USMA Admissions and Natural Language Processing

CDT Evan W. Lee
Advisors: LTC Andrew Lee & MAJ Welvin Lucero
United States Military Academy, West Point

Abstract

The United States Military Academy (USMA) at West Point is renowned for producing Army Officers entrusted with the critical mission of leading Soldiers into combat. USMA expects each graduate to serve as a leader of character, prepared to lead Soldiers in the United States Army. Through data collected at USMA, this research provides a way to analyze the character of college applicants (prior to admission) using Natural Language Processing (NLP) techniques and machine learning algorithms. We extract NLP variables from letters of recommendation that were written about college applicants in an effort to predict the number of negative Cadet observation reports (NCOR) they receive per semester, which we use as a proxy measure for poor character. We provide evidence for a positive relationship between the number of NCORs that a Cadet receives per semester and recommendations with high average words per response and a higher than average proportion of negations. However, our results demonstrate that the approach of using basic NLP techniques is insufficient for admissions departments to achieve the very difficult task of assessing college applicants for downstream character issues.

Keywords: Natural Language Processing, Letters of Recommendation, College Admissions, Character

1. Introduction

The mission of the United States Military Academy (USMA) at West Point is to "educate, train, and inspire the Corps of Cadets so that each graduate is a commissioned leader of character committed to the values of Duty, Honor, Country and prepared for a career of professional excellence and service to the Nation as an officer in the United States Army [1]." Inspiring Cadets to be "leaders of character" is part of what makes the mission of USMA unique compared to other institutions. This is because USMA graduates and commissions about 1,000 Officers each year, and military leaders are expected to possess great character, capable of leading large formations of America's sons and daughters in an ethical manner. It is typical for Senior Army Leaders to command thousands of soldiers with access to millions of dollars of resources. As such, senior Army leaders face immense scrutiny when those expectations of character are not met [2]. Leaders at West Point would benefit greatly from being able to identify which applicants possess the necessary attributes to be a successful Army Officer.

USMA is constantly innovating in its effort to enhance the character of its Cadets. For much of its history, USMA assessed Cadets based on three performance pillars: academic, military, and physical. However, in 2015, the academy unveiled a fourth pillar based on character that sought to help Cadets understand the true meaning of serving as a commissioned leader of character through education on Army ethics, honor and personal virtues [3]. Regarding the character program at USMA, the 60th Superintendent of USMA, LTG Darryl A. Williams, said, "We firmly believe character equals readiness and we are allocating resources accordingly [4]."

The mission of the West Point Admissions Department is "to recruit and enroll men and women across the nation and admit diverse, high quality candidates who meet USMA's entry qualifications and are inspired to serve as Army officers [5]." USMA evaluates five attributes in their admissions process, using the following measures of metrics:

1. Character: Recommendations, employer evaluations, interviews.
2. Academic Potential: High school rank, high school/college transcripts, test scores.

3. Leadership Potential: Extracurricular activities, athletic participation, school official evaluations (SOE).
4. Physical Aptitude: Fitness test scores, athletic participation.
5. Persistence: Completing the application process and interviews.

USMA Admissions requires Cadet Candidates to obtain a SOE from four different high school teachers in the following subjects: Math, English, Physics/Chemistry, and Physical Education. These teachers receive twelve different prompts and are asked to rate students on a scale from one to five. They also provide a written statement about the student and their performance in class.

1.1. Assessing Character

Army Doctrine Reference Publication 6-22 (ADRP 6-22) states that character helps a person determine "what is right and gives a leader motivation to do what is appropriate, regardless of circumstances or consequences [6]." ADRP 6-22 goes further and says that character is essential to successful leadership such as empathy and discipline. Cadets at the USMA are assessed in nearly every facet of their lives. Cadets are required to wear uniforms without defects, maintain a professional appearance and haircut, keep their barracks room in a standardized orderly manner, and conduct themselves professionally at all times. Cadets believe in and adhere to the Cadet Honor Code, which states, "A Cadet will not lie, cheat, steal, or tolerate those who do [1]."

