2017

Machine Learning And Natural Language Methods For Detecting Psychopathy In Textual Data

Andrew Stephen Henning
University of Mississippi

MACHINE LEARNING AND NATURAL LANGUAGE METHODS FOR DETECTING PSYCHOPATHY IN TEXTUAL DATA

A Thesis
presented in partial fulfillment of requirements
for the degree of Master of Science
in the Department of Computer Science
The University of Mississippi

by

ANDREW S. HENNING

May 2017

ABSTRACT

Among the myriad of mental conditions permeating through society, psychopathy is perhaps the most elusive to diagnose and treat. With the advent of natural language processing and machine learning, however, we have ushered in a new age of technology that provides a fresh toolkit for analyzing text and context. Because text remains the medium of choice for most personal and professional interactions, it may be possible to use textual samples from psychopaths as a means for understanding and ultimately classifying similar individuals based on the content of their language usage. This paper aims to investigate natural language processing and supervised machine learning methods for detecting and classifying psychopaths based on text. First, I investigate psychopathic texts using natural language processing to tease out major trends that appear in the classical psychological literature. I look at ways to meaningfully visualizing important features within the corpus and examine procedures for statistically comparing the use of function words of psychopaths versus non-psychopaths. Second, I use a “bag of words” approach to investigate the effectiveness of unary-classification and binary-classification methods for determining whether text shows psychopathic indicators. Lastly, I apply standard optimization techniques to tune hyperparameters to yield the best results, while also using a random forest approach to identify and select the most meaningful features. Ultimately, the aim of this research is to validate or disqualify traditional vector-space models on a corpus whose authors consistently try to hide in plain sight.

DEDICATION

This work is dedicated to Gregory, Sandra, Rafi, Kyra, and Avalie Knispel-Heyworth, my surrogate family. Without their continued support and encouragement, I'd never be pushed to pursue the craziest ideas and persevere through the hardest challenges.

ACKNOWLEDGMENTS

I would like to thank M.E. Thomas, a diagnosed prosocial sociopath and accomplished author, for helping me wade through the ins and outs of my thesis and for being the backboard upon which I pitched my strangest ideas. I would also like to thank Hannah Gurley for helping me visualize my ideas with her artistic talents in order to tell an otherwise unpleasant story. Finally, I would like to thank Dr. Naeemul Hassan, who dedicated his precious time to me when he, as a new professor, only had moments to spare – I followed his advice to the letter, and I hope I've made him proud.

TABLE OF CONTENTS

LIST OF FIGURES

SECTION I

INTRODUCTION AND BACKGROUND INFORMATION

Speaking in a broadcasted interview, neuroscientist and writer Sam Harris first qualified, then commented on the use of language to indicate a person's mental health with the following analogy:

> When people are speaking, they are thinking out loud. I am thinking out loud at this moment. If you listen to my podcast for a few hours, you know how I think. So, when people don't make sense, it's not like people are making brilliant, incisive thoughts in the privacy of their minds and they just sound like dummies when they open their mouths. Generally speaking, what you hear is what they've got… Imagine you have an urn, and every time you reach into it you pull out another piece of junk. You've got broken glass, zip-ties, and bits of bone—nothing of value. While it might seem unlikely, it's not impossible that something of tremendous value is also in there. You could pull the Hope Diamond out of there if you keep fishing around long enough. That is possible because what you pull out each round doesn't really indicate what else is in there. Minds are *not* like that. Ideas are connected. The ability to reason well is transferable from one domain to another, and so is an inability to reason. A desire not to seem incoherent, this is something that intelligent, well-informed people have [1].

Certainly, this is not to say that every illogical thinker is psychopathic or even on the spectrum, but Harris' point stands: when people use language, they demonstrate what they bring to the table intellectually. By intellectually, I do not mean some strand of brilliance and cleverness that is often misconstrued as intellectual, but rather the capacity to appear to reason well. Someone with intelligence may be devoid of any domain-specific knowledge, but when presented with evidence and supporting statement, he can supply logic to buttress an appropriate reaction. The reasoning capacity is just one example of how a person's use of language might be telling about their overall mental state, but there might be more nuanced conclusions which can be drawn. In this paper, I search for the more extreme categories of people by wading through their speech patterns. I deploy the tools of natural language processing (NLP) and supervised machine learning (SML) to first parse a mass of text, then train models based on the data. Ultimately, my aim is to visualize any correlations between language and its use with respect to psychopathy.

Definitions and Initial Positions

The most natural question to follow any thesis is one of definition. This paper is dedicated foremost to using NLP and SML to detect potential psychopathic patterns in text. The starting point, then, is to provide a proper definition for a "psychopath." As society progresses, it has gradually shifted to prefer the term "Antisocial Personal Disorder" [2]. The reason behind this shift seems to be one of ameliorating the condition. Over time, the term "psychopath" has been traded for one with a kinder connotation. Popular culture has damned those deemed psychopaths or sociopaths. It is true that some suffering from the condition express and act upon criminal intent, but many do not. There are psychopaths distributed among society, obeying laws and contributing to social betterment. Whether this is out of genuine concern for others or an act of pure self-preservation is yet to be seen. In her popular book "Confessions of a Sociopath," M.E. Thomas, herself a sociopath[1], details her own achievements as an accomplished attorney, Sunday school teacher, and community leader. Her exploits explain the functional veneer by which these individuals must live [3]. Although emotions are foreign concepts to them, their induction into a healthy societal role is not insurmountable.

It is important to note that this is not a entirely new idea, although applying it to NLP and SML may be. The field of psycholinguistics deals with precisely this issue [4]. Psycholinguists look more generally at how language influences psychological processes. It ventures into the

[1]Thomas prefers to be called a "Sociopath" over a "Psychopath" despite having been diagnosed as the latter. In her book, she explains that the term psychopath carries too much political baggage.

realms of language acquisition, language use, and language comprehension. While the first two are irrelevant to the topics discussed in this paper, the last field—language comprehension—could offer some insight, either historically or conceptually. Interestingly enough, the field of psycholinguistics came into fashion in the same time period as the seminal work *The Mask of Sanity* a book penned by Hervey Cleckley to more clearly define the conditions that must be met to classify a psychopath from other, more similar disorders, like sociopathy, narcissistic personality disorder, or antisocial behavior [5].

Another noted approach has evolved out of the field of literary criticism. Originally known as "poetic stylistics," a field known as "cognitive stylistics" took shape [6]. Like psycholinguistics, there is a strong sense of understanding psychological tendencies from language usage, but unlike psycholinguistics, cognitive stylistics exhibits more analytical fervor by drawing on literary criticism, stylometry, and general stylistics. Further, psycholinguistics looks more at content and *how* something is developed and spoken or written, whereas cognitive stylistics looks to generate a mental model by examining style, which in this sense is a word used to signify the intersection of word usage, frequency, and syntactical structure.

Generally speaking, NLP is a canopy term referring to the overlap of a number of fields. Arising from the areas of computer-assisted language analysis, NLP draws from computer science, linguistics, and artificial intelligence (or machine learning) [7][8][9]. In a narrower way, however, we may define NLP as any interaction between humans and computers where natural language is applied to computation. More specifically, natural languages in this sense are any languages that come directly from human thought and expression. Further, NLP can be separated into constituent parts that look at different parts of language. The Oxford Handbook of Computational Linguistics [8] lists no less than thirty-eight applications and areas in which NLP and computational linguistics operates. But even so, this is only a fraction of what the greater area includes, ranging from phonology to textual summarization, just to name two. In this paper, I refer to NLP as any computational method by which we can characterize an individual through their use of language. The methodology section will expand on the particular toolbox I intend to use, but NLP does aid in categorization tasks.

Finally, I must mention the specifics behind machine learning and how I use the term. The topic of machine learning has grown in popularity in the past few decades and shows no signs of slowing in development or research. There are three main areas of interest: supervised learning, unsupervised learning, and reinforcement learning (also known as semi-supervised learning). Classically, supervised learning is provided as a tool to researchers that have some set of labeled upon which a model can be trained to produce later predictions. Unsupervised learning, by contrast, does not enjoy this luxury and so must find a way around this dilemma. Reinforcement learning is the branch that rests at the intersection between the two and can be thought of as borrowing methods from both but with the added benefit of "spot-checking" for accuracy and tuning purposes [10][11].

For the purposes of this research, I put unsupervised learning and semi-supervised learning aside in favor of supervised learning for two main reasons: time and data size. The scope of this paper allows me to create labeled data required for training. Although unsupervised machine learning excels at clustering data together and clumping similarities together from a joint feature space, I hypothesize that a supervised approach may yield more accurate results. Conceptually, this makes the most sense, and I will later use SML algorithms to sift through the data and find categorical similarities between potentially psychopathic writing samples.

Scope

The scope of this paper is necessarily limited. Though strides have been made in NLP tools, such as sentiment analysis, parsing, lexxing, and stemming, much work still needs to be done to be able to analyze text in its greater context. Usually, syntactic structure is somewhat telling if it can be extracted, but in a world where language is used in multitudinous ways, there is often little or no consistent structure across source and genre. It is for this reason that I cannot reasonably expect syntactic tools to yield any reliable results given so much variety without extensive preprocessing and ontological tools. For example, any sentences devoid of grammatical indicators, such as commas, full-stops, exclamations points, and question marks render a whole host of tools mute, but it is often the case that such structures arise in quickly scrawled letters, partially given interviews, or broken English.

Moreover, for the sake of brevity in this paper, I must also limit the amount of data that I analyze. The methodology section of this paper will detail the imposed restrictions in this area, but the amount of data used here will be sufficiently large to determine some results or at least provide a mechanism by which to expand later. The era of "Big Data" has brought with it a reinterpretation of data processing and may, in fact, be the answer to such problems. However, I must sidestep this issue for this research because, as will be explained later, the acquisition of data among psychopathic writers is particularly wrought with difficulty. So, while I acknowledge that the amount of data used for this research is comparatively small with respect to what is normally expected in the era of big data, I also contend that the amount of data acquired for psychopathic writing samples and non-psychopathic writing samples is adequate given the subject matter and its inherent novelty.

Some Essential Counter-arguments

In order to establish a strong foundation upon which to base this paper, rhetorically I must visit some counter-arguments outright. Interestingly, among these exists one particular argument that is particularly biting, which I will outline and also revisit upon the concluding remarks of this paper. As Sam Harris mentions, when people speak they are essentially “thinking out loud.” What Harris is suggesting is that there are a minimum number of filters through which people naturally feed their ideas and subsequent statements. Writing may be different in that a person further edits their statements. The number of filters may increase, further hiding or softening their true intent. For some, this may cause a problem, because as will be elaborated upon in the next section, the “mask of sanity” is pulled more tightly over a person’s personality.

To this end, I actually see no problem. Modern psychological analysis is done either by a series of interviews or written personality inventories. Considering the circumstances, a person who is already presented with the opportunity to hide or soften their statements to the analyst. In other words, when a person knows that they are under the judgment of trained medical professional, they may exercise some restraint in what they say anyway. Further, in the latter case of a written personality inventory, people are presented with a series of statements. Often it is easy to choose the most positive option of a set given. For instance, given the sentence: “I work better alone,” may indicate a lack of social skill. The more intelligence test taker might detect this question’s intent and answer “Strongly Disagree” just to skirt the question and come across as more social and team-oriented. Trained psychologists are very aware of this and have built into their inventories measures to identify such deliberate avoidance [3]. My main point is not necessarily that we should ignore the chance that people will posture themselves in the most amicable light, but rather that this skewing of data is already present in methods used and accepted in common practice.

One famous account of a psychopath’s skill to avoid detection was Jeffrey Dahmer. In the

1980s and 1990s, Dahmer committed no less than sixteen murders. He managed to not only avoid detection by investigators but lure in his victims out of sheer charisma. Numerous accounts exist that describe how his victims were enticed or even charmed by his personality. This is known as "superficial charm and 'good intelligence'," and the next section will go into more detail about how this is actually within the symptom cluster of a psychopath. Again, it must be stressed that intelligence in this context is not great knowledge, but an ability to learn and reason well convincingly enough to achieve an end. Likewise, much was said about Ted Bundy, who exhibited a certain "likableness," demonstrative of the superficial charm described above.

While we are not sure where exactly from where this charm is acquired, we are getting a general idea of how it develops. As children, psychopaths learn through experience, much like non-psychopathic children; however, the main difference lies in that a psychopath is learning on a deeper level, one where specific reactions to their behavior are measured and replicated in the future. Because they cannot feel shame or remorse as non-psychopaths do, they must construct a model of the world where they survive by remembering what is positively viewed by society and acting on it.

Another counter-argument is just as compelling and difficult to deal with. In the same interview with Sam Harris, he brushes up against this problem:

> A desire not to seem incoherent is something intelligent people, well-informed people tend to have. When you hear someone speak at length at topics that are crucial to the most important enterprise they are engaging in and all they've got is bluster, bombast, and banality strewn with factual errors it is quite irrational to believe that there is a brilliant mind hiding behind that just waiting to get out [1].

The problem is simple: how do you handle someone who clearly uses, as Harris describes, "bombast" to cover up his true intentions. What may be present is some element of superficial charm, but there is lacking good intelligence. Such people have developed a framework to understand the world which hinges on delusion. Fact-checking may be the answer to this and a contributing part to any analysis, but as websites like Snopes.com have shown, this requires excessive amounts of background research and extra data not necessarily easily available. To fact check any political figure would mean having access to a number of resources that complement each other. Moreover, there are relative "gray areas," where a fact may be partially true or wholly true. There is no hit or miss option, but rather a sliding scale.

In the next section, I will discuss the formula used by Hervey Cleckley to detect a psychopath; however, I must anticipate the following: within his formula, there are elements that are, as of this writing, impossible to detect. I will elaborate on each element in the next section and touch on them in the "methodology" section. But in general, it should be stated that every analytical algorithm within the realm of NLP and SML has holes. There are data that slip through the cracks. Edge cases plague analysis, and it is up to the researcher to adequately hone their models to counteract it. The problem with this, though, is simple overfitting. Overfitting is the phenomena familiar to machine learning specialists and data scientists alike. The principle at work says that a model too well trained loses its predictive power. It cannot be depended upon to predict in which cluster data should be placed or in which category some fresh data should be sorted. Combating this means relying upon some normalizing measure, but even then this is not a cure, but a bandage to put a stop to the bleeding. Some noise is good, perhaps healthy. The main problem I encounter is how to handle a relatively small data size with respect to overfitting. With very little data to train and model on, there is great temptation not to treat subsequent validation and testing sets with the correct levels of care. In order to avoid this, I discuss the appropriate measures taken in Section III: "Methodology and Results."

Impact, Ethical Considerations, and Implications

There must be some discussion on the potential impact of this research. In essence, this paper aims to create a model that could potentially predict whether a person's writing exhibits traits of psychopathic personality disorder. In a very real way, this research is developing a means to evaluate the cognitive dissidence of another individual. Medical professionals award these types of diagnoses with extreme care, because to do otherwise is to potentially upend that individuals personal and professional lives.

As is explained later, much of the data I use comes from the public domain. Public data sets are difficult to work with for a number reasons. Ethical considerations are somewhat less clear and moral thought cannot be ignored. While there is a trend, however popular, to perform research for the sake of knowledge, research need not lose ground on intent. On one hand, there is an argument to anonymize any findings, but on the other hand it might be in the public interest to leave the data as is, showing some but not all examples of accounts that exhibit specific behavioral trends for demonstration purposes. As a personal note, the implications of certain medically-motivated research, such as any that addresses mental health, should attempt to further the "human condition," -- that which defines the existence, understanding, and prospering of man -- and be open to delicate interpretation and guided by good intent. It is the opinion of your author that while this research desires to establish a method to throw "red flags," detecting and warning against dangerous individuals, any further work moving in the same direction on the same or similar topics proceed with a careful hand. Albert Camus, in contemplating the worth of good intentions warns us: "The evil that is in the world almost always comes of ignorance, and good intentions may do as much harm as malevolence if they lack understanding."

Paper Structure

This first section has provided some cursory context to the definitions, scope, counter-arguments, approaches, and impact of using NLP and SML to model psychopathic writing styles. Section II is designed to establish a portrait of the psychopathic personality, because in order to accurately study the condition in which I aim to detect, it is necessary to outline what authorities already look for. Section III outlines the methodology for this paper and is directly drawn from the theory previously discussed in Section II. Section III also includes the preliminary results followed by subsequent attempts to tune the pruning and modeling. Section IV presents the final discussion and attempts to explain why the results follow from the method. Section V revisits the highlights of what the research yielded and contains an element of “looking forward,” whereby I make an argument for future work.

SECTION II

PORTRAIT OF A PSYCHOPATH

The late 1930s and 1940s bore an interesting set of problems for mental health professionals. Across the healthcare landscape, droves of unique yet unclassified individuals swarmed the hallways and rooms of inpatient sanatoriums and hospitals in the United States and Europe. There was something noticeably wrong -- otherwise normal individuals exhibited strange characteristics. Some of these people, often charismatic and seemingly well-meaning, were admitted based on the imploring of their family and friends. Each person carried with them eerie stories. Some of these stories involved criminal intent and behavior, while others were of a more perverse and deviant nature. This only came out under close inspection by analysts who spent great deals of time with the patients [5]. The dilemma was this: what exactly plagued the minds of these people? How were their stories all tied together?

Hervey Cleckley, a freshly minted psychiatrist from the University of Georgia Medical School, thought there must be some unifying characteristics connecting them. After years of case study and solitary research, he produced his seminal work titled *The Mask of Sanity*, which is still required reading for Harvard psychiatric residencies to this day. The work was atypical -- not produced in the classical "medical literature" fashion. Instead, Cleckley opted for a tome cataloging a series of interactions between him and specific patients. He interviewed these patients extensively, taking notes on their stories and personality traits, when he finally made a proposal for the long-awaited and much needed "psychopathic personality."

