Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 3, **Number** 2, November 2023
https://doi.org/10.47709/brilliance.v3i2.3215

Prediction of Narcissistic Behavior on Indonesian Twitter Using Machine Learning Methods

Izzatul Ummah[1*]
[1]Department of Informatics, School of Computing, Telkom University, Indonesia
[1]izzatulummah@telkomuniversity.ac.id

open access

***Corresponding Author**

Article History:
Submitted: 20-11-2023
Accepted: 25-11-2023
Published: 30-11-2023

Keywords:
classification; machine learning; narcissistic behavior; prediction

ABSTRACT

Social media's explosive expansion in recent years has changed people's behavior, leading to an increase in narcissistic tendencies. Arrogance, the need to flaunt one's accomplishments, the urge to get approval from others, and excessive dreams about success, power, intelligence, attractiveness, and other attributes are characteristics of narcissistic behavior. Social media has evolved into a platform for showcasing accomplishments, particularly for people with narcissistic tendencies. Moreover, narcissistic trait is one of the three characteristics of the dark triad personality type. Several research have demonstrated that a variety of machine learning techniques can be used to predict dark triad personality traits and narcissism from social media posts. As some studies have suggested, this narcissistic behavior can further increase the level of cyberbullying in social media, while also have strong correlation with the use of chatbot for academic cheating. This research aims to build a prediction model for behavioral symptoms of narcissism based on posts on Twitter in Indonesian language, using the natural language processing technique and several basic machine learning methods (Nearest Neighbors, Naïve Bayes, Decision Tree, and Support Vector Machine), and then compare the results. We concluded that SVM model achieved the best performance, with Accuracy = 0.72 and F1 Score = 0.725.

INTRODUCTION

The swift evolution of social media has significantly influenced shifts in individuals' behavioral habits. The bright side is that individuals are growing more receptive to new ideas and are finding it simpler to look for information online. On the other hand, social media can also have a detrimental effect on people's behavior, such as the propensity to use it as a platform to brag about one's accomplishments. This is the most typical sign of narcissistic behavior.

Narcissistic conduct is frequently discussed in conjunction with two additional behaviors: psychopathy and Machiavellism. The dark triad personality is a "dark" personality type made up of these three characteristics. Narcissistic symptoms result in a kind of self-absorption, Machiavellian symptoms manifest as a desire to rule and dominate through the means of manipulating others, and psychopathological symptoms manifest as a lack of empathy for others and a sense of remorse while harming others.

Naturally, the quick growth of social media and the widening range of platforms has also contributed to an increase in narcissistic behavior signs in society. As some studies have suggested, this narcissistic behavior can further increase the level of cyberbullying in social media. And therefore, researching the patterns of narcissistic behavior on social media and developing models to identify symptoms of narcissistic behavior on social media has become highly interesting. Another significant aspect that highlights the importance of this study is that narcissistic behavior and dark triad personality may have strong correlation with the tendencies of using chatbot-generated texts for academic cheating.

Our proposed research uses natural language processing techniques and machine learning algorithms (Nearest Neighbors, Naïve Bayes, Decision Tree, and Support Vector Machine) to build a prediction model for behavioral symptoms of narcissism based on Twitter posts in Indonesian. To the best of our knowledge, there are still very few research in this area using dataset from Indonesian language. The model we build will categorize Twitter posts into two groups: positive classes, which indicate the presence of narcissist personality traits, and negative classes, which indicate their absence. Finally, we will compare the performance of each machine learning methods that we used in our experiments, using performance metrics such as Accuracy, Precision, Recall, and F1 Score.