Character is paramount to USMA and Army Officership, thus West Point believes in developing the Cadets it admits. To do this, West Point created a character program known as the West Point Leader Development System (WPLDS) [7]. WPLDS expects each USMA graduate to:

1. Live honorably by:
   - Taking morally and ethically appropriate actions regardless of personal consequences.
   - Exhibiting empathy and respect towards all individuals.
   - Acting with the proper decorum in all environments.

2. Lead honorably by:
    - Anticipating and solving complex problems.
    - Influencing others to achieve the mission in accordance with the Army values.
    - Including and developing others.
    - Enforcing standards.
3. Demonstrate excellence by:
    - Pursuing intellectual, military, and physical expertise.
    - Making sound and timely decisions.
    - Communicating and interacting effectively.
    - Seeking and reflecting on feedback.

USMA uses different models, such as WPLDS, to measure and shape character. However, measuring character can be quite challenging. To ensure that Cadets are conducting themselves properly and to acknowledge outstanding behavior, West Point has a system for reporting observations, both good and bad, that all Cadets, instructors, and Officers at the Academy have access to. These reports are called Cadet Observation Reports (COR). There are three types of CORs: Positive (PCOR), Negative (NCOR), and Neutral (not commonly written). To write a COR, a reporter submits an electronic form that takes about a minute to complete. The report contain two key components: the type of COR (Positive, Negative, Neutral) and the comments section where the reason for writing the COR is described. CORs are not anonymous. When a COR is written, the Cadet and their Chain of Command are all notified of the type of COR, who wrote the COR, and the COR comments. In this study, we leverage the number of NCORs a Cadet received as a response variable to assess the character of Cadets at West Point 2. Due to privacy concerns, we do not have access to the comment portion of the CORs, but generally assume that a high number of NCORs indicates some sort of character flaw within a Cadet. Some common reasons for Cadets receiving NCORs are for failing to maintain room standards, missing an assignment in class, general disrespect, failure to complete one's job, or being late or absent to a class or a formation.

To measure the character of its applicants, the United States Air Force Academy distributes a test that considers 12 different dimensions[8]. The Air Force Academy has its Cadets take self-assessments on these 12 dimensions to solicit reports on the frequency of their own behavior. The 12 different dimensions are integrity, honesty, loyalty, selflessness, compassion, competency, respectfulness, fairness, responsibility, decisiveness, spiritual appreciation, co-operativeness.

2. Literature Review

2.1. Issues with Letters of Recommendation

A 2015 analysis of graduate school recommendations acknowledged that writers are more reluctant to use negative language when writing letters of recommendation for an applicant whom they have a personal connection with and when the stakes are high [9]. This same study also found that there were four main challenges in analyzing graduate school recommendations.

- Much of the text in recommendations may express information other than evaluations of the applicant (e.g., information about the recommender, discussion of the evaluation process).
- The text about the applicant may be generic, lacking specific details or examples about the applicant to indicate that he or she is well known to the recommender.
- The text about the applicant may be either positive, negative, or neutral (e.g., negative text is likely to be associated with less qualified applicants).
- If numeric ratings accompany recommendation texts, as they do in our data, then the ratings for any individual recommendation may not be consistent with its text (e.g., different recommenders may interpret rating scales differently).

An approach by Rothstein on the association between NLP and college letters of recommendation found that for underrepresented applicants (applicants from low scoring high schools, first generation high school students, low-income families), there is little difference in the quality of the recommendation [10]. He suggests that this could be due to the lack of adults at these schools who can write strong letters of recommendation.

Using a corpus of 283,676 essays submitted by 93,136 self-identified Latinx public university applicants, a 2020 study was able to calculate the reported household income and gender of these applicants with high degrees of accuracy [11]. This study found that college admissions essay content can explain about one-third more variation in household family income than a student's SAT scores [11]. This study shows that there is valuable data hidden within the writing that is presented to college admissions centers and that this data can lead to important insights on applicants.

Pennebaker conducted an NLP study amongst letters of recommendation he had written for other people [12]. He realized that, on the surface, nearly every letter of recommendation that is written is abundantly positive. He concluded that truly good letters of recommendation possessed the following qualities:

- Say more, use longer sentences and bigger words
- Use fewer positive emotion words
- Provide more detailed information
- Pay little attention to the potential reader of the letter

2.2. Natural Language Processing

Natural Language Processing (NLP) is an area of research and application that examines how computers can be utilized to gain an understanding of natural language text or speech [13].