This section is presented with dual purposes, one part psychological theory and one part computational theory, woven together to create a portrait of a psychopath and the traits he may possess. In the psychological evaluation subsection, I take a look at Dr. Cleckley's diagnostic symptom cluster in detail while compiling additional data to further hone the idea of what makes up a psychopath. At the same time, I occasionally mention what computational tools might be used in insisting with diagnosis. This section is necessary because in order to properly set out a methodology for analyzing psychopathic and non-psychopathic text, I must properly define the profile I seek to investigate.

Psychological Foundations: "The Mask of Sanity"

As previously stated, despite Cleckley's unusual approach to the source material, the *Mask of Sanity* was wildly popular at the time of its inception and writing. Its writing comes in three important sections: an outline of the problem, a series of character case studies, a profile for diagnosing like-minded individuals. For my use, the most value is to be found in the last section, where he expounds upon a method for detecting psychopaths.

Cleckley proposes that the psychopathic personality has some combination of the following traits [5]:

1. Superficial charm and good "intelligence."
2. Absence of delusions and other signs of irrational "thinking."
3. Absence of "nervousness" or psychoneurotic manifestations.
4. Unreliability.
5. Untruthfulness and insincerity.
6. Lack of remorse or shame.
7. Inadequately motivated antisocial behavior.
8. Poor judgment and failure to learn by experience.
9. Pathologic egocentricity and incapacity for love.
10. General poverty in major affective reactions.
11. Specific loss of insight.
12. Unresponsiveness in general interpersonal relations.

13. Fantastic and uninviting behavior, with drink and sometimes without
14. Suicide rarely carried out.
15. Sex life impersonal, trivial, and poorly integrated.
16. Failure to follow any life plan.

Cleckley dedicates time to each trait, describing in detail what qualifies each one. In the same way, I will briefly discuss what is meant by each item in the list while also touching on what computational tool may be helpful for identifying these traits in individuals. It should be noted that there are some items in the above list that, at the present time, are either immensely difficult or outright impossible to detect with current tools. There is not much to do in these cases, aside from listing which items these are affected and bringing them to the attention of the reader.

Superficial charm and good "intelligence"

The first item in Cleckley's diagnostic pattern is "Superficial Charm and Good 'Intelligence'." In his text, Cleckley is quick to point out that this trait in particular is not ubiquitous: "More than the average person, he is likely to seem free from social or emotional impediments, from the minor distortion, peculiarities, and awkward so common even among the successful. Such superficial characteristics are not universal in this group, but they are very common" [5]. Since *The Mask of Sanity*'s publication, other authorities in the area have parroted the same conclusion [5][12][13]. What they mean here is that the person demonstrates an outward, showy appearance. Someone bearing this trait is likely to show some level of convincing demeanor supported with the appearance of good intelligent thought.

Theoretically speaking, this are a few ways to handle this. One route might be to look at a user's use of logical words. Term frequencies including "prove," "therefore," and "conclusion," -- as well as their frame or semantic derivatives -- would show a person's attempt to apply logic, but it says nothing as to whether or not that logic is valid or sound. This does not prove to be much of a problem, because a psychopathic individual need only to appear convincing to his audience. Even demagoguery and appeals to emotion can possess some logical structure and have been in the past proven to been convincing to those unaware of their fallacious appeal. Regardless, we can still stylometrically quantify these terms against the rest of their language in their profile, as well as those in other profiles. Although such analysis is not taken in this paper, it would be interesting to see whether any of the key features (i.e. those which bear the most importance to the model) reflect any of these terms.

Absence of delusions and other signs of irrational "thinking"

Absence of delusions and other signs of irrational thinking goes hand in glove with the previous trait. Cleckley uses the traditional definition of delusion. In modern medicine, a delusion is not a misinterpretation of reality or societal norms, although I will revisit this later, but more literally an absence of a different psychosis: schizophrenia, obsessive-compulsive disorder, hallucinations. Where good intelligence exists, there must also be a lack of irrational thought by definition. It stands to reason that one way to deal with the previous issue of detection is to track whether language indicates either delusions or irrational thought. This can be accomplished by noticing antagonists to the previously mentioned terms or to detect unsubstantiated claims. Semantically speaking, finding antonyms to typical words of rational thought has value, but it may not be as sharp an analytical knife in some respects.

For instance, antonyms to logical representation may also be used to demonstrate accurate rational thought. Propositional logic can use negation to bolster an argument while avoid fallaciousness. The remedy for this would be to look at antonyms typically associated with charming terms. This means pursuing one of two approaches. First, I could track positive statements using some form of sentiment analysis, or second, I could use frame semantics to classify logical statements as likely rational or irrational.

Absence of “nervousness” or psychoneurotic manifestations

Absence of nervousness or psychoneurotic manifestations is the second of the only two defining characteristics where personality is devoid of typical, established normalcy. Even Cleckley confesses that this may not be detectable, and is, in fact, one trait easily covered up [5][14]. Someone with a psychopathic personality learns over time which responses are expected from certain situations. Coupled with their general lack of remorse and shame and sense of responsibility, there may not be much anyone can do to detect it in personal interview or by a computer making statistical estimations and inferences.

Computationally speaking, there is one tool that NLP practitioners use regularly which might provide some solace: word tagging. This is in contrast to the more popular “part of speech” tagger (POS) that is designed to identify specific parts of speech. Word tagging is designed to recognize positive and negative words outside of context. But the waters are further muddied by traditional psychological insight. Visible agitation is something well-monitored in interviews, and covering it up is a simple chore for someone trained to do so. Regardless, there might be some computational value to identifying textual snippets that display angry or nervous tones. In the next two subsections, I will discuss what resources are available to both computer scientists and psychology specialists. Of the different tagging strategies, my research uses POS tagging. Partly based in the work of Dr. James Pennebaker, the goal for choosing POS tagging is to compare the occurrences of certain parts of speech in both corpuses [15].

Unreliability

Unreliability is one of the more interesting criteria by which psychopathy is examined. On the surface, many might assume that unreliability is a simple concept, one where an individual cannot be depended upon to complete a task. However, there are deeper, much more important themes to distinguish here, only one option of which can be considered psychopathic. When a person under consideration is evaluated as unreliable by a layman, it typically means that there is some degree by which they may or may not follow through with an action where the promise to do so was made. This may be called “consistently inconsistent” [5]. A person who is consistently inconsistent is someone who may regularly miss bill payments, “stand up” a date, or fail to do homework for a necessary course. But this is not exactly what a psychopath does. Instead, a psychopath is “inconsistently inconsistent.” In other words, he exhibits the unreliability of a neurotypical individual, but only partially. To extend the examples above, this would mean that he occasionally pays his bills, may only sometimes “stand up” a date, and halfway finish his homework. The factor of unpredictability in either direction is stressed here, placing further strain on the means by which we can classify a psychopath.

A computational approach to this would mean a routine checking of not how often a person regularly confirms or does not not confirm the achieving of his goal, but rather the tracking of the number of times his language indicates that he has followed through with his promise versus the number of times his language indicates that he has not. Two approaches come to mind in order to do this: sentiment analysis, and language inquiry word count analysis. Sentiment analysis could be used to gauge a measure of positive, action-oriented statements against neutral or negative ones. Likewise, language inquiry word count can track the specific domains in which words fall [16]. By knowing in which psychological domains a word falls, it is possible to count the number of positive domains against more neutral or negative ones. Both of these methods are similar in that the computer must draw inference from the language itself and cast a verdict on how the language is used, but they vary slightly in that sentiment analysis is interested in the spectrum of positivity and negativity, while language inquiry word count speaks to categorizing individual terms across domains [16].

Untruthfulness and insincerity

Dr. Kevin Dutton, an Oxford University Psychologist, was interested in what helpful heuristics the psychopathic personality might lend those possessing neurotypical brain states. His book, *The Wisdom of Psychopaths,* describes the extent to which the professor went to in order to pursue his inquiry. In the United Kingdom, there exists a sanatorium whose inhabitants range from the estranged mentally ill to deranged serial perpetrator. This asylum, known as Broadmoor, has been the historical home of such notable patients as Christina Edmunds (the “Chocolate Cream Poisoner”) and David Copeland (the “London Nail Bomber”).

Dr. Dutton traveled to Broadmoor to interview such patients and inquire what non-violent advice they may have for the public [12]. Although few were unwilling, those among them whom were open to discussion spoke of lateral methods for achieving their ends. When asked how to eject difficult tenents their methods turned first to violence then to deception. This brings us to a problem that is particularly wrought -- the simpatico centers of the brain misfire and immediately interpret a request as an “all cost” situation as the default position. To these ends, they rely on their untruthfulness and insincerity toolkits [12].

The majority of modern-day sociopaths and psychopaths are those found prowling around “high risk - “high reward” situations. The atypical among them are the Ted Bundy and John Wayne Gacy archetypes. More likely, we are to see the CEO or Wall Street executives who, at the stroke of a pen, are willing to “ax” the wellbeing of a corporate worker and instead provide for themselves the “golden parachute.” Wall Street deals are cut with the bottom line in mind. Lying, indeed stealing, and misleading the competition across the table is commonplace. So, it is the short-sighted maneuvering of the psychopath to belie long-term goals and embrace and untruthful ploy to, in Kevin Dutton's hypothetical posed to the Broadmoor inmates, achieve their own ends [12].

Lack of remorse or shame

Psychologist Kevin Dutton also emphasizes that to be a psychopath is to have a psychopathic brain state, that is to say a state of the physical brain which would appear under testing conditions. Such images are common and can be acquired through the use of an fMRI (functional magnetic resonance imaging), PET scan, or analogous procedure. As he and the neuroimaging specialist James Fallon point out, when exposed to fMRI testing, the empathy and shame centers of the brain fail to ignite when stressed under situations where a neurotypical brain would [12][13].

This lends credence to the long-held Clecklean belief that remorse and shame are somehow numbed or toned down within a psychopath's brain. Dutton imagines the situation through the analogy of a sound recording booth, where thousands of dials and switches cover boards fashioned for different auditory purposes. In this analogy, the dials governing remorse, empathy, and shame are turned down, allowing other "sounds" to dominate in their place. Researchers, Dutton included, have taken experimentation and validation of this concept to the brink, being able to produce in a laboratory setting methods that voluntarily switch off those important centers, if only for a short time [12].

Neurotypical individuals exhibit some degree of frequency to apologize when confronted with an expectation to do so or remorseful inner-feelings. Apologies in most cultures express themselves with some element of natural language. In fact, hallmark words and phrases such as "I apologize" and "I'm sorry," are as well-established and accepted as sentiments like "I love you." It should not be hard, then, to find the semantic markers of apologetic phrasing and tracking them for further analysis computationally. Conversely, it would be interesting to see if such features in a trained model ranked highly among the most informative.

Inadequately motivated antisocial behavior

Again, there must be a differentiation between the terms “antisocial” and “atypical.” Antisocial is a term typically used to describe someone whose behavior actively works against societal norms. Ted Bundy and Jack the Ripper could be considered antisocial, but not necessarily asocial. Bundy, for example, was especially well-liked and managed to avoid his own capture for years. Instead, asocial is a word reserved for people who shy away from the public. Asocial people do not participate in community events, nor do they seek ways to circumvent isolation. In fact, asocial individuals enjoy minimum interaction with the public [2].

Since antisocial behavior carries with it such a high degree of variation from society to society and among many time-periods, it may actually be impossible to currently detect anything that is categorically “antisocial.” To do so would require tracking many other things concomitantly, thus demanding a meta-task in order to make any real progress. Oddly enough, the term “antisocial” was chosen by the American Psychiatric Association as an alternative to the term psychopathy [2]. So, by their choice in terminology, they may have inadvertently loaded a disorder with a self-defeating word. That is not to say, however, that even if someone or something is “antisocial” that it is completely impossible for to detect at all. True, psychiatrists without the aid of machines and algorithms find the task of detecting antisocial behavior difficult at times, but given enough time, patient history, and insight the job is made easier. In computer science terms, this speaks to the necessity of accessing, training, and modeling using a large enough data size, which for this project is addressed in the opening to Section III.

Poor judgment and failure to learn by experience

At first glance, it may seem counterintuitive to think how “Poor judgment” could exist in the same symptom cluster as “good intelligence,” the first trait listed by Cleckley. But upon deeper analysis, it is possible to see how the two traits could co-exist. Intelligence in the Clecklean sense refers to the calculating logic by which a psychopath seemingly operates. Indeed, this logic is said to be so convincing as to come across as “superficial charm.” The key feature that underlies the two, however, is one of time [5][17].

Judgment and experience are both tied together with a temporal relationship. When one casts judgment, often it adds to his own experience. As experience gains ground over time, these judgments should become more informed via a feedback looping mechanism. For the psychopath, this is not the case. Psychopaths are unable to play the “long game,” instead opting for an immediate, often-shallow payoff. So, while their logic and charm may seem convincing in the short term and may in fact generate benefit, it does not fit in with the longer, wiser decision. Therefore the judgment that they make informs their experience in the sense that they now know how to earn a short-lived reward, but go on uneducated as to the most gainful, sustainable path.

Pathologic egocentricity and incapacity for love

Cleckley’s original analysis of “Pathologic egocentricity and incapacity for love” has since been walked back from its original implementation in *The Mask of Sanity*. Originally, Cleckley posited that psychopaths were unable to hold anything beyond a very shallow, perhaps even surface-level, relationship with another. The first edition shows an unyielding position that no psychopath in any capacity could love another, be it familial or otherwise[2] [5]**.** Of course, this tied in strongly with the underpinning themes of egocentricity and pathological manipulation, but in recent decades, progress has been made to show otherwise.

[2] Every edition of *The Mask of Sanity* is divided into three parts. His outline of specific characteristics for the psychopathic profile is found in Section 3, part 3: “A Clinical Profile.”

As Dutton notes, some estimates of the number of psychopaths who permeate society approach nearly one-percent of the whole population, a statistic repeated by M.E. Thomas. The fact that the number is so high cannot only be due to outliers and genetic variation. If such were the case, we would expect the frequency to be much lower. Instead, it is probably the case that psychopaths, especially those who fall in the “prosocial” category, are more in tune with love than Cleckley first thought. As Dr. Fallon points out, along with the evolutionary programming of a species to survive, such would indeed provide the noticeable jump in psychopathic presence in society. When faced with the absolute knowledge of life or death, he inquires, who would you rather have on your side? Someone with abounding compassion, or a person with a palette capable of handling the unsavory decisions that must be made?

General poverty in major affective reactions

Due to the inherent interdisciplinary nature of this subject, there is an overlap of jargon in many circumstances. In order to completely understand what is meant by “major affective reactions,” we must seek an understanding for the psychiatric use of the phrase. In psychology and related fields, a major affective reaction is simply an emotional response where appropriate [18]. What Cleckley and others with psychological training mean when discussing the “general poverty” in these area, they are addressing the concern that people with this psychopathic trait are unable to respond adequately where a neurotypical person would. For example, a neurotypical patient would generally increase his own anger and passion while participating in a heated political debate. However, this is not the case for a psychopath. A psychopath, on the other hand, is able to demonstrate restraint. The archetype of the “cold-blooded” killer applies here. There is no grief, anger, screaming and yelling, or weeping in major events in the psychopath’s life.

What makes it difficult is, as we saw in the subsection discussing lack of remorse, that a psychopath can imitate these reactions when he recognizes that the situation needs one. Cleckley describes this, insisting that there is a difference between a psychopath's "readiness" to marshal an appropriate response and a neurotypical patient's default emotional projection. Along with a few of the other traits that may be difficult or impossible to process using ML or NLP, this is especially true in this instance for two reasons. First, we have the common issue of detecting when the psychopath is lying or purposefully misleading someone. Second, we also see a need for context. Without being able to detect a situation's context and predict the appropriate emotional response, we cannot weight the response in favor of or against the psychopath's possessing this trait. For these reasons, "General poverty in major affective reactions" has also been excluded in the methodology section of the paper.

Specific loss of insight

Again, we must unpack a term reserved for the psychological aspects of this paper. While the previous subsection went into a broad area discussing general areas devoid of a psychopath's consideration, this subsection and its trait focuses on the specifics. Notably, a specific loss of insight would be when a psychopath does not recognize when (or does not act in an instance where) his own future or another's future would be at stake. Indeed, this is certainly true of short term actions and their consequences, but it is especially true for instances that have long lasting consequences. To revisit a previous point established in the section discussing "Poor judgment and failure to learn from experience," this trait goes hand in glove with the psychopath's tendency not to consider the "long game" and instead defers to actions that would benefit him in the short term. Moreover, we may also consider this an expansion of a lack of the psychopath's capacity for empathy. In this way, being able to disregard the feelings of others aids and abets a predatory approach to getting what he wants.

As the Internet ages and is accepted in more areas in society, this type of behavior becomes more common[3]. Interestingly, there might be an argument to be made where the inherent lack of personal connection to those spoken to may further encourage similar behavior. When a person is anonymous and speaking to a fairly anonymous audience, he may feel more liberty to say things that might not come back to haunt him later. Compound this with the instant gratification received by people who share interests or by those performing similar actions and such psychopathic tendencies may actually become the expected behavior, instead of the encouraged behavior. Modern society boasts a make-up of about one-percent psychopaths. Could the digital age be shifting towards a more psychopathic world? Or is the current structure founded in such a way that people are only expressing what is already present in their personalities?

