Journal of Research in Personality 120 (2026) 104690

Contents lists available at ScienceDirect

Journal of Research in Personality

journal homepage: www.elsevier.com/locate/jrp

Full Length Article

Comparing machine learning methods for predicting dark triad personality traits using social media text data

Maxim Leberecht [a,*], Andre Nedderhoff [a], Steffen Zitzmann [b], Martin Hecht [a]

[a] Helmut Schmidt University, University of the Federal Armed Forces Hamburg, Germany
[b] Medical School Hamburg, Germany

ARTICLE INFO

Keywords:
Dark triad
Personality prediction
Social media text data
Machine learning
Regression models

ABSTRACT

The Dark Triad (DT) personality traits, characterized by manipulativeness, callousness, and egocentrism, are linked to both negative outcomes such as aggression and delinquency, as well as positive outcomes like career success. This study aims to compare different machine learning models for predicting DT traits – Narcissism, Machiavellianism, and Psychopathy – using social media text data from Facebook status updates and personality questionnaires. Various machine learning models were evaluated. Across traits, Random Forest achieved the lowest RMSE, outperforming most other models, followed by Support Vector Machines and Gaussian Processes. Bias was similar across all models. These findings highlight the potential of social media data to offer insights into users' personalities and carry methodological implications for future research on personality assessments.

1. Introduction

Personality is a complex construct that has been linked to many important outcomes such as academic performance (Poropat, 2009), job performance (Barrick & Mount, 1991), relationship satisfaction (Claxton et al., 2012), and well-being and health (H. S. Friedman & Kern, 2014). Given its broad associations, accurately measuring personality is essential for understanding and predicting behavior.

Traditional methods of measuring personality, such as questionnaires or interviews, are often time-consuming and require trained personnel for administration and interpretation. Consequently, there has been a growing interest in using social media text data to predict personality traits. These data are abundant, easily accessible, and offer a low-cost alternative for gaining insights into users' personalities. The present study aims to compare various machine learning methods for predicting the Dark Triad (DT) personality traits using social media text data. Before delving into this, we will first provide a brief overview of personality traits, the specific traits of the Dark Triad, general research on predicting traits from language, and previous studies focused on predicting Dark Triad traits.

1.1. Trait theory

Although there are various approaches to defining personality, one of the most widely accepted approaches is trait theory (Costa & McCrae, 1999). Trait theory suggests that personality can be described in terms of a set of stable and enduring traits (McCrae & Costa, 2008). Traits are defined as consistent patterns of thoughts, feelings, behaviors, and motivations that are relatively stable over time and across situations (Costa & McCrae, 1992). While the conceptualization of personality remains a subject of debate in psychology (see e.g., McAdams & Pals, 2006; McCrae & Costa, 2008; DeYoung, 2015), one model has gained widespread acceptance: The Five-Factor Model (FFM), also known as the Big Five (McCrae & Costa, 2008). The FFM encompasses five broad dimensions: Openness to Experience, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. These dimensions are intended to capture the most important aspects of personality. The five factors were derived primarily through lexical analysis and analysis of personality questionnaire data (McCrae & John, 1992). Although there has been some debate regarding the number of factors needed to describe personality and the labels for these factors, the FFM has consistently been shown to be a robust and valid model across different cultures and languages (McCrae & Costa, 2008). However, the FFM was not designed to be an all-encompassing model of personality but rather to capture the most important aspects. This limitation has prompted the development of other models that focus on more specific facets of personality.

* Corresponding author.
E-mail address: maxim.leberecht@hsu-hh.de (M. Leberecht).

https://doi.org/10.1016/j.jrp.2025.104690
Received 27 August 2025; Received in revised form 16 December 2025; Accepted 17 December 2025
Available online 24 December 2025

1.2. Dark triad personality traits

The Dark Triad (DT) personality trait model focuses on the negative side of personality, encompassing three distinct traits: Narcissism, Machiavellianism, and Psychopathy. While there is some overlap between these traits, they are generally considered separate constructs (Paulhus & Williams, 2002): Narcissism is mainly characterized by grandiosity, entitlement, and a lack of empathy; Machiavellianism by manipulativeness, deceitfulness, and a focus on self-interest; and Psychopathy by impulsivity, callousness, and a lack of empathy. These traits have been associated with various negative outcomes including aggression, delinquency, sociosexuality and poor interpersonal relationships (Paulhus & Williams, 2002; Jonason & Webster, 2010; Furnham et al., 2013). However, the DT has also been linked to some positive outcomes such as bravery, creativity, career success and leadership roles (Spurk et al., 2016; Paleczek et al., 2018; Kaufman et al., 2019).

Recent research has suggested expanding the DT model to include Sadism as a fourth trait, forming what is known as the Dark Tetrad (Buckels et al., 2013; Paulhus, 2014; Međedović & Petrović, 2015). Sadism is characterized by the enjoyment of inflicting pain, suffering, or humiliation on others. While the Dark Tetrad model is a more recent and more comprehensive model, the DT model remains widely used in research. Therefore, in this study, we focus on the DT traits due to the availability of data and the focus of previous research.

1.3. Predicting personality traits from language

Before predicting the dependent variables (targets), it is essential to extract features (independent variables) from the text data. There are two main approaches for this: closed-vocabulary and open-vocabulary approaches. Closed-vocabulary approaches rely on predefined dictionaries to extract features from the text data. These dictionaries contain words or phrases that are associated with specific psychological constructs. For example, the Linguistic Inquiry and Word Count (LIWC) (Pennebaker et al., 2022) is a dictionary that links words and phrases to psychological constructs such as affect and cognition. In this method, the text is tokenized into words, which are matched to the dictionary entries. The number of words associated with specific psychological constructs is then counted, and the proportion of each category within the text is calculated. These values serve as features for predicting personality traits through regression models. Closed-vocabulary approaches are straightforward to interpret and work well with small datasets (Eichstaedt et al., 2021).

In contrast, open-vocabulary approaches are data-driven and do not rely on predefined dictionaries. A common example is Latent Dirichlet Allocation (LDA), which groups words into topics based on their co-occurrence in the text data (Blei et al., 2003). Unlike closed-vocabulary approaches, the topics in open-vocabulary methods emerge from the data itself, allowing them to capture more subtle aspects of language (Eichstaedt et al., 2021). However, these methods require larger datasets and are more challenging to interpret (Eichstaedt et al., 2021).

Pennebaker and King (1999) demonstrated that the language people use can reveal insights into their personality, finding that the categories extracted with the LIWC dictionary correlated with the Big Five personality traits. Since the Big Five were derived from lexical analysis, this result was unsurprising.

Since then, there has been a growing interest in predicting personality traits from text data. Mairesse et al. (2007) used written essays to predict the Big Five traits of the writers, comparing regression models such as Linear Regression, M5′Regression Tree, M5′Rules, REPTree and a Support Vector Machine. Their results were mixed, with the Linear Regression model performing best for some traits and the M5′Regression Tree model excelling for others. Similarly, Golbeck, Robles, and Turner (2011) were among the first to use social media text data – specifically Facebook status updates – to predict personality traits using an M5′Rules model and a Gaussian Processes model. They also found mixed results, with the M5′Rules model performing best for some traits and the Gaussian Processes model performing best for others. A later study by Golbeck, Robles, Edmondson, and Turner (2011) analyzing Twitter data, which showed similar findings, with a Gaussian Processes model performing similarly to a ZeroR model.

Since these early studies, many others have used social media data to predict personality traits across various social media platforms and approaches (for an overview, see Vora et al., 2020). Commonly used regression models include M5′Rules, Gaussian Processes, Pace Regression, Support Vector Machines, and Ridge Regression.

