Proceedings of the Annual General Donald R. Keith Memorial Conference
West Point, New York, USA
May 4, 2023
A Regional Conference of the Society for Industrial and Systems Engineering

Assessing Risk with Cadet Candidates and USMA Admissions

Daniel Lee, Andrew Lee, Eric Reder, and Anna Tucker

Department of Mathematical Sciences
United States Military Academy
West Point, New York 10996

Corresponding author's Email: daniel.lee@westpoint.edu

Author Note: Daniel Lee is a 1st Class Cadet at the United States Military Academy. His advisors are LTC(P) Andrew Lee, MAJ Eric Reder, and CPT Anna Tucker. The views expressed herein are those of the authors and do not reflect the position of the United States Military Academy, the Department of the Army, or the Department of Defense.

Abstract: The United States Military Academy graduates approximately 1,000 cadets annually, though over 100 cadets from the initial cohort fail to graduate. Even with the identification of key predictor variables from current research, graduation risk has proven difficult to measure. Using data from the USMA Admissions Department, we used several machine learning algorithms to better predict which cadets would be separated or resign using potential variables that may relate to graduation risk. We found the current admissions score weights were not optimal and supplementing the current admissions criteria with data on the participation of certain extra-curricular activities improves precision.

Keywords: Admissions, USMA, Graduation Risk, Machine Learning

1. Introduction

Graduation risk, referring to the risk of a student not graduating from an undergraduate institution, is contingent upon several risk factors. Based on contemporary knowledge, both cumulative high school GPA and standardized test scores (SAT and ACT) are historical indicators of academic success (Geiser & Santelices, 2007). However, high school GPA and standardized test scores only accounted for 27% of the variation of cumulative college grades in this study, suggesting other variables are associated with academic success (Geiser & Santelices, 2007).

1.1 Admissions at the United States Military Academy

The United States Military Academy (USMA) differs from most other collegiate institutions for several reasons, most notably in the additional requirements expected of the cadets and incoming cadet candidates. In terms of admission, USMA accepts applicants based on a Whole Candidate Score (WCS) comprised of several factors related to academics, leadership, and physical fitness (Kelly et al., 2014). The academics portion of the WCS is measured through the College Entrance Examination Rank (CEER) score, weighed as 60% of the total WCS and consisting of high school GPA, class rank, and standardized test scores (Hanser & Oguz, 2015). Another 30% of the WCS is based on the Community Leader Score (CLS), consisting of the Athletic Activity Score, Extracurricular Activity Score, and the Faculty Appraisal Score (Hanser & Oguz, 2015). The remaining 10% is the Candidate Fitness Assessment (CFA), a physical fitness test that includes pull-ups, a standing basketball throw, a shuttle run, pushups, sit-ups, and a mile run (Hanser & Oguz, 2015). The WCS has proven to be the best predictor in a cadet's academic, military, and physical performance (Kelly et al., 2014).

1.2 Follow Through with Extracurricular Activities

Perhaps the most underemphasized factor contributing to graduation risk is follow-through, referring to the purposeful, continuous commitment to certain activities rather than sporadic effort in several areas (Duckworth, 2018).
In one study, researchers measured follow-through using a numerical score based on the number of extracurricular activities a student actively participates in, how many years of commitment to the extracurricular activities, and achievements earned for the activities (Duckworth, 2018). High school students with the highest follow-through scores participated in two extracurricular activities for several years and achieved acclaim or significant advancement in these activities (Duckworth,

2018). Those with lower follow-through ratings often did not participate in an activity for longer than a year before moving on to something different (Duckworth, 2018). Controlling high school GPA and standardized test scores, follow-through in extracurricular activities was the best predictor of college graduation (Duckworth, 2018). Based on separate research, students who participated in more extracurricular activities for longer were more likely to graduate within four to six years at an undergraduate institution (Gardner et al., 2020). These conclusions are congruent with previous academic literature and theoretical work that suggest that extracurricular activities offer opportunities to enhance academic values, resources, and educationally-relevant characteristics such as self-confidence, initiative, and perseverance (Gardner et al., 2020).