Unresponsiveness in general interpersonal relations

Hardly any relationships are immune to interpersonal communication. Occasionally, reports surface of a man or woman who spend years in isolation. This is very rarely the case, though. For most, interpersonal relationships begin and evolve over many years. The vast majority of people are able to navigate through them effectively, only seldomly encountering difficulties. Psychopaths, do not enjoy such a luxury. As Cleckley, and later Dr. Robert Hare, explain, they cannot be depended on to innately possesses or reciprocate a kind word or action. Further, they cannot be relied upon to carry out actions which require a large amount of trust.

There are many other traits already discussed, and a few to come, which might interweave with this one in particular, but it should not be difficult to see the natural allure of

[3] In fact, Dutton remarks that Dr. Hare, originator of the PCL-R, contributes increased Internet usage of the youth to a higher risk of emergent psychopathy in recent years [12].

such a state to a psychopath. It stands to reason that by being egotistical, misleading, and having little empathy, great “unresponsiveness” is a direct result. Because this property, much like its immediate predecessor, also necessarily involves specific context and cooperation with another person, detecting it through language will be difficult. In fact, it may even be medium-specific, in the McLuhan sense [19]. In other words, language may only be helpful here if it can give us both historical information and correspondence with the other parties involved. To this end, this paper must also exclude this trait because neither the training nor the test corpus can be depended on to have the required features.

Fantastic and uninviting behavior, with drink and sometimes without

More than the other traits, either before or following “Fantastic and uninviting behavior with drink and sometimes without,” this trait leaves an open air of uncertainty. While it is true that some psychologists and medical professionals find it difficult to quantify many of these traits, this one seems to go in a different direction. As for the other traits, an argument can be made that the analyst's position is up for debate, but in the case of particular element, it is not. For example, it is possible to know a person’s habits with drugs or alcohol, but even though we may be afforded that information, it may not tell us much. And while, some may adhere to the old *in vino veritas* adage, the research backing it in the psychopath’s case is simply not there.

Cleckley’s book is littered with the odd story of how a man or woman’s friends may insist upon a person’s institutionalization, because they believe their loved one to be expressing a psychological condition through excess use of drugs or alcohol. However, the same can be said about the amount of stories and case studies for a person’s psychopathic tendencies when they are not under the influence. Indeed, Cleckley himself is a bit suspicious that this trait has any real tie to the condition besides a pure indication that psychopathy appears in both the drinking (and drugs) demographics. It is for this reason and the insufficient supply of additional data that the trait is purposefully excluded from the project. And although I often can make an argument that detecting alcoholism or drug-abuse as an easy potential “indicator,” I must yield that without more reliable research in the area, I must err on the side of exclusion rather than inclusion.

Suicide rarely carried out

Suicide rarely carried out is another of the few traits very difficult to consider with the help of NLP or ML. If this trait were something determinable and traceable through textual analysis, it would be hugely beneficial, as it is a hallmark feature of a psychopath’s mindset. Since this is not the case, it will have limited scope in the scope of this paper; however, it is necessary to explain why it would be a valuable feature to consider if possible.

At their cores, psychopaths are narcissists, a term that speaks to the condition of self-preservation. Psychopaths are also adept at manipulation, having an acute ability to control others by knowing how to elicit a specific response from them. Neurotypical individuals demonstrate empathy towards others and tend to react when that person expresses an extreme need or cry for help. Being the expert manipulators that psychopaths are, they harness the neurotypical person’s concern for them, forcing it to its limits with suicidal threats. The suicidal threat often goes uncarried out because although psychopaths do have little feelings through which to view the world, they do have a narcissistic impulse for self-preservation. This combination, of course, means that the psychopath will learn how to signal to a victim that he is willing to go to any end to get his point across. More often than not, this is a lie.

In the foundational days of NLP, one approach to language parsing was to establish what was known as a language “ontology,” or a list of rules [20]. Practitioners in the 1960s and 1970s believed that this was the only effective way to ensure a program accurately examined a corpus was to program these long ontologies. This became problematic for two reasons. First, even though language specialists could describe these rules, when they were implemented, there existed edge cases that fell through the cracks. Second, these long lists of rules had little "age-ability." Languages differ from region to region, and as generations acquire the communicative traditions of their predecessors, they also change it slightly, meaning that any ontology developed for a specific purpose would likely need curating over a larger time-span. Insofar as detecting suicide is considered, the only real way to do this would be to develop an ontological structure that accounted for keywords or phrases that described self-harm.

Sex life impersonal, trivial and poorly integrated

While one of the over-arching themes of psychopathic behavior necessarily requires observance of a psychopath's behavior in society, there are few traits that explicitly examine the more intimate moments in a psychopath's life or lifestyle. Although other traits need some explanation as to how to use the terms in the trait's description, "Sex life impersonal" involves exactly what one would expect -- that a psychopath misunderstands the appropriate response to acceptable sexual behavior. In the extreme cases of serial killers, profilers often look for associations to other, sexually motivated crimes. Commonly, a serial killer who is also a psychopath[4] will involve himself in other crimes, such as sexual assault, rape, and child molestation. And although it goes without mentioning that these crimes carry their own set of potential psychiatric baggage, it may not always be apparent as to how an impersonal sex life may play in the grand scheme of the profile.

Since we do know that being a savvy manipulator plays a role in their societal functioning ability and that they constantly hone these skills over time, it might be the case that

[4]While the vast majority of serial killers are diagnosable psychopaths, it is possible for a serial killer not to be one. Being a serial killer does not necessarily include an automatic diagnosis of psychopathy.

having less or no practical life experience in the realm of “normal” sexual behavior could contribute to an exacerbation of long-held inner desires. Jeffrey Dahmer, the notorious serial killer who focused on homosexual victims, went on the record as saying that his primary motivation for murdering never began with an intent to kill [14]. Instead, a drive to create his own sexual “zombie” with whom he could interact to fulfill a fantasy life spurned him into action. His subsequent failure to convert them into such a mind-slave resulted in the death of his victim each time.

Robert Hare, who is fundamentally more interested in criminal psychopathy, expands Cleckley’s point by adding to his own checklist a trait that includes “many short-term marital relationships.” The success of Hare’s PCL-R on recidivism might suggest that the relationships held outside of friendship and birth family could be informative. What we see there is a psychopath’s difficulty in navigating relationships outside of shallow friendly encounters or those seen involving blood-bonds.

Failure to follow any life plan

M.E. Thomas is among the most prolific writers about sociopathy and psychopathy. Being a diagnosed sociopath herself, she makes an effort to educate the public on what to expect from the people who exhibit this kind of behavior [3]. She is quick to point out that not all people who fall within the umbrella term of “Antisocial Personality Disorder” (ASPD) go on to be prisoners or criminally motivated. On the contrary -- many such as herself circumvent their genetic predeterminations and contribute to society. Even so, there is one element that she stands firmly with Dr. Cleckley and Dr. Hare on: failure to follow any life plan.

While she acknowledges her own pride in being a well-published law professor, community leader, and Sunday school teacher with children, she admits that her life can be split into instances of three year periods [3]. In her own experiences and struggles with ASPD, Thomas describes how she is plagued by a desire to “ruin” people. By ruining people, she means that she will develop an interest in something or someone, strive to embolden the goal or relationship, then watch as it unravels. For Thomas, this is not criminal, but rather a miniature life-cycle of interest, pursuit, and decline.

Dr. Cleckley and Dr. Hare expound on this phenomenon. Because psychopaths and sociopaths are people who have desensitized feelings and are essentially “numb” to the world, they are constantly cycling through relationships and situations in order to find a fresh view on life. Moreover, their manipulative tendencies demand that they experience eventual boredom from being left too long with the familiar. Naturally, they counter this by imbuing their lives with dramatic or fantastic behavior in order to tap into whatever feelings lay buried deeply within. The penalty they pay, however, is one of constantly life implosion, meaning a necessary “restart” throughout their lives. In the most dramatic of cases, such behavior leaves their personal and professional lives tattered, but this is not always the case. Further, what constitutes a lack of following any life plan may also be subjective. What may be a three-year life-cycle in M.E. Thomas’ world may be shorter or longer in another’s.

Concluding Remarks

In order to truly investigate a subject appropriately, we must first understand the subject from multiple angles. The purpose of this section has been to introduce the traditional clinical profile of a psychopath while at the same time commenting on what computational methods may shine light into an otherwise dark area of research. It should be noted that not all the proposed computational methods are explored in this paper, while some, like POS tagging and word-frequency analysis, definitely are. In the next section, I outline the methodology for acquiring data, investigating language usage using NLP techniques, and train and test models using supervised machine learnings. Results are given at the end of each round of experimentation.

SECTION III

METHODOLOGY AND RESULTS

Section I states that there be a two-fold approach to computational methods for investigating the use of language by psychopathic individuals versus non-psychopathic individuals, also known as "neurotypical individuals," or more colloquially "empaths." In turn, this section address each of those major areas: natural language processing and supervised machine learning. For each area, I will briefly discuss the underlying processes taken to generate the results presented subsequently in the section. Before discussing the methods, however, it is necessary to discuss the tools and packages with which analyze the data how the data used for this project was acquired to begin with.

Tools and Packages

The current NLP and ML landscape is replete with tools and packages to use that streamlines the job of any researcher. Decades ago, this was not true, and the burgeoning scholars and specialists alike cut their teeth on home-made algorithms adapted from journals, textbooks, and conference papers. Today, we can sidestep much of the headache that would otherwise be encountered by leveraging some of the materials available to us.

The major contenders for use include the R programming language, MatLab / Octave, and Python. Each has its own strengths, but ultimately, choosing the most appropriate weapon for the fight grants an unyielding edge with which to tackle to harder problems most efficiently.

For this project, I chose to use Python and its associated libraries [21]. Below, I will discuss why I chose Python over the other aforementioned options and identify the areas where I think it is superior for the specific tasks this project entails.

First, we examine the R programming language [22]. R is heralded among many researchers as the "go-to" language and toolkit for statistical analysis. And while their argument is strong -- the language includes a rich library of statistical and machine learning tools -- it suffers from one major point: the R language lacks a robust combination of libraries that fluidly interconnect for natural language processing and machine learning. By this I do not mean that they lack packages altogether. Indeed, the R programming language enjoys several such open source libraries that are growing rich with features as they age (such as "OpenNLP") [23]. What I mean, however, is that the consensus among NLP researchers and ML researchers who live in the cross section between both fields utilize Python and structure their library creation and curation accordingly. For example, common naming convention among many Python libraries directed at ML tasks use the same function structure -- i.e. "fit", "transform", "predict", and "predict_proba", to name a few -- and are aware of other fields doing the same. Moreover, they design classes and functions so that data is easily convertible to the appropriate format. Where R possesses strength in statistical analysis, optimization, and, perhaps, performance, it fails in these areas. In the long run, where those elements must be considered, the case for R's appropriate use in the project can be taken more seriously, but for the scope and intent of this project, which is intended to be a "proof-of-concept" above all else, we can sacrifice those for ease of use and development.

MatLab and its open source brother "Octave" are the other contenders vying for a spot in data analysis with respect to visualization, NLP, and ML [24][25]. However, we can make a similar argument against them as we could with R. Namely, Matlab suffers from broad range community support with respect to NLP. As one UC Berkeley researcher, Siamak Faridani, notes that while MatLab excels with numerical computations, it handles text and textual preprocessing very slowly [26].Compound this with its reputation for poor memory usage on large scale projects, serious scholars must seek to offset the imbalance in creative ways. One possibility is to use a different language to preprocess the text and wrestle it into the appropriate format. For this, almost any other language would suffice, like Java, Python, or Perl. However, this could lead to unsavory and ugly code-bases, which require active maintenance over the coordination of both the preprocessor and analyzer. Having two separate code-bases is nobody's idea of convenient, so MatLab/Octave excludes itself on that alone. Moreover, a comparative analysis bolsters this conclusion. The popular data science website KDNuggets regularly investigates the popular tools being used by novices and experts alike from a range of subjects. Among the most popular are, indeed, Python, R, and MatLab/Octave; however, the battle is regularly for third place, as both R and Python tend to take the top two slots, respectively and without exception [27]. This suggests a lack of community support or interest, but in either case, it is disqualifying.

For my project, I chose to use the Python programming language over the other options. Where R and Matlab fail, Python does not excel, *per se*, but certainly pulls ahead. Python carries with it the right baggage. First, Python is very community driven. Upon encountering a dilemma, it is very likely that someone else has encountered the same problem and the community has answered the call. Sometimes, the nastier problems will even be met with a custom library designed to easily handle any issue that may arise from the given situation. Further, Python is text and number friendly. Reading in textual data, processing it, and converting it into a vector for use is a simple, straightforward task. Python also enjoys the benefits of easy prototyping.

Often, programmers will turn to Matlab/Octave for its prototyping ability. Andrew Ng, a heavy-hitting machine learning researcher from Stanford recommends Octave exclusively for this purpose, and he encourages his own students to migrate algorithms to a more efficient language only after seeing them work in the Octave form [28]. Python enjoys the same luxury, but with the added benefit in this project to be able to jump from NLP task to ML task where necessary.

Speaking specifically, I chose two main libraries with which to work: Natural Language and. These libraries are robust, feature rich, and well-documented. The creators are scholars who curate the codebase regularly aim to keep their functions and overall code inline with the latest research. In the case of NLTK, there is an entire book published as documentation, replete with examples and troubleshooting sections. Scikit-Learn's documentation comes from the popular website of the same name, which acts a learning resource ML tutorial for the newly initiated [29].

The specifics behind NLTK and Scikit-Learn's functionality will be elaborated on in this section's appropriate subsections. There are, however, a number of smaller, specialized packages that I utilized as well. These will be mentioned and documented in turn, but due to their relative lack of ubiquitousness throughout my code and analysis, their in-depth look will be delayed until their use in the later analytical sections.

Data Acquisition

The primary medium to be analyzed for this project is text. While there are a number of other neurological markers that distinguish a psychopathic personality, such as fMRI and PET scan results, this paper focuses only on those which can be represented textually. Further, in order to make a valid comparison between an ingroup and its outliers, there must also be a corpus of non-psychopathic texts upon which to compare. Ultimately, I chose collect and curate my own corpora consisting of two unique corpuses: psychopathic, and non-psychopathic. The following paragraphs detail the struggles encountered in creating each corpus.

At the outset of this project, I believed that data collection would be trivial. This, however, proved to be untrue. Over time, data acquisition for this project evolved into a chore, which bridged herculean difficulty and lateral thinking. At most, there are three areas in which to search for writing samples for analysis: medical, legal, and public access.

At its core, the diagnosing of psychopathy is a medical exercise. As the previous sections have discussed, there is a burgeoning history of analysts who have attempted to tackle the problem of determining and assessing recognizable traits. It should follow, then, that as the sciences of psychiatry and psychology develop, so, too, should their data stores. This currently holds true, as does other sciences. The problem that arises here, however, is one of access. Due to a number of limiting legal factors and federal protections against the use of psychologically related medical records, obtaining entry to any database for research is difficult, and rightly so. The current social atmosphere holds individuals with certain atypical neurological conditions in contempt, so additional protections are found layered into an already well-fortified area of legal corpus. This means that any attempt to acquire psychological medical records held by medical and federal institutions is difficult, if not impossible. Moreover, medical officials take great care to sanitize their records of many distinguishing features. Though this sanitization is in place, it is still difficult to argue for access, even with the suggestion of further anonymization. After trying to contact several such institutions, this option was nullified, and I was directed to search the remaining two options.

Oddly enough, federal and state police records enjoy fewer immediate protections, but still maintain certain “gatekeepers,” which must be confronted. After some preliminary searching, my research directed me to the records kept by the United States Federal Bureau of Investigation (FBI). State and local police were considered, but their records lack immediate structure and integrity. In other words, much of their data may be missing, duplicated, or corrupted over time.

While the FBI keeps records of high-risk individuals from previous and ongoing investigations, those records are subject to viewing by people who have higher levels of security clearance. The database of which few researchers have access is known as “Vi-CAP,” or “Violent Criminal Apprehension Program.” This database includes a large amount of personal and public data revolving around current and cold cases of importance to the FBI. Prior to the training and popularization of criminal profiling specialists like Robert Ressler, there were few attempts to collect and consolidate data of major offenders on a massive scale. Recognizing this, Ressler pushed for the creation of a database that would account for all offenders who might slip through the cracks and go undetected by detectives due to their large areas of operation. To gain access to Vi-CAP would mean gaining access to ongoing cases and their immediate developments. Hence, if a researcher were to gain access without his being properly vetted, the agency might risk the leaking of important information leading to the capture of dangerous criminals. It would be impossible to gain access at all, nevertheless so quickly.

The final place to search for data is tricky and has no shortage of tradeoffs. While finding and accessing them often involve little more than just a simple Google search, new problems arise. For one, we must consider the source of the data and its veracity. A substantial benefit that medical and police records have is being nearly indisputable as far as trust is concerned. Because the research is conducted by trained professionals, one can be confident that the data is accurate. This is not the case for the Internet. Because the Internet operates under the “Triple A Principle”

-- anyone can say anything about any topic -- precautions must be taken when considering what to add to any dataset.

