Identifying Human Dark Triad from Text Data Through Machine Learning Models

Sumona Yeasmin, Nazia Nowshin, Tasnia Afrin Chowdhury

Department of Computer Science and Engineering Faculty of Science East West University Dhaka, Bangladesh

DOI : https://doi.org/10.51584/IJRIAS.2024.906008

Received: 13 May 2024; Revised: 22 May 2024; Accepted: 27 May 2024;

Published: 28 June 2024

ABSTRACT

Machiavellian, narcissistic, and psychopathological personality characteristics are together referred to as the "Dark Triad". These characteristics include a lack of empathy, self-centeredness, and manipulative tendencies. Accurately identifying those with these features has significant ramifications for several fields, including psychology, criminology, and human resources. This study investigates the use of machine learning models and natural language processing (NLP) approaches to extract Dark Triad characteristics from text data. To assess the efficacy of several models for detecting people with Dark Triad qualities, we compare their results to ground truth obtained from survey data for Random Forest, K-Nearest Neighbors (KNN), Linear Support Vector Classifier (Linear SVC), Naive Bayes, and Neural Network models. The experiments yield Random Forest, KNN, Linear SVC, and neural network obtained testing accuracies of 83% to 84% for Machiavellianism; KNN achieved 67% for Psychopathy; and Neural Network reached 71% for Narcissism. Overall, the Linear SVC has consistently given better results between 60-80 percent for the three Dark Triad traits, showing constant precision and recall rates. Future aims include refining models, exploring alternative feature extraction methods, and expanding the dataset for improved accuracy and generalizability.

Keywords and phrases: Dark Triad, Machiavellianism, Narcissism, Psychopathy, Natural Language Processing, Machine Learning, Random Forest, KNN, Linear SVC, Naive Bayes, Neural Network.

INTRODUCTION

The development of natural language processing (NLP) tools has made it possible to analyze vast amounts of textual data and derive insightful conclusions. Through the automated study of language patterns, sentiment, and discourse made possible by NLP, it may be possible to recognize people who exhibit Dark Triad tendencies from their written or spoken language. NLP can help in the identification of language signals and linguistic markers connected to the Dark Triad features by utilizing machine learning techniques and linguistic analysis. Due to their distinct attributes and consequences for human behavior, the Dark Triad traits—Machiavellianism, Narcissism, and Psychopathy—have attracted a lot of interest in psychological study. These characteristics, while diverse, have much in common, such as a propensity for manipulation, a lack of empathy, and self-centeredness.

Finding people who exhibit Dark Triad qualities is extremely important in many different disciplines. Understanding these characteristics advances our understanding of human nature and the subtleties of personality in social psychology. The Dark Triad features and criminal behavior have been linked in criminology studies, raising the possibility of risk assessment and intervention techniques. Finding people who exhibit these attributes to a high degree can help with team composition, staff selection, and workplace management.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume IX Issue VI June 2024

This research aims to employ NLP techniques and machine learning models to identify individuals with Dark Triad traits from textual data. The analysis will be based on the Dark Triad Personality Test (SD3) [1], which provides a standardized measure for assessing these traits. By combining the SD3 test results with advanced computational approaches, such as BERT (Bidirectional Encoder Representations from Transformers) for feature extraction and Random Forest, KNN, Linear SVC, Naive Bayes, and Neural Network models for classification, we aim to develop an automated framework for identifying Machiavellianism, narcissism, and psychopathy from textual data.

The findings of this research will contribute to the understanding of Dark Triad traits and their manifestation in language, demonstrating the efficacy of NLP techniques by using transformer-based models to extract the contextual meaning of the text and machine learning models in identifying these traits. Moreover, the practical applications of this research extend to various domains, including psychology, criminology, and human resources.

BACKGROUND AND RELATED WORKS

Recent developments in machine learning and natural language processing have created new avenues for the study of personality characteristics in text data. Large corpora of text, such as emails, online forum conversations, or social media postings, can be analyzed to learn more about a person's linguistic and behavioral inclinations about the Dark Triad qualities.

Machiavellianism, narcissism, and psychopathy are three personality qualities together referred to as the "Dark Triad". These characteristics include a propensity for deceit, self-centeredness, a lack of empathy, and a readiness to take advantage of people. Due to its potential effects on a number of facets of human behavior, including interpersonal relationships, leadership, and even criminal behavior, the Dark Triad idea has attracted substantial study in psychology and social sciences.

Researchers are very interested in identifying and understanding these characteristics in people because doing so may reveal important information about their thought processes, social interactions, and possible hazards to society. Traditional techniques frequently use self-report questionnaires, interviews, and observations to evaluate these characteristics, but these methods can be biased, time-consuming, and unsuitable for large-scale research.

In recent years, numerous studies have focused on identifying Human Dark Triad traits through the analysis of textual data. To effectively categorize people based on how they use language and identify the presence of the Dark Triad feature, researchers have looked into the possibilities of machine learning and natural language processing (NLP) approaches. Among the most popular classifiers are Random Forest (RF), K-Nearest Neighbors (KNN), Linear Support Vector Classifier (Linear SVC), Naive Bayes, and Neural Networks. These models use linguistic patterns gleaned from textual data to categorize people according to their Dark Triad characteristics. Several relevant works will be discussed in this section.

A study by Mahmud et al. (2020) [2] aims to classify antisocial personality characteristics, specifically the dark triad traits using data collected from Facebook users. The study employed machine learning techniques, including Random Forest, Support Vector Machine, and Naive Bayes algorithms, to make predictions based on the SD3 model. The results indicated that Naive Bayes outperformed RF and SVM, particularly in predicting Machiavellianism. Also, while all classifiers performed well for Machiavellianism, accuracy was lower for narcissism and psychopathy due to data variance.

Smith et al. (2018) [3] used Naive Bayes and Random Forest classifiers to find Dark Triad characteristics in Twitter data. To train the models, they retrieved linguistic data including word frequencies, part-of-speech tags, and sentiment analysis results. The accuracy of the predictions for Machiavellianism, narcissism, and psychopathy was favorable.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume IX Issue VI June 2024

Similar techniques were used by Chen and Lee (2019) [4] to extract Dark Triad characteristics from online forum postings using K-Nearest Neighbors (KNN) and Support Vector Machine (SVM) classifiers. They obtained satisfactory classification accuracies for Machiavellianism, narcissism, and psychopathy by extracting lexical and syntactic elements from the textual data.

Miller et al. (2020) [5] examined different machine-learning techniques to categorize Dark Triad qualities from written replies, including Random Forest, Support Vector Machines (SVM), and Neural Networks. They used n-grams, word embedding, and bag-of-words feature extraction approaches. According to the results, Neural Networks outperformed Random Forest for psychopathy while achieving the maximum accuracy for Machiavellianism and narcissism.

Lee et al. (2022) [6] examine the connection between impulsivity, the Dark Triad personality traits, and addiction to social networking sites (SNS). The study takes a quantitative approach, using surveys to gather data and various statistical methods to analyze it. The results imply that when combined with impulsivity, people with higher levels of Machiavellianism, narcissism, and psychopathy are more likely to develop addictive behaviors associated with social networking sites.