Modern NLP began with the birth of Linguistic Inquiry and Word Count (LIWC) in 1991. LIWC was a computer program that would analyze and scan pieces of writing in an attempt to determine a person's emotional state, based on the number of feeling words a person used. To improve the capacity of LIWC, different dictionaries were added for different emotion words. In order to create these dictionaries and have them be accurate indicators of a person's emotional state, rooms full of student judges would peruse through lists of words and determine which words portrayed a certain emotion by the writer. LIWC's strengths were its ability to be consistent, its analysis speed, and its ability to compare texts.

In the 1990s, another computer program, Latent Semantic Analysis (LSA), was developed to analyze not the content of what people wrote, but their writing style. LSA determined writing style by focusing on function words, including pronouns, prepositions, articles, and a small number of similar short but common words. Pennebaker claims that in any given sentence, some words provide basic content and meaning whereas others serve quieter support functions. Ironically, the quiet words can say more about a person than the more meaningful ones [12].

There are many types of words that are considered and analyzed through NLP techniques. In this paper, we will focus on the two main types of words: content words and style words.

Content words are words that have a culturally shared meaning in labeling an object or an action. This includes:

- Nouns (table, uncle, justice, Fido)
- Regular and action verbs (to love, to walk, to hide)
- Most modifiers (adjectives and adverbs)

Style (or function) words are words that connect, shape, and organize content words. This includes:

- Pronouns (I, she, it)
- Articles (a, an, the)

- Prepositions (up, with, in, for)
- Auxiliary verbs (is, don't, have)
- Negations (no, not, never)
- Conjunctions (but, and, because)
- Quantifiers (few, some, most)
- Common adverbs (very, really)

In a study done by Pennebaker et al., a formula was created using LIWC and categorical-dynamic index (CDI), an 8-dimensional principal component analysis. This formula assigned quantitative CDI values to college admissions essays to predict academic potential. CDI is an equation composed of eight word categories that account for more than half the words used in typical English writing.[14]

$$\text{CDI} = 30 + \textit{article} + \textit{preposition} - \textit{personal pronoun} - \textit{impersonal pronoun}$$
$$- \textit{auxiliary verb} - \textit{conjunction} - \textit{adverb} \text{ - } \textit{negation} \quad (1)$$

The authors found that CDI directly correlated with the academic performance of college students and college graduates.

2.3. Institutional Measures of Character

3. The Initial Data Set

The USMA Admissions Department provided the data for this study. The data set contained 4,821 observations representing Cadets from West Point's classes of 2016 to 2020. The data encompasses three broad categories: School Official Evaluation (SOE) Information, Cadet Candidate Admissions Metrics, and Cadet Performance Metrics. As this research posed no greater than minimal

Cadets	% D1 Athletes	% USMAPS Attendees	% Both
4821	20.0%	15.5%	5.72%

Table 1: Cadet Candidates Admissions Population Demographics

	Mean	Median
NCORs (All Cadets)	5.79	4
NCORs (Separated)	9.90	6

Table 2: Negative COR Totals

risk to the subjects, it was exempt from Institutional Review Board (IRB) review by USMA.

The SOE data from Cadets includes anywhere from three to nine numerical SOE scores, one from each recommendation a Cadet Candidate receives. These scores are part of the calculation into the Whole Candidate Score. The maximum score a cadet can receive from a School Official is 740 points. Each SOE also contains a written portion, similar to that of a letter of recommendation that students from traditional colleges would receive. Within the data set, we found the average length of a single SOE to be 223 words. From the language used in the SOE responses, we extrapolate metrics of cadet performance and success.

The Cadet Candidate Admissions metrics include CEER Score, which is composed of SAT/ACT scores. They also include the population demographics of West Point attendees, whether or not they were recruited Division I athletes or United States Military Academy Preparatory School (USMAPS) attendees (1).

For this research, we use the amount of NCORs a cadet received as a proxy measure of character (3). We logically assume that the more negative CORs a cadet has, the lower the level of that cadet's character. This idea is supported in the table 2 that show the Mean and Median for NCORs amongst Cadets in our data.