2. Data

The USMA Admissions Office provided the data for this study. As this research posed no greater than minimal risk to the subjects, it was exempt from Institutional Review Board (IRB) review by USMA. The initial dataset included data on 12,640 cadets from the classes of 2012-2021, separated into three sheets. A unique numerical code replaced each cadets' name to maintain anonymity. The dataset comprised of a general admissions overview for each cadet, all high school extracurricular activities that a cadet candidate participated in, and the answers to several school official evaluation responses filled out by each cadet candidate's high school counselor for the first, second, and third sheets, respectively. The first sheet included the USMA Admissions scores that comprise the WCS.

2.1 Extracurricular Activities

To analyze the particular types of extracurricular activities, we further condensed all activities from the second sheet into sub-categories. Of note, no universally defined grouping of extracurricular activities exists, whether explicitly published by scholastic or collegiate organizations or through academic studies. However, several studies and non-academic sources have grouped extracurriculars into certain, universal categories: namely, athletics, fine arts, student government, out-of-school activities, and academic/student organizations (Haensly et al., 1985; McNeal, 1998). To maintain consistency, we grouped extra-curricular activities as closely to those of other academic studies on this topic (Table 2). The categories of "Fine Arts" and "Student/Academic Organizations" (Table 2) remained consistent with the grouping of other academic studies. As most studies either did not explicitly differentiate athletic activities in or out of school or included all athletic activities into a single category, the "Athletics" category included both high school sports and out-of-school athletic activities (Haensly et al., 1985; McNeal, 1998). Though some studies also include out-of-school activities as a category, few, if any, provide an explicit definition (Haensly et al., 1985). In the "Out of School" category, we included all activities that were not associated with school such as scouting organizations, church activities, and camp counselors. Student government, though usually referring to student council members or class officers, has also referred to extra-curricular leadership positions in general in some studies (McNeal, 1998). Due to the significant number of leadership positions listed by USMA Admissions as extra-curricular activities, we consolidated all leadership positions, including those outside of school, into a single category of "Leadership." The remaining categories – "Cadet," "Foreign Study," "Honors/Awards," and "Other Extracurricular Activities" – were designated as such as they either did not fall into a category of precedence or were of particular interest to this study as a separate category. Each of these variables were binary ('0' for no participation, '1' for participation for at least one year in at least one activity within the category).

2.2 Types of Separations

The dataset noted cadets who did not graduate with a comment stating their reason and the date for their separation. We consolidated each reason (25 reasons for separation) into three distinct categories: "Separated," "Resigned," and "Other." The "Separated" category comprised cadets separated from USMA academically or for misconduct including those suspended or enrolled in the Army Mentorship Program (AMP) - a program that sends a cadet to an active Army unit for two years as a non-promotable enlisted soldier before allowing that cadet to reapply to USMA with the goal of development. The "Separated" category also included cadets who voluntarily resigned for academic, physical fitness, honor-related, or misconduct reasons, as these cadets resigned for reasons that would likely lead to a formal separation in the near future. Cadets binned into the "Resigned" category resigned from USMA either during their initial Cadet Basic Training (CBT), for personal reasons, or for lacking motivation to continue at USMA. The final category, "Other," referred to cadets separated for medical complications, religious reasons, or other unspecified. In total, 10,439 cadets graduated, 838 (6.6%) cadets separated, 1,198 (9.5%) cadet resigned, and 165 cadets left USMA for other reasons. Resignations are often more difficult to predict than separations, considering that personal and motivational considerations are difficult to measure or quantify. Cadets may resign even if they are highly qualified, in good standing, and are performing well in the three programs at USMA. However, for the purposes of this study, we analyzed all cadets who were separated or resigned together as a combined "did not graduate," as we seek to find a general admissions model for graduation rates.

3. Methods

3.1 Measures

To explore improvements to graduation rates, we compared the current model to a model with a redistribution of weights for the five admissions scores (the CLS was broken down into its 3 sub-scores) and a model that included specific extracurricular activity involvement. The target variable was binary: '0' for those who graduated and '1' for those who did not. The primary measures for this study were Area Under the Curve (AUC), precision, and recall. The AUC provides an aggregate measure of classification performance by finding the probability that the model ranks a random positive higher than a random negative, essentially finding how accurate a model is in a score between 0 and 1. Precision refers to the percentage of true positives predicted relative to all positive predictions while recall is the percentage of true positives guessed for all actual negatives positives.