Additionally, the fundamental question of where to look holds true more so than before. A general search using the keywords “psychopath writing sample” generates a considerable amount of results. However, the majority of these results are unusable, either the result simply is not a psychopathic writing sample at all, unrelated to the general topic (i.e. the topic is too subjective), or uncited. In order to access richer and stronger data, I had to narrow my search. This brought up the question, who or what category of people are psychopaths in society? After some cursory thought, I ultimately harkened back to the words of Kevin Dutton and Robert Hare. A large percentage of incarcerated individuals are psychopathic. Of that group, members who establish an extensive career of especially heinous crimes, such as serial killing and rape, are even more likely to fit the criteria of psychopathy. So, for a subset of my dataset, I chose to search and include the writings of many serial killers and serial rapists. To bolster the case that these individuals were, indeed, psychopaths, I would further point out that of the people chose (with the exception of Jack the Ripper), each individual was interviewed *at length* by criminologist Robert Ressler, whose modus operandi included evaluating them on the Hare psychopath checklist revised. Further, I would argue that even if these criminals did not undergo rigorous psychological evaluation, which they did prior and post-trial, they at the very least include most of the elements on Cleckley’s profile for psychopathy, as discussed in Section II. These qualities, such as suicide rarely carried or egocentricity, are apparent, and in many cases indisputable.[5]

The original source of most of these samples usually comes from one of four places:

[5]As previously stated, the one exception here is Jack the Ripper. However, where he fails the requirement to be psychologically examined, he does not fail the requirement for satisfying most of the items on Cleckley’s psychopathic profile.

letters, court documents, interviews, purposefully publicized items of unique-authorship (blog posts, book chapters, detailed experimentation results, etc.). The book chapters or blog posts that are included in the psychopath corpus are exclusively from psychopaths deemed “prosocial,” meaning individuals lacking a criminal record. This enriches the dataset, because to exclude an entire echelon of people who make up the lower end of the antisocial personality spectrum, would simply just be to analyze and model the data of serial killers and violent crime offenders. A complete manifest of people who make up the dataset can be found in Appendix A, titled “Psychopath Corpus Manifest.”

The other half of my corpus consists of data taken from non-psychopathic individuals. As would be expected, this chore was significantly easier. The main challenges to overcome, however, were challenges of content and sample size. To consider who is non-psychopathic, or “neurotypical,” is to consider people who have an established publishing record lacking in any noticeable flaws. In other words, what should be excluded are samples from people who overtly demonstrate a particular writing style. Just as the psychopathic corpus includes many people with many writing styles discussing many topics, so, too, must the non-psychopathic corpus mirror the same. Even though this is the ideal, it should be noted that no dataset is perfect. While we as researchers chase the golden standard for identical data sets insofar as consistency is concerned, such may be a pipe dream -- something hoped for but unattainable. Working in that vein I attempt not to succumb to such problems, but this may, and in all likelihood *is*, still the case. To minimize them, I decided to randomly select articles of varying content and size from the New York Times database. This database includes topics on a variety of subjects extending back to the 1850s. Much of the database is from the 1850s until the 1950s remains in its imaged form, without the aid of any optimal character recognition to transcribe the text. For this reason, I chose to randomly select (with the use of python's `randInt` function and NyTimes search API) an equal number of samples as is contained in the psychopathic group. Likewise, I chose from both articles and editorials, including letters to the editor. As with the psychopath corpus, a list of data samples can be viewed in Appendix B, titled "Non-Psychopathic Corpus Manifest." There are only three exceptions, however. Three of the "long" random samples found among the non-psychopathic corpus were taken from *The Economist* database [31]. *The Economist* is an internationally renowned, weekly periodical that covers a large number of socio-economic and socio-political topics. I chose three random articles from their database because *The New York Times* is deficient in articles beyond the 2000 word wordcount. But as before, these articles were chosen at random with the aid of Python's "randint" function. The only constraint was to ensure that the articles were of the appropriate length.

Finally, as a few miscellaneous notes, I should point out the final quantity of both corpuses. The basic file structure for the Natural Language Processing element and the Machine Learning element follows this pattern: one main folder, listed as data, with two subfolders, listed as "psychopaths" and "non-psychopaths." Where necessary, this structure is extended to sub-folders indicating "training," "validation," and "test" data. The folder names are meaningful and not arbitrary, so navigating among the folders should be easy. Of each corpus, the files can be separated into three recognizable groups: short samples, including works of 999 words or less; medium samples, including works of 1000 words to 1999 words; and long samples, including works over 2000 words. Detailed information may be viewed in Figure 1 below:

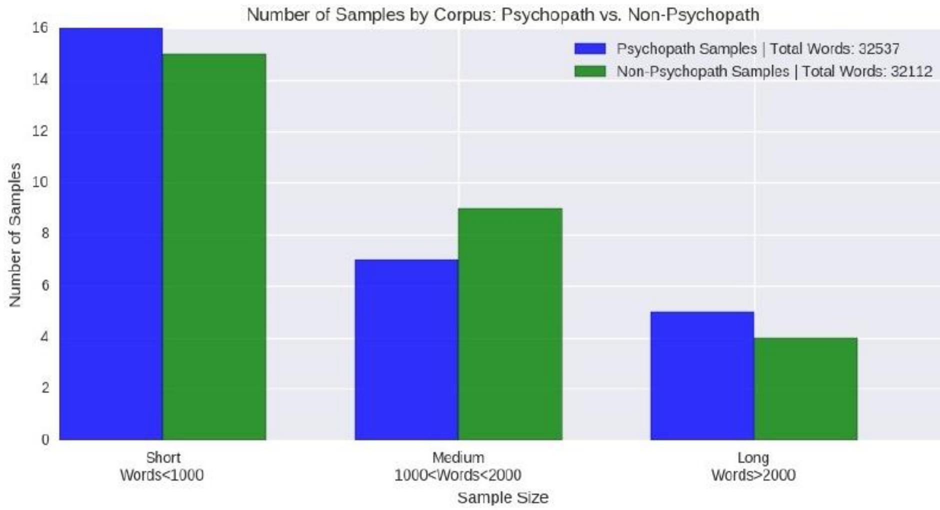

Figure 1

Natural Language Processing: Wordclouds

My first approach to analyzing the writing of psychopaths versus non-psychopaths involves the use of NLP to investigate the text before jumping into ML methods. Over the years, NLP techniques have been developed for myriad of purposes, from statistical analysis to visualization. So, it stands to reason to use NLP as a staging point for the later parts of my research. Specifically, I use NLP to follow the premise behind much of the writings of James Pennebaker, a social psychologist working at the University of Texas at Austin, who has a distinguished career in diving into people's personalities with respect to their use of language [31]. Text is inherently a rich medium, one that is naturally encoded in our everyday lives. That said, I should point out that it is impossible to view a corpus from many points of view. Such would be exhaustive and outside the scope of this study. Instead, I use NLP to approach the psychopathic and non-psychopathic corpuses from the angle of function words. In much the same way that Dr. Pennebaker interrogates text on the function word level, so too will I use NLP to piece together the portrait of a psychopath on the same function word level. According to his research, the most telling "features" of a person come out in the words most frequently used. My machine learning experiments in the last part of this section will prove either to validate or disqualify these assertions.

To begin, I chose to view this problem visually through the use of word clouds. And while word clouds are not science insomuch as they are art, they do retain significant enough information for initial research. Figures 2 and 3 show two wordclouds generated from the complete corpus of psychopathic and non-psychopathic samples. In order to remove meaningless outliers, such as punctuation and non-sense words (i.e. URLs), I cleaned each corpus accordingly. True to the spirit of Dr. Pennebaker, however, I chose to include stopwords, whereas many NLP researchers would immediately dispose of them as step one. These two visual may prove to be valuable later when comparing models and questioning on which features to focus.

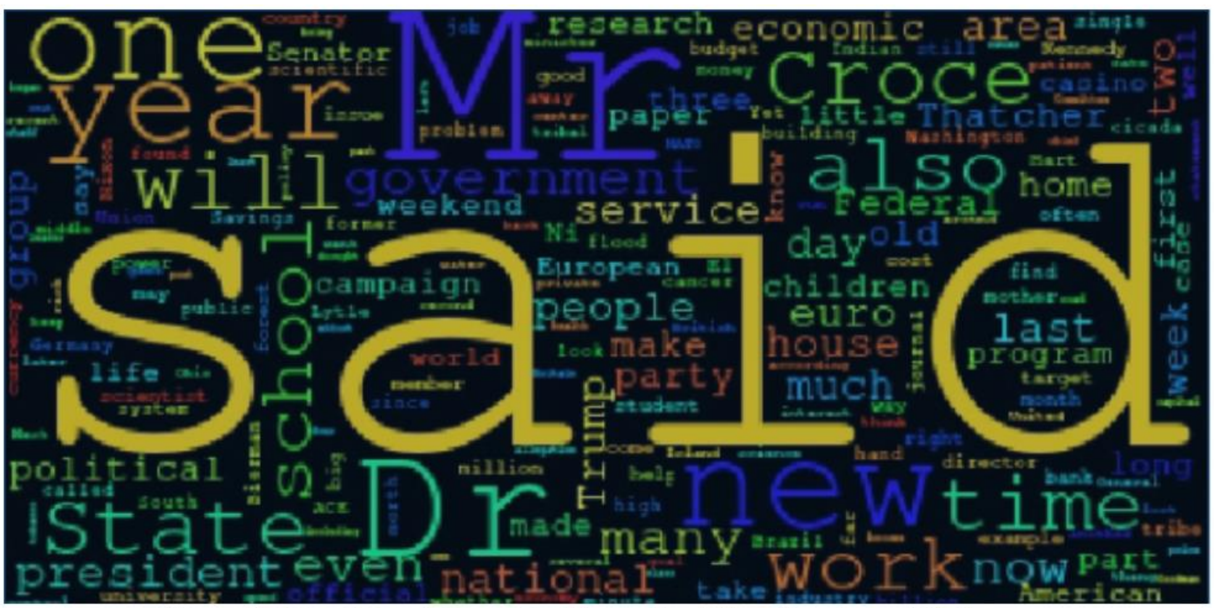

Figure 2: Wordcloud generated from the "Non-Psychopathic" corpus

Figure 2's wordcloud provides us with an initial glimpse into the topics mostly discussed in the non-psychopathic corpus. At first glance, a few things stand out. First, beyond the largest word “said,” most of the terms focus on concrete things found in reality. Nouns like “Dr.”, “State”, and “President” pop out, suggest otherness. These are not self-reflecting terms, but rather terms that actualize outward positions. Moreover, many of the words, like “Group,” “Federal,” and “Government” suggest just the opposite of selfishness, because they involve an element of inclusion by society instead of exclusion.

Figure 3 below, however, tells a different story.

Figure 3: Wordcloud generated from the “Psychopathic” corpus

Figure 3’s largest word is “People,” but beyond that we see an emergence of abstract topics. Where in figure 2 we see a group of people grounded in the concrete world, figure 3 shows us how common and important it is in psychopathic writing to consider the life of the mind. The psychopathic wordcloud references “think,” “thought,” “brain,” – all topics that feature intellectual capacity. Likewise, their verb usage entails many terms that indicate item or information possession, acquisition, or personal maintenance, such as “take”, “get”, “got”, “say”, and “said.” Moreover, the words that can be associated with nouns, are almost exclusively personal[6], mentioning “family,” “children,” and “kids,” to name a few. By keeping and counting the common words, it can be argued that we get a more holistic view of these two groups, albeit one that is rather boring. But even though these clouds do not focus on hapax legomenons which in itself may be telling, we do get to see the underpinnings of two personalities at odds: one grounded in reality whose life revolves around others, and another whose personality is inherently abstract and self-centered.

Natural Language Processing: Occurrence by Counting and Percentage

Following the initial leads brought to the surface by the wordclouds seen in Figures 2 and 3, I next chose to use NLP to see what the most common pronouns were in each corpus, first using raw counts and second using percentages, just to place the numbers in greater context. Like before, I chose to do some initial cleaning of the psychopath and non-psychopath corpuses. This cleaning stripped away any grammatical and punctuation issues, leaving both corpus files devoid of any terms that were not pronouns and punctuation. Figure 4 shows the result of the analysis in raw counting with respect to a complete list of pronouns:

(Figure 4 is on the next page)

[6]“Personal” here is a word not exactly referring to the person itself, but in contrast to “occupational,” which I use to mean an indication of a person’s job or employment.

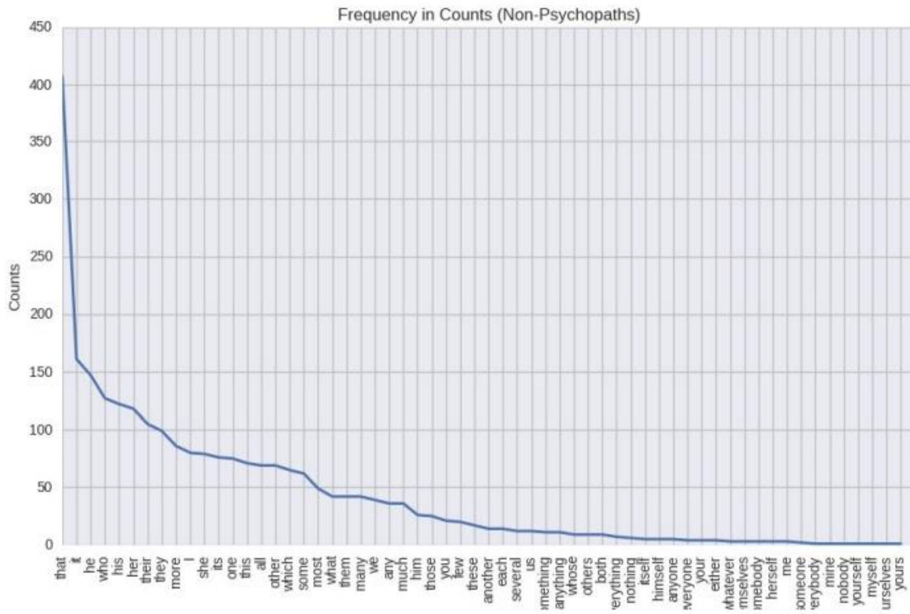

Figure 4: Pronoun Usage of Non-Psychopaths in Raw Counts

The same list of pronouns may also be seen applied in figure 5 below, as well, when applied to the psychopathic corpus.

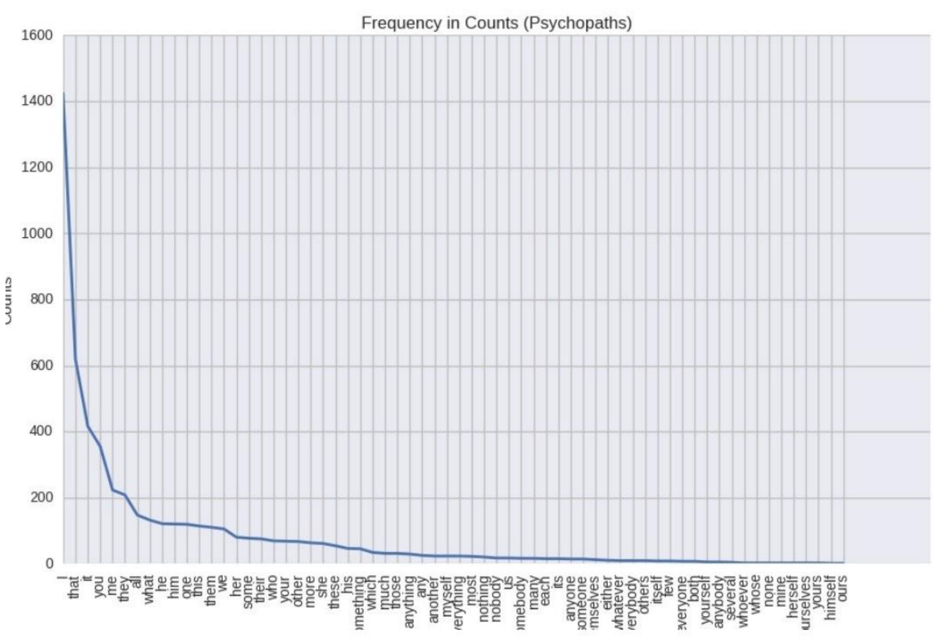

Figure 5: Pronoun Usage of Psychopaths in Raw Counts

Understanding the impact of these two plots comes with greater reward when paying particular attention to the use of personal vs. non-personal pronouns among the top twenty terms in each. In figure 5, we see that no personal pronoun is used until spot number 10, whereas in figure 5, not only do we see the word “I” ranking as the word with the highest number of occurrences, but it is also followed shortly by the word “me,” which takes the number 5 spot. In contrast, words like “he”, “she”, “it”, “her” or “his” litter the top 10 list of non-psychopaths in rapid succession, while only appearing in diminished amounts on the equivalent psychopath plot.

I also chose to represent these two plots as averages, as seen in figures 6 and 7. By representing each term with its respective average of all pronouns visually, we should be able to see just how strong the correlation between the Clecklyean psychopathic “ego” and “narcissistic” tendencies are. It should also be noted that these plots closely follow the typical Pareto distributions.