Srirejeki et al. (2023) [7] investigate academic fraud in the context of online learning during the COVID-19 pandemic. Surveying 259 students across three universities, the study reveals that Dark Triad personality traits, including Narcissism, Machiavellianism, and Psychopathy, significantly influence students' intentions to engage in academic fraud. Additionally, it underscores the importance of academic integrity culture and addresses external pressures as influential situational factors. Employing partial least squares-structural equation modeling, the research offers crucial insights into the intricate interplay between personality traits and situational factors affecting academic fraud, emphasizing the need for educational institutions to cultivate integrity and support students, particularly those with Dark Triad traits, to mitigate academic misconduct in the online learning landscape.

The authors of the study (Ahmad, et al., 2020) [8] used a deep neural network model called BILSTM to effectively identify internet users' dark triad (psychopath) personality characteristics. This study deals with the issue of text-based psychopath personality identification. to distinguish between dark triad and non-dark triad reviews. The implementation of the dark triad (psychopath) and light triad (normal) classes for the text using the deep neural network model, BILSTM, has been the goal of the authors. In a review, the long-term reliance on the BiL STM model is examined to maintain contextual knowledge of both past and subsequent states. Using a deep neural network called the BILSTM model, this study's contribution predicts dark triad (psychopath) and light triad (non-psychopath) personality characteristics from tweets.

Another study by Preotiuc-Pietro, et al. (2016) [9] authors examine the relationship between observable Twitter behavior, such as platform use, posted text, profile image preference, and the dark triad. The study involved associating various traits with psychological theories and examining how they manifest in social media use. The finding was that Narcissism tends to be expressed positively, Psychopathy is characterized by a unique pattern of online behavior, often overlapping with Narcissism and Machiavellianism. Violent and hostile posts were linked to psychopathy and reflected harmful impulses, while Machiavellianism was shown to have the least amount of activity among the three traits.

The Short Dark Triad (SD3) [1], a quick proxy measure, has been created and validated by the authors (Daniel N. and Delroy L. (2014). Four research that looked at the subscales' composition, validity, and reliability in both community and student groups. Item Selection and Reduction, Concurrent Validation Against Standard Measures, and Cross-Validation of the SD3 Scales were among the authors. According to the study, the extraction, rotation, and parameterization methods used were WLSMV for the extraction, Promax for the rotation, and Theta for the parameterization.

The authors (Michal, David, and Thore 2013) [10] showed how numerous personal characteristics, including personality traits, may be predicted from digital footprints. The study indicated that certain patterns in text data might be suggestive of Dark Triad characteristics by examining Facebook "likes" and other online actions. The

suggested technique employs dimensionality reduction to prepare the Likes data for logistic/linear regression to predict individual psych demographic characteristics from Facebook likes.

In a study (Sumner, et al. 2012) [11], researchers predicted Dark Triad features from the Twitter activity of individual participants. They used linguistic analysis to find patterns linked with Machiavellianism, Narcissism, and Psychopathy by extracting variables related to language use and mood from user tweets. This study investigated the extent to which anti-social personality characteristics might be inferred from Twitter use. The authors illustrate how machine learning may be used to make beneficial predictions but falls short when it comes to forecasting a person's Dark Triad qualities based on their Twitter behavior. Although predictive models may not be appropriate for forecasting a person's personality, they may nevertheless be useful when applied to large groups of individuals, such as improving the capacity to see whether anti-social traits are increasing or decreasing over a population.

The Dark Triad was one of the personality qualities that were studied in this study by Cristiana and Michon (2011) [12] utilizing data from Twitter. The study showed how brief text messages may be examined to find underlying personality traits by utilizing machine learning methods. Two machine learning algorithms, ZeroR and Gaussian Processes, were trained to predict scores on each of the five personality traits using the Twitter profile data as a feature set.

Several machine-learning models were used in works by Smith et al. (2018) [3], Chen and Lee (2019) [4], and Miller et al. (2020) [5] to categorize Dark Triad qualities using language data. By combining Random Forest and Naive Bayes classifiers, authors were able to predict Machiavellianism, Narcissism, and Psychopathy with promising accuracy. Chen and Lee [4] used KNN and Support Vector Machine (SVM) classifiers to obtain acceptable classification accuracies for the Dark Triad attributes. Random Forest did well for Machiavellianism and Narcissism, whereas Neural Networks showed promise for Psychopathy, according to Miller et al.'s [5] comparison of several models.

Overall, these studies demonstrate how effectively natural language processing (NLP) techniques and machine learning models can be used to extract Dark Triad characteristics from textual data. The classifier employed may vary depending on the specific trait being assessed, highlighting the need to consider multiple models for full trait identification. Researchers may get a lot of insight into people's personality traits and potentially develop practical applications for screening and intervention strategies by using these methodologies.

Some previous studies focused solely on one Dark Triad trait, manually labeled textual data, and relied on psychological and statistical methods through surveys. Our study improved upon this approach by incorporating both textual and numerical data collected from the same individuals, enhancing accuracy as the labels from numerical data have already been verified in [1]. Additionally, we applied various machine learning algorithms to validate the dataset. Most importantly, our study identified Dark Triad traits from textual data using the contextual word embedding process of the transformer-based model BERT, which is considered a state-of-the-art model in this field.

METHODOLOGY

This section of the paper presents the methodology of the experiment conducted, the data collection process, including the sources of text data, and the survey design used to gather ground truth labels for Dark Triad traits. Figure 1 shows step by step research methodology diagram for this study.

Dataset Creation

This study opted to create a unique dataset as an experiment. Both textual and numerical data are included in the collection. The objective of this study was to create an effective model by combining data in textual and numerical forms.