3.2 Machine Learning Algorithms

The four machine learning algorithms of classification used for study were logistic regression, k-nearest neighbors (KNN), random forests, and gradient boosting. Logistic regression is a commonly-used method to predict categorical dependent variables, in this case '0' and '1' representing graduation status, that can use a variety of independent variables. The K-nearest neighbors algorithm estimates the probability that a particular data point belongs to a certain classification group based on what group the data point's neighboring points belong to. The number of neighboring points that a datapoint may be compared to is indicated by the K value, which thereby compares a point to its "K" nearest neighbors and classifies the point accordingly. The random forest classifier is an ensemble learning method, using a multitude of decision trees, thereby mitigating the risk of overfitting or biases that only a single decision tree would have in classification. Like random forests, gradient boosting also uses an ensemble learning method consisting of decision trees, they differ in the construction of the trees. While the random forest classifier constructs each tree independently, the gradient boosting classifier builds each tree one after the other, correcting any errors in the previously trained tree. Random forests also combine their results at the end of the process by averaging whereas gradient boosting combines results as it continues on the process. Consequently, gradient boosting often performs better for imbalanced datasets than random forests.

3.3 SMOTE for Imbalanced Datasets

Without testing-specific data provided by USMA, each of the four methods of classification could only be trained and tested on itself. Consequently, each dataset was split into a training and a testing dataset, with the training dataset comprising 80% of all samples and the testing comprising the remaining 20%. Due to the imbalanced nature of the datasets however, simply running the four machine learning processes would result in nearly all, if not all, predictions for graduations. Generally, the two primary methods to address this issue are through undersampling and oversampling. Undersampling, specifically random undersampling, randomly deletes datapoints from the majority class to closely match or equal the number of datapoints in the minority class. However, undersampling risks deleting datapoints that are more information-rich and valuable than others and overall leads to a loss of information. The alternative is oversampling, which creates synthetic datapoints for the minority class to increase its size, preserving the already-present data in the majority class. One of the most-used methods of oversampling is the Synthetic Minority Oversampling Technique (SMOTE). SMOTE works by grabbing a random sample from the minority class and utilizing KNN to find the nearest neighbors for the data point. Using one of the neighbors, SMOTE determines the vector between that and the datapoint, then multiplies the vector by a random number between 0-1, thereby creating a synthetic data point similar to one already existing but not a duplicate.

Each of the four algorithms utilized a pipeline that included SMOTE, a scalar to normalize all the variable values, and the machine learning model itself. This pipeline was examined through stratified k-fold cross validation where instead of splitting the testing dataset randomly into 10 folds, maintains the proper ratio of '0' and '1' in terms of response variables for each fold as the whole dataset. Every model was examined to ensure that the difference between the cross-validation score of the training dataset (average of all cross-validation scores for the stratified k-fold) was within around 0.03 of the test score. Each model was also further measured based on changing certain parameters, where the most accurate parameter based on AUC score was chosen for the final model. These parameters were "C" for logistic regression (0.001, 0.01, 0.1, 1, 10, 100,

ISBN: 97819384964-4-8

1000), "K" for KNN (1-10), "n estimators" for random forests (25, 50, 75, 100), and learning rate for gradient boosting (0.05, 0.075, 0.1, 0.25, 0.5, 0.75, 1).

4. Results

4.1 Current Model

Currently, a cadet candidate is scored by USMA Admissions based on the WCS and its sub-scores. This model only uses WCS as its variable to maintain the integrity of the admission score weights for the three primary sub-scores of the WCS (CEER, CFA, and CLS).

