ISLAMIC UNIVERSITY OF TECHNOLOGY

Detecting The Dark Triads Using Social Media Data

By:

Kazi Rezoanur Rahman (144410)

Prottoy Hashem (144427)

A thesis submitted in partial fulfilment of the requirements for the degree of Bachelor of Science in Computer Science and Engineering

Academic Year: 2017-2018

Department of Computer Science and Engineering
Islamic University of Technology.
A Subsidiary Organ of the Organization of Islamic Cooperation.
Dhaka, Bangladesh.

October 2018

Declaration of Authorship

We, Kazi Rezoanur Rahman and Prottoy Hashem, declare that this thesis titled, ‘Detecting The Dark Triads Using Social Media Data’ and the work presented in it are our own. We confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.

Submitted By:

Kazi Rezoanur Rahman (144410)

Prottoy Hashem (144427)

Detecting The Dark Triads Using Social Media Data

Approved By:

MD Abed Rahman

Thesis Supervisor,

Lecturer

Department of Computer Science and Engineering,
Islamic University of Technology.

Keywords

Dark triads, Phrase-machine, Bag-of-ngrams, Psychometric tests, Narcissism, Psychopathy, Machiavennilism, Personality detection.

Abstract

This research proposes a method to detect the dark triads from a person's social media behavior. Currently there is a spread of lies and deception in the online community. Our proposed method gets a person's most recent Social media updates and predicts a person's Narcissism, Psychopathy, Machiavellianism. These three characteristics are part of **The Dark Triads.** We receive the data from Facebook and take surveys to get the class label. We treat this as a binary classification problem. We tested across several supervised machine learning algorithm and received reasonable amount of accuracy.

Table of Contents

Chapter 1: Introduction

This chapter outlines the background (section 1.1) and context (section 1.2) of the research, and its purposes (section 1.3). Section 1.4 describes the significance and scope of this research and provides definitions of terms used. Finally, section 1.5 includes an outline of the remaining chapters of the thesis.

1.1 BACKGROUND

In the digital age that we are living in it is extremely easy to be connected with people. As much helpful this connectivity is, it also brings a lot of problems to the table. People are often times fake and manipulative in their online interactions. Often times people use the anonymity given to them by the internet to manipulate people, provoke unnecessary feuds.

1.2 CONTEXT

While researching about false information spreading and controlling the toxic online environment we thought about developing a method to identify these types of peoples and their interactions in the online community. With the help supervised machine learning algorithm our research hopes to use a person's social media activity and an psychometric tests to detect the dark triads.

1.3 PURPOSES

The purpose of this research is to develop a method to correctly identify the peoples that fall under the characterises of the dark triads.

1.4 SIGNIFICANCE, SCOPE

The significance of this research is huge given the status quo. Properly identifying people that are toxic online helps the community as whole. Systems may be put in place that isolate their online interactions and help them in return. Currently most of the research focuses on The Big Five Personalities (openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism) but very few of them have worked on the dark triads. Moreover, we created our own dataset rather than depending on the datasets that have already been worked on or being worked on thus ensuring fresh research and new findings.

“**The Dark Triads**” consists of 3 types of personalities 1. Narcissism 2. Psychopathy 3. Machiavellian. Narcissism is excessive interest in or admiration of oneself. Psychopathy is antisocial behavior, impaired empathy, remorse, egotistical traits. Machiavellian is excessive interest in their own interests that they will manipulate, deceive, and exploit others to achieve their goals.

1.5 THESIS OUTLINE

Chapter 2 focuses on literature review, current state of the art works, chapter 3 focuses on research design as in how the research was conducted, chapter 4 describes the results, chapter 5 gives an analysis of the results, chapter 6 concludes the discussions.

Chapter 2: Literature Review

A key part of our work takes inspiration from the work done by (Handler, Denny, Wallach, & O'Connor, 2016).Social scientists who do not have specialized natural language processing training often use a unigram bag-of-words (BOW) representation when analysing text corpora. They offered a new phrase-based method, NPFST, for enriching a unigram BOW. NPFST uses a part of-speech tagger and a finite state transducer to extract multiword phrases to be added to a unigram BOW. They compared NPFST to both ngram and parsing methods in terms of yield, recall, and efficiency. Then demonstrate how to use NPFST for exploratory analyses; it performs well, without configuration, on many different kinds of English text. Finally, we present a case study using NPFST to analyse a new corpus of U.S. congressional bills. For their open-source implementation,