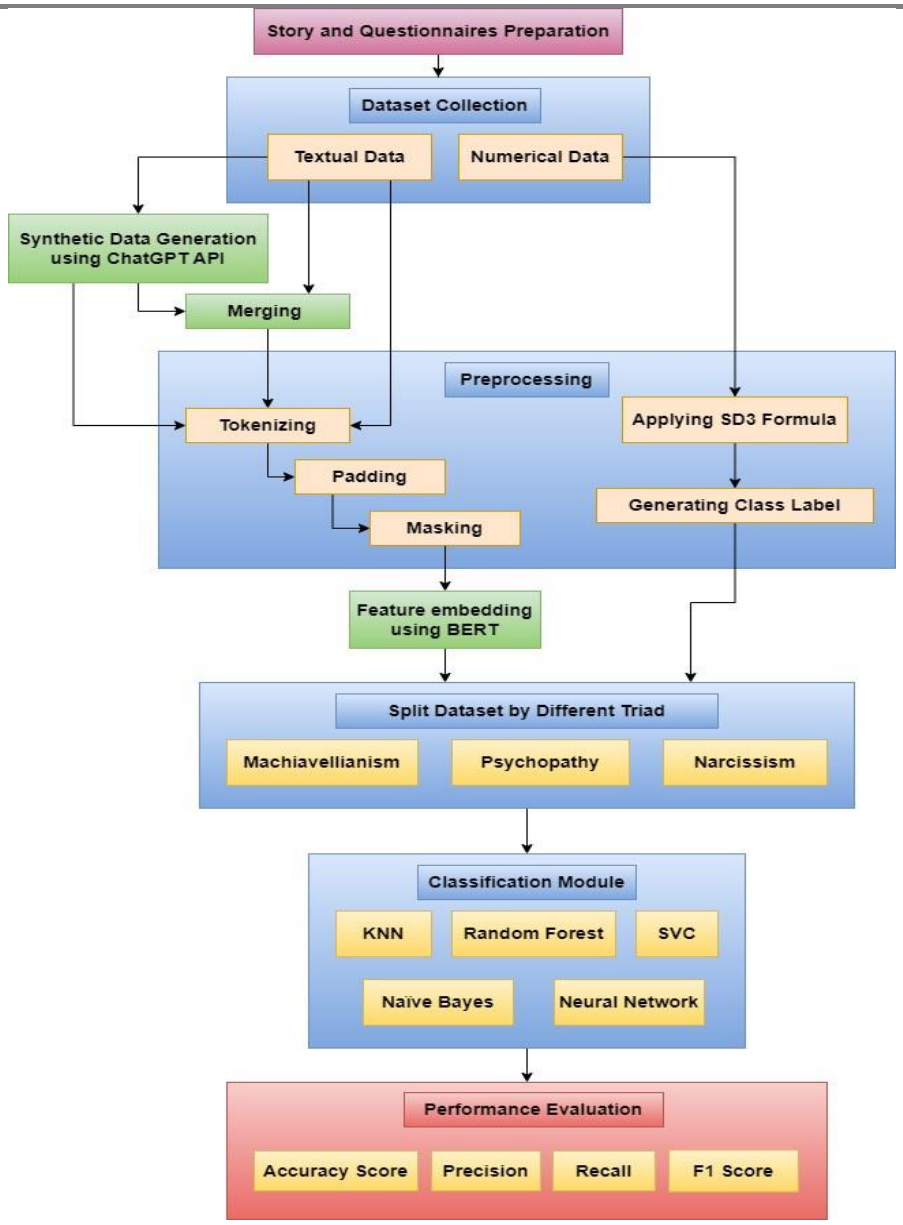

Figure 1: Methodology Diagram

Textual Data: We developed a story and three questions to gather textual data from users. The story revolves around three individuals; each embodying distinct personalities relevant to the study.

Bob, a talented software developer, faced challenges at work where his colleague Alice, despite being supportive, often took credit for his contributions. Alice, who grew up in a small village and overcame insecurities, utilized her strong communication skills to climb the corporate ladder, eventually becoming CEO. However, her success left her friendless and unhappy. Bob's personal life was also tumultuous; he dated Jenny, a calm but controlling individual. Their relationship ended due to Jenny's increasingly erratic behavior, including stalking and harassment, leading to her arrest. This left Bob to reflect on both his professional struggles and the tumultuous end of his relationship.

We formulated questions for these characters based on their personalities, and to validate these, we asked the respondents what actions they would take if placed in the characters' situations.

To understand the context, we needed a large amount of text data. We gathered this data from surveys and the responses of the respondents. We also used the Chat-GPT model to generate synthetic data, which helped to increase the size of our datasets.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume IX Issue VI June 2024

Numeric Data: To establish the class labels, we integrated the 27 statements from the SD3 [1] questionnaire into our survey form, allocating 9 statements to each dimension of the dark triad. Respondents were asked to rate their agreement level with these statements on a scale of 1 to 5.

Table 1: Examples of SD3 questionnaire

M1	It's not wise to tell your secrets.	N1	People see me as a natural leader.	P1	I like to get revenge on the authorities.
M2	I like to use clever manipulation to get my way.	N2	I hate being the center of attention.	P2	I avoid dangerous situations.

Some sample responses to the statements from three users, taken via Google Forms[13], are presented in Table 1 and Table 2.

Table 2: Sample responses to SD3 27 questions

	User1	**User2**	**User3**		**User1**	**User2**	**User3**		**User1**	**User2**	**User3**
M1	1	4	5	**N1**	4	4	2	**P1**	3	2	5
M2	4	2	4	**N2**	1	5	5	**P2**	1	4	4
M3	5	4	2	**N3**	3	3	1	**P3**	3	1	5
M4	5	5	4	**N4**	3	3	2	**P4**	3	2	2
M5	3	1	3	**N5**	5	4	1	**P5**	3	4	4
M6	5	1	5	**N6**	1	5	1	**P6**	3	3	4
M7	5	4	4	**N7**	1	4	1	**P7**	5	5	5
M8	5	3	3	**N8**	1	4	5	**P8**	5	1	3
M9	5	4	3	**N9**	5	4	4	**P9**	1	1	4

Table 3: Sample textual responses and mean of each triad responses

	User1		**User2**		**User3**	
	Textual	**Mean (M1-M9)**	**Textual**	**Mean (N1-N2)**	**Textual**	**Mean (P1-P9)**
Alice/ Machiavellianism	If I were Alice I would do exactly like Alice I'll do my work with all the hard work and keep appreciating others. I'll be calm and more	4.2	Alice is a self-centered person, only focused on her goal and she can do anything to get what she wants. Even though she has good communication skills she shouldn't take someone else's credit for her own benefit. I wouldn't	3.11	Alice was wrong because Alice often took credit for his work though she was a great supportive colleague which Bob Believed. She shouldn't do that.	3.67

	composed. I'll be happier with what I got.		do that. She shouldn't shove everyone away for her goal.			
Bob/ Anti-Narcissism	If I were Bob I wouldn't do what Bob did. I'll appreciate others whenever it is their time. This is the behavior of a selfish person which I never encourage.	2.67	Bob is a skilled and hardworking person with a lack of communication skills. He also can't handle the reality and failure and blame others. He needs to work on his lacking and focus on improving himself. I would work on improving my lacking and won't let anyone take credit for my hard work.	4	If I were in Bob's situation, I would communicate with my colleagues and superiors about the issue and document my work to prove my contributions. Bob made a mistake when he decided to end their relationship.	2.44
Jenny/ Psychopathy	If I were Jenny I would leave Bob in his own space and concentrate on my life rather than doing all this stupid stuff. Life should be amazing with good deeds.	3	Jenny is an obsessed and controlling type of person. Even though someone is loving her and give her importance she shouldn't try to control their lives. I won't let these things happen to me or anyone cause these are very annoying things. She should take therapy and work on her personality.	2.56	Jenny might seem irrational, but it's normal to expect equal effort and respect in return for what you give. Bob's lack of response made her feel insulted, prompting her to resolve the issue immediately out of frustration. If I were in her position, I would feel the same because mutual effort in relationships are essential.	4

From Table 3 we can observe user 1, the information obtained from the text indicates that Alice, who embodies a Machiavellian trait, is generally supported by this user. However, this reaction seems to contradict Bob and Jenny. In this case, the user agreed more with the SD3 questions related to Machiavellianism, and the mean score for all the nine questions (M1-M9) is 4.2 which is higher than the rest of the mean values. This response indicates that the user's personality leans toward Machiavellianism.

Going on to test the scenario for user 2, the user believes opposes Bob. Since Bob is a symbol of anti-narcissism, the final text in response 2 shows a significant disagreement with Bob, which may indicate that the user may be narcissistic. We can verify this further by looking up this user's responses to the SD3 questions about narcissistic traits. In this case, the respondent largely agrees, and the SD3 questioned (N1-N9) mean value is 4 suggesting that the user may be associated with narcissistic characteristics.

For user 3's test case, the respondent agrees with Jenny who represents Psychopathy. The user's SD3 question tends to agree more with the Psychopathy side of the questions. Additionally, the mean score for Psychopathy's SD3 questions (P1-P9) is higher than the other two dark triads' mean value which gives a clear hint that this user has a Psychopathy characteristic.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume IX Issue VI June 2024

In conclusion, this research [1] can be used to validate the class levels utilized in this study. This study bridges the gap between human psychology and the predictions of machine learning models. The responses to the DS3 and textual questions determined the personality types of each user based on human psychology, and machine learning confirmed these results. This method also demonstrated the accuracy of our models.

