U.S. Army Recruiting:
Applications for Data Analytics and Machine Learning

Introduction

In the U.S. Army's 244-year history there has been much discussion on how it recruits Soldiers. However, today's data-rich environment moves the conversation in new directions. With new tools, new techniques, and most importantly, a new generation of digitally connected recruits, the Army must further leverage data-driven outreach and recruiting strategies if it is to remain a competitive option in the race for human capital. The Army has been very adaptive and innovative in this digital era, to include leveraging third-party data to improve lead generation, but more must be done. I believe that in the evolving age of big data, leveraging machine learning is a key enabler in helping the Army meet its required force structure with America's best and brightest. Unlike other literature that speaks broadly about *why* the Army should use machine learning, this paper centers on what machine learning is and *how* the Army can use it. More specifically, it examines a sample of machine learning techniques at a depth that makes clear the capabilities and limitations these tools offer. It concludes by offering five use case examples of how machine learning techniques could be employed to help improve Army recruiting outcomes.

Recruiting is an Issue of National Security

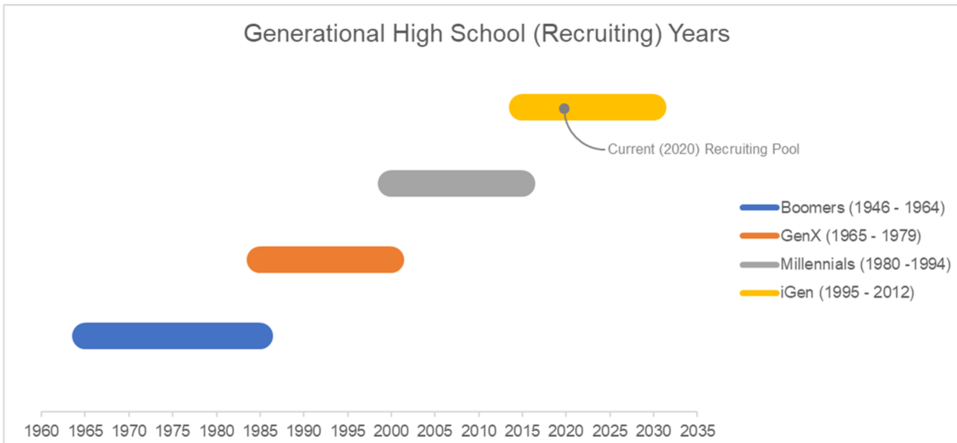
The Department of Defense (DoD) has reported that more than a decade of conflict, budget uncertainty, and reductions in force structure have degraded military readiness. In response, the DoD has made rebuilding the readiness of the military forces a top priority. However, the Army, which is the largest service in the DoD, has struggled recently in meeting its overall personnel end strength authorizations. The Army fell short by 0.38 percent in 2017 and fell short again by 2.56 percent in 2018 (Pendleton, 2019). This equates to about 6,500 recruits, or the equivalent of Brigade Combat Team. According to the 2019 Government Accountability Office report on Army Readiness, Army officials have stated that the primary reason it has struggled to meet its authorized personnel end strength is because it has had difficulty meeting recruiting goals (Pendleton, 2019). Failure to maintain adequate personnel end strength reduces Army readiness and puts its ability to execute the National Defense Strategy at risk.

The Recruiting Landscape

The goal of Army leadership is to fill operational units to 105 percent by the end of fiscal year 2020 (Pendleton, 2019). However, size alone is not all that matters. The Army requires having the right talent for the right job at the right time. With each day's advancement in technology and growing sophistication in warfighting, the talent required to fulfill the Army's war time functions becomes harder to find.

With the expansion of cyber, electronic warfare, and the ubiquitous proliferation of technological devices and sensors on the battlefield, in many cases, the Army is competing for the same talent as tech behemoths like Google and Facebook. However, what makes Army recruiting even more challenging than it is for its big-tech competitors is that more than two-thirds of American youth between 18 and 23 do not meet Pentagon recruiting standards (Grady, J.). Additionally, Military.com reports that 15% of those qualified for service fail to meet graduation requirements and return home after attempting initial entry training (McHugh, M.). These statistics, combined with the fact the size of this age-group population has remained roughly the same size over the last decade, brings to focus the challenge of successfully identifying and connecting with recruits that have the capacity and capability to serve.