1.4. ***Predicting*** dark triad personality traits

Most previous research has focused on predicting the Big Five personality traits, while studies on predicting the DT traits from social media data remain limited. Unlike the Big Five traits, the DT traits are not derived from lexical analysis, creating uncertainty about whether the same language-based prediction methods can be effectively applied.

One of the first studies to explore predicting DT traits from social media was conducted by Sumner et al. (2012). Using a dataset of tweets, they extracted features with the LIWC dictionary and built four different classification models: a Support Vector Machine, a Naive Bayes classifier, a J48 classifier, and a Random Forest to classify individuals as either above or below the median on each of the DT traits. The authors also hosted a competition using the same dataset, with the best performing model being a combination of multiple models. Garcia and Sikström (2014) predicted the DT traits using Facebook status updates with an unspecified regression model. Instead of a closed-vocabulary approach, they applied Latent Semantic Analysis (LSA), an open-vocabulary approach for feature extraction. Preotiuc-Pietro et al. (2016) took a different approach, using a Twitter dataset to predict the DT traits with a Linear Regression model with Elastic Net regularization. Their model incorporated text data, image data, and Twitter usage data, finding that text data was the most informative data source for predicting the DT traits.

1.5. Purpose and scope

Applied teams increasingly work with short, naturalistic text from everyday online behavior, where small samples, fixed lexicon features, and limited tuning are the norm. A controlled head-to-head comparison on the same features yields actionable insight. Practically, it shows which models are most plug-and-play to deploy and clarifies trade-offs among error, systematic bias, interpretability, and maintenance in real workflows (e.g., triaging large volumes of posts, prioritizing outreach, or routing cases for human review). Scientifically, it indicates when DT-language relations in online behavior are essentially linear versus meaningfully non-linear, and whether gains from flexible models are consistent across traits, information that refines theory about how these traits surface in everyday communication. Importantly, given the nature of DT traits, understanding and predicting patterns of online behavior associated with them can be useful: some online behavior can be harmful to others or to the community; being able to extend our conceptualization of DT traits to include their online verbal expressions helps researchers and practitioners anticipate risks, tailor interventions, and communicate uncertainty responsibly. The goal is to answer the following research question:

Which machine learning regression model provides the most accurate predictions of DT traits from social media text data?

To our knowledge, this is the first study to benchmark a set of widely used regression algorithms on continuous DT outcomes from Facebook text using a common feature set and evaluation protocol, enabling a controlled comparison of performance across traits. The study was exploratory and was not preregistered.

To answer the research question, we will evaluate the performance of different regression models in predicting the DT traits using social media text data. First, we will describe the study participants, the measures used to assess the DT traits, the methods for feature extraction from text data, and the regression models employed. We will then present the results of the regression analyses and compare the performance of the different models. Finally, we will discuss the implications of these findings for future research on predicting personality traits from social media data.

We emphasize that any applied use should be privacy-preserving, consent-based, and framed as probabilistic signal detection, a decision aid for human judgment, not a diagnostic label.

2. Methods

2.1. Dataset

This study uses the same dataset analyzed by Garcia and Sikström (2014), making this a secondary analysis of their dataset. The dataset comprises the 15 most recent Facebook status updates from each individual alongside three personality questionnaires.

2.2. Participants

Initially, 304 participants were recruited through Amazon's Mechanical Turk. Participants were excluded based on the following criteria: (1) being younger than 18 years old, (2) not speaking English, and (3) providing fewer than 100 words across all status updates. The age restriction was applied because the questionnaires were designed for adults. Speaking English was necessary as the feature extraction dictionaries were based on English language. The word count criterion ensured that the text data was sufficient for feature extraction.

After applying these exclusions, the final dataset consisted of 266 participants, with an average age of 26.6 years (SD = 7.41), ranging from 18 to 62 years. Of these, 103 participants were male and 163 participants were female.

2.3. Measures

The DT traits were measured using three different questionnaires: The short version of the Narcissistic Personality Inventory (NPI-16) for Narcissism (Ames et al., 2006), the Machiavellianism Scale (MACH-IV) for Machiavellianism (Christie & Geis, 1970), and the short version of the Eysenck Personality Questionnaire Revised (EPQR-S) for Psychopathy (Eysenck et al., 1985).

2.3.1. NPI-16

The NPI-16 is a 16-item forced-choice questionnaire measuring Narcissism unidimensionally, where participants choose between a narcissistic and non-narcissistic statement. Scores are determined by counting the number of narcissistic choices.

2.3.2. MACH-IV

The MACH-IV is a 20-item questionnaire that assesses Machiavellianism, with participants rating their agreement with statements on a 7-point Likert scale. Scores are derived by calculating the mean of the ratings.

2.3.3. EPQR-S

The EPQR-S is a three-dimensional questionnaire measuring Extraversion, Neuroticism and Psychoticism. It consists of 12 forced-choice items per trait, asking whether a statement describes the participant or not. In this study, only the Psychoticism scale (relevant to Psychopathy) was used, with scores based on the number of "yes" responses.

2.4. Extracted features

For the resulting sample, the mean word count was 437.29 words (SD = 289.71, median = 360.50), ranging from 103 to 2131 words. A closed-vocabulary approach was used for feature extraction, with tokenized text analyzed using the following dictionaries:

2.4.1. Linguistic Inquiry and Word Count 2022 (LIWC-22)

Using the LIWC-22 (Pennebaker et al., 2022) dictionary, 120 features were extracted, including standard counting metrics such as the total word count or words per sentence, as well as features related to psychological processes (e.g. affect and cognition) and features related to personal concerns (e.g. work, home, or leisure). A full list of features is available in the LIWC-22 manual (Boyd et al., 2022).

2.4.2. NRC Emotion Lexicon (EmoLex)

The NRC EmoLex (Mohammad & Turney, 2013) dictionary associates English words with eight basic emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust), as well as two sentiments (negative and positive). Using this dictionary, 10 features were extracted by counting matches in each participant's text.

2.4.3. AFINN-111

AFINN (Nielsen, 2011) is a dictionary that rates English words for valence on a scale from -5 (negative) to 5 (positive). Using the AFINN-111 dictionary, eleven features were extracted.

2.4.4. Mean segmental type-token ratio (MSTTR)

In addition to the dictionary-extracted features, the MSTTR was included as a feature. It measures lexical diversity by dividing the number of unique words by the total word count in a text, adjusted by averaging across multiple segments.

2.5. Feature space

All downstream models used the same combined feature set consisting of LIWC-22 categories, ten NRC EmoLex scores (eight emotions + two sentiments), eleven AFINN-111 polarity scores, MSTTR, age, and gender (total: 144 features).

2.6. Regression approaches

In regression analysis, the goal is to predict continuous outcomes based on predictor variables, while minimizing the error between the predicted and observed values. The following approaches were included to compare the performance of various regression approaches, which have been used in previous research to predict personality traits from social media data.

2.6.1. ***Gaussian** processes **(GP)***

GP models are non-parametric regression models that analyze the relationship between the predictor variables and the target variable by modeling it as a multivariate Gaussian distribution (Williams & Barber, 1998).

2.6.2. Generalized linear model with elastic net regularization (GLMNet)

GLMNet is a regression model that combines the Lasso and Ridge regression methods. It helps select the most important features and reduces overfitting by assigning weights to features and penalizing both highly correlated features and those with weak correlations to the target variables (J. H. Friedman et al., 2010).

2.6.3. K-Nearest-Neighbor regression (kNN)

kNN regression is a non-parametric regression model that predicts the target variable by averaging the values of the k closest data points (Aha et al., 1991).

2.6.4. M5′Rules (M5R)

M5R is a regression model that constructs a decision tree and then converts it into a set of rules, which are then used to predict the target variable (Holmes et al., 1999).