LITERATURE REVIEW

Numerous researchers have studied this subject using different machine learning techniques. Sumner et al. carried out the first extensive study, predicting dark triad personality behavior based on Twitter posts. using the Naïve Bayes, J48, Support Vector Machine, and ensemble learning (Random Forest) (Sumner et al., 2012). Wald and Sumner used ensemble learning, which integrates four techniques, i.e. Support Vector Machine, Multi-Layer Perceptron, Logistic

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 3, **Number** 2, November 2023
https://doi.org/10.47709/brilliance.v3i2.3215

Regression, and ensemble learning (Random Forest), to predict psychopathy tendencies in Twitter users (Wald et al., 2012). Pietro developed a prediction model for dark triad behavior based on publicly shared tweets by analyzing the connection between social media behavior and the three elements of dark triad behavior (Pietro et al., 2016). Ahmad used the Bidirectional LSTM (Long Short-Term Memory) to distinguish between psychopathic (dark tried) and non-psychopathic (light triad) personalities (Ahmad et al., 2020). Asghar used Bidirectional LSTM to classify psychopath vs non-psychopath classes (Asghar et al., 2021). Hassanein predicts dark triad personality based on social media traits using the Linear Regression method with a net regularizer, Normal Linear Regression, Logistic Regression, and ensemble learning (Random Forest) (Hassanein et al., 2021). Using two machine learning methods, namely Naïve Bayes and Support Vector Machine, Haz divides Spanish Twitter posts into groups based on whether the content is narcissistic or not (Haz et al., 2022). Savci used machine learning to create a prediction model for problematic social media use (PSU), which includes (amongst other symptoms) symptom of narcissistic personality (Savci et al., 2022). The range of accuracy for all those researches mentioned above is 61-85%. Meanwhile, the most recent studies that has been conducted by Greitemeyer, examined the connection between personality characteristics (which includes narcissist behavior and dark-triad personality) and the desire to use texts created by chatbots for academic fraud (Greitemeyer & Kastenmüller, 2023). Another study conducted by Mereu, used artificial intelligence to maximize the one-way random intraclass correlation coefficient (ICC) in predicting dark triad personality trait (Mereu, 2021).

METHOD

This is the methodology that we used in our experiment:

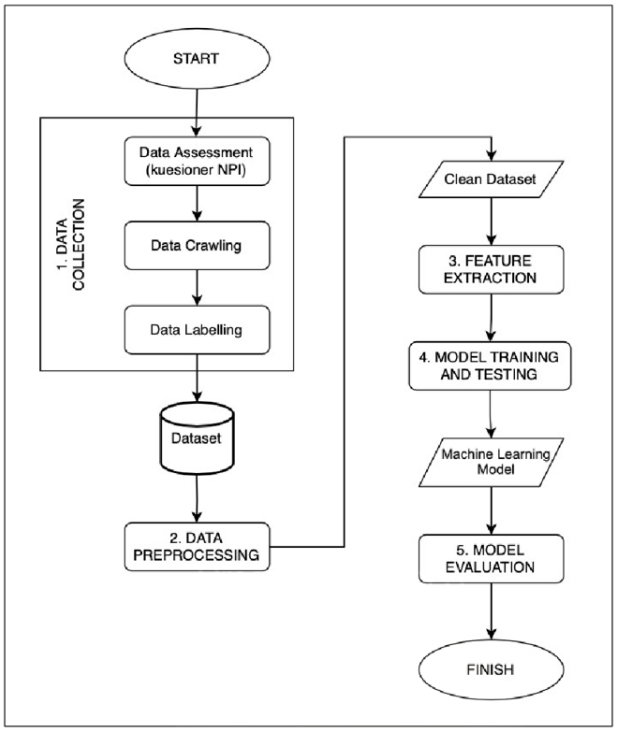

Figure 1. Research methodology

We explain the details of our research methodology below:

Data Collection

The process of data collection is divided into three tasks:

1. Data Assessment: We distributed the Narcissistic Personality Inventory (NPI) questionnaire which consists of 40 questions to measure the level of narcissistic behavior of our respondents (Raskin & Terry, 1988). All respondents were asked to fill a form of agreement which gives permission to the author to take data from their social media posts.
2. Data Crawling: We crawled the respondents' Twitter posts using Twitter API (application programming interface). The data that we crawled were taken from those respondents who had filled in the questionnaire and had given their agreement in data assessment phase before, and we also crawled data from public accounts in Twitter who had not filled in the questionnaire.

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 3, **Number** 2, November 2023
https://doi.org/10.47709/brilliance.v3i2.3215

3. Data Labelling: We labelled our data using two ways, i.e., questionnaire-based labelling and manual labelling which is done by an expert in psychology. We used these two ways because in our opinion, only using questionnaire-based labelling is not enough for this research.