Today's recruits come from a generation having never known life without the internet, and have lived much of their lives online with a smartphone being the last thing they see each night before bed, and first thing they check in the morning when they awake. Long-time generational scientist Doctor Jean M. Twenge categorizes as them as the "iGen" (Jean M. Twenge, PhD, 2017). She calculates that even with the competing demand of schoolwork, sleep and other life sustaining functions, the iGen is clocking more than six hours a day on digital media (text, internet, etc.). Online is where this generation lives, and that creates interesting opportunities for data-driven outreach and recruiting. However, this generation that prefers screen time to face-to-face interaction is comprised of sophisticated consumers that have been marketed to their enter life. This means that the Army must come armed with sophisticated data-driven marketing and outreach to break through the noise and appear relevant when competing for this generation's attention.



Data Source: Jean M. Twenge, PhD, 2017. iGen. ATRIA, New York, NY

Manning an Army

Recruiting the right talent begins with building broad awareness and generating demand to serve one's country through effective marketing at the national level. The concept of marketing in the Army is not new. Perhaps the most iconic and ubiquitous Army marketing tagline came from the World War One slogan "I want YOU for the U.S. Army". Another popular effort was the work centered around "Be All That You Can Be" originating in the 1980s and helped transition the Army from the draft era of Vietnam to the all-volunteer force (AVF) the U.S. Army remains today.

The Army has currently transitioned to a campaign centered on "Warriors Wanted" which has tested well with thrill-seekers who might not otherwise think about joining the Army but the approach is misaligned with the "stay safe" characteristics of the iGen. Dr. Twenge concludes that "IGen'ers' risk aversion goes beyond their behaviors toward a general attitude of avoiding risk and danger" and points out that the percent of high school students that report being comfortable with "taking risk" has dropped over 15% in recent years (Jean M. Twenge, PhD, 2017).

Another missed opportunity is that the Army's latest campaign delivered content that is already in line with what many iGens perceived about the U.S. Army through Hollywood. As a result, it did little to pique the interest of most iGens. Furthermore, centering a campaign on this "Hollywood" image is disproportionate to the real need the Army is recruiting for considering it only comprises these "action" jobs approximate to only 15% of the total Army enterprise. Recognizing these shortfalls, the Army is working to better communicate the value proposition of the many other ways a person can serve in the Army.

Leading this effort is the newly established Army Enterprise Marketing Office. The AEMO replaces the Army Marketing and Research Group that was located in D.C. and has an expanded mission responsible for coordinating the Army's national marketing and advertising strategy, developing and maintaining relationships with the marketing and advertising industry, and developing marketing expertise and talent within the Army. The AEMO is currently being established in Chicago and is partnered with Doyle, Dane, and Bernbach (DDB), a creative and marketing juggernaut with offices worldwide serving customers such as McDonalds, Uber, Kellogg's, and Johnson and Johnson just to name a few. The partnership is resourced through a $4 billion contract with a completion date of November 2028. The AEMO will work in coordination with Headquarters Department of the Army (HQDA), Army Human Resources Command, U.S. Army Recruiting Command and their partner DDB to create and deliver a marketing and outreach strategy designed to competitively position the Army for America's young talent while achieving end strength goals that meet the requirements of the National Defense Strategy.

Data Analytics and Machine Learning

According to some researchers, it is estimate that our digital universe is growing by 40% a year. While there's some debate on the true rates of growth, one fact can be agreed on: with more devices, connected to more people, more often, it's clear that the amount of data we collect, store and process will continue to grow. The conditions to apply data-driven techniques and machine learning in helping the Army achieve its recruiting goals have never been better. The connectivity, access to data, processing capability and now a generation of recruits living and sharing their lives online is unprecedented. Unfortunately, as a function of its recency and fluid evolution, many marketers and recruiting practitioners lack an appreciation of how machine learning can enable their efforts. There is much surface level talk about the importance of using machine learning, but very little discussion on what it is and how it can be applied. The following sections attempt to shed more light on this.

What is machine learning and how it can help?

Put simply, machine learning is focused on teaching computers to solve a problem. This is accomplished through the construction of computer programs and algorithms that automatically improve with experience. Machine learning concepts draw on statistics, computer science, artificial intelligence and other information theory-based study (Mitchell, 1997). Because of this broad composition, machine learning is very interdisciplinary.