2.6.5. Random Forest (RF)

RF is a method that constructs multiple decision trees and then averages their predictions to produce a final result (Breiman, 2001). This type of approach, which combines the predictions from multiple models, is an ensemble method.

2.6.6. Support Vector Machine (SVM)

SVR minimizes ε-insensitive loss around a flat tube and can capture non-linear relations via kernel mappings (e.g., RBF). In our implementation we used the default settings provided by mlr3 for the SVM learner.

2.6.7. Linear Regression (LR)

A simple regression model, linear regression, was used as the baseline model. Linear regression models the relationship between the predictor variables and the target variable as a linear function.

2.7. Model implementation and evaluation

The regression models were implemented using the mlr3 (Lang et al., 2019) and mlr3benchmark (Casalicchio & Burk, 2024) packages in R (R Core Team, 2023). To ensure reproducibility, default parameters were applied for all models. Prior to training, all data were scaled between 0 and 1 using the Min-Max scaling method to ensure uniformity across features and targets. This scaling prevented features with larger values from dominating the models and allowed for comparable evaluation metrics.

The performance of the regression models was evaluated using root mean squared error (*RMSE*) and *bias*. The RMSE is a measure of prediction accuracy that combines both bias and variability (e.g., Zitzmann et al., 2021). It is calculated by taking the square root of the mean across the squared differences between the predicted and observed values. A lower RMSE indicates that the predicted values have a higher likelihood of being close to the observed value, meaning the model performs better. For a detailed discussion, see, for example, Pargent et al. (2023). Bias measures the difference between the predicted and observed values, with a positive bias indicating the model overestimates the target variable, and a negative bias indicating underestimation. A bias of zero suggests that the model predicts the target variable accurately.

We modeled three continuous outcomes: Psychopathy, Narcissism, and Machiavellianism**.** Crucially, every algorithm saw the same combined predictor set, totaling 3 traits × 7 models per trait (total 21 models). To quantify uncertainty, we used a case-bootstrap with nested K-fold cross-validation**:** For each trait and each algorithm, we drew B = 1,000 bootstrap samples of participants with replacement. Within each bootstrap sample, we ran 10-fold cross-validation and aggregated RMSE and bias across folds, yielding one RMSE and one bias per bootstrap replicate. To make algorithm comparisons paired, we made sure that the same bootstrap index was used across algorithms for a given draw. Uncertainty was summarized by using 95% CIs, taken from the 1,000 bootstraps, for each algorithm × trait.

We report trait-wise point estimates with 95% CIs for RMSE and bias for each model. For each trait, we computed paired differences of RMSE across the same bootstrap replicates for every learner pair (Δ = RMSE_A − RMSE_B; lower favors A). For bias we compared |bias| (smaller is better). We formed Bonferroni-adjusted percentile CIs for the paired differences with global familywise $\alpha = 0.05$ across all pairs for the metric (i.e., RMSE and |bias| analyzed separately). A difference was considered statistically significant if the simultaneous CI excluded 0. We visualize results with significance matrices where an "x" indicates that the row model significantly outperformed the column model.

3. Results

3.1. Descriptive statistics

The mean scores were 3.08 (SD = 1.84, range = 0 – 10) for Psychopathy, 5.74 (SD = 3.52, range = 0 – 16) for Narcissism, and 3.16 (SD = 0.53, range = 1.75 – 4.55) for Machiavellianism. The correlations between the DT traits were $r = 0.25$ ($p < 0.001$) between Psychopathy and Narcissism, $r = 0.26$ ($p < 0.001$) between Psychopathy and Machiavellianism, and $r = 0.28$ ($p < 0.001$) between Narcissism and Machiavellianism.

3.2. Prediction accuracy

Table 1 reports RMSE (M, 95% CI) by approach and trait; Fig. 1 shows means with 95% CIs. For Psychopathy, RF achieved the lowest mean RMSE (0.116, 95% CI [0.102, 0.130]), followed by SVM (0.127, 95% CI [0.113, 0.142]) and GP (0.137, 95% CI [0.125, 0.149]). For Narcissism, RF (0.147, 95% CI [0.130, 0.163]) ranked best, followed by SVM (0.168, 95% CI [0.149, 0.187]) and GP (0.177, 95% CI [0.161, 0.193]). The same pattern emerged for Machiavellianism, with RF (0.127, 95% CI [0.113, 0.140]) performing best, followed again by SVM (0.137, 95% CI [0.122, 0.151]) and GP (0.148, 95% CI [0.135, 0.161]).

3.3. Bias

Table 2 reports bias (M, 95% CI) by approach and trait; Fig. 2 shows means with 95% CIs. For Psychopathy, both kNN (0.000, 95% CI [-0.020, 0.017]) and GLMNet (0.000, 95% CI [-0.004, 0.004]) performed similarly well, followed by RF (-0.001, 95% CI [-0.008, 0.005]). For Narcissism, GLMNet (0.001, 95% CI [-0.004, 0.007]) achieved the lowest bias, followed by M5R (-0.002, 95% CI [-0.020, 0.017]) and RF (-0.003, 95% CI [-0.011, 0.005]). For Machiavellianism, the lowest bias was found for RF (0.000, 95% CI [-0.007, 0.008]), GLMNet (-0.001, 95% CI [-0.006, 0.003]) and M5R (-0.001, 95% CI [-0.016, 0.015]).

3.4. Model comparison

Table 3 and Table 4 list all paired ΔRMSE and Δ|bias| with simultaneous CIs (FWER = 0.05 across all pairs). Table 5, Table 6, and Table 7 summarize which models significantly outperformed each other for each trait, when comparing RMSE. Generally, almost all models outperformed LR on each trait. RF won the most comparisons, significantly outperforming other models 15 out of 18 times. GP was the only other model beating models beside LR. When comparing bias, no model outperformed another significantly.

3.5. Exploratory feature importance

We ran model-agnostic permutation feature importance (PFI) on the held-out folds of each approach (10-fold CV; 5 permutations per feature) in order to explore which inputs most consistently drove predictions. To avoid model-specific idiosyncrasies, we summarize "stable" signals as features appearing in the Top-10 for at least three of the seven approaches. Percentages refer to the share of Top-10 slots occupied by each feature family, aggregated across approaches for the given trait.

Across traits, textual LIWC categories accounted for most importance in Psychopathy (≈ 88 %), with recurrent signals from perception/action language (e.g., perception, motion) and stylistic markers (e.g., swearing). Narcissism drew on a mix of demographics (Age, Gender; ≈ 51 %) and textual cues (≈ 49 %), especially interaction/expressivity topics (e.g., conversation, risk, tech, exclamation use). Machiavellianism showed substantial contributions from demographics (≈ 45 %) alongside textual features (≈ 35 %), with additional affective signal from NRC/AFINN (≈ 20 % combined; e.g., disgust, negative-valence bins) and acquisition-related language.

Table 1
Predictive accuracy (RMSE) by approach and trait, M [95% CI].

Trait	RF	SVM	GP	kNN	GLMNet	M5R	LR
Machiavellianism	0.127 [0.113, 0.140]	0.137 [0.122, 0.151]	0.148 [0.135, 0.161]	0.148 [0.124, 0.173]	0.170 [0.152, 0.190]	0.180 [0.156, 0.223]	0.452 [0.303, 0.766]
Narcissism	0.147 [0.130, 0.163]	0.168 [0.149, 0.187]	0.177 [0.161, 0.193]	0.200 [0.166, 0.233]	0.200 [0.181, 0.223]	0.210 [0.180, 0.257]	0.516 [0.347, 0.846]
Psychopathy	0.116 [0.102, 0.130]	0.127 [0.113, 0.142]	0.137 [0.125, 0.149]	0.150 [0.123, 0.177]	0.159 [0.141, 0.178]	0.179 [0.144, 0.196]	0.374 [0.249, 0.620]

Note. Values are means with 95% confidence intervals from 1,000 resamples. Lower RMSE indicates better performance.