Data Preprocessing

The data preprocessing tasks which we conducted consists of:

1. Lowercasing (changing all Twitter posts into lower case).
2. Removing HTML tag and URL.
3. Removing punctuation and emoji.
4. Chat word treatment (normalization).
5. Removing stop words.
6. Tokenization.
7. Stemming/lemmatization.

Feature Extraction

In this research, we used Word2Vec to extract features from Twitter posts, and to apply word embedding to change the text representations into vectors. This technique is based on a simple, basic neural network model using only one input layer, one output layer, and one hidden (projection) layer, which then be trained by iterating a corpus. Word2Vec has two variants of model architectures, i.e., CBOW (Continuous Bag of Words) and Skip-Gram. CBOW model predicts word target using a given context, whereas Skip-Gram model predicts a context using a given word. Below is the CBOW dan skip-gram model, as explained by Mikolov (Mikolov et al., 2013).

Model Training and Testing

Here we trained the dataset using several basic machine learning methods:

1. *K-Nearest Neighbors (kNN)*

K-nearest neighbor algorithm aims to find the k nearest points near a certain point in a certain set (Peng et al., 2022). The step-by-step of kNN classification algorithm is described as follow (Sun et al., 2018):

a) Computing distance: Given a data point as a data test, compute the distance from this data point (data test) to all other data points in the training dataset.
b) Finding neighbors: determine the k nearest neighbors for the given data test.
c) Classification: apply the weighted voting method to classify the data test.

2. *Naïve Bayes (NB)*

Naive Bayes is a supervised learning algorithm which is developed upon Bayes theorem, using a naive assumption that says that each feature inputs are independent of each other (Lakshmi & Kumari, 2018). Below is the formula of Bayes theorem:

$$P(h|d) = \frac{P(d|h) * P(h)}{P(d)} \tag{1}$$

Where:
h is a hypothesis that a data can be classified into a class, so if we have 2 classes (Yes or No) then we will have 2 hypotheses, i.e., h_{YES} and h_{NO}; if we have 3 classes (Cat, Dog, Horse) then we will have 3 hypotheses, i.e., h_{CAT} and h_{DOG} and h_{HORSE}.
P(h|d) is the probability of a given data d can be classified into each class.
P(d|h) is the probability of data d, within each class in the training data.
P(h) is the probability of each class throughout the whole dataset.
P(d) is the probability of the data throughout the whole dataset.

Therefore, for each class, we will have to calculate P(h_{CLASS}|d), and then we choose the class in which the calculation of P(h_{CLASS}|d) yields the biggest amongst all classes. This is formally called the maximum a posteriori (MAP) hypothesis, as described below:

$$MAP(h) = \max(P(h|d)) = max\left(\frac{P(d|h) * P(h)}{P(d)}\right) = max(P(d|h) * P(h)) \tag{2}$$

3. *Decision Tree*

Decision Tree (C4.5) is one of the popular algorithms for decision tree induction, which was proposed by Ross Quinlan in 1994 as an extension of ID3 algorithm. C4.5 algorithm can manage continuous attributes by putting

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 3, **Number** 2, November 2023
https://doi.org/10.47709/brilliance.v3i2.3215

forward two distinct tests depending on the kind of value for each property (Cherfi et al., 2018). Below are the step-by-step of C4.5 algorithm:

1) Look for base cases, such as when every sample in the list is from the same class, when no feature offers any further information, or when a previously unknown class is found.
2) Determine the normalized information gain ratio from splitting on a for each attribute a
3) Assign the highest normalized information gain to the attribute a_{best}
4) Establish a split decision node based on a_{best}
5) Add those nodes as children of node by recursing over the sub lists produced by splitting on a_{best}

4. *Support Vector Machine (SVM)*

SVM algorithm aims to use a separator that maximizes the margin between many classes in the training set. In other words, SVM's goal is to maximize the ability to generalize in a model (Cervantes et al., 2020). The SVM algorithm can be described as follow:

Given a linearly separable data $X = \{x_i, y_i\}_{i=1}^{N}$ where $x_i \in R^d$ and $x_i \in (+1, -1)$, then:

$$\langle w \cdot x^{+} \rangle + b = +1$$
$$\langle w \cdot x^{-} \rangle + b = -1 \quad (3)$$

We run separate training with each learning machine methods above using the same dataset. The model that is built will classify Twitter posts into 2 classes: the positive class that has identified symptoms of narcissism behavior, and the negative class that has not identified symptoms of narcissism behavior. The data is then split into training set (80%) and testing set (20%). The training phase will be carried out repeatedly, using the k-fold cross validation technique.