The goal of machine learning is to assist analysts with making sense of massive datasets by using algorithms that can transform data into actionable knowledge. Like all learning, machine learning involves the abstraction of data into a structured representation that is generalized to create knowledge and inference in new contexts (Mitchell, 1997). Machine learners learn from examples and features in the data, and then summarize the data in the form of a model, which is then used in support of descriptive or predictive tasks (Lantz, 2015). The type of task the learning algorithm performs determines whether it is considered Unsupervised Learning or Supervised Learning.

Unsupervised Learning

The process of training a descriptive model is known as Unsupervised Learning. This type of learning is exploratory by nature as the analysts leverages powerful machine learning capabilities to summarize the data and uncover insights that would be difficult to find using non-automated techniques (Lantz, 2015). With unsupervised learning no single feature of the data is more important than another, and a target variable, or label, is unknown to the analyst. This "unknown" aspect is why it is called unsupervised—the analysts does not provide supervision of any particular target outcome. Unsupervised learning is particularly popular in data mining applications where the analysts is trying to discover unknown relationships in the data. Two very powerful and highly used unsupervised learning techniques are Association Rules Mining and Clustering.

Association Rules Mining

Association Rules Mining, often referred to as Market Basket Analysis, is a machine learning technique that identifies associations among features in a dataset to help identify useful and actionable patterns. One of the most widely used algorithms for Association Rules Mining is the Apriori algorithm (Lantz, 2015). The Apriori algorithm uses a simple, but computationally intensive, approach to reduce the association rule space by requiring that the subset of a frequent feature (or combination of features) that imply some other feature's outcome, must also be a frequent feature (or features).

Analyzing a dataset with Aprioir results in a collection of association rules that specify relationships among features in the dataset. From a technical aspect, the rules are built as two-sided relationships that help show what group of features (left-hand side) often imply some other feature (right-hand side). Organizing the association in this way allows an analyst to identify a specific right-hand side feature, say enlistment, and then try to find what left-hand side feature (or features) lead to a specific nominal outcome of the right-hand side (e.g. "yes", "no").

The association rules that are generated are constrained, and later refined by three primary evaluation parameters: Support, Confidence and Lift. Support of a rule measures how frequently the features, or combination of features, occur in the dataset. Confidence of a rule is a measure of its predictive power or accuracy. Lift of a rule measures how much more likely a feature (or combination of features) is likely to lead to some other feature's outcome relative to the rate of the other feature's outcome within the dataset absent the association (Lantz, 2015).

The strengths of this algorithm include its applicability to large datasets, its usefulness in datamining and discovering unknow knowledge in datasets, and perhaps most importantly, its generation of rules that are intuitive and easy to understand (Lantz, 2015). Some drawbacks however are that the algorithm is not very useful on small datasets, and it is easy to draw the wrong conclusions from random patters. Additionally, association rules are not used for prediction but rather for unsupervised knowledge discovery. However, once the rules are established, organizations can use this new understanding of their data and apply classification and segmentation insights to their current marketing processes to improve outcomes.

Clustering: K-means

Clustering is closely related to Association Rules Mining and is perhaps best described using the popular adage "birds of a feather flock together." Humans naturally like to group things together based on shared values, attributes, etc. However, doing this with datasets containing vast amounts of features and records is nearly impossible as it is often too difficult to see the underlying patterns and relationships that would group data instances together. But though the help of machine learning and algorithms like K-means, this grouping, or clustering, can be done near instantaneously.

Clustering is a very mature machine learning task that generates new data by describing patterns in existing data. One thing that makes it so helpful is that clustering is an unsupervised classification where unlabeled data examples are given a cluster label that has been inferred entirely from what was learned from the relationships and patterns in the existing data (Lantz, 2015). In more simpler terms, clustering neatly groups data into clusters that share similarities.

These similarities are determined by mathematical distance between data points. A common algorithm to use in clustering is K-means. Generally, the goal of the algorithm is to take each of the "n" data examples, and group it to one of the k clusters, such that the distance between data points within a cluster are minimized and the distances between the clusters is maximized. K-means starts by determining a "center point" for each cluster. The number of clusters to make is determined ahead of time, again, often assisted by domain knowledge. With initial cluster center points created, K-means treats the values for each feature of the data being modeled as coordinates in a multi-dimensional space and uses this "location" to determine proximity to its newly assigned "center point" and to the "location" of other data points thereby allowing it to determine distances and create the boundaries of each cluster. This process is repeated as many times as necessary until the model achieves a satisfactory level of stabilization where all data points are assigned to a cluster in way that creates the least amount of error for the model.