Due to the formal nature of the NCOR and the notification of the command-

	Mean	Median
NCORs	5.79	4
PCORs	6.97	6

Table 3: Cadet Observation Report Statistics (Neutral CORs are excluded due to the small quantity of them)

ing Officer and chain-of-command, Cadets might confront an individual about their behavior rather than writing a NCOR. So, for the NCORs written about Cadets, although they only represent one mistake each that a Cadet has made, there typically are many more instances of observed substandard behavior behind each one.

When a negative COR is written, there is a potential consequence that comes from it. Individuals write NCORs when they want to notify a Cadet's leadership that there is a behavior issue occurring. On the other hand, PCORs are written for all sorts of different reasons (academic performance, outstanding spirit, volunteering, etc.). Thus, we do not believe that PCORs are associated with good behavior as much as NCORs are associated with substandard behavior. For the sake of this study, we assume that all Cadets have generally good character and that PCORs, due to the nature of the COR system, are not necessarily indicators of good character. We do not use the number of PCORs as a response variable.

At USMA, a turn-back is a Cadet who does not graduate with their class on their designated graduation date. A Cadet can become a turn-back for four different reasons: academics, honor, medical, or physical fitness. For example, if a Cadet fails a class and is not able to meet their graduation requirements on time, they may need to repeat a semester in order to meet their graduation requirements. Another example is if a Cadet gets injured and is unable to complete their physical fitness requirements on time. A turn-back still graduates from USMA, but they typically either graduate in December or in May the following year. In this data set, we identified 107 cadets as turn-backs.

Proportion	Mean	Median
Nouns	22%	22%
Verbs	16%	16%
Adjectives	8.4%	8.5%
Adverbs	4.8%	4.7%
Pronouns	9.4%	9.3%
Negations	0.62%	0.58%

Table 4: Calculated Mean & Median of parts of speech

A separated Cadet is a Cadet from USMA that does not graduate and leaves the academy. Cadets can be separated for many reasons, and the most common reasons are academic, honor, medical, physical, or loss of motivation. In this data set, 870 Cadets have been separated and of these Cadets, 24 of them were readmitted and graduated. We noticed an association between higher NCOR totals and lower CQPAs among the population of separated Cadets (**??**).

3.1. Data Extracted for the Study

3.1.1. Natural Language Tool Kit

To extract NLP metrics from the written part of the SOEs, we utilized Natural Language Processing with Python (NLTK), written by Bird, Klein, and Loper [15].

Once the data was prepared, we itemized each word in a given SOE and then classified them into different grammatical parts of speech. For every single written SOE, we computed the number of nouns, verbs, adjectives, adverbs, and pronouns and then converted those numbers into proportions based on the number of total words in each SOE. We also used a similar method to count the proportion of negations within written SOEs. Table 5 shows the mean and median of each part of speech used 4

Some other variables we computed using the NLTK were Sentence Count, Sentence Length, Average Stemmed Word Length, and Average Words per Response.

- Sentence Count: Total number of sentences from an aggregated set of SOEs for a given Cadet Candidate.
- Sentence Length: Average number of words within a sentence written about a Cadet Candidate. Pennebaker predicted that the longer sentences written about a person, the higher regard for the person from the writer[12].
- Average Stemmed Word Length: This is the average length of stemmed words written on a Cadet Candidate. In simplistic terms, stemmed words are words that have both their prefix and suffix removed. We chose to use stemmed words for this category to better measure the length of the root of words. Had we not done this, we could mistake less sophisticated words as words that deceivingly count for a greater length based on how the writer chose to structure their sentence. Pennebaker also claims that when sophisticated language is used by the writer of a letter of recommendation, then typically the candidate is of better quality. [12].
- Average Words per Response: This is the sum of all words written about a Cadet Candidate divided by the number of SOE writers they had (number of SOE writers varies between candidates with a range of 3-9).

