *percent* *passed weapons tests* were created using accessible data. The unit identification code (UIC) was used to join data from different sources. Utilizing joins and the contour, Vantage produced a complete table of twelve variables from the variable catalog.

The final dataset – after eliminating NA's and standardizing the variables – is comprised of ten predictor variables each with 1,056 observations. The selected variables were chosen due to their predictive influence on the model, availability of the data, and desire to include new variables that are currently not measured. The study also avoided data relating to personally identifiable information and SHARP due to sensitivity. Using Python, NA's were identified and filled using the *Fill NA* function with the mean or median of that variable. *Average overseas deployment quantity* and *average combat tours quantity* both had 500 NA values. Similarly, *pacing FMC GCSS actual percent* had 266 NA values. Instead of dropping the NA variables for these variables, effectively cutting the data set in half, the values were filled with the mean or median of that variable. Using a correlation matrix, variables with high correlation were removed to prevent collinearity among independent variables. This procedure removed two variables, from twelve to ten, shown in the heat map in figure 1.

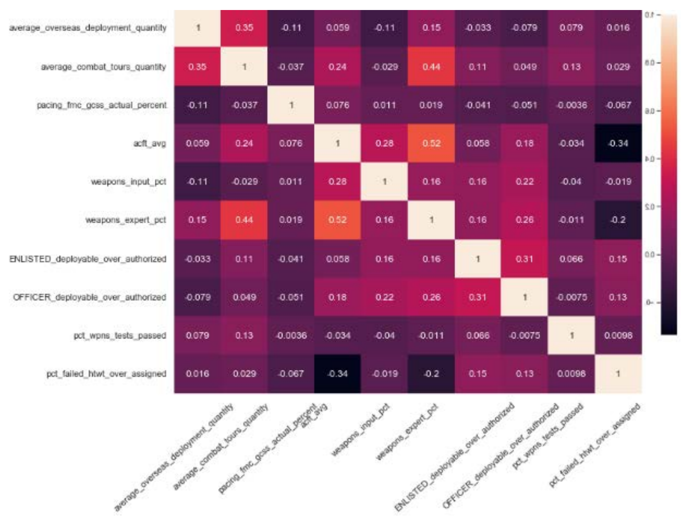

Figure 1. Variable Heat Map

4. Modeling

4.1 K-Means Clustering

The beginning portions of CRISP-DM helped understand what the Army needed and what the Amy currently has as a model. With a clean dataset, we began the process of creating a viable model. The unsupervised machine learning model takes readiness data from battalion-sized active-duty Army units and creates various readiness groupings. For security purposes the Unit Identification Code (UIC) was removed to maintain classification.

K-means clustering was chosen because of its simplicity paired with its ability to create a specific number of output classifications. The desire to employ a simple model is based in the fact that this model must be expandable across more branches of the Army and DoD. Additionally, the model must be easy to comprehend for decision makers. Using models like Neural networks or other unsupervised machine learning methods provide a relatively more complex model that is much more dependent on the data itself. Instead of giving the model the autonomy to decide on the number of classifications, K-means clustering allows analysts to predetermine the amount. Without this ability it is possible that a neural network approach would create six clusters for one model while providing three for another. This would make the comparison between units and overall Army readiness difficult to compare. K-means clustering provides simple classification that should appeal to decision makers who may not be well versed in the methods of unsupervised machine learning.

5. Results and Analysis

5.1 Evaluation

Within Vantage it became clear that there is a lack of continuity among various battalions reporting procedures. Some battalions report statistics within different time frames, there were some data integrity issues, and there were many incomplete fields. Additionally, there is no data dictionary associated with Vantage which left the interpretation of data and its column headers to the team, though most of the variables were self-explanatory.