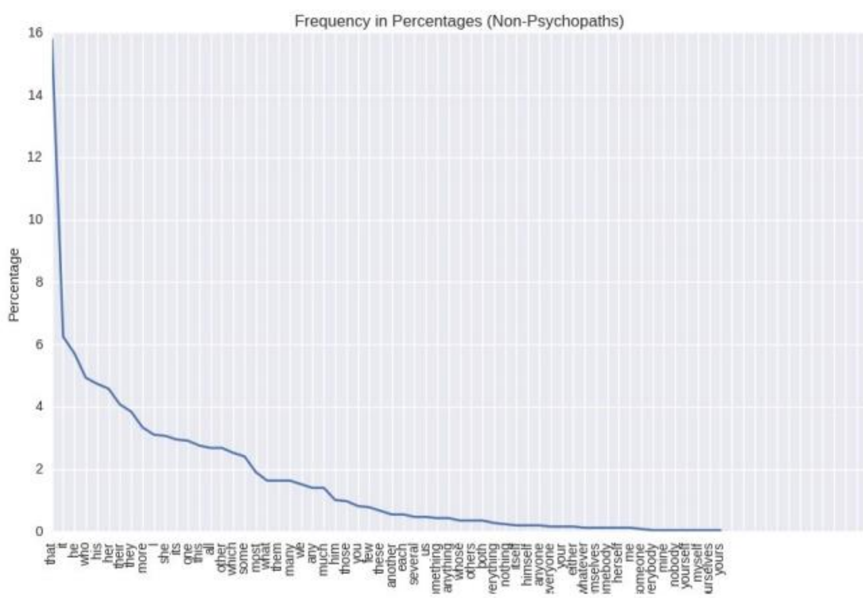

Figure 6: Pronoun Usage of Non-Psychopaths in Percentages

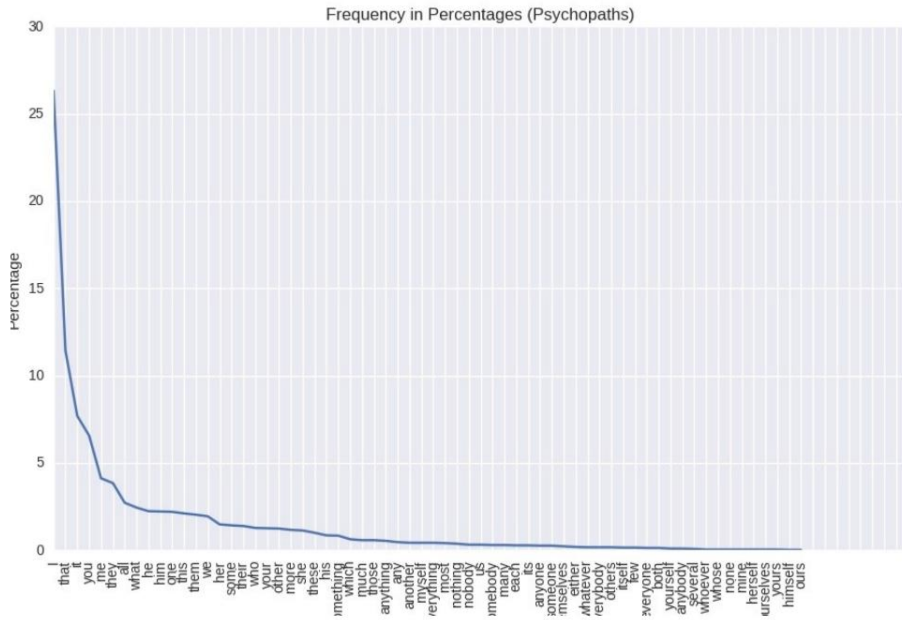

Figure 7: Pronoun Usage of Psychopaths in Percentages

Natural Language Processing: Part of Speech Frequency

Building further by using Dr. Pennebaker's approach and focusing on function words, I chose to use a long-standing NLP tool known as "Part of Speech Tagging." Basically, I wanted to know if the part of speech occurrences between psychopaths and non-psychopaths differed substantially. In order to make this comparison, I first had to combine the samples in each corpus as before, but I also had to deploy the NLTK's POS tagging capabilities to do so. POS tagging is an evolving field, and new, more robust taggers come out from time to time. The specific tagger used for this analysis was `nltk.pos_tag(Text)`, which returns a list of tuples containing the word and its tagged part of speech.[7] After the list was generated for each corpus, I simply tracked each part of speech in a counting dictionary, which was finally graphed. Figure 8 shows the result, with two different-colored bars for easy comparison.

[7] If never before witnessed, the tags used can look foreign, and "Appendix C" has been added to help decipher the more arcane tags.

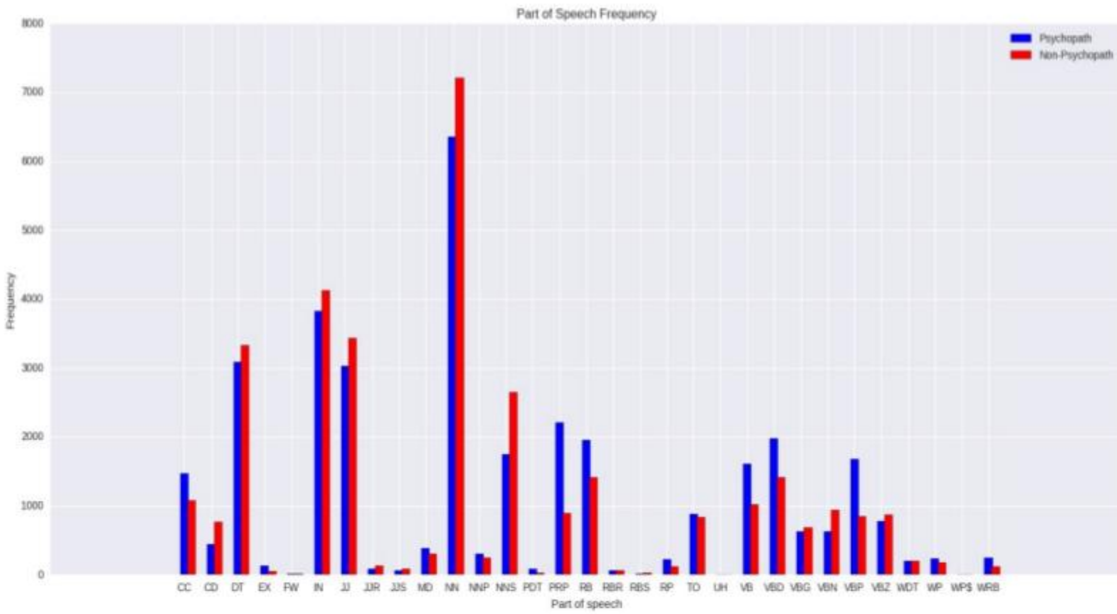

Figure 8: Comparison in POS Occurrences between Psychopaths and Non-Psychopaths

Natural Language Processing: Lexical Dispersion Analysis

The final analysis I performed on both corpuses was to generate lexical dispersion plots. Lexical dispersion plotting can be hugely revealing if done properly or be a devastating disappointment. The principle at work when examining a lexical dispersion plot is to see how often instances of a word arise as the corpus size grows. In other words, a lexical dispersion plot tracks the "word offset" of a term against the whole corpus. It is possible to create a dispersion plot for a corpus's entire vocabulary set, but doing so is both impractical and inelegant. To get the most use out of it, however, is to pick a subset of words that may have a correlation. After plotting the offset and viewing the "dispersion," it is possible to see indicators of which direction to take further research.

Based on the information thus far, I chose to take a look at personal pronouns versus non-personal pronouns in both corpuses. Both pronoun frequencies, like the ones shown in figures 4 – 7, and the POS analysis, shown in figure 8, suggest that examining the dispersion between personal versus non-personal pronouns may be the way to go. As with some previous cleaning strategies, punctuation was removed in an effort to pick up every instance needed for the `nltknltk.draw.dispersion.dispersion_plot(corpus)` to work accurately. I chose to look at 2 dispersion plots. The first dispersion plot examined “ego-inclusive” pronouns: “I”, “me”, “we”, “us”, and “mine.” The second dispersion plot examined “ego-exclusive” pronouns: “you”, “your”, “his”, “her”, and “their.” The results are shown in figures 9 – 12.

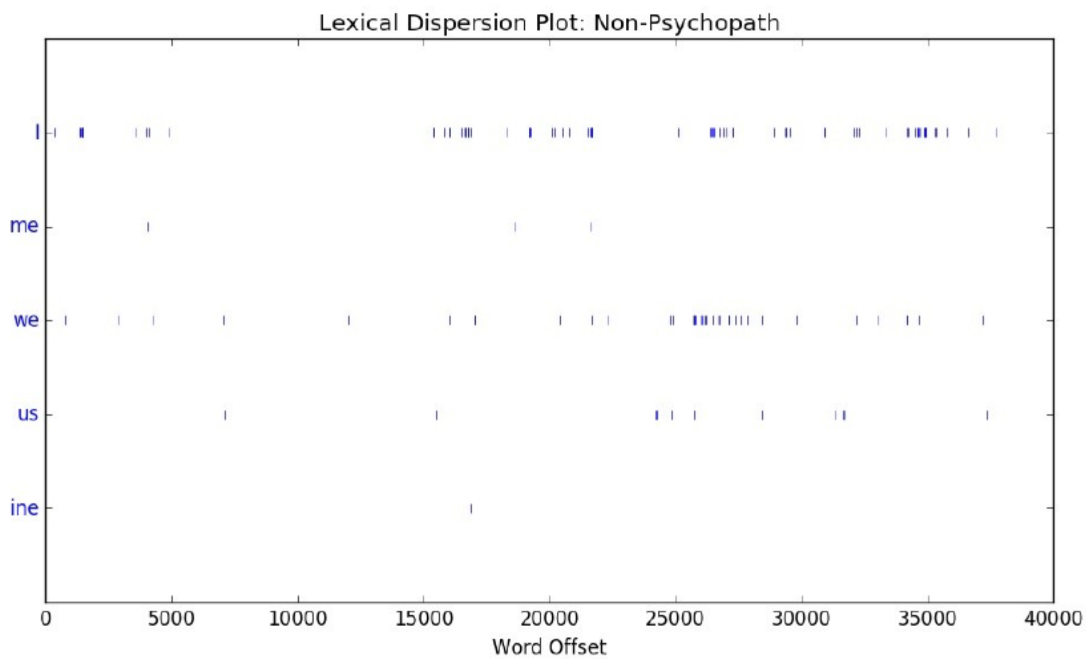

Figure 9: Non-Psychopath Ego-Inclusive Dispersion Plot

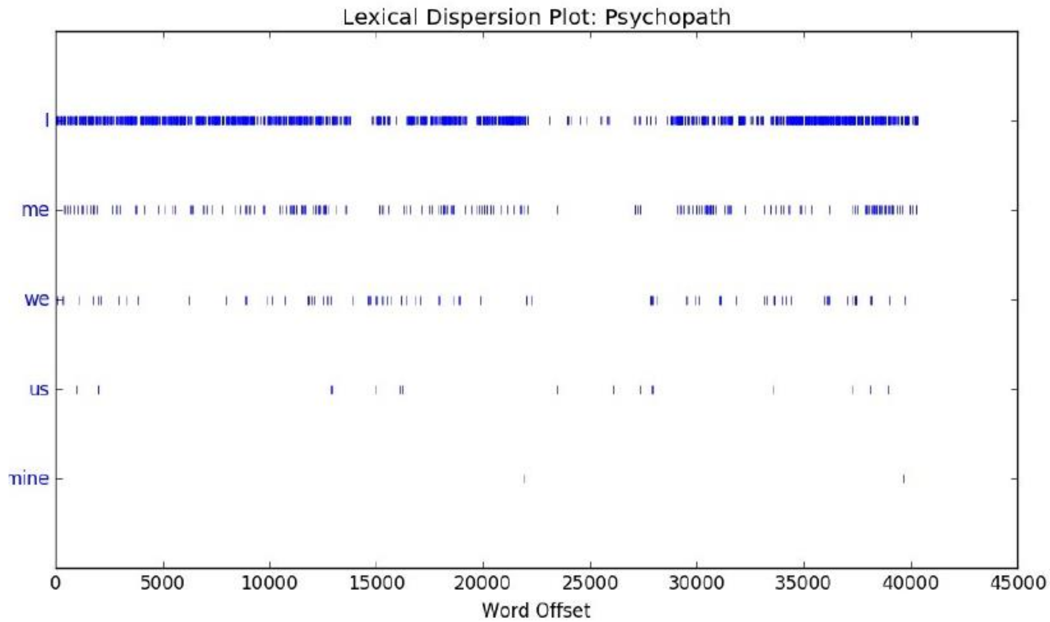

Figure 10: Psychopathic Ego-Inclusive Dispersion Plot

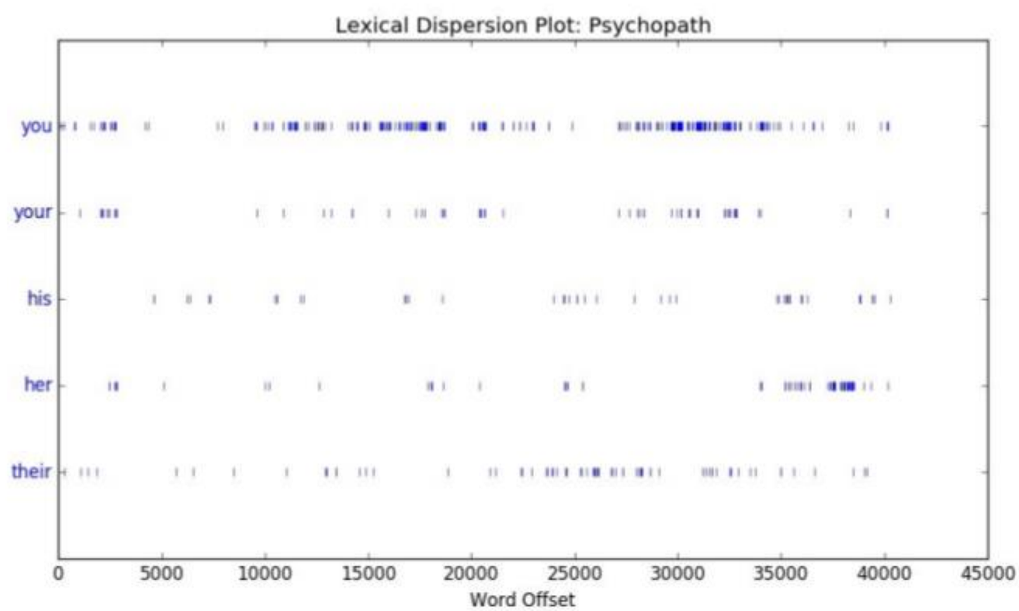

Figure 11: Non-Psychopath Ego-Exclusive Dispersion Plot

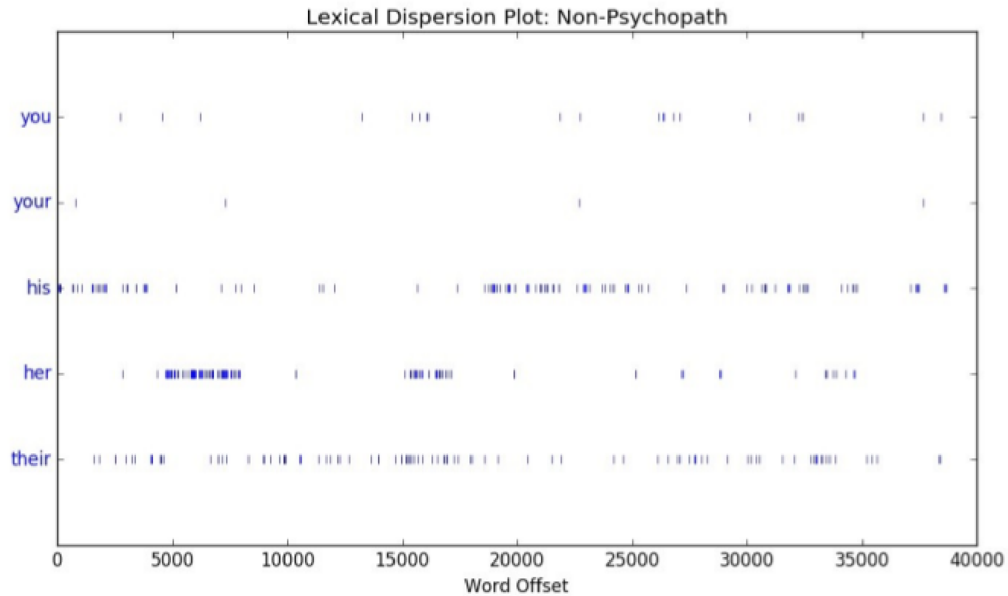

Figure 12: Non-Psychopath Ego-Exclusive Dispersion Plot

Supervised Machine Learning: General Overview

For the second part of this project, I chose to analyze and model the writing of psychopaths using two different machine learning approaches. For the first approach, I chose to view the problem of whether or not a particular writing sample was written by a psychopath as a one-class classification problem. Second, I expanded upon the first approach by choosing to view the problem as a binary classification problem. For both approaches, I used Python's Scikit-Learn package, which includes a number of different implementations of several popular classification algorithms. Moreover, Scikit-Learn is very customizable, allowing for the tuning of many important hyperparameters. As will be shown, I took a threefold plan of attack to garner the best results. Each of those will be discussed in turn along with the respective graphs of accuracy.

For each classifier, I used the Scikit-Learn `load_files(directory)` to read in all the files. Second, I used the `train_test_split(data)` method to split the data into appropriately sized training and testing set if needed, and for each classifier in approach two, I gaged the accuracy by using k-fold cross validation via the `cross_val_score(classifier, X, y, cv)`, where classifier is the specific classifier being tested along with its preset parameters, X is the data to fit, y is an array of the labels, and cv is the number of folds. If no cv value is specified, Scikit-Learn uses a 3-fold strategy by default. Each classifier underwent the procedure a total of 15 times, having the number of folds equal the current iteration. So, for instance, on the 15^{th} run of the algorithm, cv = 15. To check the accuracy of the algorithm, Scikit-Learn's independent `accuracy_score(y, predicted)` was used to verify the results. This extra layer of checking was needed because each Scikit-Learn `classifier.score(X,y)` may have a different scoring strategy. Regardless, unless otherwise specified, the scores printed in this section are accurate from comparing the two aforementioned scoring methods.