We also conducted Semantic Analysis using the Natural Language Tool Kit [15]. From the package "nltk.sentiment.vader," we imported "SentimentIntensityAnalyzer." VADER stands for Valence Aware Dictionary and Sentiment Reasoner. It is a lexicon and rule-based feeling analysis instrument that excels at drawing positive and negative sentiment out of all types of text. VADER is its own lexicon, thus it does not require any training data. VADER is also highly functioning on many different types of text, so it fits in perfectly with this research. VADER reads written text and gives higher positive and negative scores based on different factors. The positive and negative scores given based on sentiment fit the information we wanted to pull out of the SOEs perfectly. These factors include text written in all caps, punctuation, emoticons, booster words (i.e. very, extremely, etc.), slang, as well as negations. For example, for the sentences “He is a good student,” and “He is a very GOOD student!” The semantic analysis would score the latter higher due to the booster word ‘very”, the capitalization of the word “good”, and the punctuation.

From sentiment analysis, we extracted four variables: Compound Sentence Score, Positive Sentence Score, Negative Sentence Score, and Neutral Sentence Score.

- Compound Sentence Score: Sum of the positive, negative, and neutral score within each sentence of a Cadet Candidate’s written SOEs and then normalized to stay between the value range of (-1,1).

- Positive Sentence Score: Calculates the positive sentiment within a Cadet Candidate’s written SOEs using the VADER lexicon and then normalizes the score between (0,1).

- Neutral Sentence Score: Calculates the neutral sentiment within a Cadet Candidate’s written SOEs using the VADER lexicon and then normalizes the score between (0,1).

- Negative Sentence Score: Calculates the negative sentiment within a Cadet Candidate’s written SOEs using the VADER lexicon and then normalizes the score between (0,1).

  The lowest compound sentence score recorded by a Cadet Candidate was in the positive direction (5). Table 6 shows the mean, median, mode, and the lowest semantic analysis score for compound, positive, negative, and neutral score. This helps us conclude that for our data set, written letters of recommendation (SOEs) are extremely positively skewed. Overly positive letters of recommendation have been noted in other academic works [9] and these data are no different.

We also took the numerical SOE scores given to each Cadet Candidate and calculated the SOE Maximum Score, the SOE Low Score, and the SOE average

	Mean	Median	Mode	Lowest
Compound Sentence Score	.41	.41	0.90	.018
Positive Sentence Score	.21	.21	.61	.059
Negative Sentence Score	.023	.022	.108	0
Neutral Sentence Score	.77	.77	.89	.38

Table 5: Mean, Median, Mode, and Lowest Score of Sentiment Analysis

	Max	Average	Median	Low
Lowest SOE Score	740	684	698	323
Average SOE Scores	740	714	720	526
Highest SOE Score	740	735	740	640

Table 6: Numerical SOE Score statistics for all Cadets

Score. The scores in Table 6 illustrate the issue with SOE scores. The majority of SOE scores are very high scores. A reason for this could be because the SOEs we analyzed were from applicants that did receive entry to USMA.

4. Methodology

To begin analyzing the relationship between the explanatory variables and the response variables, we utilized scatter plots and calculated lines of best fit between the data and different subsets of the data. This was a prima facie approach to analyzing the data and from these observations, we were able to predict which explanatory variables had a higher chance of being associated with our response variable (NCORs per semester).

We hypothesized that different subgroups of the data set would exhibit different relationships with resulting different lines of best fit. We analyzed both the "at-risk" and separated on honor subgroups.

For our models, we chose linear and Poisson regression. We compare the linear and Poisson regression models using two common penalized-likelihood information criteria: the Log-Likelihood and the Akaike Information Criterion

(AIC). We can define these as $AIC = -\ln L + P$ where $\ln L$ is the log-likelihood of the model and P is the number of estimated parameters. We show the results of our analysis in Table 7; we found that linear regression creates models that account for 7.7% of variation in the number of NCORs per semester without NLP variables included and 8.6% with NLP variables included. We found that the p-value for both our linear models was $2.2e^{-16}$ 7. In our Poisson models, the AIC is lower when NLP variables are included, indicating that the Poisson model with NLP variables gives a better balance of model fit with generalizability.