Dataset Pre-processing

Textual Data: We comprised three sets of textual answers from survey questions. These sets were combined to form a consolidated survey dataset containing 120 actual data entries. We used OPENAI to generate synthetic data. This approach helped us to increase the size of the dataset. We combined our dataset with model generated dataset to create a larger dataset for this study.

Figure 2 and Table 4 show the distribution of actual data, synthetic data, and merged data for three personality traits: Machiavellianism, psychopathy, and narcissism. For each trait, the table shows the number of data points in each category. We collected 28 data points in the Machiavellianism category for actual data but we generated more data by creating 60 synthetic datasets and finally, we merged the datasets for the Machiavellianism category which gave us a total of 88 datasets. Similarly, For Psychopathy, we were able to collect 39 datasets but we increased the dataset by 81 through the synthetic data generation process. For Narcissism, we collected 50 datasets and improvised the data again which gave us a total of 151 text results. Finally, we were able to generate enough data to study the personality traits of Machiavellianism, psychopathy, and narcissism. This data can be used to develop new insights into these traits and their relationship to other variables. We generated different amounts of synthetic data to increase the size of the dataset. We did over-sampling, but the amounts varied because of the quality of some paraphrases was ranging good and poor.

Table 4: Data Distribution

-	Actual Data	Synthetic Data	Merged Data
Machiavellianism	28	60	88
Psychopathy	39	81	120
Narcissism	50	101	151

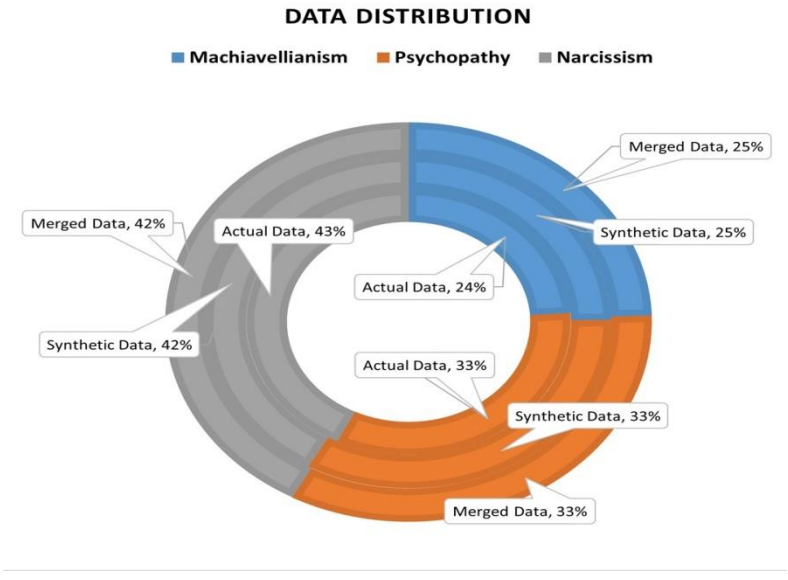

Figure 2: Data Distribution

Next, we took these three types of datasets (real, synthetic, and combined) to get them ready for subsequent processing. There were numerous crucial phases in the processing of the dataset.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume IX Issue VI June 2024

1. **Tokenization:** Textual data from each dataset was tokenized into smaller units called tokens. This process involved breaking down the text into meaningful segments, which could be words or sub-words. Tokenization is a crucial step for text processing.

2. **Padding:** To standardize input sizes for the subsequent model, padding was applied. This involved adding special tokens (e.g., [PAD]) to texts with fewer tokens so that they matched the length of the longest text in the dataset.

3. **Masking:** To enhance the model's understanding of context, masking was utilized. Certain tokens within the text were replaced with a special [MASK] token. This technique helps the model predict the original tokens and learn contextual relationships.

4. **Pretrained BERT Model:** The pre-trained BERT model was employed for its ability to capture contextual information from text. Tokenized, padded, and masked data were inserted into the model to obtain contextualized embedding, which represents the nuanced meaning of the text.

After these preprocessing steps, numerical features were extracted using the pre-trained BERT model. These contextualized embeddings captured semantic information from the text and were employed as inputs for downstream tasks.

Numerical Data: The processes involved in this study's numerical data preparation include applying the SD3 formula to get the mean values for each trio of responses, creating class labels based on thresholds, and consulting the SD3 formula's source. Here is a succinct justification: for processing the numerical data, the SD3 formula was applied, as outlined in [1]. With the use of this formula, the dark triad qualities of Machiavellianism, Narcissism, and Psychopathy are represented by each set of nine responses.

The mean values of the triad answers were compared to predetermined threshold levels to produce class classifications. The user was considered to have characteristics related to the dark triad feature Machiavellianism if the mean of responses to the M1–M9 statements was more than 3.86. Similar to this, 3.68 and 3.40 were employed as the thresholds [1] for narcissism and psychopathy, respectively.

The thresholds were defined based on the known literature and empirical data, and this method allowed for categorizing users into various dark triad qualities based on their mean answer values.

To prepare our data for machine learning analysis, we did several preprocessing. We created separate datasets for each of the three dark triad traits: Machiavellianism, Narcissism, and Psychopathy. Each dataset was carefully designed to capture the specific features of that trait. To do this, we assigned class labels to each element at a vital encoding stage. Machiavellianism, for example, was represented by class label 1, whereas users without that particular feature were represented by class label 0. Within each dataset, this deliberate labeling made it easier to distinguish between individuals who possessed the specified features and those who did not.

DESIGN AND IMPLEMENTATION

Train Test Split

To ensure that our machine learning models were accurate, we used an 80-20 train-test split. 80 percent of the data was used to train the models and the rest was used to test them.

The training set allowed the models to learn the complex patterns that characterize each dark triad attribute. The test set allowed us to evaluate the models' accuracy in classifying people into predefined characteristic groups. In other words, we used 80 percent of the data to teach the models how to identify the dark triad attributes, and then we used the remaining 20 percent of the data to see how well the models could identify those attributes in new data.

Applying Classifying Algorithm

With our well-structured datasets primed for analysis, we embarked on the next phase of our study—applying a diverse set of machine learning models to gain a comprehensive understanding of the distinct dark triad traits. Leveraging the feature vectors extracted from textual data as input and coupling them with the class labels derived from numerical data, we employed five prominent machine learning algorithms: Random Forest, K-Nearest Neighbors (KNN), Support Vector Classifier (SVC), Naive Bayes, and a Neural Network.

We carefully trained and adjusted each of these models using our special datasets, making sure they understood the unique parts of Machiavellianism, Narcissism, and Psychopathy. We used special numbers to help the models learn the differences between these traits. The feature vectors, which are special pieces of information from the text, helped the models learn and find the hidden patterns and connections in the words.

Our models were able to learn more about people's personalities and the distinctive qualities that make each unique thanks to the combination of information. It's similar to putting jigsaw pieces together to acquire a better understanding of how individuals think and what characteristics they have.

Performance Evaluation

To evaluate the effectiveness of our models in predicting the different dark triad traits, we used a performance evaluation metric called the accuracy score. The accuracy score is a measure of how well the models were able to correctly classify users into their respective Machiavellianism, Narcissism, and Psychopathy categories.

Specifically, the accuracy score is calculated by dividing the number of correctly classified users by the total number of users. A higher accuracy score indicates that the models are more effective in predicting the dark triad traits.