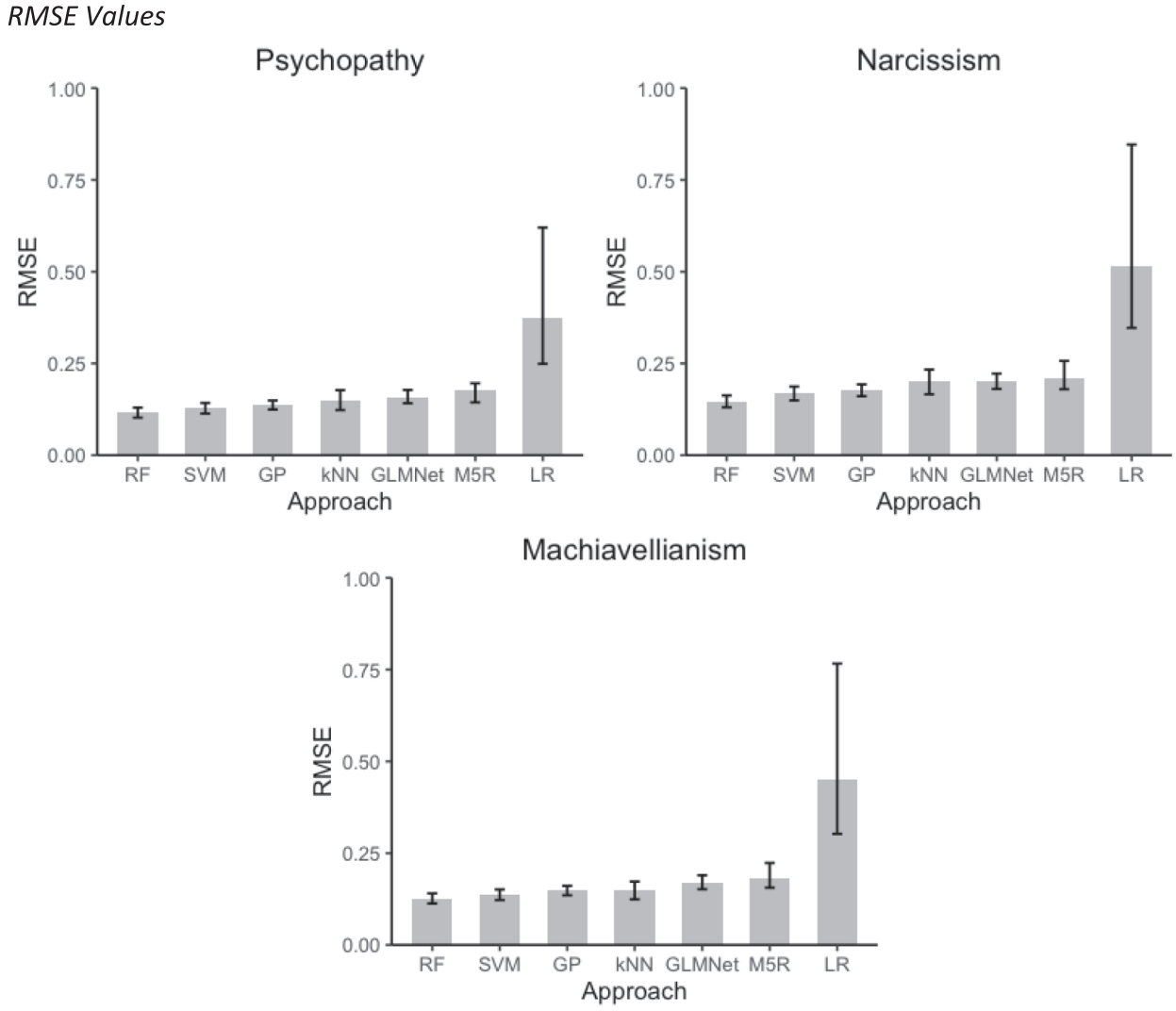

Fig. 1. *Note*. RMSE for all regression models for all traits. RF = Random Forest; SVM = Support Vector Machine; GP = Gaussian Processes; kNN = k-Nearest Neighbor; GLMNet = Generalized Linear Model with Elastic Net regularizer; M5R = M5′Rules; LR = Linear Regression.

Table 2
Prediction bias by approach and trait, M [95% CI].

Trait	RF	SVM	GP	kNN	GLMNet	M5R	LR
Machiavellianism	0.000 [-0.007, 0.008]	-0.002 [-0.012, 0.008]	0.002 [-0.003, 0.008]	-0.003 [-0.023, 0.017]	-0.001 [-0.006, 0.003]	-0.001 [-0.016, 0.015]	-0.001 [-0.061, 0.057]
Narcissism	-0.003 [-0.011, 0.005]	0.017 [0.006, 0.029]	0.004 [-0.002, 0.010]	0.008 [-0.019, 0.035]	0.001 [-0.004, 0.007]	-0.002 [-0.020, 0.017]	0.003 [-0.066, 0.071]
Psychopathy	-0.001 [-0.008, 0.005]	0.007 [-0.001, 0.014]	0.005 [0.001, 0.011]	0.000 [-0.020, 0.017]	0.000 [-0.004, 0.004]	-0.003 [-0.016, 0.013]	-0.004 [-0.052, 0.047]

Note. Bias > 0 indicates overprediction; Bias < 0 indicates underprediction. Intervals are 95% CIs from 1,000 resamples.