Table 7: Summary of Models for predicting NCORs per semester

Distribution Method	Linear	Linear (NLP variables)	Poisson	Poisson (NLP variables)
(Intercept)	8.17	7.84	8.73	8.36
	(0.37)	(0.381)	(0.434)	(.452)
SOE.Average.Score	-0.011	-0.009	-0.013	-0.0112
	(0.0005)	(0.0005)	(0.0006)	(0.000646)
Average.Words.Per.Response		-0.0005		-0.00104
		(0.00009)		(0.0001)
Negation.Percentage		-10.2		-15.9
		(2.99)		(4.53)
Log-Likelihood	-5914	-5890	-4971	-4939
Multiple R-squared	0.077	0.086		
F-Statistic	407	153		
Degrees of freedom	4819	4817	4819	4817
AIC			9946.3	9886.4
p-value (χ^2)	2.2e-16	2.2e-16		

Based on the analysis, average word count and negation percentage in SOE written responses provide a better predictor of the amount of NCORs a Cadet will receive per semester. However, the improvement is marginal and there is still a lot more work that would need to be done.

4.1. Principal Component Analysis (PCA)

To reduce the amount of noise in our data set, we utilized Principal Component Analysis (PCA). PCA is a dimensionality-reduction method that is used to reduce the dimensionality of large data sets [16]. This is done by reducing variables that are redundant in the information they provide and leaving significant variables that account for more of the variance in the data.

There is a trade-off that occurs when PCA is performed. When reducing the number of variables within a data set, the disadvantage is that there is a loss of information (accuracy). The advantage is that there is greater simplicity within the data set, thus, it is easier to explore, visualize, and manipulate the data.

We begin by removing variables that we deemed redundant. For example, we removed "SOE Word Count" and "SOE Sentence Count" and chose "Average Words Per Response" because all of these variables contributed strongly in the same dimension and were colinear by nature. We also created dot plots with lines of best fit for each of these variables. If we were able to conclude that the explanatory variable was insignificant based on the shape of the scatter plot and the direction of the line of best fit, we removed the variable from the PCA. After removing redundant data, and data with weak eigenvalues (Number of Nouns, Number of Verbs, Number of Adjectives, Number of Adverbs, and Number of Pronouns in particular), we ended up with 7 focused variables. The Eigenvalues and Variance Percentages can be found in Table 8.

From PCA we found that the NLP variables that explained the most variability in the data set were average words per response, sentence length, negation percentage, average stemmed word length, positive sentence score, negative sentence score, and SOE low score (See Tables 8 & 9).

4.2. Random Forests

Next, we used random forest models to determine which variables were most important in predicting our response variable. Random forest models get their name, "forests" because they are composed of many decision "trees" that are

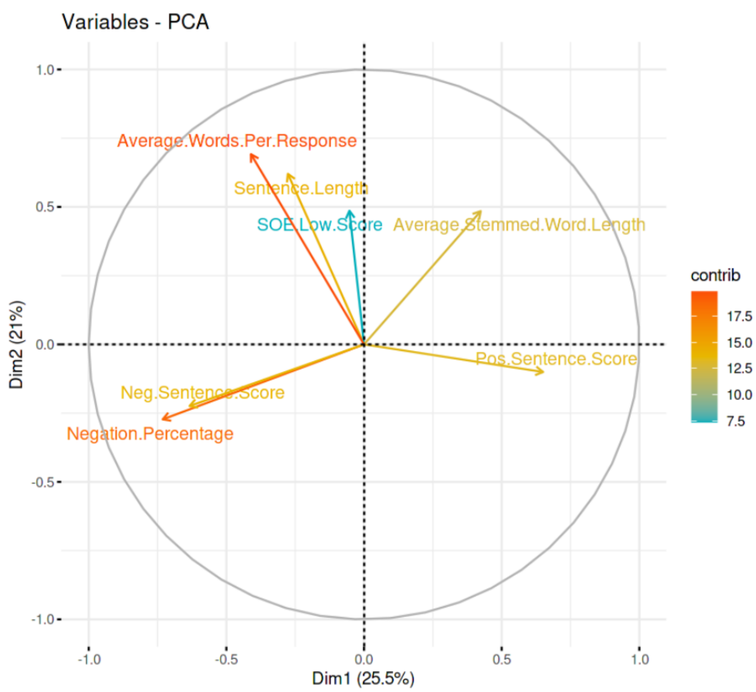

Figure 1: Principal Component Analysis with 7 Focused Variables

	Eigenvalue	Variance Percent	Cumulative Variance Percent
Dim 1	1.787	25.5	25.5
Dim 2	1.468	20.1	46.5
Dim 3	1.109	15.8	62.4
Dim 4	0.853	12.1	74.5
Dim 5	0.695	9.9	84.5
Dim 6	0.581	8.3	92.8
Dim 7	0.503	7.2	100.0