Clustering is best used for undirected data mining because the clustering technique finds patterns in data with no target feature. In other words, it doesn't need to know what it is looking for, nor is it trying to predict the outcome of a feature. In this exploratory way, clustering can help analysts better understand their data and use the insights from the clusters to direct further analysis, to include predictive analysis. The tricky part is that the clusters often lack any sort of intrinsic or obvious value, they are just groups of data that share some degree of similarity with each other, and differences with the rest of the data. As such, having domain knowledge is a critical prerequisite to effective clustering as the analysts will be required to determine actionable and meaningful labels to the clusters.

Supervised Learning

The process of training a predictive model is known as Supervised Learning. This type of machine learner is used when trying to predict the value of a particular feature by using the other features in the dataset. For example, trying to predict the weight of a person by knowing their age, gender and height. The feature that is trying to be predicted is refed to as the target feature, or class. This type of learning is considered supervised because the analyst knows what he or she is looking for and as a result, gives the model clear instructions on what they need to learn. More specifically, the supervision comes from the fact that since the target feature is known, the learner has a way to evaluate how well it has learned and can make adjustments as needed (Lantz, 2015).

Target features can be both numeric, where regression type learning occurs, or they can be categorical, where classification type learning occurs. Several powerful and highly used supervised learning techniques are explored below.

Decision Trees and Random Forest

Decision trees are a machine learning method that makes complex decisions from a set of simple choices. As such, decision trees are one of the most widely used machine learning technique and they can easy be applied to almost any type of data. As the name implies, decision trees utilize a tree like structure to present their knowledge in a flow-chart methodology making them easily understood without any previous statistical knowledge (Lantz, 2015). Decision trees are intuitively understood as most people have performed the same logic when weighing the pros and cons of any decision. They function much in the same way by simply mapping out all the decision nodes and branches one may consider when making decisions. A decision tree can be built using a variety of algorithms, but generally they all follow a technique called recursive partitioning (Mitchell, 1997). This technique uses a divide and concur approach to splitting data into subsets, which are then further split into additional subsets, and so on and so forth. A tree begins with the root node where no splitting has occurred. Next the algorithm chooses a feature to split on that is most predictive of the target class. The algorithm continues to partition off from features, or feature values, until a final "leaf" is created. Leaves are a collection of homogenous data, such as a "enlisted" or "did not enlist", and they represent the predicted decision.

The key to an effective tree is the ability of the algorithm to choose the best split upon. The intent here is split the data into as few subsets as possible that contain only a single class. This is referred to as purity and decision tree algorithms have different techniques at achieving this. For example, the C5.0 algorithm developed by computer scientist John Ross Quilan, uses an entropy calculation and selects splits that reduce entropy to increase homogeneity in the subsets (C4.5 algorithm, 2020). The algorithm calculates the change in homogeneity that would result from a split in each feature, and then chooses the split that maximizes information gain. The higher the information gain, the better the feature (and split) are at creating homogenous data groups, and ultimately, improving target class predictions. Once the trees are created using training data, rules are generated and applied to the test data to determine how well the tree does at predicting the target class. From here, an analyst can adjust certain tuning parameters to increase the performance of the decision tree.

A variation on the decision tree is the Random Forest. Random forests are no different than decision trees except that it is an ensemble type modeling where the dataset is split into many subsets of data and then multiple decision tree algorithms are run. This method combines the base principles of bagging with random feature selection to add diversity to the typical decision tree model. After several trees are ensembled, the forest is created, and the model uses a majority vote to combine the trees' predictions combining versatility and power into a single decision tree learner.

Naïve Bayes

The Naïve Bayes algorithm uses basic statistical and probability ideas that have existed for centuries. It is named after mathematician Thomas Bayes whose research developed foundational principles to describe the probability of an event, or more specifically, how the probability of an event should be adjusted if new information becomes available. Put more simply, Bayesian methods update the probability of an event occurring as new information is learned. For example, if a person were inside a building without access to the current weather conditions and was asked about the probability of rain, this person would the estimate with no additional information. Now if this same person were then brought outside and she could see the sun was shining without a cloud in sight, she would most likely change her estimated probability. This is the basic intuition of Bayesian methods. The probability of something happening conditioned on some other probability. For example, the probability of enlistment would change, if it were learned that the enlistee had a family member that retired from military service. The Naïve Bayes algorithm applies this technique, know as Bayes Theorem, in its classification modeling.