In paper done by (Abell & Brewer, 2014) they have investigated whether Machiavellian men and women employ self-presentation tactics (i.e. self-monitoring and self-promotion) and how honest they are in their interactions on Facebook. The study investigated the potential relationship between Machiavellianism and self-promotion, self-monitoring, dishonest self-promotion and relational aggression via Facebook separately for male and female participants. The Results did not differ between online and offline participants so they were analysed together. Participants first provided preliminary demographic information (age, gender) and then completed a series of items (devised by the researchers) assessing Facebook activity (e.g. frequency of viewing friend's activity). The Mach IV (Christie, 1970) measures the cynicism, morality and manipulative behaviour which constitute Machiavellianism. Facebook self-promotion was measured using five statements derived from (Carpenter & J, 2012) Items included 'How often do you post status updates to Facebook' and 'How often do you tag pictures of yourself on Facebook?' Participants responded on a 5-pont Likert scale from 1 = never to 5 = all the time. In the present study, 16 statements were selected and adapted to measure participants self-monitoring of behaviour on Facebook. Adapted statements included 'When I am uncertain of what to put as a status update, I look at the updates of my Facebook friends' and 'Even if I am not enjoying myself, I often pretend on Facebook that I am'. The honesty of self-promotion behaviours was measured using 14 statements. These items (devised by the researchers) included: ‘I often update my status saying I am doing something exciting even though this is not true’ and ‘I often send friend requests to people I don’t know in order to increase my number of Facebook friends’. Participants responded on a 5-point Likert Scale from 1 = strongly disagree to 5 = strongly agree and seven items were reverse coded to create a total dishonesty self-promotion score. Relational aggression specific to Facebook activity was measured using 19 statements developed by the researchers. Participants were asked to respond to this with reference to a close friend whom they interact with both offline and via Facebook. These statements include: ‘I often ignore my friend when they try to speak to me on Facebook chat’ and ‘I often write something embarrassing about my friend in my Facebook status’. Participants responded on a 5-point Likert scale from 1 = strongly disagree to 5 = strongly agree). In the present study, all scales demonstrated acceptable reliability: Higher scores represent higher levels of Machiavellianism, self-promotion, self-monitoring, relational aggression and dishonest self-promotion which contains a greater amount of dishonesty. Correlation analyses revealed significant positive and negative relationships between Machiavellianism and self-promotion, self-monitoring, relational aggression and dishonest self-promotion. Standard regressions were conducted to investigate the influence of Machiavellianism on Facebook self-promotion, self-monitoring, dishonest self-promotion and relational aggression. Participants are likely to over-estimate possession of socially desirable qualities and the amount of time spent on Facebook. Therefore, future research should consider the direct evaluation of Facebook profiles, such as the frequency and type of information shared in Facebook updates and the descriptions and sharing of photographs. Future Studies should also focus on cross cultural evaluations and also with large number of male participants. Not only focusing on Facebook but also other online contexts such as dating sites can also be interesting targets. Machiavellianism women engaged in dishonest self-promotion and relational aggression on Facebook whilst Machiavellianism men engaged in self-promoting behaviour. Both high Machiavellian men and women engaged in greater self-monitoring on Facebook than those with lower levels of this personality trait.

Work done by (Rubin, Conroy, Chen, & Cornwell, 2016) focus on providing a conceptual overview of satire and humour, elaborating and illustrating the unique features of satirical news, which mimics the format and style of journalistic reporting. Satirical news stories were carefully matched and examined in contrast with their legitimate news counterparts in 12 contemporary news topics in 4 domains (civics, science, business, and "soft" news). Several factors contribute to the believability of fake news online. Recent polls have found that only 60% of Americans read beyond the headline Unless a user looks specifically for the source attribution, an article from The Onion looks just like an article from a credible source, like The New York Times. In this study they collected and analysed a dataset of 360 news articles as a wide-ranging and diverse data sample, representative of the scope of US and Canadian national newspapers. The dataset was collected in 2 sets. The first set was collected from 2 satirical news sites (The Onion and The Beaverton) and 2 legitimate news sources (The Toronto Star and The New York Times) in 2015. The 240 articles were aggregated by a 2 x 2 x 4 x 3 design (US/Canadian; satirical/legitimate online news; varying across 4 domains (civics, science, business, and "soft" news) with 3 distinct topics within each of the 4 domains. For each of the 12 topics, 5 Canadian (from The Beaverton) and 5 American (from the Onion) satirical articles were collected. Each satirical piece was then matched to a legitimate news article that was published in the same country, and as closely related in subject matter as possible. A trained linguist content-analysed each pair (legitimate vs. satirical), looking for insights on similarities and differences as well as trends in language use and rhetorical devices. For machine learning they used the combined set of 360, reserving random 25% of the combined 2 sets data for testing, and performing 10-fold cross-validation on the training set. They collected features like Absurdity and Humour, Sentence Complexity. They propose and test a set of 5 satirical news features: Absurdity, Humour, Grammar, Negative Affect and Punctuation. The method begins with performing a topic-based classification followed by sentiment-based classification, and feature selection based on absurdity and humour heuristics. The training and evaluation of their model uses a state-of-the-art method of support vector machines (SVMs). They combined cross-validation with the holdout method in reporting overall model performance. Cross validation on their 75% training produced a performance prediction on incoming data. They then confirmed this prediction in a second stage using their 25% hold out testing set. This allowed them to investigate which records from the set were incorrectly predicted by the model.