Table 8: Eigenvalues and Variance: 7 Variables

	Dim 1	Dim 2
Sentence Length	4.281	26.195
SOE Low Score	0.157	16.078
Negation Percentage	29.926	5.059
Average Stemmed Word Length	10.052	15.988
Pos Sentence Score	23.676	0.686
Neg Sentence Score	22.465	3.419
Average Words per Response	9.439	32.572

Table 9: Weight of PCA Variables in Dimensions 1 & 2

used to classify data. In our study, we utilized regression random forests and these models combine predictions from the random trees and average them in order to output which variables from our data set contribute most significantly to the variance.

Using a random forest model, we used the seven focused variables from our PCA results to find out which variables provided more importance to the variance in our model to predict NCORs.

What was most surprising is that we found that NCORs per semester is best predicted by average words per response. This follows directly with Dr. Pennebaker's idea that the more somebody is written about in a letter of recommendation, they are more likely to be stronger candidates[12]. We saw this many times while reading through SOEs. Figure 2 shows IncNodePurity- a function that determines how the best splits are chosen. The variables of higher importance within the model have higher IncNodePurity.

Figure 3 shows the relationship between the NCORs a Cadet earned at the academy and the Average Words per Response. The blue line is the line of best fit and helps illustrate the relationship we analyzed.