To perform its classification, the Naïve Bayes algorithm constructs tables of probabilities based on the features it was trained on in order to make predictions on the target class. Bayesian classifiers are best used in classification models in which information from numerous features should be considered simultaneously in calculating the final probability of the target class. Many other machine learning algorithms typically ignore features that have little impact on the overall classification, but Bayesian methods are capable of accounting for even the smallest probabilities that each feature has in determining the target feature. Adding up all these small probabilities can lead to significant impact. This makes Naïve Base algorithms very powerful, and often a smart first choice when conducting classification learning tasks.

Kth Nearest Neighbor (kNN)

Although it is one of the simplest classification techniques available, k-nearest neighbors (kNN) is extremely powerful and widely used today. It is a classifier that is particularly well-suited for classification tasks where relationships among the features and the target class are numerous, complicated or extremely difficult to understand. In simple terms, kNN classifiers are defined by their characteristic of classifying unlabeled examples by assigning them to the class they are most like, or "nearest", in similarity. The idea of "nearest" relates to distance in that each feature in the dataset gets treated as coordinates in multidimensional feature space (Tan et al., 2006). Using distance formulas like Euclidean distance or Manhattan distance, kNNs evaluate the distance between features of data instances in order to classify them with the target they are most like (Lantz, 2015).

The trick is to use the right number of ks to determine how man neighbors to compare each instance to. If more than one k is used, say for example 3, each data instance gets its features compared to others and a vote is taken among the three to determine how to classify a particular instance. There are many techniques used in choosing the appropriate k for a given scenario or dataset, but generally, the decision involves a bias-variance tradeoff. A large k will reduce the variance caused by noisy data and will probably generalize a little better than a model with fewer ks, but important features and patters in the data may get lost in the broader voting. Conversely, too few K's will most likely result in an overfitting of the data.

Another interesting note on KNNs is that they are often referred to as "lazy Learners." kNNs apply instance-based learning where rather than abstracting a model, they store the training data. This results in quick training, but making predictions later requires more time. However, unless very large datasets are being analyzed, this differentiation is not very significant.

Support Vector Machines (SVMs)

Support Vector Machines (SVMs) are in a category of classification algorithms sometimes referred to as "black box" algorithms. They get this name as a function of the complexity in their inner workings and mathematics. They are extremely powerful, but also extremely difficult to understand and must be applied smartly. SVMs use multidimensional surfaces to define the relationship between features and outcomes. SVMs can be imagined as a surface, which is a multidimensional hyperplane, that creates a boundary between points plotted in multidimensional space. These points represent examples of the data and their features from the dataset. By dividing the space with this hyperplane, SVMs establish a boundary between the target classes. The boundary is created by using select data points, which become known as support vectors. Support vectors are representative of each class and are the points that from each class that are closest to another class. As the SVMs are established and the boundary built, the algorithm attempts to thread the hyperplane between the support vectors in a way that maximizes the margin between the vectors (Lantz, 2015). The maximization of this margin optimizes the model by reducing the probability of being over fit to the training data and thereby expanding the model's generalizability.

Support vectors rely on vector geometry and involves sophisticated math that can distinguish between linearly separable and non-linearly separable data. Linear separable data can be divided into target classes much like other machine learning classifiers by creating a linear boundary. But when data is not linear separable, for example, classifying an eye on the picture of a person's face, a linear line will not suffice in classification. Therefore, a slack variable is introduced that modifies the margin between support vectors allowing for misclassified examples. However, the misclassified examples are kept to a minimum as each one comes at a cost to the algorithm. Slack variables, and techniques such as "kernels", allow SVMs to map data to higher dimensional space by transforming non-linear relationships into linear relationships through a change in data perspective.

SVMs combine the aspects of both kNN learning and that of linear regression modeling. This combination is incredibly powerful and results in SVMs being able to classify highly complex relationships, including both classification and numeric prediction.

Artificial Neural Networks (ANNS)

Similar to how a person's brain responds to and processes stimulus, so too does a machine learning technique called Artificial Neural Networks (ANNs). This machine learner models relationships between a set of input signals and through a network of nodes connects them to output signals. These nodes are called artificial neurons and perform similarly to a brain's neurons.