“Automatic deception detection: Methods for finding fake news” a paper done by (Conroy, Rubin, & Chen, 2016) surveyed the available methodology for detecting fake news online. They divided the types in two major categories 1) Linguistic Approaches 2) Network Approaches

Linguistic Approaches focus on liars using their language strategically to avoid being caught. In spite of the attempt to control what they are saying, language “leakage” occurs with certain verbal aspects that are hard to monitor such as frequencies and patterns of pronoun, conjunction, and negative emotion word usage (Feng & Hirst, 2013). Data Representation Perhaps the simplest method of representing texts is the “bag of words” approach, which regards each word as a single, equally significant unit. In the bag of words approach, individual words or “n-grams” (multiword) frequencies are aggregated and analysed to reveal cues of deception. Many deception detection researchers have found this method useful in tandem with different, complementary analysis (Zhang, Fan, Zheng, & Liu, 2012) and (Ott, Cardie, & Hancock, 2013).Semantic Analysis As an alternative to deception cues, signals of truthfulness have also been analysed and achieved by characterizing the degree of compatibility between a personal experience (e.g., a hotel review) as compared to a content “profile” derived from a collection of analogous data. This approach extends the n-gram plus syntax model by incorporating profile compatibility features, showing the addition significantly improves classification performance (Feng & Hirst, 2013). The intuition is that a deceptive writer with no experience with an event or object (e.g., never visited the hotel in question) may include contradictions or omission of facts present in profiles on similar topics. Classifiers Sets of word and category frequencies are useful for subsequent automated numerical analysis. One common use is for the training of “classifiers” as in Support Vector Machines (SVM) (Feng, Banerjee, & Banerjee, 2012) and Naïve Bayesian models (Ciampaglia, et al., 2015). Naïve Bayes algorithms make classifications based on accumulated evidence of the correlation between a given variable (e.g., syntax) and the other variables present in the model (Mihalcea & Mihalcea, 2009). Amongst the Network Approaches checking “Social Network Behaviour” stands out. The recent use of twitter in influencing political perceptions (Cook, Waugh, Waugh, Hashemi, & Rahman, 2014) is one scenario where certain data, namely the inclusion of hyperlinks or associated metadata, can be compiled to establish veracity assessments. Centering resonance analysis (CRA), a mode of network-based text analysis, represents the content of large sets of texts by identifying the most important words that link other words in the network.

Since we were looking for persons online activity work done on Facebook's own **myPersonality** dataset seemed like a good place to start. (Ferwerda, Tkalcic, & Schedl, 2017) focused on using this dataset to investigate relationship between personality and music genre preferences. Next to users' personality scores, this subset consists of the listening history of Last.fm users. The myPersonality app was launched on Facebook that collected users' information and their friends' information This dataset also had users listening history. For each user, the artists that were listened to were aggregated by the indexed genre with their play-count. The genre play-count for each user was then normalized to represent a range of $r\epsilon[0,1]$, this in order to be able to compare users with differences in the total amount of listening events. Users' personality in the myPersonality application was assessed using the Big Five Inventory to measure the constructs of the Five factor model: openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism. While analyzing these data zero play count artists were removed as they did not hold any value to the study. Spearman's correlation was computed between personality traits and the genre play-count. Alpha levels were adjusted using the Bonferroni correction to limit the chance on a Type I error. The reported significant results comply with alpha levels of $p <.001$. Results varied for different types of personality among the Big Five personalities.……The literature review chapter should demonstrate a thorough knowledge of the area and provide arguments to support the study focus. The aim of the literature review chapter is to delineate various theoretical positions and from these to develop a conceptual framework for generation of hypotheses and setting up the research question.

.

Chapter 3: Research Design

This chapter describes the design adopted by this research to achieve the aims and objectives stated in section 1.3 of Chapter 1. Section 3.1 discusses the methodology used in the study, the stages by which the methodology was implemented, and the research design; section 3.2 details the participants in the study; section 3.3 lists all the instruments used in the study and justifies their use; section 3.4 outlines the procedure used and the timeline for completion of each stage of the study; section 3.5 discusses how the data was analysed; finally, section 3.6 discusses the ethical considerations of the research and its problems and limitations.

3.1 METHODOLOGY AND RESEARCH DESIGN

3.1.1 Methodology

Our methodology includes collecting direct Facebook status of the user. Processing these raw texts (removing punctuation marks, words written in another language but in English etc) and applying Bagofngrams model over these texts to find most frequent words against each user. Also using ‘Phrasemachine’ (Handler, Denny, Wallach, & O’Connor, 2016) to get relevant phrases and finding out their frequency for each user. Bagofngrams model and these phrase frequencies gave us the feature set. We got the class labels by taking 3 tests for each personality of the dark triads. Openpsycometrics.com has universally accepted standard tests for all these three but for sake of privacy we made our own HTML page following the rules these standard tests are based on (as in how many points for each question). After taking the test we did a median spilt on the gathered test scores. Users who got lower values than the median were assigned false for that following test (i.e. if the median for Machiavellian was 60 then users below score 60 got labelled as false for having Machiavellian and vice versa) essentially making this a binary classification problem. This was done across all three tests (Machiavellian scored out of 100, Narcissism scored out of 40 and Psychopathy). Psychopathy tests gives 2 scores. Scores range from 1 (low) to 5 (high). Primary psychopathy(lsrp1) is the affective aspects of psychopathy; a lack of empathy for other people and tolerance for antisocial orientations. Secondary psychopathy(lsrp2) is the antisocial aspects of psychopathy; rule breaking and a lack of effort towards socially rewarded behaviour. We then applied correlation-based feature section. And applied SVM and Random forest on the top one fifth of the ranked features returned from the feature selection process. Cross validation on 80% training produced a performance prediction on incoming data and then confirmed this prediction in a second stage using 20% hold out testing set. This allowed us to investigate which records from the set were incorrectly predicted by the model.