Table 1. AUC, Precision, and Recall for the Current Model

	AUC	Precision	Recall
Logistic Regression	0.619	0.24	0.56
KNN	0.541	0.21	0.44
Random Forests	0.542	0.20	0.43
Gradient Boosting	0.605	0.25	0.49

4.2 Updated Model: New Admission Score Distributions

The updated model with new admission score distributions analyzes the WCS by the individual component scores to create a model using the ideal weights determined by the algorithm. These individual scores are the CEER Score, the CFA Score, the Athletic Activity Score, the Extracurricular Activity Score, and the Faculty Appraisal Score. We used logistic regression to determine the statistical significance and the coefficient for each component score, determining which variables showed a relationship with graduation and the relative importance of each variable. Using the coefficients, we can use the ratios compared to the coefficients of the other scores to determine a better weight. Since the Faculty Appraisal Score had a positive coefficient (while the others had a negative coefficient), we only compared the ratios of the other four scores to each other and assigned no weight to the Faculty Appraisal Score, considering that this score had a statistically significant relationship with not graduating and would likely be unhelpful in an admissions score. The corresponding coefficients, new weights, and statistical significance based on logistic regression are shown in Table 2.

Table 2. Coefficients and P-value of updated model for score distributions (* For $p < 0.05$)

	Coefficient	Ratio	New Weight	P-value
CEER Score	-0.0036	0.0036 / 0.0072	50%	0.000*
CFA Score	-0.0022	0.0022 / 0.0072	30%	0.000*
Athletic Activity Score	-0.0006	0.0006 / 0.0072	10%	0.009*
Ex. Curricular Act. Score	-0.0008	0.0008 / 0.0072	10%	0.001*
Faculty Appraisal Score	0.0037	0	0%	0.000*

Table 3 shows the AUC, precision, recall scores of the machine learning algorithms based on the Updated Model with new admission score distributions.

Table 3. AUC, Precision, and Recall for the Updated Model

	AUC	Precision	Recall
Logistic Regression	0.633	0.24	0.54
KNN	0.549	0.21	0.40
Random Forests	0.568	0.22	0.27

ISBN: 97819384964-4-8

Gradient Boosting	0.623	0.25	0.53

4.3 Extracurricular Activities Model

Because the Athletic Activity and Extracurricular Activity Scores are essentially numerical interpretations of categorical variables with fairly arbitrary weights attached to each activity, we wanted to look into a different method of measurement with follow-through in mind. From the original scores that comprised the WCS, the Extracurricular Activities (EA) Model consists of the CEER Score, CFA Score, and the Faculty Appraisal Score. However, it substitutes the Athletic Activity and Extracurricular Activity Scores with the individual extracurricular activity variables: Athletics, Cadet, Fine Arts, Foreign Study, Honors/Awards, Leadership, Other Extracurricular Activities, Out of School, and Student/Academic Organizations (reference 2.1). We also added additional variables representing the total number of activities for athletic activities and non-athletic activities to this model, differentiating athletics from other activities since athletic activities are the most popular and seemingly the most prestigious of all activities (McNeal, 1998). The total number of activities were based on the years of participation (1,2,3,4) for each type as well as total overall counts for each type.

All extracurricular activity-related variables were also checked for multicollinearity, as it can undermine the statistical significance of independent predictor variables. Multicollinearity was determined through Variation Inflation Factor (VIF), which determines correlation in variables. Generally, a VIF of five or greater indicates high multicollinearity. All variables related to total athletic activities (aggregate count and count of sports by years played) indicated high multicollinearity, likely a result of the inclusion of "Athletics" (which provided a binary output of participation in sports). In addition, the "Total Activities" a cadet candidate participated in and Total Activities for 1 Year also showed a high VIF. These variables were subsequently removed from the final model.

Table 4 shows the AUC, precision, and recall scores of the machine learning algorithms based on the EA Model.

Table 4. AUC, Precision, and Recall for the EA Model

	AUC	**Precision**	**Recall**
Logistic Regression	0.643*	0.24	0.58*
KNN	0.562	0.26	0.17
Random Forests	0.580	0.37*	0.07
Gradient Boosting	0.636	0.37	0.05

5. Conclusion

Of the four machine learning algorithms, logistic regression performed the best in nearly all categories. Using the logistic regression scores, we found that the Updated Model with the individual sub-scores WCS had a higher accuracy (through AUC) than the Current Model while the EA Model performed the best overall in accuracy. Though the precision for all three models remained fairly constant, the EA Model also had the highest recall score. Based on these results, we believe that our Updated and EA Models are better than the Current Model for assessing graduation risk. We assert that the Athletic Activity and Extracurricular Activity Scores should be updated, particularly to better reflect certain aspects of follow-through and to adjust which activities should be ranked higher in precedence. Future research into this subject should focus on the improvement of the CLS, perhaps in other methods to better quantify extracurricular activity involvement, achievement, and leadership.

