JOURNAL OF COMMUNICATION, LANGUAGE AND CULTURE

Linguistic Indicators of Narcissistic Tendencies for Predicting Narcissistic Traits in Malaysians Through Social Media Content

Siti-Soraya Abdul-Rahman[1]*, Amira An-Nur Rusli[2], Rafidah Aga Mohd Jaladin[3]

[1,2]Department of Artificial Intelligence, Faculty of Computer Science and Information Technology, Universiti Malaya, Kuala Lumpur, Malaysia
[3]Department of Educational Psychology and Counselling, Faculty of Education, Universiti Malaya, Kuala Lumpur, Malaysia

*Corresponding author: siti_soraya@um.edu.my; ORCID iD: 0000-0002-9945-1826

ABSTRACT

Narcissistic tendencies are increasingly observable in online interactions, particularly on social media platforms. X (formerly Twitter) has been identified as a popular platform among individuals exhibiting narcissistic traits. Predicting these traits based on language use remains an essential area of study. This research examines narcissistic tendencies among Malaysian social media users by analysing 2,129 posts on Twitter. Utilising natural language processing (NLP) methods and machine learning algorithms, we assess linguistic patterns, sentiment, and engagement metrics to identify indicators of narcissism. Four machine learning algorithms, Support Vector Machine (SVM), Naïve Bayes, Logistic Regression, and Gradient Boosting, were assessed based on multiple performance metrics. Results indicate that SVM is the most effective model, achieving 80% accuracy with 10-fold cross-validation, demonstrating its reliability in predicting narcissistic traits. These findings contribute to computational psychology and social media analytics, offering insights into the psychological dimensions of digital self-presentation and the implications of narcissistic behaviours in online communities.

Keywords: narcissism, social media, X (formerly Twitter), machine learning, natural language processing

Received: 30 March 2025, **Accepted:** 1 July 2025, **Published:** 30 July 2025

Introduction

Narcissism, as defined in the Oxford Dictionary, is a type of personality disorder where a person has an excessive interest or admiration towards themselves. According to Narcissistic Personality Disorder Facts (2018), 6.2% of people have narcissism at least once in a lifetime, and 7.7% of men are said to develop more narcissism than 4.8% of women. There are both advantages and disadvantages to possessing narcissistic traits. On the positive side, such traits can contribute to an individual's self-

Journal of Communication, Language and Culture
doi: https://doi.org/10.33093/jclc.2025.5.2.6

Published by MMU Press. URL: https://journals.mmupress.com/jclc

confidence. However, on the negative side, narcissistic individuals often exhibit manipulative and controlling behaviours towards others.

Narcissism has become a growing concern in contemporary society. Macenczak et al. (2016) highlighted that extensive research examined narcissism as a significant factor influencing individual behaviour, particularly within organisational contexts. Similarly, Ismail (2018) emphasised that narcissistic personality disorder has predominantly negative effects, impacting attitudes, behaviours, social relationships, and job performance. Moreover, studies suggest that women with narcissistic traits exhibit higher prevalence rates of major depressive disorder and anxiety disorder. Meanwhile, men with narcissistic traits are more prone to developing substance use disorders and antisocial personality disorders. ("Narcissistic Personality Disorder Facts," 2018).

Holtzman and Strube's (2013) study provides insights into the connection between physical attractiveness and dark personality traits, emphasising how people with these traits effectively adorn themselves to improve their appearance. Although they are adept at strategically enhancing their appearance, dark personalities—especially psychopaths and narcissists—are not naturally more physically attractive. This could help them develop and possibly take advantage of social connections more successfully by facilitating their manipulative social strategies. While numerous studies have explored the identification of narcissism through personality tests, behavioural assessments, and language analysis, relatively few have focused on detecting narcissistic traits through social media usage (Sumner et al., 2012). On social media, selfies are becoming increasingly common. By coding participants' selfies uploaded on a social networking site and assessing their Big Five personality traits, Qiu et al. (2015) investigated the relationship between personality and selfies. Their research opens a new line of inquiry into the relationship with narcissistic traits because it is the first to identify personality-related cues in selfies and to offer a picture-coding scheme for selfie analysis. According to previous research (Ksinan & Vazsonyi, 2016), grandiose narcissists frequently use social media to post updates, statuses, and photos as a means of promoting themselves and seeking attention from others. Excessive narcissism is commonly viewed as problematic, both for individuals displaying these traits and for those in their social circles ("Narcissistic Personality Disorder Facts," 2018). The use of Twitter has experienced a significant increase, and Davenport et al. (2014) observe that the platform has gained popularity among researchers studying narcissism. Similarly, Twitter can serve as a medium that amplifies the expression of narcissistic traits ("Social media and narcissism," n.d.).

Text analytics and machine learning techniques have become integral tools for extracting insights, identifying key terms, and categorising content into thematic topics or sentiment classes such as positive, neutral, or negative. Furthermore, these advanced methods are increasingly leveraged for detection and predictive modelling tasks, including the identification of cyberbullying, suicide risk, and psychopathic behaviour within social media platforms (Anandarajan et al., 2019).