Supervised Machine Learning: One-Class Classification (Unary Classification)

The unary classification approach is fairly straightforward. The idea is to take a corpus consisting of just one class and split it into a training, validation, and "unseen" datasets accordingly. The classifier must learn from the data and then predict whether the unseen data is either in the class or is not. This is in contrast to a multi-label problem where a classifier has many classes to learn. Scikit-Learn handles a one-class classification problem with the `OneClassSVM` class. After feeding the classifier the training data, Scikit-Learn's `OneClassSVM` will make a prediction on any new data presented to it by returning a -1, 0, or +1. By returning a -1, the classifier is indicating that the fresh data is *not* a part of the class. By returning a 0, the classifier is indicating that the fresh data is indeterminate, and by returning a +1, the classifier is indicating that the fresh data is, indeed, part of the class. The `OneClassSVM` does not return an accuracy measure in the way that multi-class classifiers do. So in order for me to measure the classifier’s progress, I tracked the number of accurately predicted labels versus inaccurately predicted labels. At the end of the program, I printed out a percentage (the average) of how often the classifier correctly predicted whether the writing was of the psychopath class. In order to see the trends over time, ran the algorithm 5 distinct times with the number of iterations increasing each times. The number of iterations were: 250, 500, 1000, 2500, and 5000. I chose to vectorize the data using Scikit-Learn’s `CountVectorizer` and `TfidfVectorizer` classes. For the one-class classification only, I chose to run the experiments without cleaning the data and with partial cleaning of the data, which included only removing punctuation and nonsense characters. Due to time constraints, the only kernel test was linear. The results of each run (250, 500, 1000, 2500, and 5000) are shown in figures 13 – 17. The tests were run on train/test split with the four following configurations: 80/20, 85/15, 90/10, and 95/5.

(Figures 13 – 17 begin on the next page)

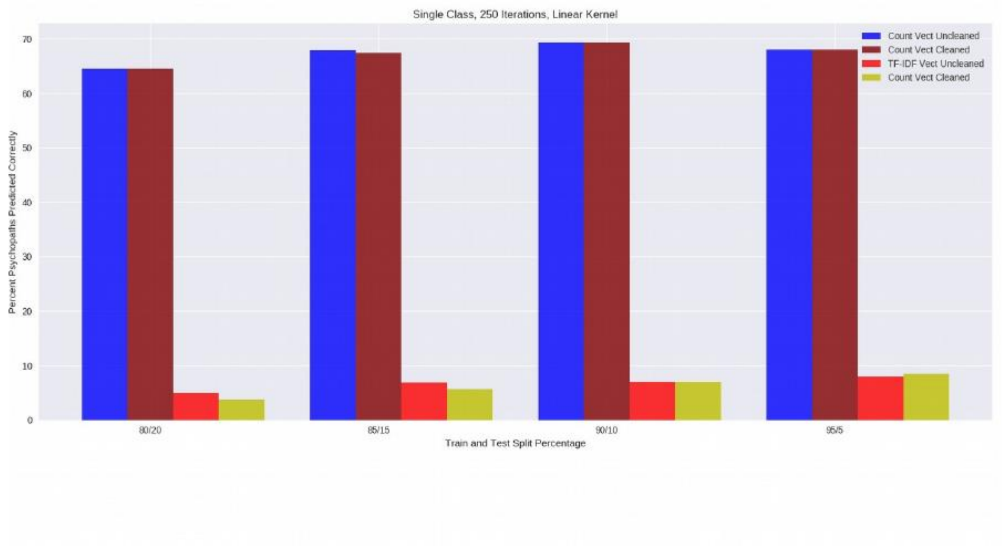

Figure 13: One-class Classification with 250 iterations

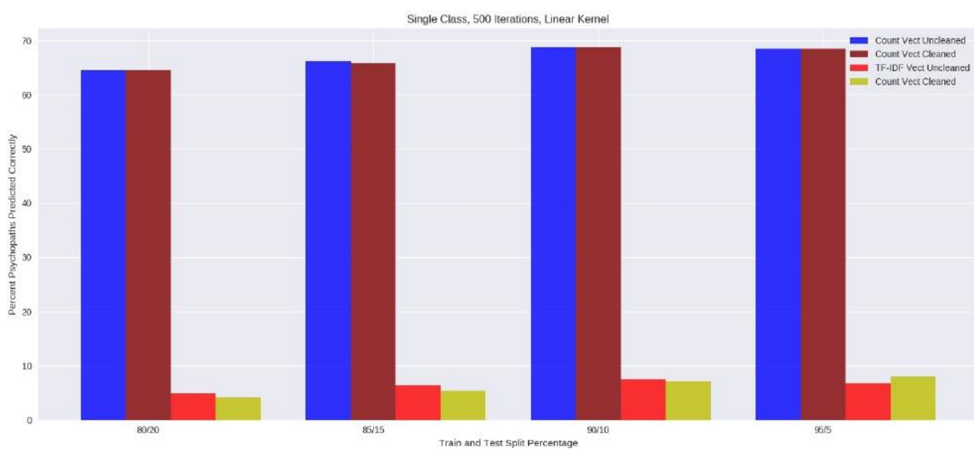

Figure 14: One-class Classification with 500 iterations

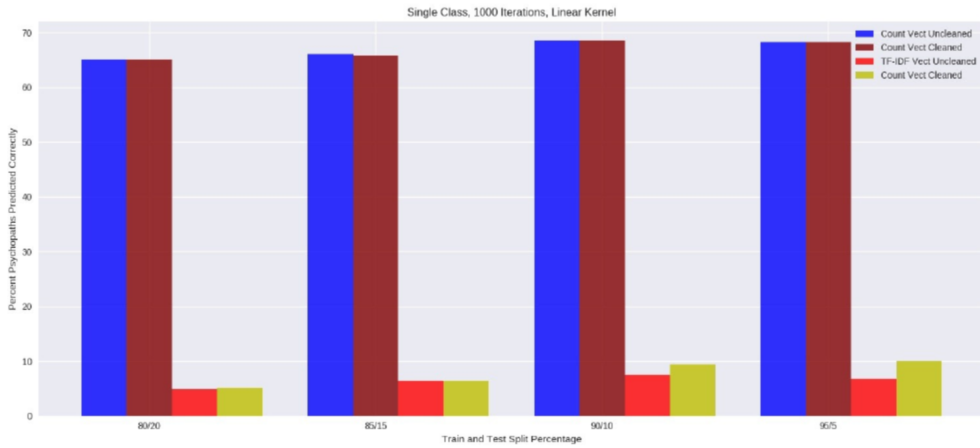

Figure 15: One-class Classification with 1000 iterations

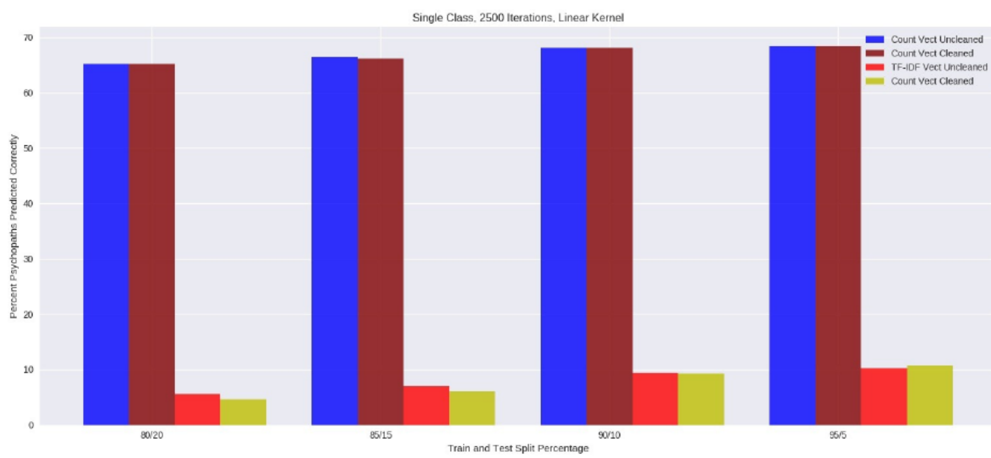

Figure 16: One-class Classification with 2500 iterations

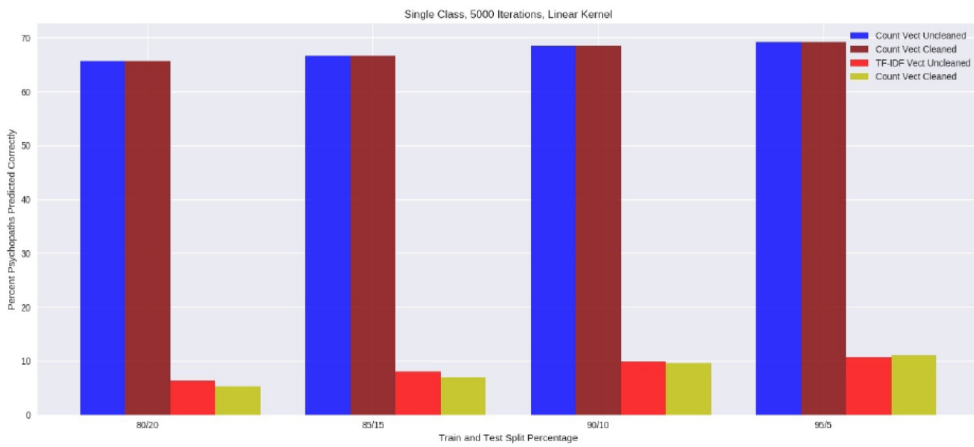

Figure 17: One-class Classification with 5000 iterations

It is interesting to note here that for both “cleaned” and “uncleaned” data, the count vectorizer outperforms tf-idf. This holds true for all test/train splits and all iterations upon those splits from 250 to 5000. However, if we look at the rate of improvement over time, tf-idf experience much faster improvement while count vectorization holds steady. At 5000 iterations, the experiments took nearly 30 minutes a piece, but given enough time and processing power, it would be interesting to see how long before tf-idf caught up to its count vectorization counterpart.

Supervised Machine Learning: Binary Classification

For the binary classification approach, I tested four different classifiers, three of which leveraged Scikit-Learn and its optimization abilities: LinearSVM, Stochastic Gradient Descent, and Gaussian Naive Bayes. The four classifier tested was “word2vec” and leveraged “FastText,” the word2vec package created by the Facebook machine learning team.

The process for evaluating the classifiers followed the same procedure outlined in “Supervised Machine Learning: General Overview” with two additional points added. First, like the unary classification approach described above, the binary classification approach utilized both count vectorization and tf-idf vectorization. The results shown in figures 20 – 25 show the varied outcomes given different testing conditions. Further, each set of experiments were run a total of 3 times. For the first run, an unaltered classifier was applied "fresh out of the box" from Scikit-Learn. Second, I employed the use of the `GridSearchCV` class, which is an exhaustive search algorithm that finds the best parameters for each classifiers (i.e. the combination that yields the highest accuracy score). On the third run, I leveraged the `ExtraTreesClassifier` to find, select, and retrain the models on the 100 most informative features. Figures 18 and 19 are the respective graphs which show the top 10 features from the count vectorization and tf-idf vectorization accordingly.

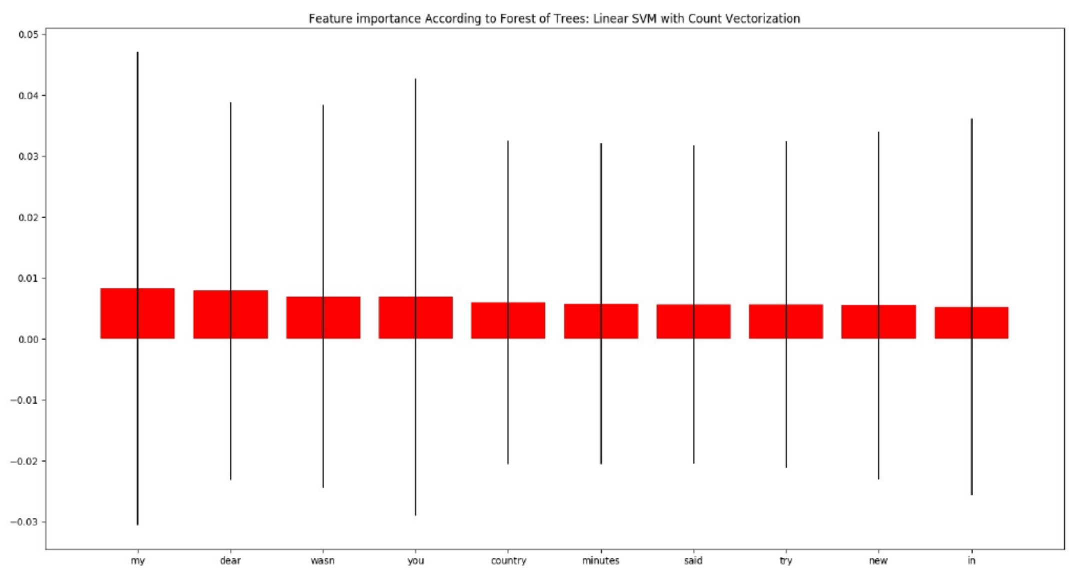

Figure 18: Feature Importance with Random Forest and Count Vectorization

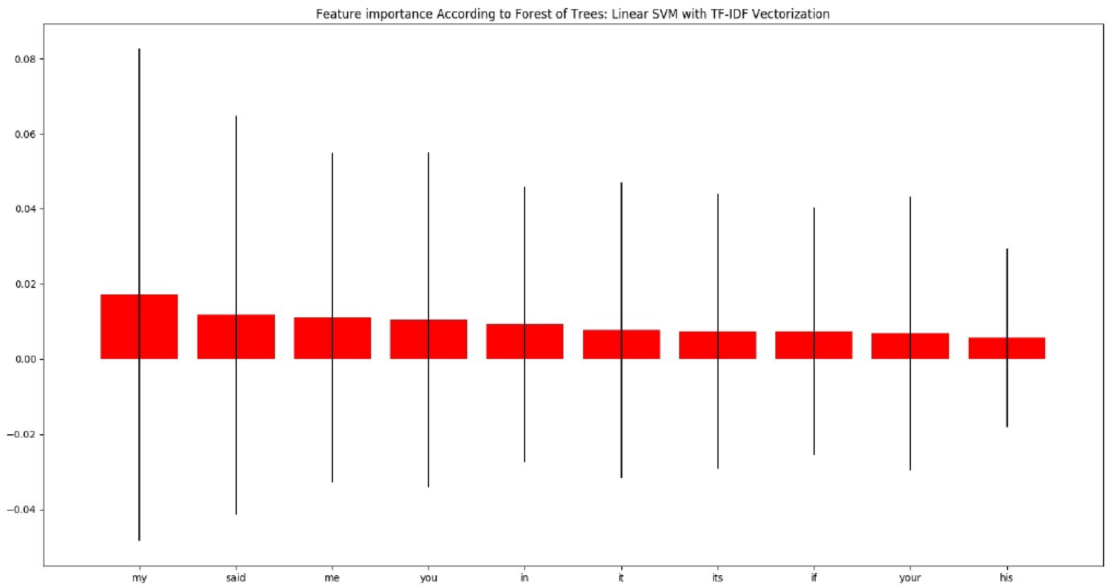

Figure 19: Feature Importance with Random Forest and TF-IDF Vectorization

Importantly, it should be noticed that while figures 19 and 20 are titled "Linear SVM" that all three classifiers will use the same 100 important features generated from the vectorization of the text. In other words, the subset of 100 important features applied to the Stochastic Gradient Descent classifier will be the same as the same subset of 100 important features applied to the Gaussian Naive Bayes.

The results from each experiment (both with count vectorization and tf-idf vectorization) are compiled into singular graphs shown below in figures 20 – 25.