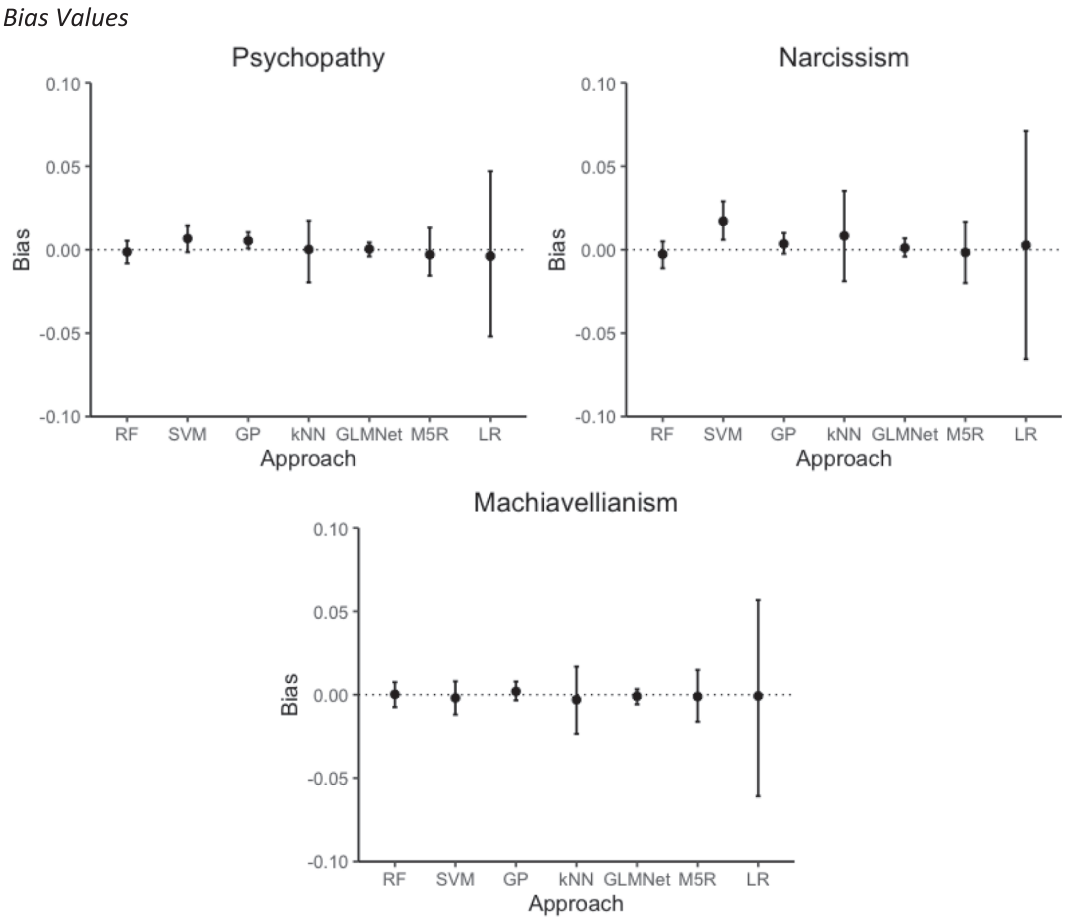

Fig. 2. *Note.* Bias for all regression models for all traits. RF = Random Forest; SVM = Support Vector Machine; GP = Gaussian Processes; kNN = k-Nearest Neighbor; GLMNet = Generalized Linear Model with Elastic Net regularizer; M5R = M5′Rules; LR = Linear Regression.

These results are exploratory: PFI reflects performance drops under permutation and can diffuse importance across correlated features; small negative values can occur from sampling noise. Nevertheless, the patterns suggest that content-level semantics and style jointly support prediction, with demographics contributing more strongly for Narcissism and Machiavellianism than for Psychopathy.

4. Discussion

The aim of this study was to compare different methods for predicting the DT traits from social media text data by extracting features using a closed-vocabulary approach and applying various machine learning regression models. The performance of these models was then evaluated and compared in terms of their ability to predict the DT traits.

The regression analysis results showed that regarding RMSE, the RF model performed best for predicting all three traits. When evaluating the models based on bias, the GLMNet model performed best for Psychopathy and Narcissism, and the RF model for Machiavellianism. However, the differences in bias between the models were small, indicating that all models predicted the target variables accurately. Using paired bootstrap comparisons with Bonferroni-adjusted simultaneous CIs (global FWER = 0.05), we found RF most often outperformed alternatives (Tables 3–7). For bias, no pairwise differences were significant (Table 4).

The results of this study align with previous research demonstrating that personality traits can be predicted from social media text data using machine learning models (Golbeck, Robles, & Turner, 2011; Golbeck, Robles, Edmondson, & Turner, 2011; Farnadi et al., 2016; Azucar et al., 2018). However, this study is among the first to specifically focus on predicting the DT traits from social media text data. Previous research has shown that ensemble learning methods can be more effective than other regression models for predicting personality traits (Sumner et al., 2012). This study underlines those previous findings, given that RF was the model that outperformed most others, across all DT traits.

5. Limitations

The results of this study should be interpreted with several limitations in mind. First, the features extracted from the text data were based on only three dictionaries, which may not capture the full range of language used on social media. Second, the regression models used in this study were implemented with default parameters and may not be optimal for predicting the DT traits. Because tuning sensitivity differs across algorithms, using defaults can advantage some approaches (e.g., simpler linear models) and disadvantage others (e.g., kernel or ensemble methods), potentially attenuating or inflating between-learner differences. Third, the results do not allow for the interpretation of the models' absolute performance, the RMSE values are meaningful only in comparison to each other and are not interpretable in an absolute sense. Fourth, the questionnaires used to measure the DT traits are not the most up-to-date. Newer measures, such as the Short Dark Triad (SD3; Jones & Paulhus, 2014) assess all three DT traits in a single questionnaire. Fifth, the feature space was highly dimensional, with 144 features, some of which may have been highly correlated and not particularly useful for predicting the DT traits. Finally, we did not include Sadism, so our scope is the Dark Triad rather than the full Dark Tetrad.

Table 3
Bootstrap paired differences of RMSE with global Bonferroni-simultaneous CI (familywise α = 0.05 across all pairs; CI level = 0.9992).

Trait	Comparison	Δ (a − b)	CI excludes 0
Machiavellianism	GP vs GLMNet	-0.022 [-0.063, 0.005]	No
Machiavellianism	GP vs RF	0.021 [0.008, 0.044]	Yes
Machiavellianism	GP vs SVM	0.011 [-0.018, 0.022]	No
Machiavellianism	GP vs kNN	-0.000 [-0.035, 0.039]	No
Machiavellianism	LR vs GLMNet	0.282 [0.091, 0.939]	Yes
Machiavellianism	LR vs GP	0.304 [0.112, 0.966]	Yes
Machiavellianism	LR vs M5R	0.271 [-0.098, 0.941]	No
Machiavellianism	LR vs RF	0.325 [0.127, 0.996]	Yes
Machiavellianism	LR vs SVM	0.315 [0.120, 0.981]	Yes
Machiavellianism	LR vs kNN	0.304 [0.097, 0.963]	Yes
Machiavellianism	M5R vs GLMNet	0.011 [-0.037, 0.276]	No
Machiavellianism	M5R vs GP	0.032 [0.000, 0.300]	Yes
Machiavellianism	M5R vs RF	0.054 [0.023, 0.318]	Yes
Machiavellianism	M5R vs SVM	0.044 [-0.015, 0.312]	No
Machiavellianism	M5R vs kNN	0.032 [-0.015, 0.296]	No
Machiavellianism	RF vs GLMNet	-0.043 [-0.089, -0.018]	Yes
Machiavellianism	RF vs SVM	-0.010 [-0.061, 0.003]	No
Machiavellianism	RF vs kNN	-0.021 [-0.056, 0.005]	No
Machiavellianism	SVM vs GLMNet	-0.033 [-0.077, 0.012]	No
Machiavellianism	kNN vs GLMNet	-0.021 [-0.071, 0.024]	No
Machiavellianism	kNN vs SVM	0.011 [-0.037, 0.045]	No
Narcissism	GP vs GLMNet	-0.023 [-0.063, -0.004]	Yes
Narcissism	GP vs RF	0.031 [0.016, 0.052]	Yes
Narcissism	GP vs SVM	0.010 [-0.023, 0.020]	No
Narcissism	GP vs kNN	-0.022 [-0.064, 0.024]	No
Narcissism	LR vs GLMNet	0.316 [0.079, 1.255]	Yes
Narcissism	LR vs GP	0.339 [0.099, 1.279]	Yes
Narcissism	LR vs M5R	0.307 [0.044, 1.250]	Yes
Narcissism	LR vs RF	0.370 [0.130, 1.314]	Yes
Narcissism	LR vs SVM	0.349 [0.104, 1.295]	Yes
Narcissism	LR vs kNN	0.317 [0.069, 1.271]	Yes
Narcissism	M5R vs GLMNet	0.009 [-0.038, 0.236]	No
Narcissism	M5R vs GP	0.032 [-0.006, 0.271]	No
Narcissism	M5R vs RF	0.063 [0.026, 0.300]	Yes
Narcissism	M5R vs SVM	0.042 [-0.005, 0.279]	No
Narcissism	M5R vs kNN	0.010 [-0.047, 0.255]	No
Narcissism	RF vs GLMNet	-0.053 [-0.100, -0.027]	Yes
Narcissism	RF vs SVM	-0.021 [-0.068, -0.004]	Yes
Narcissism	RF vs kNN	-0.053 [-0.093, -0.010]	Yes
Narcissism	SVM vs GLMNet	-0.033 [-0.076, 0.011]	No
Narcissism	kNN vs GLMNet	-0.000 [-0.063, 0.055]	No
Narcissism	kNN vs SVM	0.032 [-0.016, 0.073]	No
Psychopathy	GP vs GLMNet	-0.022 [-0.070, -0.002]	Yes
Psychopathy	GP vs RF	0.020 [0.007, 0.038]	Yes
Psychopathy	GP vs SVM	0.010 [-0.021, 0.019]	No
Psychopathy	GP vs kNN	-0.013 [-0.045, 0.022]	No
Psychopathy	LR vs GLMNet	0.215 [0.042, 0.695]	Yes
Psychopathy	LR vs GP	0.237 [0.064, 0.711]	Yes
Psychopathy	LR vs M5R	0.195 [-8.039, 0.686]	No
Psychopathy	LR vs RF	0.258 [0.084, 0.733]	Yes
Psychopathy	LR vs SVM	0.247 [0.077, 0.719]	Yes
Psychopathy	LR vs kNN	0.224 [0.031, 0.702]	Yes
Psychopathy	M5R vs GLMNet	0.020 [-0.047, 8.172]	No
Psychopathy	M5R vs GP	0.042 [0.003, 8.215]	Yes
Psychopathy	M5R vs RF	0.063 [0.024, 8.237]	Yes
Psychopathy	M5R vs SVM	0.052 [-0.010, 8.225]	No
Psychopathy	M5R vs kNN	0.029 [-0.023, 8.196]	No
Psychopathy	RF vs GLMNet	-0.042 [-0.095, -0.020]	Yes
Psychopathy	RF vs SVM	-0.011 [-0.055, 0.003]	No
Psychopathy	RF vs kNN	-0.034 [-0.062, -0.002]	Yes
Psychopathy	SVM vs GLMNet	-0.032 [-0.082, 0.011]	No
Psychopathy	kNN vs GLMNet	-0.009 [-0.060, 0.032]	No
Psychopathy	kNN vs SVM	0.023 [-0.027, 0.054]	No