We acquired a total of 15 accuracy ratings after applying our trained models to the datasets—5 models were evaluated across the 3 different triad features. Each accuracy score indicated the proportion of times the models correctly predicted the outcome given the input data. These findings provided insightful information about each model's capabilities and ability to identify users with various psychological dispositions.

Also, along with the accuracy score we have used precision, recall, and F1 score to measure the performance of different five machine learning models on three different triads. Precision is a vital metric in machine learning, focusing on the accuracy of positive predictions. It quantifies the percentage of true positive predictions among all positive predictions, including both true positives and false positives. Recall, also known as sensitivity, assesses the model's ability to correctly identify all relevant instances in the dataset. It measures the percentage of true positive predictions among all actual positive instances, including both true positives and false negatives. Lastly, the F1 score represents a balanced metric that combines precision and recall by calculating their harmonic mean. It provides a compromise between precision and recall, accounting for false positives and false negatives.

With the accuracy score, precision, recall, and F1 score, we can evaluate classification models as they provide a comprehensive understanding of a model's performance. Accuracy offers an overall measure, while precision, recall, and the F1 score offer insights into the model's ability to make correct positive predictions, crucial when differentiating between real-world consequences of false positives and false negatives.

RESULT AND DISCUSSION

Classification Results

This section (Figure 3 and Table 5) presents the accuracy result analysis details on identifying human dark triad traits, using machine learning models on text data. The classification results showed good accuracy in revealing individuals with these traits based on language patterns.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume IX Issue VI June 2024

Table 5: Performance Evaluation (Accuracy)

		Random Forest	KNN	SVC	Naïve Bayes	Neural Network
Actual Data	Machiavellianism	0.83	0.83	0.84	0.74	0.83
	Psychopathy	0.58	0.67	0.54	0.54	0.58
	Narcissism	0.63	0.63	0.67	0.54	0.71
Synthetic Data	Machiavellianism	0.76	0.76	0.78	0.76	0.73
	Psychopathy	0.76	0.67	0.80	0.47	0.76
	Narcissism	0.53	0.49	0.59	0.57	0.57
Merged Data	Machiavellianism	0.76	0.79	0.83	0.74	0.82
	Psychopathy	0.68	0.65	0.59	0.60	0.60
	Narcissism	0.57	0.56	0.57	0.60	0.60

Machiavellianism

The models were able to accurately detect Machiavellianism in all datasets. The Linear Support Vector Classifier (SVC) had the highest testing accuracy of 84 percent, followed by Random Forest, KNN, and Neural Network at 65-83 percent accuracy. These results suggest that the models were able to distinguish individuals with Machiavellian tendencies from their text data responses.

Narcissism

The models exhibited varying levels of accuracy in the identification of Narcissism. The Neural Network had the highest testing accuracy of 71 percent, followed by Linear SVC (57-67 percent), Random Forest, and Naiıve Bayes (53-67 percent). KNN had the lowest accuracy of 49-63 percent. These results suggest that the Neural Network and Linear SVC were particularly effective in identifying individuals with Narcissistic traits based on text data, while KNN performed less optimally.

Psychopathy

The KNN and SVC model was the most effective in detecting Psychopathy, with a testing accuracy of 67-80 percent. The Neural Network had an accuracy of 58-73 percent, followed by Random Forest (58-76 percent), Linear SVC (54-80 percent), and Naive Bayes (47-60 percent). These results suggest that the KNN model was the constant best at distinguishing individuals with Psychopathic traits from text data.

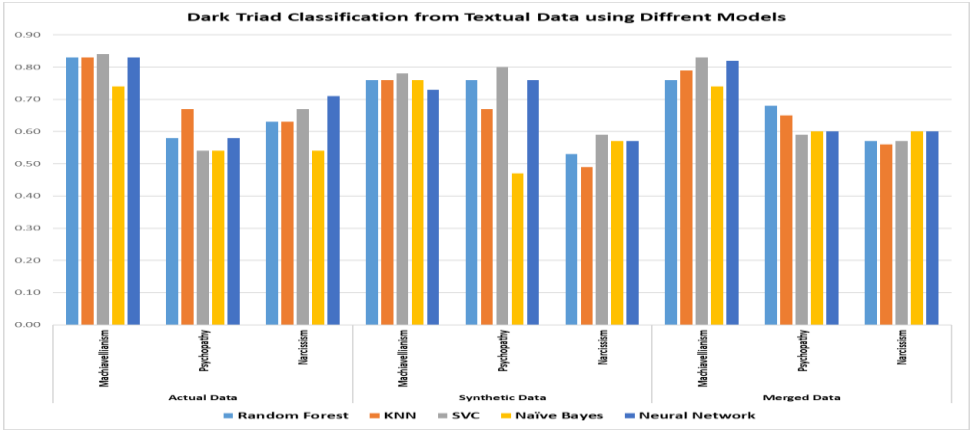

Figure 3: Performance Evaluation (Accuracy Score)

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume IX Issue VI June 2024

Comparison of Models

The Linear Support Vector Classifier (Linear SVC) performed consistently well across all three Dark Triad traits, achieving high accuracy in all cases. The Random Forest and KNN models performed well in identifying Machiavellianism and Psychopathy. The Neural Network scored the highest accuracy in detecting Narcissism, also performing well on the other traits.

Table 6 presents the precision, recall, and F1 scores for machine learning algorithms applied in this study. Precision represents the positive predictions, recall measures the true positive rate, and the F1 score is a ratio metric for precision and recall.

Table 6: Performance Evaluation (Precision, Recall, F1 Score)

		Not Machiavellianism	**Machiavellianism**	**Not Psychopathy**	**Psychopathy**	**Not Narcissism**	**Narcissism**
Random Forest	**Precision**	0.87	1.00	0.61	0.50	0.65	0.57
Random Forest	**Recall**	1.00	0.25	0.79	0.30	0.79	0.40
Random Forest	**Fl-Score**	0.93	0.40	0.69	0.37	0.71	0.47
KNN	**Precision**	0.90	0.50	0.65	0.75	0.71	0.60
KNN	**Recall**	0.90	0.50	0.93	0.30	0.71	0.60
KNN	**F1-Score**	0.90	0.50	0.76	0.43	0.71	0.60
SVC	**Precision**	0.90	0.50	0.62	0.50	0.70	0.50
SVC	**Recall**	0.90	0.50	0.71	0.40	0.50	0.70
SVC	**F1-Score**	0.90	0.50	0.67	0.44	0.58	0.58
Naïve Bayes	**Precision**	0.86	0.33	0.60	0.44	0.71	0.60
Naïve Bayes	**Recall**	0.90	0.25	0.64	0.40	0.71	0.60
Naïve Bayes	**F1-Score**	0.88	0.29	0.62	0.42	0.71	0.60
Neural Network	**Precision**	0.83	0.44	0.29	0.34	0.58	0.33
Neural Network	**Recall**	1.00	0.25	0.50	0.58	1.00	0.25
Neural Network	**F1-Score**	0.91	0.32	0.37	0.43	0.74	0.29