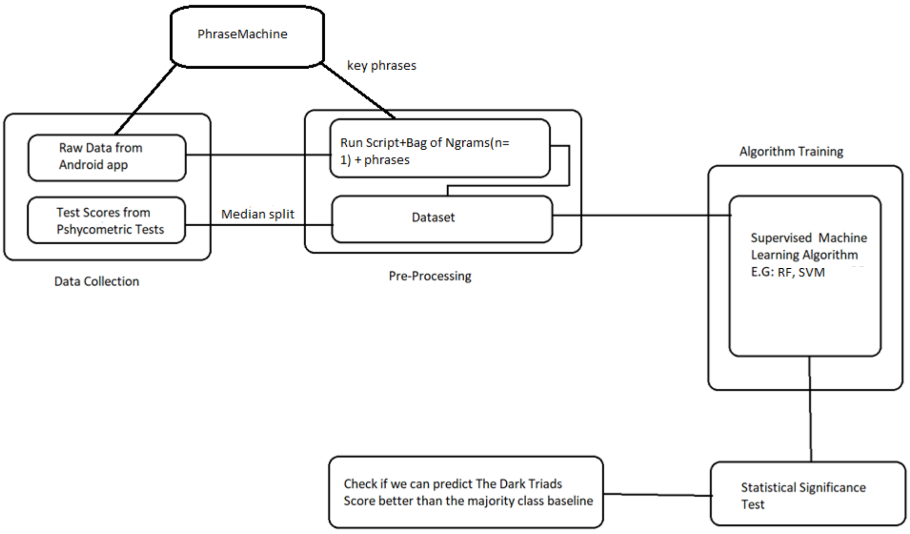

PhraseMachine
key phrases
Raw Data from Android app
Test Scores from Pshycometric Tests
Data Collection
Median split
Run Script+Bag of Ngrams(n= 1) + phrases
Dataset
Pre-Processing
Algorithm Training
Supervised Machine Learning Algorithm E.G: RF, SVM
Check if we can predict The Dark Triads Score better than the majority class baseline
Statistical Significance Test

Figure: Methodology Overview

3.2 PARTICIPANTS

Participants included 100 male students from the Islamic university of technology. Age ranged from 21-24. Since IUT students come from various parts of the country and family background and are one of the most active communities in Facebook the choice was justified. Also, the availability of the students was a key part in deciding from where we gather data.

3.3 INSTRUMENTS

The instruments or software we used the in the study included:

1. An android based app to implement Facebook’s Login API in order to securely and privately fetch the statuses.

2. Google script and google sheets to store the status found against each user.

3. Three HTML pages that included the three psychometric tests.

3. Rstudio to run ‘phrasemachine’ and extract key phrases.

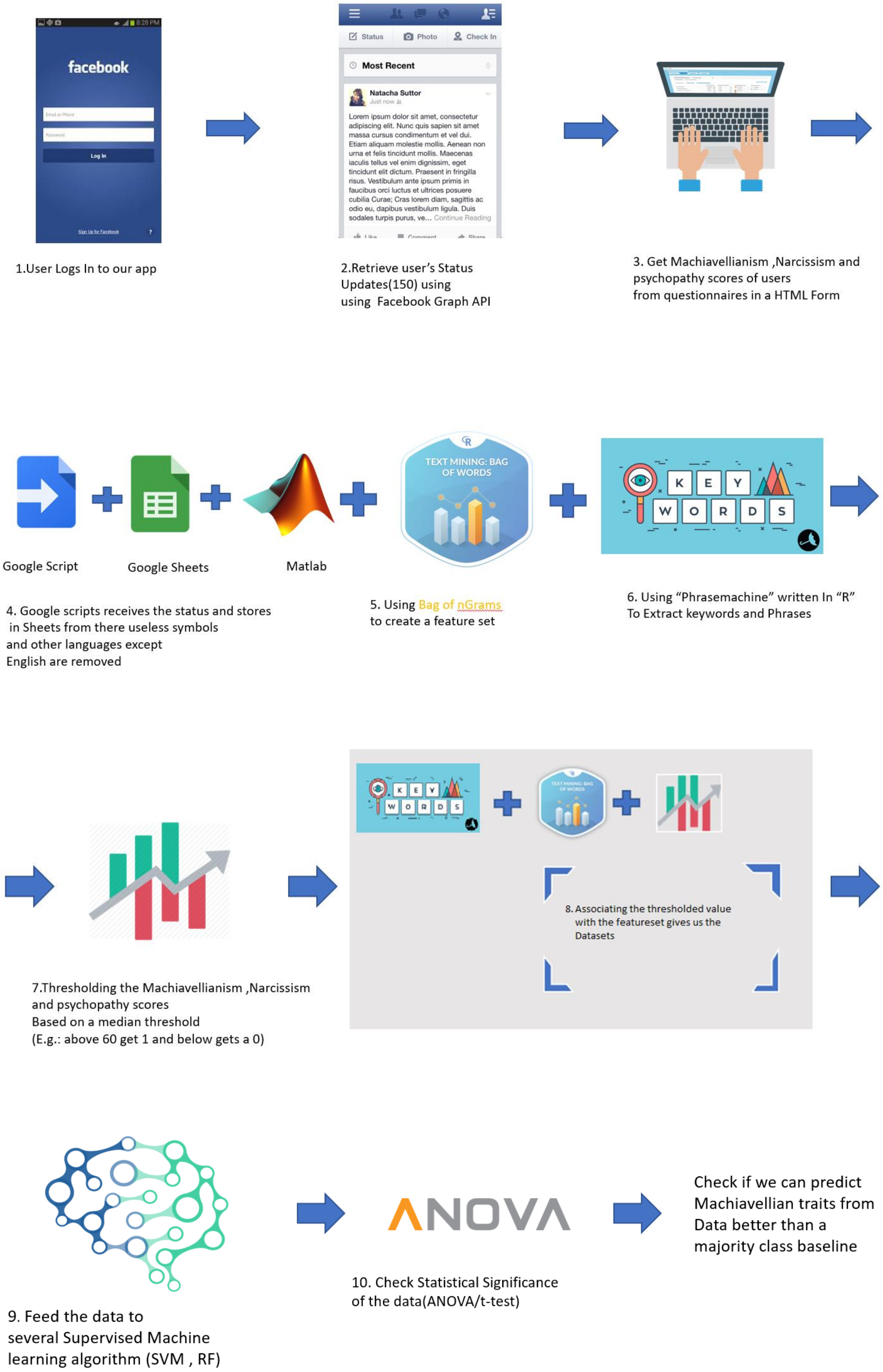
4. Matlab to clean the data, create Bagofngrams model, clean phrases, create featureset and run supervised machine learning algorithm SVM and Random forest.

3.4 PROCEDURE AND TIMELINE

At first the user logs in to Facebook using our app meaning he has given our app the permission to fetch his first 150 Facebook posts. These posts can last his entire lifetime or even span only one day. Our app sends the data to google script to google sheets. The user while data is being fetched sits down to do the 3 psychometric tests. The results are stored manually in another excel file. These data are stored in an excel file.

3.5 ANALYSIS OF THE DATA

We manually download the excel file containing the posts and process using Matlab. Meanwhile this same excel file containing the Facebook posts is fed into phrasemachine to extract key phrases. After that bagofngrams model is created using ngram = 1 and the counts of phrase frequency for each user. As we are getting each word frequency as well as important phrases it is redundant to create another featureset where ngram will be higher than 1. These two concatenated give us the featureset where each row contains users and columns contain words and phrases. Position (i.j) in the matrix gives us how many times the user ‘i’ used the word/phrase j in his 250 posts. The data from the psychometric tests give the class labels by doing a median split on them. Where each row is true/ false for each characteristic of each user. The dataset is then used to run svm and random forest. We had 4 types of datasets. 1. Dataset with Machiavellian values as class labels and ngrams + phrases as features (Termed machivelli full, lsrp1 full, lsrp2 full, narc full) 2. Dataset with only ngram values (Termed machivelli ngram, lsrp1 ngram, lsrp2 ngram, narc ngram). 3. Dataset with only phrases (Termed machivelli phrases, lsrp1 phrases, lsrp2 phrases, narc phrases) 4. Dataset without doing any sort of processing to check whether words in dataset are actually valid English words or not (Termed Machiavelli raw, lsrp1 raw, lsrp2 raw, narc raw). The Detailed procedure can be seen from the given illustration.



Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nunc quis sapien sit amet massa cursus condimentum et vel dui. Etiam aliquam molestie mollis. Aenean non urna et felis tincidunt mollis. Maecenas iaculis tellus vel enim dignissim, eget tincidunt elit dictum. Praesent in fringilla risus. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Cras lorem diam, sagittis ac odio eu, dapibus vestibulum ligula. Duis sodales turpis purus, ve... Continue Reading

1.User Logs In to our app

2.Retrieve user's Status Updates(150) using using Facebook Graph API

3. Get Machiavellianism ,Narcissism and psychopathy scores of users from questionnaires in a HTML Form

Google Script

Google Sheets

Matlab

4. Google scripts receives the status and stores in Sheets from there useless symbols and other languages except English are removed

5. Using Bag of nGrams to create a feature set

6. Using "Phrasemachine" written In "R" To Extract keywords and Phrases

7.Thresholding the Machiavellianism ,Narcissism and psychopathy scores Based on a median threshold (E.g.: above 60 get 1 and below gets a 0)

8. Associating the thresholded value with the featureset gives us the Datasets

9. Feed the data to several Supervised Machine learning algorithm (SVM , RF)

10. Check Statistical Significance of the data(ANOVA/t-test)

Check if we can predict Machiavellian traits from Data better than a majority class baseline

Figure: Step by step method from getting the data to getting results.