(Figures 20 – 25 Begin on Next Page)

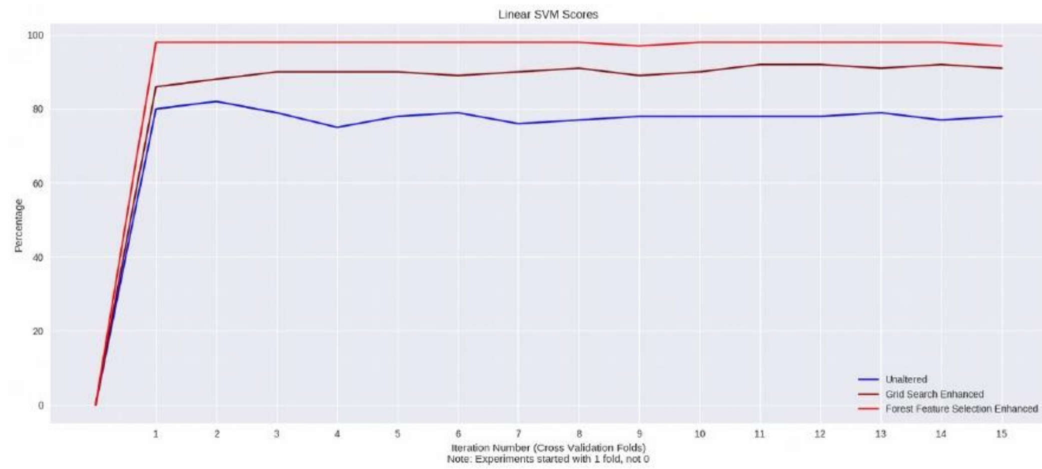

Figure 20: Linear SVM Accuracy Percentages with Count Vectorization

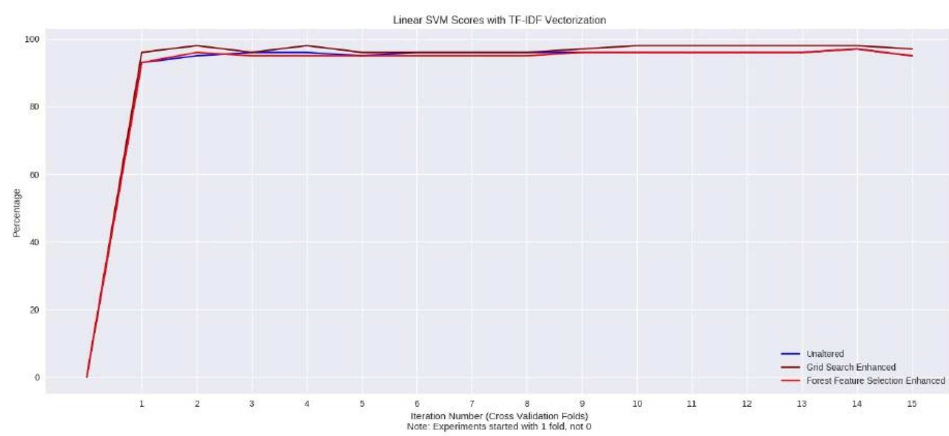

Figure 21: Linear SVM Accuracy Percentages with TF-IDF Vectorization

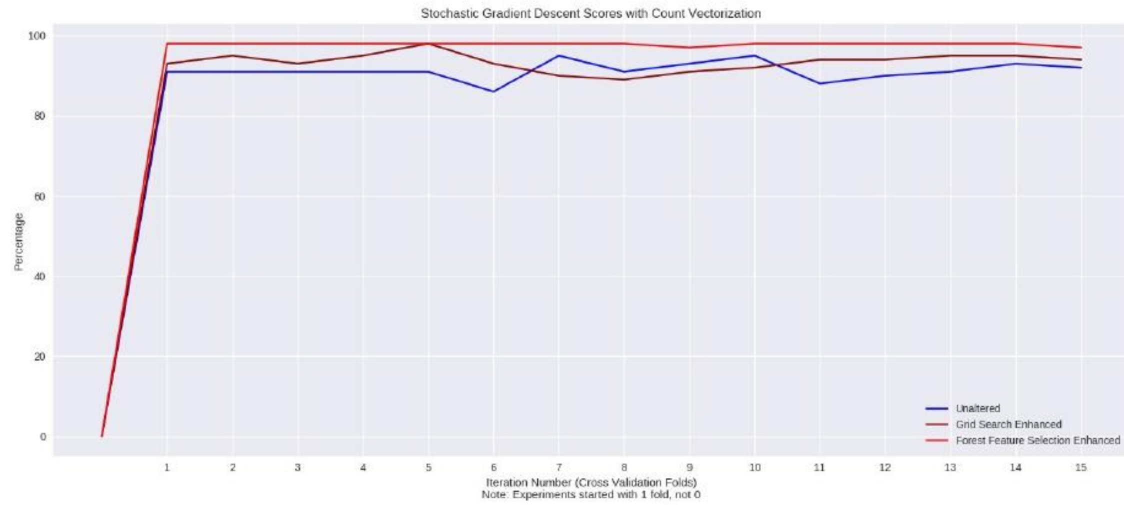

Figure 22: Stochastic Gradient Descent Accuracy Percentages with Count Vectorization

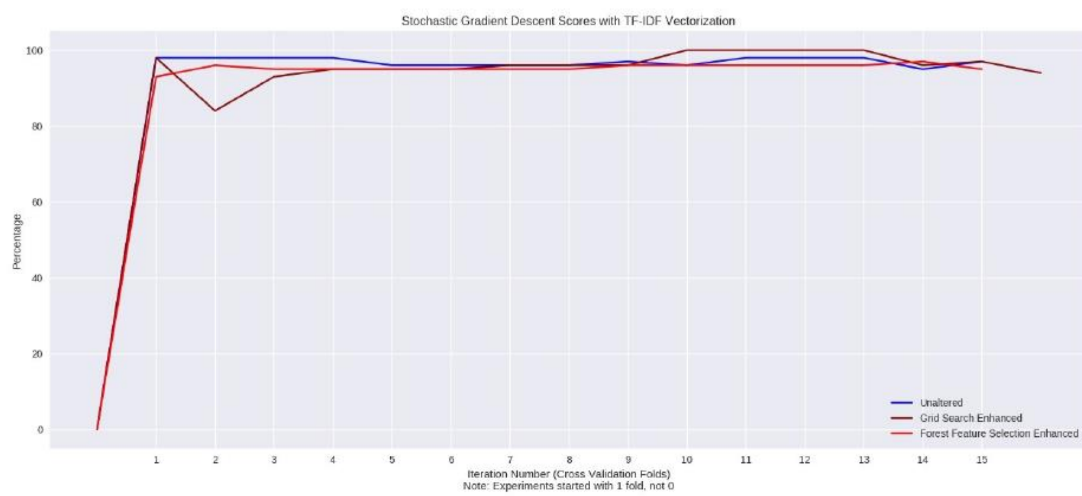

Figure 23: Stochastic Gradient Descent Accuracy Percentages with TF-IDF Vectorization

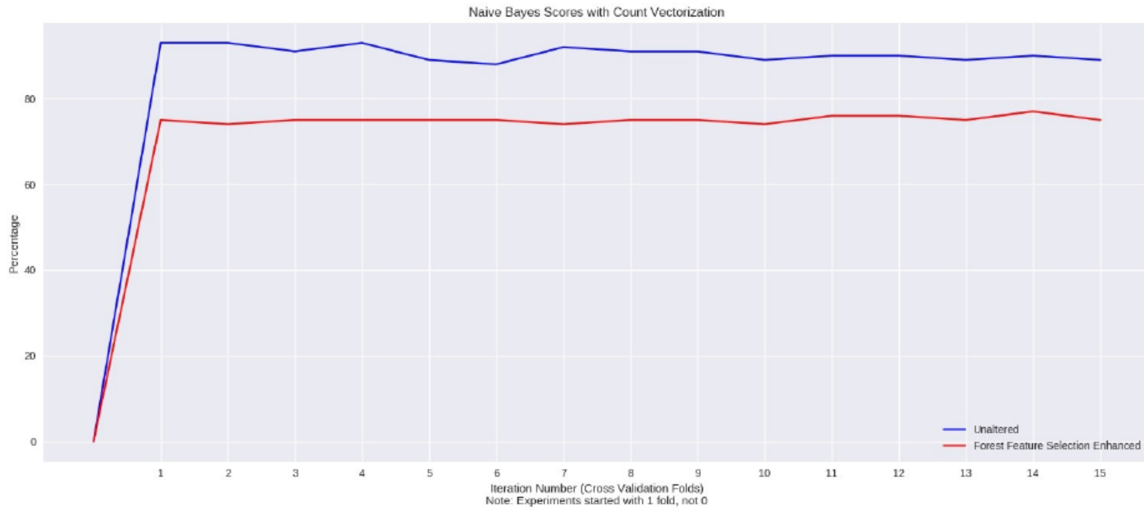

Figure 24: Gaussian Naive Bayes Accuracy Percentages with Count Vectorization

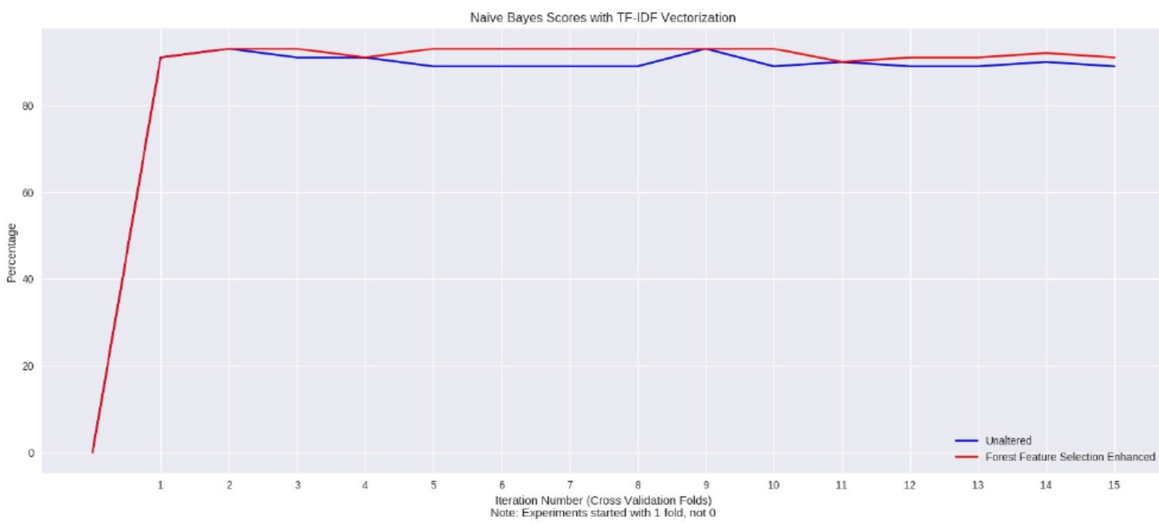

Figure 25: Gaussian Naive Bayes Accuracy Percentages with TF-IDF Vectorization

Supervised Machine Learning: Word2Vec

As opposed to the other classification models used so far in this paper, Word2Vec is slightly different, and thus requires further explanation. First, it should be noted that unlike Gaussian Naive Bayes, Stochastic Gradient Descent, and LinearSVM, Word2Vec is not just one self-contained algorithm. Moreover, Scikit-Learn has no implementation of it. Instead, the implementation used for analysis in this paper is FastText, which provides a python Word2Vec wrapper.

FastText was originally designed by the Facebook machine learning team to churn through mountains of textual data found on Facebook. Since its becoming opensource, many researchers have begun to use it for classification tasks. FastText's Word2Vec model works by taking as input one large text file as training data. From that file, it builds its own vector representation from each word in the vocabulary and outputs both a model.bin and model.vec for persistence purposes and to reload and evaluate models created by others.

One helpful way to think of Word2Vec is to understand that it uses the beneficial parts of other popular algorithms, such as CBOW and Skipgram, to churn through the text and "normalize" the data into similar vectors, which it then combines with a pre-sharpened pre-processing pipeline to great effect. Omar Levy, a prolific Word2Vec investigator, insists that by choosing to represent words in a vector space built off the backbone of some shallow neural networks and a "honed" pipeline, that Word2Vec is accomplishing rates of success that the NLP community could only dream to see, despite its relative youth compared to algorithms 20 years and older.

Because the Word2Vec implementation comes as an alternative to Scikit-Learn's package set, it is difficult to make an accurate comparison between the two fundamentally different implementations of the modeling process. I can, however, describe the relatively simple process and preliminary results upon which there are some positive findings. I should note, that while the it took several optimization and feature selection steps to get an acceptable algorithm from Scikit-Learn's set, Word2Vec managed to produce similar, albeit very slightly lower results, without much training at all.

To prepare the text for Word2Vec, I took a random sample of the complete data from both the psychopath and non-psychopath corpuses. For each line, I prepended the correct label – psychopath or non-psychopath. Following this step, I took the newly produced, singular training file and fed it to FastText to vectorize and create its own model. Once that was done, the randomly "left out" text files were then fed to the model for testing purposes. Most of the testing was done by hand, requiring me to change the text files manually each time. For this reason, the tests were run only five times under the same model. The numbers recorded below represent the 10 test runs made and the produced "precision measure" generated as output following training. The graph to display the results has been kept on the same axis for comparison to previous Scikit-Learn results. The exact numbers of the returned from the trials are found in figure 27.

(Figures 26 and 27 are on the following page)

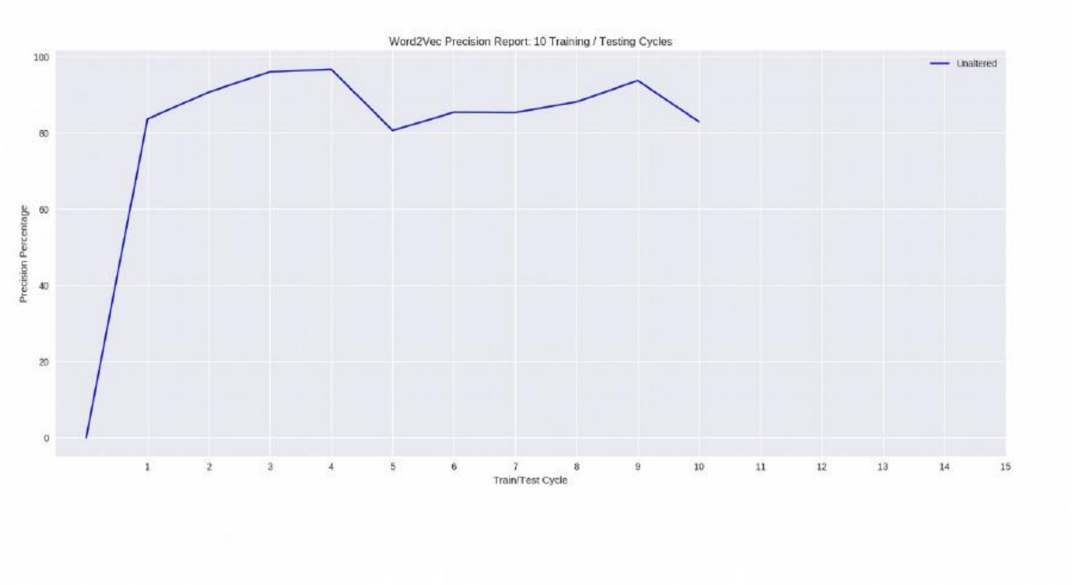

Figure 26: Word2Vec Precision Percentage over 10 Train / Test Cycles

Trial Number:	Precision:
1	83.64 %
2	90.69 %
3	96.05 %
4	96.69 %
5	80.63 %
6	85.41 %
7	85.32 %
8	88.14 %
9	93.72 %
10	82.93 %

Figure 27: Table of Precision Percentages

SECTION IV

DISCUSSION

When Hervey Cleckley published the first edition of The Mask of Sanity, he made sure to stress the point that the challenge he was undertaking was both difficult and arcane. At the time, there was no common theme tying together a class of individuals who clearly needed help and actively sought to hide from the spotlight. Because no one had made the push to outlining such an elusive profile, he started by listening and learning from the people whose lives had most been impacted. By the end of the book, he noticed that the only way to distill a diagnosis was to take into account of number of factors, both in their history and personality. Still he lamented that coming to the correct conclusion that an individual was, indeed, a psychopath was non-trivial. In fact, it took his entire career and the supplemental insight of other psychologists and criminologists to give us the ubiquitous tools used today. Regardless, however, any practicing physician or clinician hands down such a verdict only after knowing and reviewing individual case files over a long period of time. Granting such a diagnosis is one that could end a person's career, ruin their chances at parole, and upend their personal lives.

I reminded myself of this throughout the thesis research and implementation. By their very nature, psychopaths are people self-trained over a lifetime to hide in plain sight. From childhood through adolescence and into adulthood, psychopaths hone their skills to appear normal, even banal. Despite that, Cleckley believed that they would occasionally "slip up," leaving a few noticeable markers in the wake of their actions. With this project, I am forced to reconcile that the path I ventured down was one still very difficult for a trained human being to travel on, never mind that of a computer model. The best effort that I or anyone could have made to begin with is one that takes into account the character traits of psychopaths as outlined by Cleckley and his successors. To that end, I made a twopronged attack, using NLP and ML, respectively. Just as it takes a keen eye and careful ear for a psychiatrist to make a sound judgment, it also took me a concerted effort not to look singularly at the 59 problem, but to investigate clues to a person's personality by building upon pieces taken from the whole. To put it all together, then, I must assess the effectiveness of each prong and relate how it fits into the greater profile that Cleckley provided for us.

I should note that at the outset of this project, I strongly believed and expected a greater failure. It was my understanding that there is a strong difference between content and structure in writing style. My hypothesis was that by examining content in (such as the "bag of words" approach leveraged the machine learning subsection), we would see very few recognizable results, and thus the true research to be performed would be breaking ground on structural models. While this turned out to be partly true, the results are surprising. From the first approach (NLP), a few details can be stitched together. First, in the same vein that Pennebaker pursued, we can discern that like most people, the markers of a psychopathic personality rise to the surface through their use of function words – those small, often unnoticed words that few people remember to filter in everyday contact, and those very same terms that are reinforced in our own style throughout the day.

From the earliest wordclouds through to the analysis of pronoun type and frequency, there is a noticeable shift between both datasets. For example, the frequent use of abstract verbs and nouns in the psychopathic wordcloud is further reflected by the immense egotism teased out in the frequency distribution found later. The words “I” and “me” make up nearly a third of all pronoun usage in the psychopathic corpus, while terms directed at others, like “his”, “her”, “you”, and “yours,” fall much further down the line.

The inverse is true for non-psychopaths, however. The non-psychopathic corpus exhibit terms grounded in reality and focused more on others. Moreover, the words “his”, “her”, “you”, and “yours” pop-up more frequently for non-psychopaths, suggesting less self-centeredness. When considering the use of function words on the level of part of speech, the expected results show again. From my POS tagging experiment, it is clear that while both groups use comparable numbers of nouns and verbs, the psychopath corpus showed over twice as many personal pronouns when compared to the non-psychopaths. Additionally, psychopathic writing shows a spike in the use of all verb forms, which might be an indicator of increased grandiosity or speak to manipulation and megalomania. The final brick to be laid by NLP, however, is shown in the lexical dispersion analysis. If what Pennebaker and Cleckley assert to be true holds, we would expect to see certain distinguishing markers shown as word offsets in a corpus. Having gleamed some insight from the previous experiments, we see exactly what we would expect: a high use of personal pronouns dispersed throughout the corpus. In other words, if psychopathic writing is to carry a theme of narcissism, we should see a fairly dense bar on its corresponding lexical dispersion chart for the words “I” and “me,” which is exactly what is shown. Further, we would expect to see a deficiency in ego-exclusive words to mirror the inclusive ones, which is also the case. In contrast, the lexical dispersion analysis for non-psychopaths shows the reverse: a deficiency in ego-inclusive terms, and a spike in ego-exclusive terms.