Table 4
Bootstrap paired differences of |bias| with global Bonferroni-simultaneous CI (familywise α = 0.05 across all pairs; CI level = 0.9992).

Trait	Comparison	Δ (a − b)	CI excludes 0
Machiavellianism	GP vs GLMNet	0.001 [-0.013, 0.028]	No
Machiavellianism	GP vs RF	-0.000 [-0.009, 0.028]	No
Machiavellianism	GP vs SVM	-0.002 [-0.016, 0.025]	No
Machiavellianism	GP vs kNN	-0.006 [-0.034, 0.027]	No
Machiavellianism	LR vs GLMNet	0.020 [-0.005, 0.142]	No
Machiavellianism	LR vs GP	0.020 [-0.027, 0.143]	No
Machiavellianism	LR vs M5R	0.016 [-0.050, 0.139]	No
Machiavellianism	LR vs RF	0.019 [-0.009, 0.141]	No
Machiavellianism	LR vs SVM	0.018 [-0.015, 0.141]	No
Machiavellianism	LR vs kNN	0.014 [-0.034, 0.144]	No
Machiavellianism	M5R vs GLMNet	0.005 [-0.008, 0.053]	No
Machiavellianism	M5R vs GP	0.004 [-0.025, 0.054]	No
Machiavellianism	M5R vs RF	0.003 [-0.009, 0.052]	No
Machiavellianism	M5R vs SVM	0.002 [-0.013, 0.052]	No
Machiavellianism	M5R vs kNN	-0.002 [-0.033, 0.041]	No
Machiavellianism	RF vs GLMNet	0.001 [-0.013, 0.010]	No
Machiavellianism	RF vs SVM	-0.001 [-0.012, 0.009]	No
Machiavellianism	RF vs kNN	-0.005 [-0.034, 0.009]	No
Machiavellianism	SVM vs GLMNet	0.002 [-0.012, 0.016]	No
Machiavellianism	kNN vs GLMNet	0.006 [-0.010, 0.037]	No
Machiavellianism	kNN vs SVM	0.004 [-0.012, 0.030]	No
Narcissism	GP vs GLMNet	0.002 [-0.008, 0.023]	No
Narcissism	GP vs RF	-0.000 [-0.013, 0.021]	No
Narcissism	GP vs SVM	-0.013 [-0.029, 0.006]	No
Narcissism	GP vs kNN	-0.009 [-0.048, 0.018]	No
Narcissism	LR vs GLMNet	0.023 [-0.007, 0.210]	No
Narcissism	LR vs GP	0.021 [-0.018, 0.207]	No
Narcissism	LR vs M5R	0.018 [-0.039, 0.194]	No
Narcissism	LR vs RF	0.021 [-0.013, 0.208]	No
Narcissism	LR vs SVM	0.008 [-0.028, 0.195]	No
Narcissism	LR vs kNN	0.012 [-0.048, 0.203]	No
Narcissism	M5R vs GLMNet	0.005 [-0.010, 0.053]	No
Narcissism	M5R vs GP	0.004 [-0.021, 0.055]	No
Narcissism	M5R vs RF	0.004 [-0.012, 0.052]	No
Narcissism	M5R vs SVM	-0.010 [-0.031, 0.042]	No
Narcissism	M5R vs kNN	-0.006 [-0.052, 0.049]	No
Narcissism	RF vs GLMNet	0.002 [-0.012, 0.014]	No
Narcissism	RF vs SVM	-0.013 [-0.031, 0.013]	No
Narcissism	RF vs kNN	-0.009 [-0.051, 0.014]	No
Narcissism	SVM vs GLMNet	0.015 [-0.004, 0.032]	No
Narcissism	kNN vs GLMNet	0.011 [-0.008, 0.053]	No
Narcissism	kNN vs SVM	-0.004 [-0.028, 0.038]	No
Psychopathy	GP vs GLMNet	0.004 [-0.008, 0.023]	No
Psychopathy	GP vs RF	0.002 [-0.009, 0.023]	No
Psychopathy	GP vs SVM	-0.002 [-0.014, 0.023]	No
Psychopathy	GP vs kNN	-0.002 [-0.027, 0.019]	No
Psychopathy	LR vs GLMNet	0.017 [-0.005, 0.153]	No
Psychopathy	LR vs GP	0.014 [-0.018, 0.151]	No
Psychopathy	LR vs M5R	0.011 [-1.595, 0.152]	No
Psychopathy	LR vs RF	0.016 [-0.008, 0.153]	No
Psychopathy	LR vs SVM	0.012 [-0.016, 0.149]	No
Psychopathy	LR vs kNN	0.011 [-0.026, 0.151]	No
Psychopathy	M5R vs GLMNet	0.007 [-0.010, 1.603]	No
Psychopathy	M5R vs GP	0.003 [-0.023, 1.601]	No
Psychopathy	M5R vs RF	0.006 [-0.008, 1.608]	No
Psychopathy	M5R vs SVM	0.002 [-0.020, 1.600]	No
Psychopathy	M5R vs kNN	0.001 [-0.027, 1.604]	No
Psychopathy	RF vs GLMNet	0.001 [-0.009, 0.009]	No
Psychopathy	RF vs SVM	-0.004 [-0.023, 0.010]	No
Psychopathy	RF vs kNN	-0.005 [-0.026, 0.007]	No
Psychopathy	SVM vs GLMNet	0.005 [-0.008, 0.022]	No
Psychopathy	kNN vs GLMNet	0.006 [-0.008, 0.029]	No
Psychopathy	kNN vs SVM	0.001 [-0.016, 0.025]	No

5.1. Implications and future research

The results of this study have several implications for future research. First, the findings indicate that most models were able to predict the DT traits from social media text data. This suggests that social media data can provide useful insights into users' personalities, which may be valuable for employers, psychologists, and other professionals in making informed decisions related to hiring, treatment, advertising, and more. For example, employers could potentially use social media data (with consent) to screen job applicants and identify individuals with high levels of Psychopathy, Narcissism, or Machiavellianism, helping reduce the risk of hiring individuals prone to unethical behavior or harm. However, the use of social media data to predict personality traits involves important ethical and legal considerations and should be approached with caution.