5.1.1 Results

The result of the K-means algorithm provides a set of four clusters based on ten continuous predictor variables. The clusters represent four classifications of like units. Rather than stating which unit is ready or not ready, it depicts a group of units whose characteristics allow them to be grouped into one of four clusters. Unlike the existing DRRS model, this model depicts where units may stand out from others, while informing commanders on where to apply resources. The model produced four clusters with the following results: cluster 0 with 408, cluster 1 with 307, cluster 2 with 279, and cluster 3 with 62 battalions respectively. Cluster 0 units perform lower than other clusters in all categories to include ACFT, Weapons Qualifications, Height/Weight, and number of deployments. Cluster 1 units are characterized by high deployable over assigned personnel numbers, however they appear to have equipment and height/weight issues. Cluster 2 represents battalions that may be inexperienced with low numbers of previous deployments but perform well on personnel and equipment. Cluster 3 performs well across all categories.

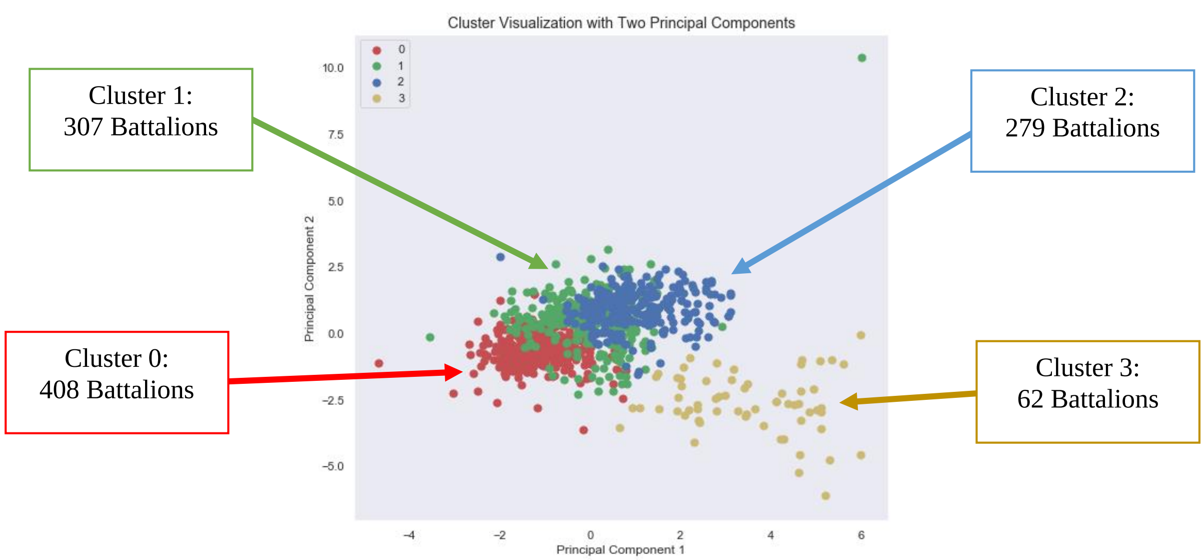

Figure 2. Cluster Visualization

Once complete the model can be rerun on the new data points. Analyst can either refine the clusters or assign new units to the previously defined clusters. Due to the high dimensionality of the model (ten predictors means ten dimensions are required to represent this visually), principal components analysis was used to create a two-dimensional visualization. The principal component analysis visualization in Figure 2 explains variation within the model and provides a visual cue for decision makers. The clusters with more spread suggest that the model accounts for more variation within that cluster, which is a positive characteristic.

Additional analysis of the model consisted of pulling the model output into Excel to determine the average values for each variable within their respective clusters as seen in Figure 3. This model suggests that cluster 3 with 62 of the 1,056 battalions analyzed are the readiest. The diagram suggests that cluster 1 and 2 are the least ready due to the associated battalions lower metrics in most categories. The excel analysis is focused on the spread or inclusivity of the model. Cluster 3 appears to be the readiest since it contains those units that have the highest potential and thus perform the best in each variable category. Conceptually it makes sense that ready units are more proficient in a greater number of categories while those that are not ready will fail to reach more favorable values. Conditional formatting allows for easier visual analysis of higher performing variables.