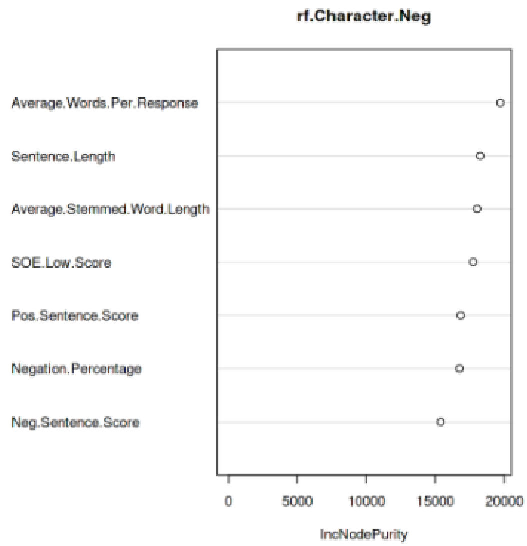

Figure 2: NCORs Important Variables

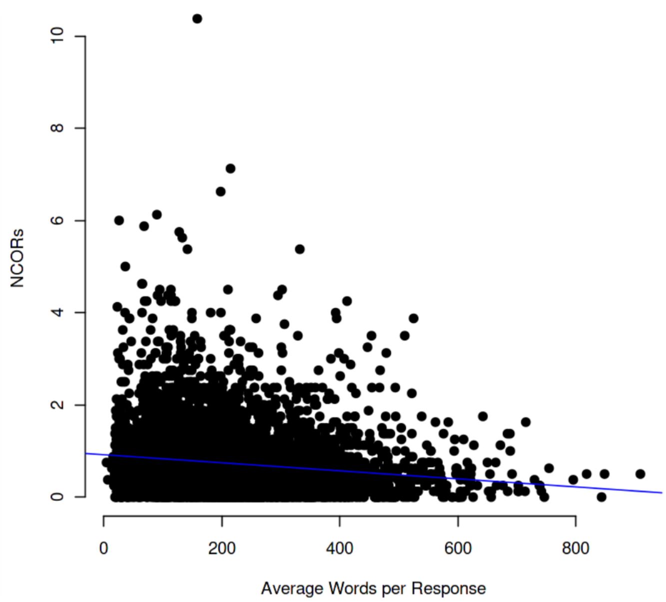

Figure 3: NCORs per Semester vs. Average Words per Response

5. Discussion/Conclusion

5.1. Limitations and Future Work

There are many limitations to our data that need to be considered. First, there was data that was lost during the IRB approval process. For each SOE entry, we did not have access to the title of the evaluator who wrote it even though that information exists. It could have been worthwhile to analyze which SOE evaluator was more significant in determining our explanatory variables (Math teacher, English teacher, team leader, etc.). We also did not have access to the state, hometown, or high school of the Cadet Candidates, and we were not given the predicted family income of a Cadet, nor the education level of their parents. More educated areas of the nation may produce more elaborately written SOEs with larger word counts. We also did not have access to whether a Cadet was prior service military, which may impact separation rates. Further research can be done on identifying other important SOE variables and taking into consideration more variables from the candidates. For example, recognizing and adjusting for confounding variables such as sex and home residence zip code could potentially help explain the differences in NCORs and CQPA. Looking into different subgroups of Cadets, such as Cadets that are division I athletes, Attendees of USMAPS, found on honor violations, etc. could also provide some interesting insights.

Futhermore, we did not have data for the amount of NCORs earned each year as a Cadet over time. Having the NCOR amounts per year could potentially show character growth throughout a Cadet's four years at the academy. It would be interesting for a future researcher to look into not just the amount of NCORs, but if they look at the rate of change of NCORs from year to year. A study analyzing the rate of NCOR changes would potentially show who is going through more character growth. This idea may tie directly to NLP statistics; for example, maybe we can determine which Cadets have more grit [17] and which Cadets will work harder based on the words that are used in SOEs to describe them.

5.2. *Applicability of Work*

What we uncovered is a small but reliable relationship of NLP measures to NCORs, which is important in its own right. This shows that there exists some merit in using NLP methods to measure the character of college applicants. More work is needed to build useful predictive models.

Despite not having NCORs to measure, any college admissions department could use similar techniques on letters of recommendation to find potentially more honorable applicants, to reduce cheating and expulsion rates.

College admissions departments could look into similar processes to highlight the profiles of applicants to find suitable applicants. Using NLP techniques has the potential to help college admissions departments better assess applicants and provides oddity information extracted from letters of recommendation.

References

[1] Simon Center for Professional Military Ethic, Gold Book, The United States Military Academy, West Point, NY 10996, 2019.

[2] R. Chandrasekaran, G. Jaffe, Petraeus scandal puts four-star general lifestyle under scrutiny, The Washington Post.

[3] United States Military Academy, Gold Book, The United States Military Academy, West Point, NY 10996, 2015.

[4] T. Brading, Culture of character growth will build readiness, says west point superintendent, Darryl A. Williams.

[5] West Point Admissions, Admissions Slides Overview, West Point, NY, USA, 2020.

[6] Army leadership, Army Doctrine Reference Publication 6-22, Headquarters Department of the Army.

[7] West Point, The West Point Leader Development System, The United States Military Academy, West Point, NY 10996, 2018.

[8] C. C. Strange, Measuring up: Defining and assessing outcomes of character in college, New Directions For Institutional Research (122) (2004) 25–36.

[9] M. Heilman, F. J. Breyer, F. Williams, D. Klieger, M. Flor, Automated analysis of text in graduate school recommendations, ETS Research Report Series 2015 (2) (2015) 1–12.

[10] J. Rothstein, Qualitative information in undergraduate admissions: A pilot study of letters of recommendation.

[11] A. Alvero, N. Arthurs, A. L. Antonio, B. W. Domingue, B. Gebre-Medhin, S. Giebel, M. L. Stevens, Ai and holistic review: informing human reading in college admissions, in: Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society, 2020, pp. 200–206.

[12] Pennebaker, JW, The Secret Life of Pronouns: What Our Words Say About Us, Bloomsbury Press, NY, USA, 2011.

[13] G. G. Chowdhury, Natural language processing, Annual Review of Information Science and Technology 37 (1) (2003) 51–89.

[14] James W. Pennebaker, Cindy K. Chung, Joey Frazee, Gary M Lavergne, David I. Beaver, When small words foretell academic success: The case of college admissions essays.

[15] S. Bird, E. Klein, E. Loper, Natural language processing with Python: analyzing text with the natural language toolkit, " O'Reilly Media, Inc.", 2009.

[16] Z. Jaadi, A step-by-step explanation of principal component analysis (pca), Retrieved June 7 (2021) 2021.

[17] A. L. Duckworth, C. Peterson, M. D. Matthews, D. R. Kelly, Grit: perseverance and passion for long-term goals., Journal of personality and social psychology 92 (6) (2007) 1087.