The machine learning approaches require a bit more unpacking to understand. Interestingly, though, some of the analysis from the NLP approach bleeds over, specifically with respect to examining feature selection using random forests.

The unary-classification problem demonstrated results that were, on average, better than random guess under certain circumstances. Notably, the type of vectorization used was a major contributing factor. Figures 13 – 17 show that a unary-classifier built with a count vectorized was able to predict inliers to the psychopath corpus about 60 – 70 percent of the time. The same unary classification problem approached with tf-idf showed results that were less promising, but over extended durations and iterations of the model, actually showed the most growth over long periods of time. This anomaly could be due to the nature of the data being modeled, since the sources for each vary in size, time-period, and content, meaning that the mechanism by which term-frequency is related 6 In figure 8, the tag "PRP" stands for personal pronoun. 61 to importance is skewed. So, an over-performing count vectorizer may be exactly what is to be expected, because the underlying principle at work is much simpler and based off of pure word counts, making it less sensitive to the nuances built into a tf-idf vector.

Interestingly, though, both the count vectorization and tf-idf vectorization performed comparably in the binary-classification approach. When given two different datasets, psychopathic and non-psychopathic writings, each of the three Scikit-Learn models started at a base accuracy rate of no lower than 70%. After some initial optimizations using gridsearch, that number climbed even higher. The most telling improvement made, however, was the leveraging of random forests in feature selection. By picking the top 100 features for both count and tf-idf vectorizations, we got to see first-hand which features the had the most impact. It is here we found a resurgence of details from the NLP analysis. When shown graphically, we see that many of the words that carried the most "weight" in analysis were indeed function words. Further, several of these words were words that appeared in both wordclouds and in the pronoun usage analysis. This is especially interesting, because it could signify that the same traits a model found impactful were traits formulated by Cleckley and Hare and teased out by NLP investigation previously in the research. Moreover, these traits were discovered and utilized by the models themselves automatically, without any pruning on my part. This in itself lends credibility to the claims made by psychologists for decades and bolstered by researchers doing linguistic analysis presently. So, where I had originally though that a vector-space model might be inherently useless, it would seem that that is not necessarily the case. Not only can vector-space modeling adapt to a psychopathic corpus, but it can independently dredge up the very same traits profiled by Cleckley, Hare, and Ressler generations before.

SECTION V

LOOKING FORWARD

At its core, this research was, at best, a compromise. From the beginning, I held no delusions that it would be possible to completely remove a psychopath's "mask of sanity" by closely examining their use of language. At most, I expected to see a few positive results that indicated potential to classify an individual as psychopathic or non-psychopathic. To that end, there were some successes. However, much progress is needed to completely remove the mask.

First, the data collection is admittedly smaller than is desirable. The biggest direct improvement to be made would be to gain access to one of the aforementioned datasets currently unavailable. The data used for this project was collected by hand from a variety of sources, which in the long run may have belabored certain aspects of the process, such as dramatic differences seen between the unary-classification and binary-classification. A richer dataset may present a smoothing mechanism that could force the model to conform to norms. As it stands now, because there is such sparse data, any outlier is felt with much greater effect. More data would dampen the detrimental features that were incorrectly calculated, thus improving overall accuracy.

Second, this research took a "bag of words" approach to the machine learning models. Whenever a bag of words is used, much of the information surrounding the words is removed. This trimming of context cold be inadvertently removing much of the valuable information and features indicative of a person's overall style. Therefore, one potentially fruitful investigative route would be to look more deeply into structural models over bag of words. That approach could look very different from normal practice. For example, tracing a person's overall syntactic patterns and use (or misuse) of grammar might be as revealing of a person's mental state as the content is, if not more. Part of this thesis was intended to either validate or disqualify a bag of words approach to modeling the nuances of the human mind. What was ultimately learned is that while the bag of words model is not entirely helpful under every set of circumstances, it still has value. The most successful future developments might actually arise from a blending of content and structural approaches.

Third, there is much left to investigate with respect to Word2Vec and its derivatives (Word2Vec, Sent2Vec, and Para2Vect all belong to the same algorithmic "family"). If Omer Levy is to be believed, Word2Vec is ushering in a renaissance of text processing and classification. The progress and potential Word2Vec has shown in this paper is quite remarkable, considering no hyperparameter tuning or optimizations were performed on it. Despite lacking some of the richer features that come in concert to the Scikit-Learn models, it performed comparably well. It would be worth researching to see how it stands up with added hyperparameter tuning or feature selection algorithms. Likewise, it would be valuable to see multiple experiments run concomitantly with a larger dataset and against Sent2Vec and Para2Vec. The amount of times Word2Vec was run in this paper was limited due to time constraints, so an individual study narrowing the scope may be in order for a second glance.

The human mind is one of the most complex structures created in nature. Within the confounds of a three-pound mass of organic tissue, a wealth of memories, personality, and emotions coordinate with each other to produce the sum total of a human being. Finding ways to detect specific aspects of the human mind is already a challenge – modeling it to diagnose a poorly understood mental condition is even more so. Regardless, this research was intended to chip away at a mask that so many people continue to use throughout their lives. While there is doubt to be cast with respect to data, sources, and modeling techniques, the over-arching lesson here is that the human mind is imperfect and does, on occasion, release indicators for problems simmering below the surface. And while this study is far from exhaustive and definitive, it has teased out a few points of entry for later probing. It took Cleckley a lifetime to develop and test his own theories about psychopathy, and the profile he originally created is still evolving to this day. Perhaps by taking the first steps to prying away the mask of sanity is also the first steps into venturing down a path chasing an ever-moving target in plain sight.

BIBLIOGRAPHY

[1]J. Rogan, "Sam Harris", *The Joe Rogan Experience*, 2016. .

[2]*The Diagnostic and Statistical Manual of Mental Health Disorders*, 5th ed. Arlington, Virginia: American Psychiatric Association, 2013.

[3]M. Thomas, *Confessions of a Sociopath: A Life Spent Hiding In Plain SIght*, 1st ed. New York, New York: Broadway Books, 2013.

[4]M. Traxler Jr., *Introduction to Psycholinguistics: Understanding Language Science*, 1st ed. Sussex, England: John Wiley and Sons Ltd, 2012.

[5]D. Cleckley, *Mask Of Sanity*, 1st ed. Mansfield Centre, Connecticut: Martino Publishing, 2015.

[6]E. Semino and J. Culpeper, *Cognitive stylistics*, 1st ed. Amsterdam: J. Benjamins Pub. Co., 2002.

[7]S. Bird, E. Klein and E. Loper, *Natural language processing with Python*, 1st ed. Beijing, China: O'Reilly, 2009.

[8]R. Mitkov, *The Oxford handbook of computational linguistics*, 1st ed. Oxford, England: Oxford University Press, 2003.

[9]E. Kumar, *Natural Language Processing*, 1st ed. New Delhi, India: I.K. International Publishing House Pvt. Ltd., 2011.

[10]C. Bishop, *Pattern Recognition and Machine Learning*, 1st ed. New York, New York: Springer, 2006.

[11]S. Raschka, *Python Machine Learning*, 1st ed. Birmingham, England: Packt Publishing, 2015.

[12]K. Dutton, *The Wisdom of Psychopaths: What Saints, Spies, and Serial Killers can teach us about Success*, 1st ed. New York, New York: Scientific American/Farrar, Straus and Giroux.

[13]J. Fallon, *The Psychopath Inside*, 1st ed. USA: Penguin Books, 2013.

[14]R. Ressler and T. Schachtman, *I Have Lived in the Monster: Inside the Minds of the World's most Notorious Serial Killers*, 1st ed. London: Pocket, 1998.

[15]J. Pennebaker, *The Secret Life of Pronouns: What our Words say about us*, 1st ed. New York: Bloomsbury Press, 2013.

[16]"LIWC | Linguistic Inquiry and Word Count", *Liwc.wpengine.com*, 2017. [Online]. Available: https://liwc.wpengine.com/. [Accessed: 24- Mar- 2017].

[17]R. Hare, *Without Conscience*, 1st ed. New York, New York: Guilford Publications, 2012.

[18]J. Kantor, "The Psychology of Feeling or Affective Reactions", *The American Journal of Psychology*, vol. 34, no. 3, p. 433, 1923.

[19]M. McLuhan and W. Gordon, *Understanding media*, 1st ed. New York, New York: Gingko Press, 2013.

[20]T. Gruber, "A translation approach to portable ontology specifications", *Knowledge Acquisition*, vol. 5, no. 2, pp. 199-220, 1993.

[21]"Welcome to Python.org", *Python.org*, 2017. [Online]. Available: https://www.python.org/. [Accessed: 24- Mar- 2017].

[22]"R: The R Project for Statistical Computing", *R-project.org*, 2017. [Online]. Available: https://www.r-project.org/. [Accessed: 24- Mar- 2017].

[23]K. Hornik, "Apache OpenNLP Tools Interface [R package openNLP version 0.2-6]", *Cran.r-project.org*, 2017. [Online]. Available: https://cran.r-project.org/web/packages/openNLP/index.html. [Accessed: 24- Mar- 2017].

[24]"MathWorks - Makers of MATLAB and Simulink", *Mathworks.com*, 2017. [Online]. Available: https://www.mathworks.com/. [Accessed: 24- Mar- 2017].

[25]"GNU Octave", *Gnu.org*, 2017. [Online]. Available: https://www.gnu.org/software/octave/. [Accessed: 24- Mar- 2017].

[26]S. Faridani, "An NLP library for Matlab", *AMPLab - UC Berkeley*, 2017. [Online]. Available: https://amplab.cs.berkeley.edu/an-nlp-library-for-matlab/. [Accessed: 24- Mar- 2017].

[27]"Four main languages for Analytics, Data Mining, Data Science", *Kdnuggets.com*, 2017. [Online]. Available: http://www.kdnuggets.com/2014/08/four-main-languages-analytics-data-mining-data-science.html. [Accessed: 24- Mar- 2017].

[28]A. Ng, "Machine Learning - Stanford University | Coursera", *Coursera*, 2017. [Online]. Available: https://www.coursera.org/learn/machine-learning. [Accessed: 24- Mar- 2017].

[29]"scikit-learn: machine learning in Python — scikit-learn 0.18.1 documentation", *Scikit-learn.org*, 2017. [Online]. Available: http://scikit-learn.org/stable/. [Accessed: 24- Mar- 2017].

[30]R. Terman, "Scraping New York Times Articles with Python: A Tutorial | D-Lab", *Dlab.berkeley.edu*, 2017. [Online]. Available: http://dlab.berkeley.edu/blog/scraping-new-york-times-articles-python-tutorial. [Accessed: 24- Mar- 2017].

[31]"The Economist", *Economist.com*, 2017. [Online]. Available: http://www.economist.com/. [Accessed: 24- Mar- 2017].

[32]C. O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, 1st ed. [S.l.]: Penguin Books, 2016.

[33]D. West, *I.A. Richards and the Rise of Cognitive Stylistics*, 1st ed. London, England: Bloomsbury, 2013.

[34]"Penn Treebank Part of Speech Project", 2017. [Online]. Available: https://www.ling.upenn.edu/courses/Fall_2003/ling001/penn_treebank_pos.html. [Accessed: 24- Mar- 2017].

APPENDICES

APPENDIX A: PSYCHOPATH CORPUS MANIFEST

FILE NAME	DOCUMENT TYPE	SOURCE
Bundy_Final_Interview	Interview	Pureintimacy.org/f/fatal-addiction-ted-bundys-final-interview
Bundy_Letter_1	Correspondence	www.f4.ca/text/pen-pals.html
Bundy_Letter_2	Correspondence	www.f4.ca/text/pen-pals.html
Bundy_Letter_3	Correspondence	www.f4.ca/text/pen-pals.html
Dahmer_Interview_1	Interview	[14]
Dahmer_Interview_2	Interview	[14]
Dahmer_Interview_3	Interview	[14]
Dahmer_Interview_4	Interview	[14]
Dahmer_Interview_5	Interview	[14]
Dutton_Make_Me_a_Psychopath	Book Chapter	[12]
Fallon_Atlantic_Interview	Interview	www.theatlantic.com/health/archive/2014/01/life-as-a-nonviolent-psychopath/282271/
Fallon_Huff_Post_Interview	Interview	www.huffingtonpost.com/nick-simmons/interview-with-the-psycho_b_5534192.html
Gacy_Letter_1	Correspondence	www.huffingtonpost.com/2012/08/15/john-wayne-gacy-letter-serial-killer-luis-kutner-photo_n_1783489.html
Hinckley_Letter_1	Correspondence	www.f4.ca/text/pen-pals.html
Hinckley_Letter_2	Correspondence	www.f4.ca/text/pen-pals.html
Hinckley_Letter_3	Correspondence	www.f4.ca/text/pen-pals.html
Jack_Letter_1	Correspondence	www.jack-the-ripper.org
Jack_Letter_2	Correspondence	www.jack-the-ripper.org

Jack_Letter_3	Correspondence	www.jack-the-ripper.org
Jack_Letter_4	Correspondence	www.jack-the-ripper.org
Jack_Letter_5	Correspondence	www.jack-the-ripper.org
Jack_Letter_6	Corresponden-ce	www.jack-the-ripper.org
James_Fallon_Chapter_Sampl e	Book Chapter	[13]
Kaczynski_Letters	Correspondence	Archive.org/stream/al_Ted_Kaczynski_Letter_to_a_Turkish_anarchist_a4/Ted_Kaczynski_Letter_to_a_Turkish_anarchist_a4_djvu.txt
Manson_Interview_1	Interview	MansonDirect.com
Manson_Interview_2	Interview	MansonDirect.com
Rader_Confession	Legal Documents	Edition.cnn.com/2005/LAW/06/27/rader.transcript/
Thomas_Chapter_Sample	Book Chapter	[3]

APPENDIX B: NON-PSYCHOPATH CORPUS MANIFEST

FILE NAME	**DOCUMENT TYPE**	**SOURCE**
Rand_Samp_Long_1	Article	NYT Database
Rand_Samp_Long_2	Article	*The Economist* Database
Rand_Samp_Long_3	Article	*The Economist* Database
Rand_Samp_Long_4	Article	*The Economist* Database
Rand_Samp_Medium_1	Article	NYT Database
Rand_Samp_Medium_2	Article	NYT Database
Rand_Samp_Medium_3	Article	NYT Database
Rand_Samp_Medium_4	Article	NYT Database
Rand_Samp_Medium_5	Article	NYT Database
Rand_Samp_Medium_6	Article	NYT Database
Rand_Samp_Medium_7	Interview	NYT Database
Rand_Samp_Medium_8	Article	NYT Database
Rand_Samp_Medium_9	Article	NYT Database
Rand_Samp_Small_1	Article	NYT Database
Rand_Samp_Small_2	Op-Ed	NYT Database
Rand_Samp_Small_3	Op-Ed	NYT Database
Rand_Samp_Small_4	Article	NYT Database
Rand_Samp_Small_5	Sports	NYT Database
Rand_Samp_Small_6	Sports	NYT Database
Rand_Samp_Small_7	Sports	NYT Database
Rand_Samp_Small_8	Sports	NYT Database

Rand_Samp_Small_9	Sports	NYT Database
Rand_Samp_Small_10	Sports	NYT Database
Rand_Samp_Small_11	Article	NYT Database
Rand_Samp_Small_12	Article	NYT Database
Rand_Samp_Small_13	Article	NYT Database
Rand_Samp_Small_14	Op-Ed	NYT Database
Rand_Samp_Small_15	Article	NYT Database

APPENDIX C: PART OF SPEECH TAGGER CHART

TAG	**PART OF SPEECH**
CC	Coordinating conjunction
CD	Cardinal number
DT	Determiner
EX	Existential *there*
FW	Foreign Word
IN	Preposition or subordinating conjunction
JJ	Adjective
JJR	Adjective, comparative
JJS	Adjective, superlative
LS	List item marker
MD	Modal
NN	Noun, singular or mass
NNS	Proper noun, plural
NNP	Proper noun, singular
NNPS	Proper noun, plural
PDT	Pre-determiner
POS	Possessive ending
PRP	Personal pronoun
PRP$	Possessive pronoun
RB	Adverb
RBR	Adverb, comparative

RBS	Adverb, superlative
RP	Particle
SYM	Symbol
To	*to*
UH	Interjection
VB	Verb, base form
VBD	Verb, past tense
VBG	Verb, gerund or present participle
VBN	Verb, past participle
VBP	Verb, non-3rd person singular present
VBZ	Verb, 3rd person singular present
WDT	Wh-determiner
WP	Who-pronoun
WP$	Possessive wh-pronoun
WRB	Wh-adverb

This chart was taken as an adaptation from the Penn Treebank Project [34].

VITA

Born in 1989, Andrew Henning began his formal university education at the age of 17 at the University of Mississippi, studying English and mathematics. After participating in a successful research project involving an investigation into an alleged seventh Shakespeare signature, he chose to continue his work by enrolling as a graduate student in digital humanities at King's College London. Upon completion of his graduate work, Mr. Henning returned to work as the Barksdale Fellow of Digital Humanities at the Sally McDonnell Barksdale Honors College at the University of Mississippi, helping honors students develop their own projects and ideas for their thesis research. Realizing that his interests were becoming more aligned with the technological aspect of text analysis, he pursued a second graduate degree in computer science while simultaneously teaching courses in programming, digital media, and computer organization and assembly language. Currently, Mr. Henning's research interests include text processing, machine learning, and teaching.