Model Evaluation

We evaluated our model using confusion metrics below:

Table 1. Confusion Matrix

Classification		Prediction	
		+	-
Actual	+	TP	FN
	-	FP	TN

Where:

True Positive (TP): if the model classifies a data to be positive, and this is true according to the actual
True Negative (TN): if the model classifies a data to be negative, and this is true according to the actual
False Positive (FP): if the model classifies a data to be positive, and this is false according to the actual
False Negative (FN: if the model classifies a data to be negative, and this is false according to the actual

The parameters that we used in this research are:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$
$$Precision = \frac{TP}{TP + FP}$$
$$Recall = \frac{TP}{TP + FN}$$
$$F1\ Score = \frac{2 * Precision * Recall}{Precision + Recall}$$

RESULT

The data used in this study is a dataset taken from Twitter social media posts in Indonesian language. After data preprocessing and cleaning, we obtained 972 posts from 36 user accounts (25 of them has filled in the NPI questionnaire). The data is labeled using two ways: automatic labeling from the result of NPI questionnaire, and manual labeling for those who did not fill in the NPI questionnaire form.

Below is the result of NPI questionnaire filled by 25 users:

Table 2. Result of NPI questionnaires

NPI Score	Number of users
0-10	2
11-20	8
21-30	12
31-40	3

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 3, **Number** 2, November 2023
https://doi.org/10.47709/brilliance.v3i2.3215

The result from Table 2 is then used to label the tweets posted by the users. We classify all tweets posted by users with NPI scores 0-20 as negative class, and all tweets posted by users with NPI score 21-40 as positive class.

Below is an example of our manual labeling result:

Table 3. Example of labeling on tweets

Tweets	Class
"Gue habis beli tas CK nih, kalian pasti gak mampu beli ini kan wkwkwk"	Positive
"Lo lulusan kampus mana sih kok kampungan banget? Kampus antah berantah ya?"	Positive
"Tadi ke rumah temen, macet banget di jalan"	Negative
"Gue heran sama temen gue, ngitung kyk gitu aja ga bisa, bodo banget deh"	Positive
"Makan di warung perempatan jalan tadi enak banget deh, murah lagi"	Negative
"Yuk kumpulin donasi…"	Negative

We implemented our model using Python 3.11.4, Scikit library (version 1.3.0), and NLTK library (3.8.1). We build four models using each method (kNN, Naive Bayes, Decision Tree, SVM), and we used kNN model as a baseline. For each model, we run the experiment using k-Fold cross validation technique (with k=5), and we split the data into training set (80%) and testing set (20%).

For kNN model, we first experimented using Manhattan distance and Euclidan distance. For each experiment, we used elbow method to find the best value for hyperparameter k. For kNN-Manhattan, we find that the best value is k=16, which obtains the lowest Error Rate (0.38), or highest Accuracy (0.62). For kNN-Euclidan, we find that the best value is k=13, which obtains the lowest Error Rate (0.37), or highest Accuracy (0.63). Therefore, we used Euclidan distance for our kNN model because it yields the best result.

Figure 2. Elbow method for kNN

For Naive Bayes model, we first experimented using Gaussian NB, Multinomial NB, Complement NB, and Bernoulli NB, and we obtain the accuracy for each experiment is 0.55, 0.59, 0.56, and 0.61 respectively. And therefore, we conclude that by using Bernoulli NB, we can obtain the highest accuracy for Naive Bayes model, which is 0.61.

For Decision Tree model, we first experimented with various parameters and hyperparameters. We then found that the best value that obtains the highest accuracy are criterion='gini', splitter='best', ccp_alpha = 0.05. The last parameter is used for pruning. For Decision Tree model, we obtain accuracy = 0.68.