3.6 ETHICS AND LIMITATIONS

The main challenge of this approach was how to keep the user’s privacy in mind and also to get just enough data to do the research but not enough that it causes privacy issues. We developed our own app and used facebook’s own login api ensuring the most secure information retrieval way possible supported by facebook. We also implement our own standardized tests so that our user data does not get transmitted to outside servers. Also even if anyone gets a copy of the dataset there’s no way to tell which status was given by which user and which user has which scores because of how the data is arranged.

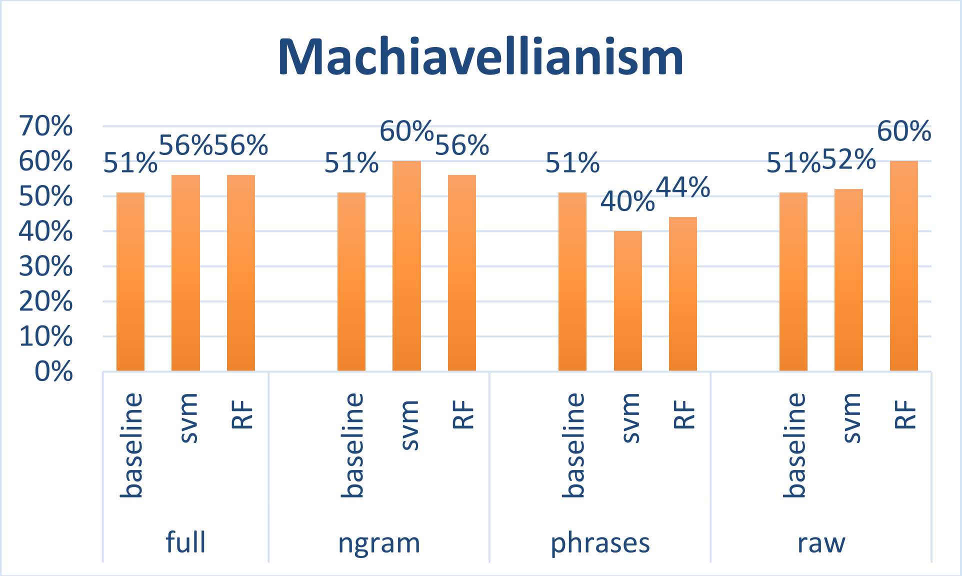
Chapter 4: Results

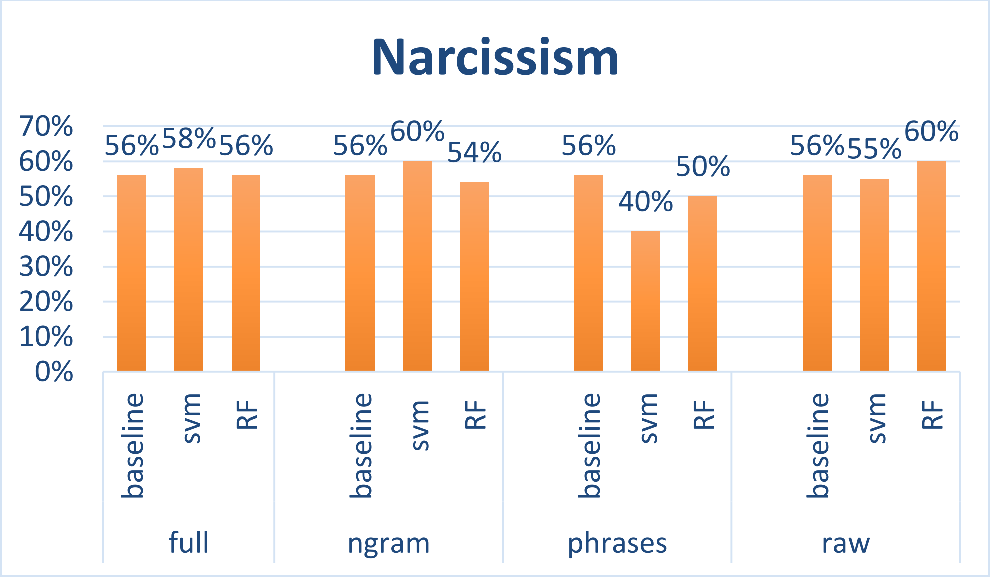
The results varied across machine learning algorithm that we implemented such as svm, random forest and Ensemble (subspace discriminants). We did holdout validation (80/20) across svm and random forest and 10-fold cross validation on ensemble (subspace discriminants) Classifying Machiavellianism was done 65% time accurately by svm highest amongst all while Random forest did a better while classify phrases for lsrp2 personality. While the lowest across all the scores was while classifying with Machiavellianism with phrases, only reaching 40-45% accuracy.