ANNs are very complex and sophisticated machine learners capable of conducting classification and numeric predication tasks, as well as unsupervised pattern recognition. They are best applied when input and output data are well-defined but the process that relates them is extremely complex. These learners have been around for nearly half a century but with the growing power of today's computing systems, ANNs have seen increased use as they have been applied to problems such as speech recognition, handwriting, the automation of smart devices (to include the Internet of Things), photographic interpretation and labeling, and many other sophisticated modeling tasks (Lantz, 2015).

Although there are many variations of neural networks, most share the following characteristics: an Activation function, an Architecture, and the Training Algorithm. The activation function is the result of a weighted sum of inputs (features) that get transformed into a single output signal to be passed along in the network if the summed values meet a firing threshold. Each input node is responsible for processing a given feature in a dataset. The value of the feature is weighted and passed to the output node to be transformed by the activation function. If the summed values fail to meet a threshold value, they do not get passed on.

An ANN's architecture, or topography, is the pattern of the network defined by the number of layers, the number of neurons in each layer, and the way in which they are connected within a specific model. The input and output nodes of a neural network form its layers. A network can be single, or multi-layered with hidden layers that process signals from inputs before reaching the output nodes. The number of input nodes is determined by the number of features in the dataset, and the number of output nodes is determined by the number of outcomes to be modeled. However, by adding additional layers and additional nodes between the inputs and outputs, a user can increase the complexity of the ANN. The value of having additional layers and nodes is to expand the complexity of the model and thereby expand the capability of the learner and its ability to handle increasingly complex tasks. Networks with multiple hidden layers are often referred to as Deep Neural Networks and they have made popular the idea of deep learning.

Lastly, the training of the neural network is where the real interesting things happen. This is where the network learns and assigns weights throughout the network based on patters seen between the data features. Today's ANNs use a technique called backpropagation to accomplish this. Simply put, this is an interactive process that initializes the network with weights assigned as a random baseline and as the algorithm iterates back and forth through the network, it adjusts the weights of all the input nodes until it has reduced the total error in the network (Lantz, 2015).

These powerful ANN learners can be applied to classification or numeric prediction efforts and are capable of modeling extremely complex patterns. One major downside however is that these learners are very much a "black box" solution in that they are not intuitive like a decision tree and they are grounded in complex calculus the make them difficult to interpret and difficult to explain.

Text Mining and Natural Language Processing

Data growth over the last five years has been overwhelmingly in unstructured data which accounts for almost 80% of total enterprise data today ("In Cloud We Trust," n.d.). As a result, some organizations are searching for ways to leverage this data, specifically, written and spoken text. Specifically, they are searching for ways to improve their text mining and natural language processing capabilities.

Although there are small differences in how text mining and natural language processing work and are applied, generally, they both include a range of computational techniques for analyzing and representing naturally occurring text, both written and spoken, for the purpose of achieving human-like language processing. This analysis can support many different applications from condensing written reports and books, reading resumes, or conducting market intelligence.

Applications for Machine Learning Techniques

Use Case 1: *Micro-targeting of population segments to improve recruit lead generation.*

K-means clustering could help Army marketers design and execute personalized marketing based on a recruit's location, age, sex, and a variety of other characteristics and features. This targeted approach would provide recruits a more personalized experience and may improve efficiency of the recruiting process as information asymmetry is reduced and recruits are offered more relevant information to their decision-making process. Additionally, as the K-means algorithm learns and creates improved subsets of the target population, the Army can direct recruiting efforts to fill specific jobs by surgically marketing to niche audiences with relevant messaging.

Use Case 2: *Better identify recruits that will successfully enlist and potentially serve beyond their first enlistment by modeling currently serving Soldiers that meet this criteria.*

To perform analysis of this sort would require data from systems of record that record initial recruiting data, personal data, and the professional data each service member has generated throughout their career. Most of this data is maintained by the Army's Human Resource Command (HRC). However, the data would require supplemental information be added to improve usefulness. The data should also include commentary, perhaps collected through a survey, on the overall satisfaction each service member had with each job they held, e.g., the team she worked with, her commander, the installation, the happiness of her spouse with the unit, and alike. These features, among others, help determine the level of satisfaction a soldier has with a given Army assignment, and in turn, informs her decision on whether to continue to serve. Learning from this data, a model can be built that predicts Service member's that enlist beyond their first term. This model could then be used to help identify recruits that are more likely to fulfill a similar career pattern. A Naive Bayes classifier would be an excellent tool in this use case. Using this technique, the learner could help predict the probability of a recruit making military service a career by leveraging data the Army already has on Service members that exhibited the same behavior.