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume IX Issue VI June 2024

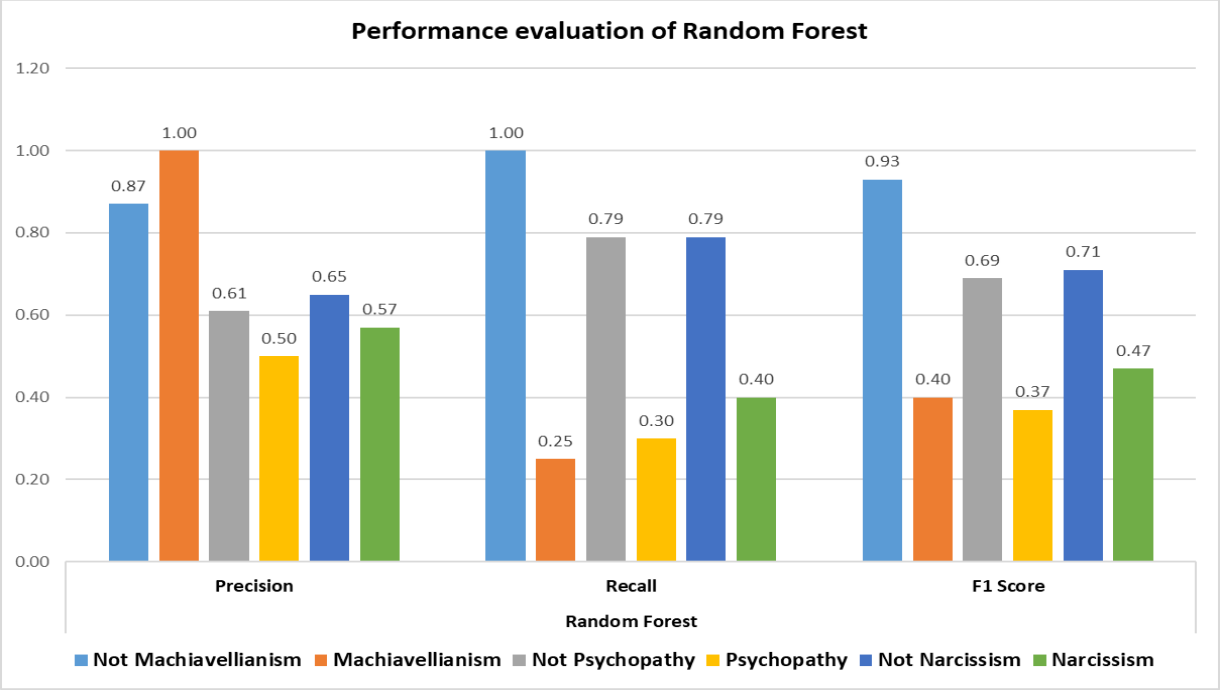

Figure 4: Performance Evaluation of Random Forest (Precision, Recall, F1 Score)

From Figure 4 we can see for Machiavellianism, Random Forest achieved high precision (1.00) but lower recall (0.25), resulting in an F1 Score of 0.40. It had modest performance for Psychopathy (precision: 0.50, recall: 0.30, F1 Score: 0.37) and Narcissism (precision: 0.57, recall: 0.40, F1 Score: 0.47).

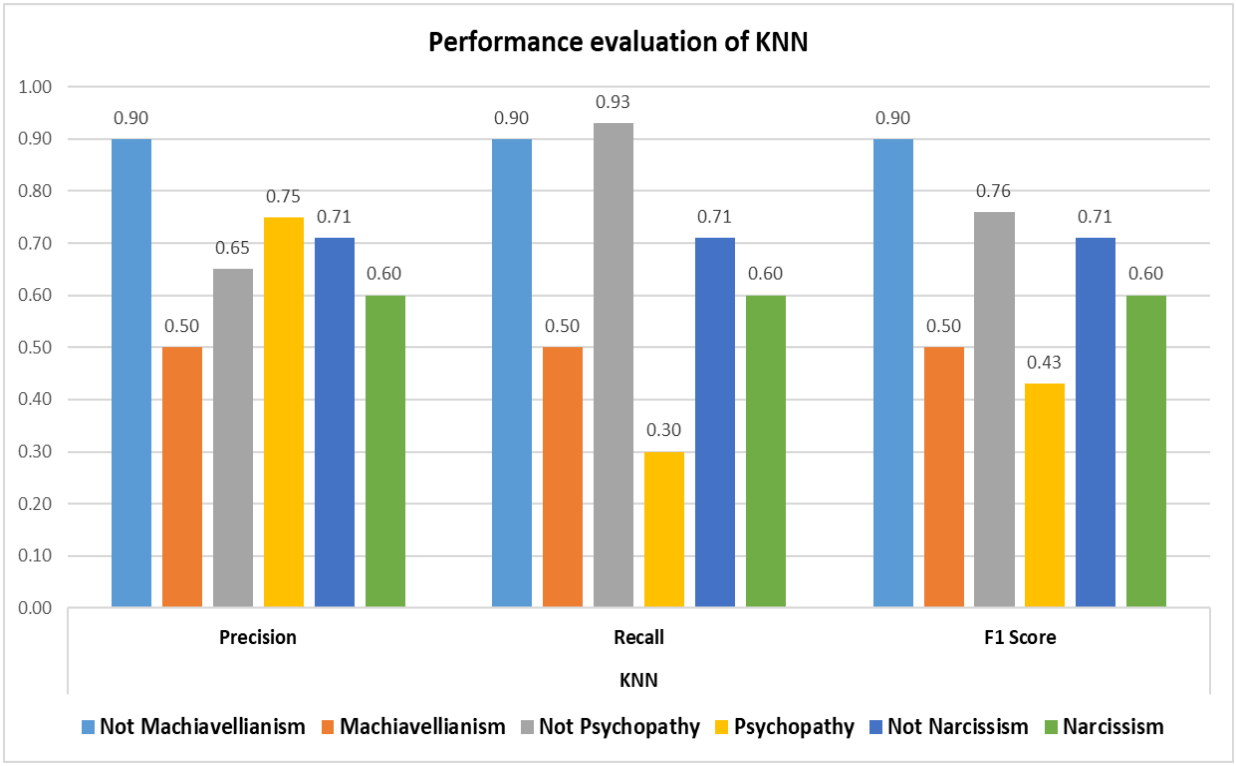

Figure 5: Performance Evaluation of KNN (Precision, Recall, F1 Score)

KNN performed well for detecting Psychopathy (precision: 0.75, recall: 0.30, F1 Score: 0.43) and showed balanced performance for Machiavellianism (precision: 0.50, recall: 0.50, F1 Score: 0.50) and Narcissism (precision: 0.60, recall: 0.60, F1 Score: 0.60), which we can see in Figure 5.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume IX Issue VI June 2024