In terms of the weights of admission scores, the current distribution is not optimal as determined by the coefficients of logistic regression. Based on the coefficients, a better distribution of the scores would reflect the ratios of the coefficients for each score. Since Faculty Appraisal Score shows a positive coefficient, a high score in this category actually predicts a higher likelihood of not graduating. The Faculty Appraisal Score provides limited feedback in the success of a cadet and its removal from the WCS may provide no change, if not a benefit. For the other four scores, a possible distribution could be 50% for the CEER Score, 30% for the CFA Score, and 10% for both the Athletic Activity Score and Extracurricular Activity Score. These distributions were based on the ratios of each score to each other based on the coefficients and are found in Table 2, providing the relative importance and weight of each score on predicting separations. We, however, do not advocate for the complete removal of the Faculty Appraisal Score, especially since this score could provide valuable information on a candidate's character and work ethic – important factors to graduating. Rather, the Faculty Appraisal Score could be used as a non-scored portion of a candidate's application or be partially absorbed into the other sub-scores of the WCS.

We also noted the marked increase in precision for the Random Forest and Gradient Boosting algorithms in the EA Model at the cost of a significantly lower recall score. As the Random Forest and Gradient Boosting algorithms are better suited for imbalanced datasets (compared to logistic regression or KNN), such an increase could be a result of using both these algorithms with SMOTE (both compensating for imbalanced data) or due to the use of multiple binary variables for extracurricular activities in the EA Model. Such oddity, however, did not impact the general findings of this study as logistic regression proved to be the best machine learning algorithm and was consequently used instead.

One of the main limitations of this study was the lack of insight into the exact process of scoring the CLS sub-scores, preventing specific analysis into the methods. Another limitation was the scope of this research. Though a more robust prediction model would include various demographic variables, we only sought to look into the aspects of the WCS, as demographic variables would be inappropriate in determining an admissions score through the WCS. The objective of this study was not to predict graduations, but rather use the current admissions criteria to discover potential avenues for improvement in the WCS in terms of graduation risk.

6. References

Duckworth, A. (2018, August 21). *Grit: The Power of Passion and Perseverance* (Reprint). Scribner.

Gardner, M., Hutt, S., Kamentz, D., Duckworth, A. L., & D'Mello, S. K. (2020). How Does High School Extracurricular Participation Predict Bachelor's Degree Attainment? It is Complicated. *Journal of Research on Adolescence*. https://doi.org/10.1111/jora.12557

Geiser, S., & Santelices, M. V. (2007). Validity Of High-School Grades in Predicting Student Success Beyond The Freshman Year: High-School Record vs. Standardized Tests as Indicators of Four-Year College Outcomes. *Research & Occasional Paper Series: CSHE.6.07*, 1–35. https://cshe.berkeley.edu/publications/validity-high-school-grades-predicting-student-success-beyond-freshman-yearhigh-school

Haensly, P. A., Lupkowski, A. E., & Edlind, E. P. (1985). The Role of Extracurricular Activities in Education. *The High School Journal*, *69*(2), 110–119. http://www.jstor.org/stable/40365131

Hanser, L., & Oguz, M. (n.d.). *United States Service Academy Admissions: Selecting for Success at the Military Academy/West Point and as an Officer*. RAND Corporation. Retrieved December 13, 2022, from https://www.rand.org/pubs/research_reports/RR723.html

Kelly, D. R., Matthews, M. D., & Bartone, P. T. (2014, July). Grit and Hardiness as Predictors of Performance Among West Point Cadets. *Military Psychology*, *26*(4), 327–342. https://doi.org/10.1037/mil0000050

McNeal, R. B. (1998). High School Extracurricular Activities: Closed Structures and Stratifying Patterns of Participation. *The Journal of Educational Research*, *91*(3), 183–191. https://doi.org/10.1080/00220679809597539