Second, as hyperparameter tuning was disabled in this study, future

Table 5
Psychopathy — RMSE (x = row significantly better than column; Bonferroni-simultaneous bootstrap CI, global FWER α = 0.05).

	RF	SVM	GP	kNN	GLMNet	M5R	LR
RF	—		x	x	x	x	x
SVM		—					x
GP			—		x	x	x
kNN				—			x
GLMNet					—		x
M5R						—	
LR							—

Table 6
Machiavellianism — RMSE (x = row significantly better than column; Bonferroni-simultaneous bootstrap CI, global FWER α = 0.05).

	RF	SVM	GP	kNN	GLMNet	M5R	LR
RF	—		x		x	x	x
SVM		—					x
GP			—			x	x
kNN				—			x
GLMNet					—		x
M5R						—	
LR							—

Table 7
Narcissism — RMSE (x = row significantly better than column; Bonferroni-simultaneous bootstrap CI, global FWER α = 0.05).

	RF	SVM	GP	kNN	GLMNet	M5R	LR
RF	—	x	x	x	x	x	x
SVM		—					x
GP			—		x		x
kNN				—			x
GLMNet					—		x
M5R						—	x
LR							—

research could focus on optimizing the parameters of the regression models to further enhance predictive accuracy. It would be interesting to see if optimization provides indeed a measurable advantage. Additionally, other methods for predicting DT traits from social media data could be explored, such as deep learning models or other ensemble models that combine different approaches. Promising results have been found using deep learning models like Long Short-Term Memory (LSTM) networks (Kosan et al., 2022) and Bidirectional Encoder Representations from Transformers (BERT; Arijanto et al., 2021). Ensemble models that integrate different regression models have also proven effective for predicting personality traits from social media data (Sumner et al., 2012). These models may outperform the basic regression models used in this study. Furthermore, since the DT traits are correlated (Paulhus & Williams, 2002), using a multivariate model that captures these correlations might improve the predictive accuracy. Indeed, Farnadi et al. (2016) demonstrated that multivariate models can be more effective for predicting the Big Five traits from social media data than univariate models. Future research could explore whether the same holds true for the DT traits.

Third, future research could explore alternative methods for extracting features from social media text data. In this study, a closed-vocabulary approach was used, which has a long history in personality research but may not capture all nuances of language used on social media. Several studies have employed open-vocabulary approaches, such as LSA (Garcia & Sikström, 2014) or Differential Language Analysis (DLA; Schwartz et al., 2013), to extract features from social media text data. These data-driven, bottom-up approaches may be more effective as they do not rely on predefined dictionaries. However, they typically require larger datasets and can be more challenging to interpret (Eichstaedt et al., 2021). Finally, future research could expand its focus to include the Dark Tetrad traits. While this study concentrated on predicting Psychopathy, Narcissism, and Machiavellianism, the Dark Tetrad model also includes Sadism as a fourth trait (Buckels et al., 2013). Incorporating Sadism would require a larger, new dataset but would offer a more comprehensive understanding of users' personalities on social media.

In conclusion, this study demonstrates that Dark Triad traits can be predicted from social media text data with meaningful accuracy using machine learning approaches. Across all three traits, Random Forest consistently achieved the lowest prediction error and outperformed most alternative models in paired comparisons, indicating that flexible ensemble methods are particularly well suited for this task. Nevertheless, no single model dominated all others under all evaluation criteria, and differences in bias were small across approaches. These findings suggest that while Random Forest currently represents a strong baseline for applied prediction of Dark Triad traits from text, there remains substantial room for future work to further refine modeling strategies, optimize hyperparameters, and explore alternative feature representations to gain deeper, more nuanced insights into the complex personalities behind digital profiles.

Open Science Statement

The study materials, data and analysis scripts used for this article will not be made publicly available as the authors did not collect data, but rather used a dataset provided by another group of researchers (see Methods section).

Preregistration

The study was not preregistered.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

References

Aha, D. W., Kibler, D., & Albert, M. K. (1991). Instance-based learning algorithms. *Machine Learning, 6*(1), 37–66. https://doi.org/10.1007/BF00153759

Ames, D. R., Rose, P., & Anderson, C. P. (2006). The NPI-16 as a short measure of narcissism. *Journal of Research in Personality, 40*(4), 440–450. https://doi.org/10.1016/j.jrp.2005.03.002

Arijanto, J. E., Geraldy, S., Tania, C., & Suhartono, D. (2021). Personality prediction based on text analytics using bidirectional encoder representations from transformers from english twitter dataset. *International Journal of Fuzzy Logic and Intelligent Systems, 21*(3), 310–316. https://doi.org/10.5391/IJFIS.2021.21.3.310

Azucar, D., Marengo, D., & Settanni, M. (2018). Predicting the big 5 personality traits from digital footprints on social media: A meta-analysis. *Personality and Individual Differences, 124*, 150–159. https://doi.org/10.1016/j.paid.2017.12.018

Barrick, M. R., & Mount, M. K. (1991). The big five personality dimensions and job performance: A meta-analysis. *Personnel Psychology, 44*(1), 1–26. https://doi.org/10.1111/j.1744-6570.1991.tb00688.x

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *The Journal of Machine Learning Research, 3*, 993–1022.

Boyd, R. L., Ashokkumar, A., Seraj, S., & Pennebaker, J. W. (2022). *The development and psychometric properties of LIWC-2022*. University of Texas at Austin.

Breiman, L. (2001). Random Forests. *Machine Learning, 45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Buckels, E. E., Jones, D. N., & Paulhus, D. L. (2013). Behavioral confirmation of everyday sadism. *Psychological Science, 24*(11), 2201–2209. https://doi.org/10.1177/0956797613490749

Casalicchio, G., & Burk, L. (2024). Evaluation and benchmarking. In B. Bischl, R. Sonabend, L. Kotthoff, & M. Lang (Eds.), *Applied machine learning using mlr3 in R*. CRC Press.

Christie, R., & Geis, F. L. (1970). *Studies in Machiavellianism*. *Elsevier*. https://doi.org/10.1016/C2013-0-10497-7

Claxton, A., O'Rourke, N., Smith, J., & DeLongis, A. (2012). Personality traits and marital satisfaction within enduring relationships: An intra-couple discrepancy approach. *Journal of Social and Personal Relationships, 29*(3), 375–396. https://doi.org/10.1177/0265407511431183

Costa, P. T., & McCrae, R. R. (1992). The five-factor model of personality and its relevance to personality disorders. *Journal of Personality Disorders, 6*(4), 343–359. https://doi.org/10.1521/pedi.1992.6.4.343

Costa, P. T., & McCrae, R. (1999). A five-factor theory of personality. *The Five-Factor Model of Personality: Theoretical Perspectives, 2*, 51–87.

DeYoung, C. G. (2015). Cybernetic big five theory. *Journal of Research in Personality, 56*, 33–58. https://doi.org/10.1016/j.jrp.2014.07.004

Eichstaedt, J. C., Kern, M. L., Yaden, D. B., Schwartz, H. A., Giorgi, S., Park, G., Hagan, C. A., Tobolsky, V. A., Smith, L. K., Buffone, A., Iwry, J., Seligman, M. E. P., & Ungar, L. H. (2021). Closed- and open-vocabulary approaches to text analysis: A review, quantitative comparison, and recommendations. *Psychological Methods, 26*(4), 398–427. https://doi.org/10.1037/met0000349

Eysenck, S., Eysenck, H., & Barrett, P. (1985). A revised version of the psychoticism scale. *Personality and Individual Differences, 6*(1), 21–29. https://doi.org/10.1016/0191-8869(85)90026-1

Farnadi, G., Sitaraman, G., Sushmita, S., Celli, F., Kosinski, M., Stillwell, D., Davalos, S., Moens, M.-F., & De Cock, M. (2016). Computational personality recognition in social media. *User Modeling and User-Adapted Interaction, 26*(2), 109–142. https://doi.org/10.1007/s11257-016-9171-0

Friedman, H. S., & Kern, M. L. (2014). Personality, Well-being, and Health. *Annual Review of Psychology, 65*(1), 719–742. https://doi.org/10.1146/annurev-psych-010213-115123

Friedman, J. H., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software, 33*(1), 1–22. https://doi.org/10.18637/jss.v033.i01.