This method removes the dimensionality of the model, uses the standardized variables, and allows for individual variable performance to be ranked against one another within each cluster.

	Cluster 0	Cluster 1	Cluster 2	Cluster 3
average_overseas_deployment_quantity	0.3229205	0.3525913	0.2233964	0.7225077
average_combat_tours_quantity	0.329775	0.6295064	0.4322597	1.8638313
pacing_fmc_gcss_actual_percent	0.7921484	0.6815288	0.873179	0.7649541
acft_avg	0.4012809	0.4013825	0.4617029	0.4874396
weapons_input_pct	0.1743115	0.275453	0.4408898	0.2427634
weapons_expert_pct	0.0936483	0.169484	0.3298637	0.6932062
ENLISTED_deployable_over_authorized	0.7088557	1.0971129	0.9598312	0.8892843
OFFICER_deployable_over_authorized	0.7438549	1.1818372	1.2199757	1.046283
pct_wpns_tests_passed	0.9911606	0.9952914	0.9881296	0.9981189
pct_failed_htwt_over_assigned	0.6324502	0.5468499	0.6182435	0.669095

Figure 3. Excel Cluster Analysis

5.1.2 Conclusion

The work conducted on this project identified several administrative and technical challenges related to data collection. The 53 variables identified in the variable catalog proved to be extremely difficult to collate into one dataset, and there were data integrity issues to handle and clean.

As hypothesized, unsupervised machine learning can be used to model Army readiness. The work presented in this report verifies the methodology as a credible way to measure and classify readiness groupings. However, until applied to the DOD on a larger scale and implemented in a real-time environment, it cannot be validated. Validation would show that the proposed K-means clustering method is an improvement in measuring readiness over the current method.

This study constrained the model by Active Component battalion-sized Army units and limited the number of variables to ten due to data accessibility and correlation between variables. In future work, analysts could adjust the echelon level to provide more precise recommendations for decision makers. The model should then be applied to other services outside of an Army case study. Initial analysis generated a total of 53 possible readiness metrics in the variable catalog. The additional 43 variables should be analyzed and applied to the model to provide an increased depth of analysis and include other unit attributes such as command climate, operational environment, and unit discipline.

The proposed K-means model informs commanders on the readiness classification of their unit by using raw data and assigning units to unique clusters of like units. Though it does not provide factual assessments on readiness with a readiness classification, the model provides insightful information on which battalions should be selected for certain missions. This provides more insight to decision makers than the current readiness C-ratings allowing for informed decisions about unit employment and resourcing. The application of unsupervised machine learning, and specifically K-Means Clustering, cannot be overlooked as it is a viable model for interpreting Army readiness data.

6. References

Betts, R. K. (1995). *Military Readiness.* Washington D.C.: The Brookings Institution.

Caddell, J. (2022, January 28). MAJ.

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). CRISP-DM 1.0. *SPSS Inc.*

Headquarters, D. o. (2010). *Army Unit Status Reporting and Force Registration* . AR220-1.

Parnell, G. S., Driscoll, P. J., & Dale, H. L. (2011). *Decisoin Making in Systems Engineering and Management* . New Jersey: Wiley.

Suominen, A., Toivanen, H., & Seppanen, M. (2017). Firms' Knowledge Profiles: Mapping Patent Data With Unsupervised Learning. *Elsevier*.

Usama, M., Qadir, J., Raza, A., Arif, H., Yau, K.-l. A., Elkhatib, Y., . . . Al-Fuqaha, A. (2019). Unsupervised Machine Learning for Networking: Techniques, Applications and Research Challenges. *Institute of Elictrical and Electronics Engineers*.

Zheng, W., Yan, L., Gou, C., Zhang, Z.-C., Zhang, J. J., Hu, M., & Wang, F.-Y. (2021). Learning to learn by yourself: Unsupervised meta-learning with self-knowledge distillation for COVID-19 diagnosis from pneumonia cases. *Wiley Online Library*.