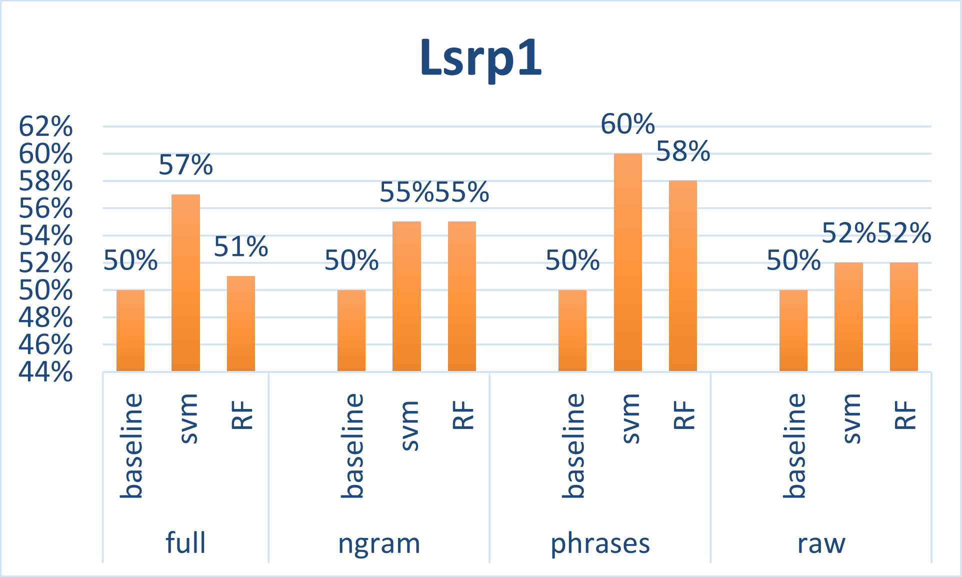
We performed statistical tests on the “Full” dataset for all of the 4 classification targets. 4 Separate one-way ANOVA were performed with Bonferroni Correlation where Independent variables were the classifiers and dependent variables were the accuracy. We found for all of the dark triads traits SVM beats the baseline, in some cases It even beats Random forest.

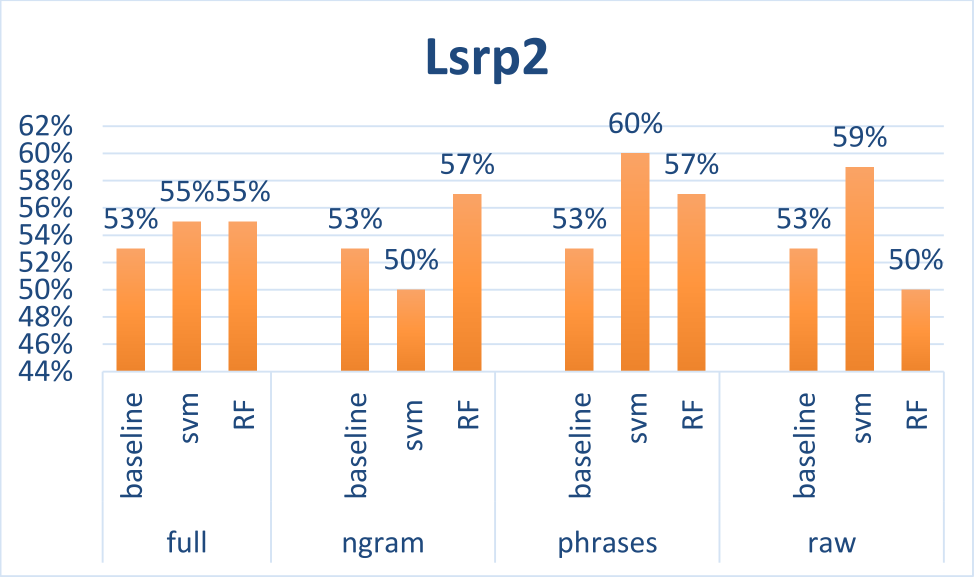
Datasets used were:

- Full = Ngrams + key phrases (Dictionary Implemented)
- Ngrams = only Ngram value where N=1(Dictionary Implemented)
- Phrases =Only key phrases (Dictionary implemented)
- Raw = Ngrams + Key phrases (without any modification)
- Correlation Based feature selection


Machiavellianism
70%
60%
50%
40%
30%
20%
10%
0%
51%
56%
56%
51%
60%
56%
51%
40%
44%
51%
52%
60%
baseline
svm
RF
baseline
svm
RF
baseline
svm
RF
baseline
svm
RF
full
ngram
phrases
raw


Narcissism
70%
60%
50%
40%
30%
20%
10%
0%
56%
58%
56%
56%
60%
54%
56%
40%
50%
56%
55%
60%
baseline
svm
RF
baseline
svm
RF
baseline
svm
RF
baseline
svm
RF
full
ngram
phrases
raw