Use Case 3: *Nurturing relationship between recruit and recruiter to better meet the needs of both parties.*

One opportunity it to leverage the Naïve Bayes strength as a text classifier. For example, many recruits spend a lot of time with a recruiter before finally deciding to enlist. Sometimes they will visit local military units together, go on hikes, a variety of activities to help inform the recruit on what life will be like in the military. During this period there is much conversation between the recruiter and the recruit, and much of this dialogue happens over text messages. At the macro level, transcripts from these conversations across the country could be uploaded to a single database repository where a Naïve Bayes learner could conduct text classification. Through this analysis, much could be learned about the motivations of recruits, as well as how certain strings of text indicate a trend that needs to be addressed. At the micro level, recruiters could then use this tool to identify themes in conversations between him and his recruits that he was unaware of. This could help him provide answers to questions the recruit has failed to ask, or is perhaps not comfortable asking. Additionally, this algorithm could be used to prioritize a recruiter's time with candidates that have the highest probability of committing to service by making probabilistic predictions based on demographic and recruiting data collected.

***Use Case 4**: Reducing initial entry training costs and improve basic training outcomes by prioritizing recruits most likely to succeed.*

Similar to the statistic mentioned early, a Heritage Foundation report published in February of 2018 found that 71% of Americans ages 17 - 24 cannot qualify for military service due to health problems, physical fitness, education, or criminality (Spoehr, T.). Association rule mining could provide U.S. Army recruiting an opportunity to optimize recruiting efforts. This machine learner could help improve the identification of recruits that will complete the initial phase of enlistment successfully. For example, if the target feature was whether a recruit completed the transition from "Recruit" to "Soldier" by successfully completing basic training, association rule mining could help determine the additional features (demographic, social, physical, mental aptitude, medical history, extra-curricular activities, etc.) that lead to successful outcomes. Recruiters could then use these rules to better identify and screen potential candidates before making heavy recruiting investments. Association rule mining can leverage data to develop proven, fact-based solutions that improve the efficiency in the recruiting process offering the targeted approach the U.S. Army needs to meet its recruiting numbers without increasing resource expenditures.

Using decision trees, the Army has already demonstrated progress in the area of improving basic training outcomes as seen in a study published in 2013 titled "Cost-Effectiveness Analysis of the U.S. Army Assessment of Recruit Motivation and Strength (ARMS) Program." The study was conducted at six Military Entrance Processing Sites (MEPS) during 2005–2006. The objective was to compare morbidity and attrition of Army recruits who exceeded body fat with those who met body fat standards and deemed physically fit as measured by a 5-minute step test, and those who met body fat standards but deemed unfit (Niebuhr et al, 2013). The study developed a decision tree for U.S. Army applicants and whether they were accessed, experienced musculoskeletal injury and/or attrited. The decision trees used in this study were generated in such a way as both probability of an event (an applicant being accessed, experiencing musculoskeletal injury and/or attrite) and the costs associated with each outcome were represented by variables. A similar approach could be used to predict certain outcomes on recruiting efforts, or outcomes on recruits by using decision tree machine learning techniques.

***Use Case 5:** Leveraging Social Media to assist with recruiting efforts.*

A great use case example of how to leverage social media can be illustrated by the Social Networking Service (SNS) analysis Brett Lentz performed while conducting sociological research on teenage identities at the

University of Notre Dame (Lantz, B., 2015). Using an automated web crawler, Lentz captured the full text of SNS (Facebook, Twitter, etc.) user profiles from users across four high school graduation years (2006 to 2009). The data also included each teen's gender, age, and number of SNS friends. From there, Lantz extracted each user's page content into individual words where the top 36 were chosen to represent five categories of interest. Using this data, and formatted to support machine learning clustering, Lentz was able to extract incredible insights from the analysis. Most relevant to AEMO is that he was able to group teenagers with others who have similar interest and develop a typology of teen identities that was predictive of personal characteristics. It is easy to see how valuable this sort of analysis can be to AEMO efforts at learning about the interests of potential candidates and developing precision messaging to speak to each cluster. In summary, the power of clustering lies in this: Beginning with massive amounts unlabeled data, an analyst not only learns interesting things about her data, but she can also use clustering to create class labels for her data, and then apply a supervised learning technique to find the most important predictors of her labels. This is certainly a capability that could be very successfully leveraged by the AEMO.