For SVM model, we first experimented using several kernels and different values for hyperparameters and regularization parameter (C). Our experiment for SVM model is presented in Table 4 below. From Table 4 we see that the best result is achieved by using RBF (radial basis function) kernel.

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 3, **Number** 2, November 2023
https://doi.org/10.47709/brilliance.v3i2.3215

Table 4. Performance evaluation of each experiment for SVM model

Kernel	optimum C	optimum degree (for poly)	optimum gamma (for rbf, poly, sigmoid)	Accuracy
linear	0.81	-	-	0.580
poly	0.93	4	scale	0.625
rbf	0.85	-	scale	0.720
sigmoid	0.78	-	auto	0.703

DISCUSSION

We selected the best result from each of the four models, and then we evaluated and compared their performances to see which model suits the best with our dataset. The results of the performance evaluation of each machine learning algorithms is presented in Table 5 dan Figure 3.

Table 5. Performance evaluation of each machine learning methods

Algorithm	Accuracy	Precision	Recall	F1 Score
kNN	0.63	0.59	0.56	0.572
Naive Bayes	0.61	0.65	0.58	0.613
Decision Tree	0.68	0.68	0.66	0.670
SVM	0.72	0.73	0.71	0.725

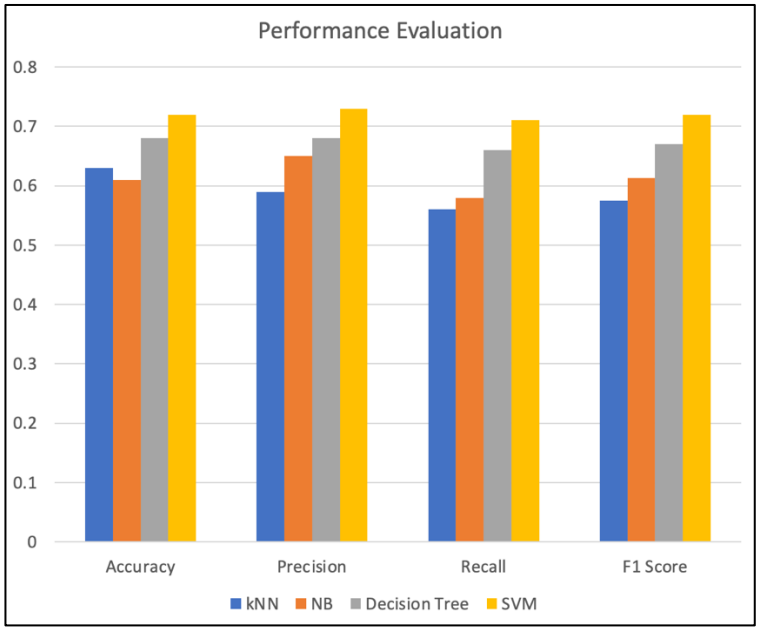

Figure 3. Performance evaluation of each machine learning methods

Based on Table 5 and Figure 3, we find that SVM model obtains the best score overall, with Accuracy, Precision, Recall, and F1 Score ≥ 0.70. Meanwhile, Decision Tree obtains modest performance, and kNN and Naive Bayes obtains the lowest performance. Our result is quite modest, compared to those result achieved by Sumner, Wald, Hassanein, and Haz; but still very low compared to the result achieved by Ahmad. Another limitation of our research is that the size of our dataset is much smaller than those researches mentioned above.

CONCLUSION

Based on dataset that we have taken from Twitter posts and NPI questionnaire, we conclude that SVM has performed better than the other machine learning methods to predict narcissistic behavior in social media. However, it should be noted that we have not weighted the labeling from both manual labeling and questionnaire-based labeling. For our further research, we suggest weighted labeling, as well as implementing more sophisticated machine learning methods such as deep neural network and ensemble learning, and we also would like to collect larger dataset to obtain a better result.