Lsrp1
62%
60%
58%
56%
54%
52%
50%
48%
46%
44%
50%
57%
51%
50%
55%
55%
50%
60%
58%
50%
52%
52%
baseline
svm
RF
baseline
svm
RF
baseline
svm
RF
baseline
svm
RF
full
ngram
phrases
raw


Lsrp2
62%
60%
58%
56%
54%
52%
50%
48%
46%
44%
53%
55%
55%
53%
50%
57%
53%
60%
57%
53%
59%
50%
baseline
svm
RF
baseline
svm
RF
baseline
svm
RF
baseline
svm
RF
full
ngram
phrases
raw

Chapter 5: Analysis

From the result it can be concluded that the more data we received about the users the better the accuracy was. In some cases, removing redundant words were helpful such as “machivelli full” dataset. But in some cases, removing the useless words actually hurt the accuracy such as “machivelli phrases” dataset. This maybe due to the fact that the features that meant more to the classifiers were removed while we got rid of unrecognizable words. People write more important phrases in Bangla rather than in English. The raw datasets got the better scores all around meaning. Some only “ngram” datasets did reasonably better than the other such as “lsrp1 ngram” dataset which may be because of how people write. People with psychopathic tendency tend to write as little as possible and are generally anti-social so people with high lsrp1 scores wrote less words which in itself contributed to the better results.

Chapter 6: Conclusions

As we have seen people’s tendency to be antisocial, to be narcissistic and Machiavellian is co-related with what he writes. The more data there is the better the accuracy will be. Personality is not just a binary value, except for some cases (bi polar disorder) people can not just decide to be narcissistic or not be. There are levels of narcissism as well as other 2 dark triad characteristics.

Future work can focus on getting more data and various age groups. We only included male participants which can bias the results to some extent. Research shows women tend be more Machiavellian (Abell & Brewer, 2014). So, diversifying the data can lead to a better outcome or if not else a close to real world one. Future work can also be done by not treating this as a binary classification problem rather than a multiclass one. Multiclass will not only make the data more real world like also make the “levels” in personality that we talked about previously. We ran only 3 supervised machine learning algorithms future work can implement more up to date algorithms such as naïve Bayes or logic boost.

Bibliography

Abell, L., & Brewer, G. (2014). Machiavellianism, self-monitoring, self-promotion and relational aggression on Facebook. *Computers in Human Behavior*, 258-262.

Carpenter, & J, C. (2012). Narcissism on Facebook: Self-promotional and anti-social. *Personality and Individual Differences*, 482-486.

Christie, R. F. (1970). Studies in Machiavellianism. . *London: Academic Press.*

Ciampaglia, G. L., Shiralkar, P., Luis M. Rocha, Johan Bollen, Filippo Menczer, & Flammini, A. (2015). Computational Fact Checking from Knowledge Networks. *PLoS ONE*.

Conroy, N. J., Rubin, V. L., & Chen, Y. (2016). Automatic deception detection: Methods for finding fake news. *Proceedings of the Association for Information Science and Technology*, 1-4.

Cook, D. M., Waugh, B., Waugh, B., Hashemi, O., & Rahman, S. A. (2014). Twitter Deception and Influence:Issues of Identity, Slacktivism, and Puppetry. *BePress*.

Feng, S., Banerjee, R., & Banerjee, R. (2012). Syntactic stylometry for deception detection. *ACL '12 Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*, (pp. 171-175).

Feng, V., & Hirst, G. (2013). Detecting deceotive opinion with profile compatability.

Ferwerda, B., Tkalcic, M., & Schedl, M. (2017). Personality Traits and Music Genres: What Do People Prefer to Listen To? *UMAP*. Bratislava, Slovakia.

Handler, A., Denny, M. J., Wallach, H., & O’Connor, B. (2016). Bag of What? Simple Noun Phrase Extraction for Text Analysis. *NLP+CSS@EMNLP*.

Mihalcea, R., & Mihalcea, R. (2009). The lie detector: explorations in the automatic recognition of deceptive language. *ACLShort '09 Proceedings of the ACL-IJCNLP 2009 Conference*, (pp. 309-312). Suntec, Singapore.

Ott, M., Cardie, C., & Hancock, J. T. (2013). Negative Deceptive Opinion Spam. 497-501.

Rubin, V. L., Conroy, N. J., Chen, Y., & Cornwell, S. (2016). Fake News or Truth? Using Satirical Cues to Detect Potentially Misleading News., (pp. NAACL-CADD2016).

Wang, N., Kosinski, M., Stillwell, D. J., & Rust, J. (2014). Can Well-Being be Measured Using Facebook Status Updates? Validation of Facebook’s Gross National Happiness Index. *Social Indicators Research*, 483–491.

Zhang, H., Fan, Z., Zheng, i., & Liu, Q. (2012). An Improving Deception Detection Method in Computer-Mediated Communication . *JOURNAL OF NETWORKS*.