The rest of this paper is structured as follows: Section 2 discusses related work, and we present research methodology in Section 3. In Section 4, we highlight results and discussion, and finally, in Section 5, we conclude our research and propose future work.

Literature Review

Narcissistic Personality Disorder (NPD) is a clinical term that refers to the act of feeling arrogant, superior, self-centred, lacking empathy, and seeking attention (Ismail, 2018). A narcissistic personality can be presented in two types or categories: grandiose and vulnerable narcissism (Ksinan & Vazsonyi, 2016). A grandiose narcissist is more likely to have feelings of being superior, self-assured, extroverted, and attention-seeking. In contrast, a vulnerable narcissist tends to have feelings towards neglect, introversion, and a lack of self-esteem and confidence (Hart et al., 2017). Grandiose narcissism is a dominant form of narcissism defined by an exaggerated self-focus at the expense of others. Individuals with grandiose narcissism tend to make decisions that primarily benefit themselves. This form of narcissism is strongly linked to characteristics such as egoism, envy, and a heightened sense of entitlement. Moreover, grandiose narcissists often refuse to acknowledge their mistakes or learn from them (Thorpe, 2016). Vulnerable narcissism is the opposite of grandiose narcissism. Individuals with a

vulnerable narcissistic sense of superiority respond differently because of their need for attention; they will act differently if their validation is unmet. Moreover, vulnerable narcissists more often experience depression, a sense of worthlessness, or humiliation when their feelings are ignored or deprived of the attention they seek (Thorpe, 2016). Vulnerable narcissism, when compared to its grandiose counterpart, is a particularly concerning form of narcissism, as it can pose a significant risk to an individual's well-being. As noted by the Mayo Clinic Staff (2017), there are notable contributing factors, such as suicidal ideation or associated prevalent behaviours that exhibit vulnerable narcissism, which often stem from the underlying form of depression and anxiety. The emergence of narcissistic traits can be shaped by an individual's interactions with both themselves and others.

To analyse narcissistic traits through language, James W. Pennebaker developed software called The Linguistic Inquiry and Word Count (LIWC) for text analysis that quantifies word usage across various categories. Many researchers and data analysts have leveraged LIWC for years. LIWC includes multiple classifications across domains, including linguistic dimensions, psychological processes, personal concerns, and spoken language patterns. According to Rathner et al. (2018), many researchers in the past have leveraged LIWC to analyse narcissistic traits through language used by calculating the percentage of words that could potentially be related to narcissism. Several reviews of the literature in the past have concluded that first-person singular pronouns, social processes, swear words, and affective process words exhibit a strong association with narcissistic traits.

A narcissist is often characterised as an individual who frequently engages in promoting themselves. As described previously, researchers have suggested that individuals with narcissism tend to use a higher frequency of first-person singular pronouns while using fewer plural pronouns in communication. According to DeWall et al. (2011), an individual exhibiting narcissistic traits may be unaware of their frequent use of first-person singular pronouns in conversation, which serves as a means of attracting attention to themselves. First-person pronouns refer to the speaker. An example of the singular form of the first-person is "I", "me", "my", and the plural form of the first-person is "we", "our", "us" (Carey et al., 2015). Although there are many studies done by researchers that show the correlation between first-person singular pronouns and narcissistic personality, Holtzman et al. (2010) mentioned in their research that they only found minimal proof that shows a strong relationship between them.

Kacewicz et al. (2014) found that Individuals in lower-status positions exhibited a greater frequency of first-person singular pronoun usage (e.g., "I," "me," "my"), suggesting an increased self-focus and heightened social awareness. Conversely, individuals in higher-status roles demonstrated a reduced use of first-person singular pronouns, opting instead for third-person references and plural pronouns, indicating a stronger emphasis on others and the broader social context. Additionally, their research indicates that status is linked to attentional biases, with a higher rank associated with other-focused behaviour and a lower rank with self-focused behaviour.

Words related to relationships or behaviours towards other individuals are known as social processes or interactions. In the LIWC tool, social processes have a few categories: friends (e.g., "friend", "buddy", "coworker", "ex"), family (e.g., "mother", "father") and humans (e.g., "man", "woman", "boy"). A study by Holtzman et al. (2010) found a strong association between narcissism and extraverted behaviour, specifically the use of "friend" words. Moreover, McGregor (2010) asserted that "friend" words exhibit a stronger correlation with narcissism compared to other terms within the social processes category. This suggests that narcissists are inclined to socialise with others and seek to be the centre of attention within social groups.

As stated by DeWall et al. (2011) in their study on linguistic analysis and narcissism, narcissists often use swear words as a means of expressing anger or hostility towards others. Swear words are rude and offensive terms commonly used across the globe. Examples of swear words include "crap," "damn," "fuck," "piss," and others. Additionally, Holtzman et al. (2010) revealed that a study on narcissistic tendencies in daily life reveals a positive correlation between the use of profanity and narcissistic traits. The study also noted that narcissists often combine swear words with anger-related language to express their frustration.