Furnham, A., Richards, S. C., & Paulhus, D. L. (2013). The dark triad of personality: A 10 year review. *Social and Personality Psychology Compass*, *7*(3), 199–216. https://doi.org/10.1111/spc3.12018.

Garcia, D., & Sikström, S. (2014). The dark side of Facebook: Semantic representations of status updates predict the dark triad of personality. *Personality and Individual Differences, 67*, 92–96. https://doi.org/10.1016/j.paid.2013.10.001

Golbeck, J., Robles, C., Edmondson, M., & Turner, K. (2011). Predicting personality from twitter. *2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing*, 149–156. https://doi.org/10.1109/PASSAT/SocialCom.2011.33.

Golbeck, J., Robles, C., & Turner, K. (2011). Predicting personality with social media. *CHI '11 Extended Abstracts on Human Factors in Computing Systems*, 253–262. https://doi.org/10.1145/1979742.1979614.

Holmes, G., Hall, M., & Prank, E. (1999). Generating rule sets from model trees. In G. Goos, J. Hartmanis, J. Van Leeuwen, & N. Foo (Eds.), *Advanced topics in artificial intelligence* (pp. 1–12, Vol. 1747). Springer Berlin Heidelberg. https://doi.org/10.1007/3-540-46695-9_1.

Jonason, P. K., & Webster, G. D. (2010). The dirty dozen: A concise measure of the dark triad. *Psychological Assessment, 22*(2), 420–432. https://doi.org/10.1037/a0019265

Jones, D. N., & Paulhus, D. L. (2014). Introducing the short dark triad (SD3): A brief measure of dark personality traits. *Assessment, 21*(1), 28–41. https://doi.org/10.1177/1073191113514105

Kaufman, S. B., Yaden, D. B., Hyde, E., & Tsukayama, E. (2019). The light vs. dark triad of personality: Contrasting two very different profiles of human nature. *Frontiers in Psychology, 10*, 467. https://doi.org/10.3389/fpsyg.2019.00467

Kosan, M. A., Karacan, H., & Urgen, B. A. (2022). Predicting personality traits with semantic structures and LSTM-based neural networks. *Alexandria Engineering Journal, 61*(10), 8007–8025. https://doi.org/10.1016/j.aej.2022.01.050

Lang, M., Binder, M., Richter, J., Schratz, P., Pfisterer, F., Coors, S., … Bischl, B. (2019). Mlr3: A modern object-oriented machine learning framework in R. *Journal of Open Source Software, 4*(44), 1903. https://doi.org/10.21105/joss.01903

Mairesse, F., Walker, M. A., Mehl, M. R., & Moore, R. K. (2007). Using linguistic cues for the automatic recognition of personality in conversation and text. *Journal of Artificial Intelligence Research, 30*, 457–500. https://doi.org/10.1613/jair.2349

McAdams, D. P., & Pals, J. L. (2006). A new big five: Fundamental principles for an integrative science of personality. *American Psychologist, 61*(3), 204–217. https://doi.org/10.1037/0003-066X.61.3.204

McCrae, R. R., & Costa, P. T. (2008). The five-factor theory of personality. In *Handbook of personality: Theory and research, 3rd ed.* (pp. 159–181). The Guilford Press.

McCrae, R. R., & John, O. P. (1992). An Introduction to the five-factor model and its applications. *Journal of Personality, 60*(2), 175–215. https://doi.org/10.1111/j.1467-6494.1992.tb00970.x

Međedović, J., & Petrović, B. (2015). The dark tetrad: Structural properties and location in the personality space. *Journal of Individual Differences, 36*(4), 228–236. https://doi.org/10.1027/1614-0001/a000179

Mohammad, S. M., & Turney, P. D. (2013). Crowdsourcing a word-emotion association lexicon. *Computational Intelligence, 29*(3), 436–465.

Nielsen, F. Å. (2011). A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. https://doi.org/10.48550/ARXIV.1103.2903.

Paleczek, D., Bergner, S., & Rybnicek, R. (2018). Predicting career success: Is the dark side of personality worth considering? *Journal of Managerial Psychology, 33*(6), 437–456. https://doi.org/10.1108/JMP-11-2017-0402

Pargent, F., Schoedel, R., & Stachl, C. (2023). Best practices in supervised machine learning: A tutorial for psychologists. *Advances in Methods and Practices in Psychological Science, 6*(3), Article 25152459231162559. https://doi.org/10.1177/25152459231162559

Paulhus, D. L. (2014). Toward a taxonomy of dark personalities. *Current Directions in Psychological Science, 23*(6), 421–426. https://doi.org/10.1177/0963721414547737

Paulhus, D. L., & Williams, K. M. (2002). The dark triad of personality: Narcissism, machiavellianism, and psychopathy. *Journal of Research in Personality, 36*(6), 556–563. https://doi.org/10.1016/S0092-6566(02)00505-6

Pennebaker, J. W., Boyd, R. L., Booth, R. J., Ashokkumar, A., & Francis, M. E. (2022). Linguistic Inquiry and Word Count: LIWC-22.

Pennebaker, J. W., & King, L. A. (1999). Linguistic styles: Language use as an individual difference. *Journal of Personality and Social Psychology, 77*(6), 1296–1312. https://doi.org/10.1037/0022-3514.77.6.1296

Poropat, A. E. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychological Bulletin, 135*(2), 322–338. https://doi.org/10.1037/a0014996

Preotiuc-Pietro, D., Carpenter, J., Giorgi, S., & Ungar, L. (2016). Studying the dark triad of personality through twitter behavior. *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, 761–770. https://doi.org/10.1145/2983323.2983822.

R Core Team. (2023). *R: A language and environment for statistical computing* (Version 4.3.1) [Computer Software]. Vienna, Austria. https://www.r-project.org/.

Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Ramones, S. M., Agrawal, M., Shah, A., Kosinski, M., Stillwell, D., Seligman, M. E. P., & Ungar, L. H. (2013). Personality, gender, and age in the language of social media: The open-vocabulary approach. *PLoS ONE, 8*(9), Article e73791. https://doi.org/10.1371/journal.pone.0073791

Spurk, D., Keller, A. C., & Hirschi, A. (2016). Do bad guys get ahead or fall behind? Relationships of the dark triad of personality with objective and subjective career success. *Social Psychological and Personality Science, 7*(2), 113–121. https://doi.org/10.1177/1948550615609735

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*, 386–393. https://doi.org/10.1109/ICMLA.2012.218.

Vora, H., Bhamare, M., & Kumar, D. K. A. (2020). Personality prediction from social media text: An overview. *International Journal of Engineering Research and Technology, V9*(05). https://doi.org/10.17577/IJERTV9IS050203

Williams, C., & Barber, D. (1998). Bayesian classification with gaussian processes. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 20*(12), 1342–1351. https://doi.org/10.1109/34.735807

Zitzmann, S., Lüdtke, O., Robitzsch, A., & Hecht, M. (2021). On the performance of Bayesian approaches in small samples: A comment on smid, mcneish, miocevic, and van de schoot (2020). *Structural Equation Modeling: A Multidisciplinary Journal, 28*(1), 40–50. https://doi.org/10.1080/10705511.2020.1752216