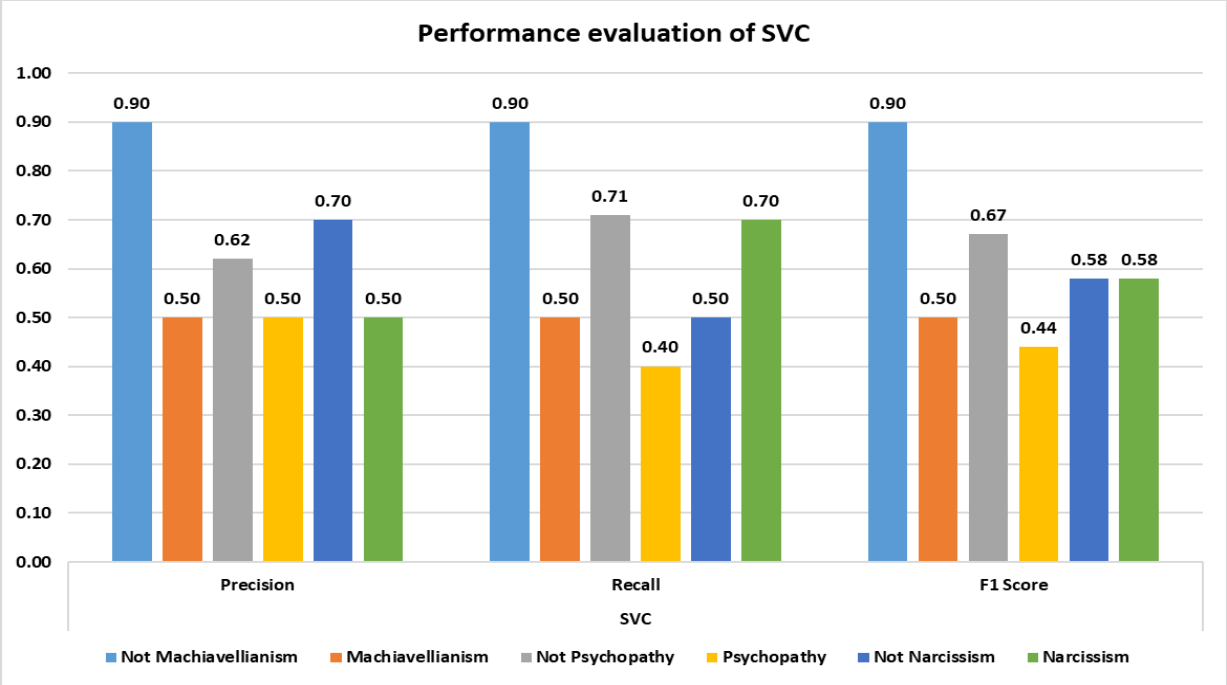

Figure 6: Performance Evaluation of SVC (Precision, Recall, F1 Score)

Also, Figure 6 shows SVC delivered a balanced precision and recall for Machiavellianism (0.50) resulting in an F1 Score of 0.50. It had a high precision and lower recall pattern for Psychopathy (precision: 0.50, recall: 0.40, F1 Score: 0.44) and Narcissism generated the opposite with a good recall value (precision: 0.50, recall: 0.70, F1 Score: 0.58).

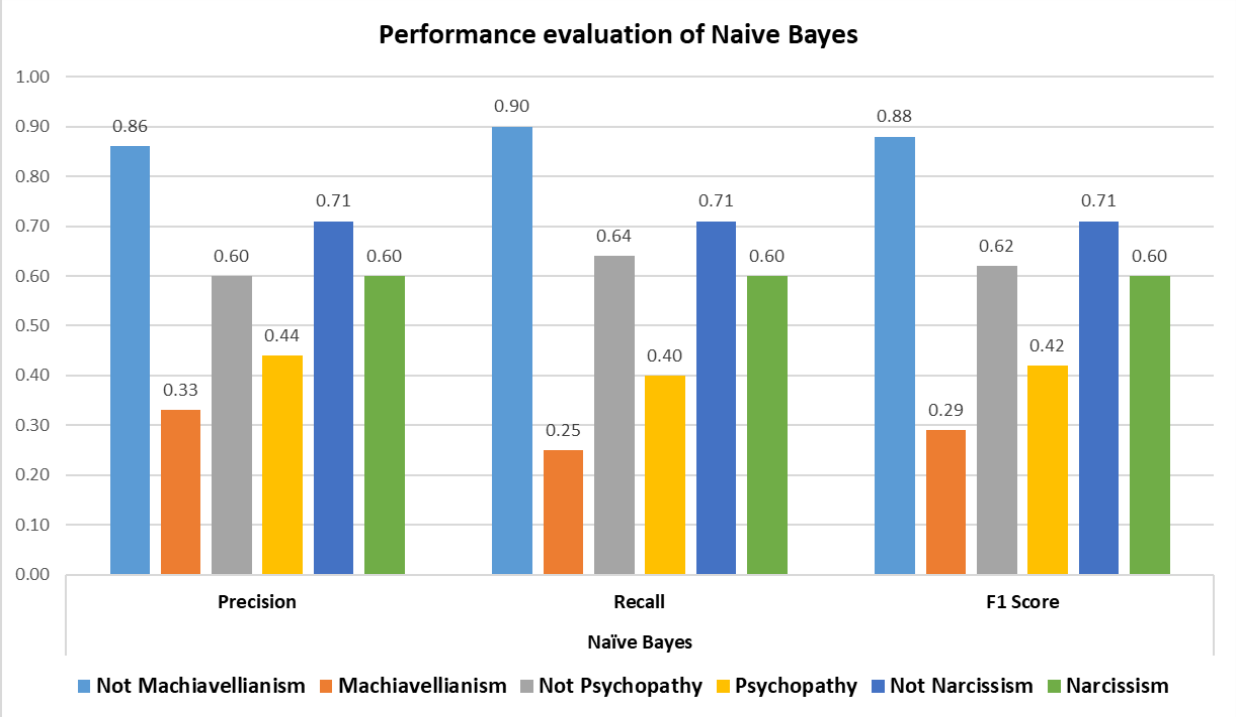

Figure 7: Performance Evaluation of Naive Bayes (Precision, Recall, F1 Score)

In figure 7, Naive Bayes showed balanced precision and recall for Machiavellianism (precision: 0.33, recall: 0.25, F1 Score: 0.29) and Psychopathy (precision: 0.44, recall: 0.40, F1 Score: 0.42), while it performed well for Narcissism (precision: 0.60, recall: 0.60, F1 Score: 0.60).

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume IX Issue VI June 2024

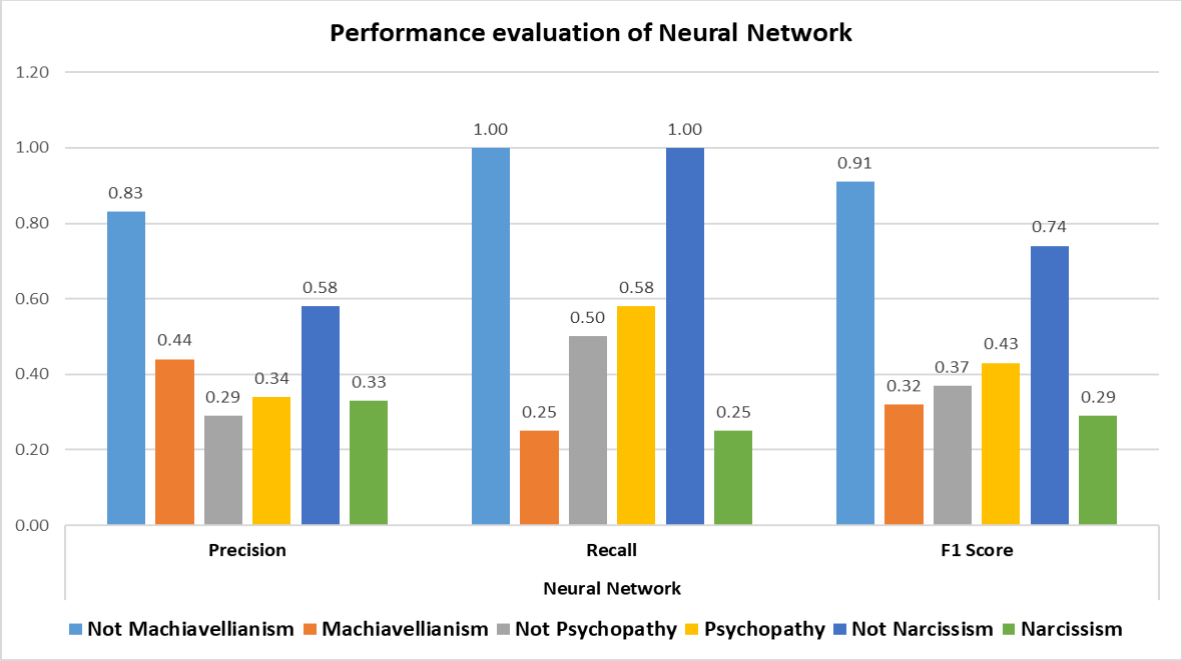

Figure 8: Performance Evaluation of Neural Network (Precision, Recall, F1 Score)

The Neural Network model achieved high recall for Psychopathy (0.58) but lower precision (0.34), resulting in an F1 Score of 0.43. It performed moderately for Machiavellianism (precision: 0.44, recall: 0.25, F1 Score: 0.32) and had a lower performance for Narcissism (precision: 0.33, recall:

0.25, F1 Score: 0.29), which is shown in Figure 8.