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 3, **Number** 2, November 2023
https://doi.org/10.47709/brilliance.v3i2.3215

REFERENCES

Ahmad, H., Arif, A., Khattak, A. M., Habib, A., Asghar, M. Z., & Shah, B. (2020). Applying Deep Neural Networks for Predicting Dark Triad Personality Trait of Online Users. *2020 International Conference on Information Networking (ICOIN)*, 102–105. https://doi.org/10.1109/ICOIN48656.2020.9016525

Asghar, J., Akbar, S., Asghar, M. Z., Ahmad, B., Al-Rakhami, M. S., & Gumaei, A. (2021). Detection and Classification of Psychopathic Personality Trait from Social Media Text Using Deep Learning Model. *Computational and Mathematical Methods in Medicine*. https://doi.org/10.1155/2021/5512241

Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L., & Lopez, A. (2020). A comprehensive survey on support vector machine classification: Applications, challenges and trends. *Neurocomputing*, *408*, 189–215. https://doi.org/https://doi.org/10.1016/j.neucom.2019.10.118

Cherfi, A., Nouira, K., & Ferchichi, A. (2018). Very Fast C4.5 Decision Tree Algorithm. *Applied Artificial Intelligence*, *32*(2), 119–137. https://doi.org/10.1080/08839514.2018.1447479

Greitemeyer, T., & Kastenmüller, A. (2023). HEXACO, the Dark Triad, and Chat GPT: Who is willing to commit academic cheating? *Heliyon*, *9*(9), e19909. https://doi.org/https://doi.org/10.1016/j.heliyon.2023.e19909

Hassanein, M., Rady, S., Hussein, W., & Gharib, T. (2021). Predicting the Dark Triad for Social Network Users using Their Personality Characteristics. *International Journal of Computers and Their Applications*, *28*, 204–211.

Haz, L., Rodríguez-García, M. Á., & Fernández, A. (2022). Detecting Narcissist Dark Triad Psychological Traits from Twitter. *Proceedings of the 14th International Conference on Agents and Artificial Intelligence - Volume 2: ICAART*, 313–322. https://doi.org/10.5220/0010839100003116

Lakshmi, K. V, & Kumari, N. S. (2018). Survey on Naive Bayes Algorithm. *International Journal of Advance Research in Science and Engineering (IJARSE)*, *7*(3), 240–246. http://ijarse.com/images/fullpdf/1520763647_OUCIP486ijarse.pdf

Mereu, A. (2021). Dark triad personality traits prediction with AI. *European Psychiatry*. https://doi.org/https://doi.org/10.1192/j.eurpsy.2021.386

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. *CoRR*, *abs/1301.3781*. http://dblp.uni-trier.de/db/journals/corr/corr1301.html#abs-1301-3781

Peng, H., Xu, Z., Mo, W., Wang, Y., & Huang, Q. (2022). Survey on kNN. *CAIBDA 2022; 2nd International Conference on Artificial Intelligence, Big Data and Algorithms*, 1–7.

Pietro, D. P., Carpenter, J., Giorgi, S., & Ungar, L. (2016). Studying the Dark Triad of Personality through Twitter Behavior. *2016 25th ACM International on Conference on Information and Knowledge Management*, 761–770. https://doi.org/10.1145/2983323.2983822

Raskin, R., & Terry, H. (1988). A principal-components analysis of the Narcissistic Personality Inventory and further evidence of its construct validity. *Journal of Personality and Social Psychology*, *54*(5), 890–902. https://doi.org/10.1037//0022-3514.54.5.890

Savci, M., Tekin, A., & Elhai, J. (2022). Prediction of problematic social media use (PSU) using machine learning approaches. *Current Psychology*, *41*. https://doi.org/10.1007/s12144-020-00794-1

Sumner, C., Byers, A., Boochever, R., & Park, G. J. (2012). Predicting Dark Triad Personality Traits from Twitter Usage and a Linguistic Analysis of Tweets. *2012 11th International Conference on Machine Learning and Applications*, *2*, 386–393. https://doi.org/10.1109/ICMLA.2012.218

Sun, J., Du, W., & Shi, N. (2018). A Survey of kNN Algorithm. *Information Engineering and Applied Computing*, *1*. https://doi.org/10.18063/ieac.v1i1.770

Wald, R., Khoshgoftaar, T. M., Napolitano, A., & Sumner, C. (2012). Using Twitter Content to Predict Psychopathy. *2012 11th International Conference on Machine Learning and Applications*, *2*, 394–401. https://doi.org/10.1109/ICMLA.2012.228