Another way to leverage social media is through the text mining and Natural Language processing. The Army could gain insight on sentiment analysis from posts it makes on social media. For example, the U.S. Army Chief of Staff, GEN McConville has a public affairs team managing his social media account and they actively post across several social media platforms (Facebook, Twitter, etc.) daily. With hundreds of thousands of followers, the comments and reactions to topics mentioned should be text mined. Compiling this data would provide insight on topics of interest and the public's perception of these topics allowing the Army to amplify positive reactions and address negative reactions. Activities like this assist with brand management and make telling the Army story to potential recruits much more effective.

Conclusion

Recruiting the right talent is vital to maintaining Army readiness and keeping it postured to meet the National Defense Strategy. But competition for talent and scarcity of qualified individuals means the Army must employ the right tools to remain competitive as it continues marching into the digital era. Through machine learning techniques, the Army can deploy data-driven recruiting and outreach strategies that meet a generation of young adults living digitally connected lives with smart, targeted, and relevant messaging. This precise and personalized connection is the critical first step in delivering the correct value proposition that will mobilize talent to join the Army team.

Bibliography

C4.5 algorithm, 2020. Wikipedia. https://en.wikipedia.org/w/index.php?title=C4.5_algorithm&oldid=943211739 (accessed 4.18.20.)

David W. Niebuhr, MC USA, William F. Page, PhD, David N. Cowan, PhD, Nadia Urban, MHS, Marlene E. Gubata, MC USA, Patrick Richard, PhD, Cost-Effectiveness Analysis of the U.S. Army Assessment of Recruit Motivation and Strength (ARMS) Program, *Military Medicine*, Volume 178, Issue 10, October 2013, Pages 1102–1110, https://doi.org/10.7205/MILMED-D-13-00108

Grady, J., 2019. Panel Says U.S. Military Recruitment Pool Must Broaden. USNI News. URL https://news.usni.org/2019/06/17/panel-says-u-s-military-recruitment-pool-must-broaden (accessed 4.21.20).

In Cloud We Trust [WWW Document], n.d. . Credit Suisse. URL https://www.credit-suisse.com/eg/en/articles/asset-management/in-cloud-we-trust-uk-201905.html (accessed 4.26.20).

It's Not the Economy: Why the Army Missed its Recruitment Goals and What to Do About It, 2019. Modern War Institute. URL https://mwi.usma.edu/not-economy-army-missed-recruitment-goals/ (accessed 3.31.20).

Jean M. Twenge, PhD, 2017. iGen. ATRIA, New York, NY.

Lantz, B., 2015. Machine learning with R: discover how to build machine learning algorithms, prepare data, and dig deep into data prediction techniques with R, Second edition. ed, Community experience distilled. Packt Publishing, Birmingham Mumbai.

Lim, N., Orvis, B.R., Curry Hall, K., 2019. Leveraging Big Data Analytics to Improve Military Recruiting: [WWW Document]. URL https://www.rand.org/pubs/research_reports/RR2621.html (accessed 3.31.20).

McHugh, M., n.d. The Top 3 Reasons You Could Fail Basic Training [WWW Document]. Military.com. URL https://www.military.com/join-armed-forces/the-top-3-reasons-you-could-fail-basic-training.html (accessed 4.19.20).

Mitchell, T.M., 1997. Machine learning, International ed., [Reprint.]. ed, McGraw-Hill series in computer science. McGraw-Hill, New York, NY.

Pendleton, J., 2019. ARMY READINESS: Progress and Challenges in Rebuilding Personnel, Equipping, and Training (Testimony No. GAO-19-367T). Government Accountability Office.

Spoehr, T., n.d. The Looming National Security Crisis: Young Americans Unable to Serve in the Military [WWW Document]. The Heritage Foundation. URL https://www.heritage.org/defense/report/the-looming-national-security-crisis-young-americans-unable-serve-the-military (accessed 3.31.20).

Tan, P.-N., Steinbach, M., Kumar, V., 2006. Introduction to data mining, 1st ed. ed. Pearson Addison Wesley, Boston.

Why Machine Learning is the Future of Recruitment and Retention in the Military [WWW Document], n.d. URL https://blog.datarobot.com/why-machine-learning-is-the-future-of-recruitment-and-retention-in-the-military (accessed 3.31.20).