Overall Classification results showed good accuracy in identifying human dark triad traits based on language patterns. The results of our study suggest that machine learning models, combined with natural language processing techniques, can effectively identify individuals with Dark Triad traits from textual data. The varying accuracy scores across different models and traits highlight the importance of model selection for specific trait identification.

- Linear Support Vector Classifier (Linear SVC) consistently performed well across all three Dark Triad traits.
- Random Forest and KNN models performed well in identifying Machiavellianism and Psychopathy.
- Neural Network had the highest accuracy in detecting Narcissism and also performed well in identifying Psychopathy.
- KNN and SVC were the most effective models for detecting Psychopathy, while Naive Bayes performed well for Narcissism.
- Precision, recall, and F1 scores varied for different models and traits, with some models performing better in specific areas for certain traits.
- Linear Support Vector Classifier (Linear SVC) tends to provide balanced precision and recall for all three Dark Triad traits.
- Random Forest demonstrates high precision, particularly for Machiavellianism.
- KNN shows balanced precision and recall for Psychopathy and Machiavellianism.

The algorithms' performance varied across different personality traits, and no single algorithm consistently outperformed the others for all traits. The choice of algorithm may depend on the specific trait of interest and the balance between precision and recall required for a particular application. However, the comparatively consistent performance of Linear SVC indicates its suitability for identifying individuals with Dark Triad traits, reinforcing its potential as a reliable classifier for personality assessment. These findings are consistent with previous research efforts that have used machine learning and NLP approaches to uncover underlying personality traits from text data.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume IX Issue VI June 2024

CONCLUSION

This study presented the potential of machine learning models and natural language processing (NLP) to identify human dark triad traits (Machiavellianism, narcissism, and psychopathy) from text data. The results analysis showed that machine learning models can be used to identify these traits with promising accuracy. This study has shown a better accuracy in identifying Machiavellianism rather than Narcissism and Psychopathy using all. The state of art machine learning models used in this research. However, there is still room for improvement, and future research should focus on improving the feature extraction techniques and expanding the dataset.

The findings of this study have possibilities for a variety of fields, including psychology, criminology, and human resources. In psychology, the ability to predict these complex characteristics known as dark triad traits could be used to develop interventions for individuals experiencing specific needs or challenges. This method can prevent people from engaging in criminal activities. In criminology, the ability to identify dark triad traits could be used to assess the risk of relapse or to identify potential criminals. In human resources, recognizing dark triad traits could be used to screen job applicants or to identify employees who are at risk of engaging in workplace violence. This study also shows a possibility of preventing Scholl or institutional violence among students as machine learning can now detect potential dark personalities.

This study is a first step towards understanding the relationship between language and personality. As technology improves and our understanding of human behavior deepens, machine learning and NLP can be used to make discoveries, develop treatments, and improve our understanding of the human psyche. A future study on ethical guidelines is required as it raises concerns regarding data privacy, informed consent, and responsible use of machine learning models. A multimodal approach will be adopted to understand personality traits, datasets will be expanded for wider applicability, and model accountability will be ensured for ethical AI practices in the future. Future directions include refining feature extraction with advanced models like BERT and GPT. Hopefully, this research will pave the way for better understanding and effective interventions for Dark Triad traits.

REFERENCES

1. D. N. Jones and D. L. Paulhus, “Introducing the short dark triad (SD3) a brief measure of dark personality traits,” Assessment, vol. 21, pp. 28–41, 2014.
2. S. Mahmud, M. Rana, F. R. Zahir and M. R. Huq, “Detection of antisocial personality based on social media data,” in ICT Analysis and Applications: Proceedings of ICT4SD 2020, Volume 2, Springer, 2021, pp. 651–659.
3. L. N. Smith, K. D. Williams and B. Cyberpsychology, “Dark Personality Traits and Twitter: A Macro-Social Psychological Perspective,” Social Psychological and Personality Science, vol. 9, pp. 758–768, 2018.
4. C. Chen and K. Lee, “Exploring the Dark Triad on Social Media: An Analysis of Online Forum Discussions,” Journal of Research in Personality, vol. 80, pp. 84–93, 2019.
5. J. D. Miller, M. L. Crowe, B. Weiss, J. L. Maples-Keller and D. R. Lynam, “Predicting the Dark Triad of Personality from Twitter Usage and a Linguistic Analysis of Tweets,” Journal of Research in Personality, vol. 87, p. 103999, 2020.
6. S. L. Lee, Y. E. Tan, C. L. Tam and J. Ahn, “The Facilitative Effect of Impulsiveness on the Dark Triad and Social Network Sites Addiction: The Dark Triad, Impulsiveness, SNS Addiction,” International Journal of Technology and Human Interaction (IJTHI), vol. 18, pp. 1–15, 2022.
7. K. Srirejeki, P. Kiky, I. Agung and S. Bambang, “Understanding Academic Fraud: The Role of Dark Triad Personality and Situational Factor,” Journal of Criminal Justice Education, vol. 34, no. Taylor & Francis, pp. 147–168, 2023.
8. H. Ahmad, A. Arif, A. M. Khattak, A. Habib, M. Z. Asghar and B. Shah, “Applying deep neural networks for predicting dark triad personality trait of online users,” in 2020 International Conference on Information Networking (ICOIN), IEEE, 2020, pp. 102–105.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume IX Issue VI June 2024

9. D. Preotiuc-Pietro, J. Carpenter, S. Giorgi and L. Ungar, "Studying the Dark Triad of personality through Twitter behavior," in Proceedings of the 25th ACM international on conference on information and knowledge management, 2016, pp. 761–770.
10. M. Kosinski, D. Stillwell and T. Graepel, "Private traits and attributes are predictable from digital records of human behavior," Proceedings of the national academy of sciences, vol. 110, pp. 5802–5805, 2013.
11. C. Sumner, A. Byers, R. Boochever and G. J. Park, "Predicting Dark Triad Personality Traits from Twitter Usage and a Linguistic Analysis of Tweets," in 2012 11th International Conference on Machine Learning and Applications, 2012, pp. 386-393.
12. J. Golbeck, C. Robles, M. Edmondson and K. Turner, "Predicting personality from twitter," in 2011 IEEE third international conference on privacy, security, risk and trust and 2011 IEEE third international conference on social computing, IEEE, 2011, pp. 149–15
13. https://docs.google.com/forms/d/1C_gnwk6zIK73U2pzUk0yO94uUcN9fGfPXOHaAQrx9Q4/viewform?edit_requested=true