Affective processes are words that usually deal with feelings or emotions. In the LIWC Dictionary, affective processes are divided into several sub-classes called positive emotions (e.g., "happy", "great", "admire"), negative emotions (e.g., "hate", "enemy", "sad"), anxiety (e.g., "nervous", "distraught", "afraid"), anger (e.g., "annoyed", "mad", "jealous") and sadness (e.g., "sad", "regret", "heartbroken"). According to Czarna and Zajenkowski (2018), to gain a deeper understanding of narcissism, researchers emphasised that affective processes are crucial hints of narcissistic personality traits.

A study conducted by Holtzman et al. (2019) provides a comprehensive analysis across 15 independent samples in characterising linguistic markers closely associated with grandiose narcissism. By leveraging the LIWC tool, their study, which used the LIWC tool, demonstrated that people with high levels of grandiose narcissism had unique linguistic patterns, including greater use of first-person singular pronouns and words related to achievement and negative emotions.

Casale and Banchi (2020) found evidence linking narcissism to Problematic Social Media Use (PSMU) that has not yet been systematised, even though many meta-analyses have been carried out to synthesise empirical evidence on the relationship between narcissism and common online behaviours (such as uploading photos and usage frequency). Their study is the first systematic review of its kind. Grandiose narcissism and Problematic Facebook Use (PFU) were found to be positively and significantly correlated, according to consistent findings. Overall, the findings suggested that narcissism may play a role in PFU, though its effects may vary depending on the social media platform. These findings are also consistent with the research conducted by Ksinan and Vazsonyi (2016).

Anggarawati et al. (2023) examined the mediating influence of narcissistic behaviour on the association between the context of materialism and online consumption on social media platforms, which confirmed that narcissistic traits could influence online behaviours.

Various machine learning algorithms are well-suited for text analytics and have been leveraged by many researchers. Among the notable ones are decision trees, logistic regression, K-Nearest Neighbors (KNN), Naïve Bayes, Support Vector Machines (SVM), Gradient Boosting, Artificial Neural Networks (ANN), Random Forest, and others. Mariconti, Suarez-Tangil, Blackburn et al. (2019) analyse and model YouTube videos along multiple axes (metadata, audio transcripts, and thumbnails) using a ground truth dataset of videos. Their study suggests an automated method for identifying YouTube videos. They examine comment sections using an NLP and an ensemble of classifiers to find trends that point to organised harassment. Their work offers a crucial first step in implementing proactive systems to identify and stop coordinated hate attacks on websites such as YouTube.

A commonly used supervised machine learning model called Support Vector Machine (SVM) is used for regression, classification, and anomaly detection tasks. Texts converted into vectors can be used for classification or text analysis tasks, like determining whether a person is narcissistic. Al-Garadi et al. (2016) used the SVM classifier to predict psychopathy on social media. Their results demonstrated that SVM performed better than linear regression in this regard. Burnap et al. (2015) conducted a study on classifying suicide-related content on Twitter, demonstrating that SVM performed well with short text and significantly outperformed other machine learning classifiers in terms of accuracy metrics.

Logistic regression is known for its effective model, a widely utilised algorithm in NLP. It can capture the relationship between input features and categorical outcomes, making it particularly suitable for tasks such as sentiment analysis and text classification. Its ability to provide probabilities for class membership and its interpretability make it particularly valuable for categorising or predicting discrete outcomes based on linguistic features. Besides, logistic regression's features on scalability and its efficiency in handling large datasets contribute to its widespread use in various classification problems, such as spam detection (Milosevic et al., 2017). According to Brindha et al. (2016), logistic regression is a commonly used machine learning classifier for fundamental binary classification tasks, given its effectiveness in distinguishing between two distinct classes.

On the other hand, Naïve Bayes is a probabilistic classifier that assigns data to categories based on Bayes' Theorem under the assumption of independence between variables or features, meaning that one feature's presence does not influence another's presence. According to Al-Garadi et al. (2016), Naïve

Bayes is one of the machine learning classifiers proven effective in social media studies. For example, research conducted by Pratama and Sarno (2016) utilised Multinomial Naïve Bayes (MNB) to classify personality traits on Twitter text. In their study, the weight of each word was calculated using a multinomial distribution. The results demonstrated that MNB surpassed other machine learning classifiers, such as SVM and KNN, in terms of performance accuracy metrics during cross-validation.

Gradient Boosting is a machine learning classifier that connects multiple weak learners to construct a strong predictive model. The classifier gradually improves its performance by iteratively increasing these weak learners' accuracy, making it highly effective at classifying complex data, including textual information. Regression tasks, multi-class classification, and binary classification issues can benefit from gradient boosting. However, there is a notable distinction between bagging and boosting. While boosting adjusts the distribution of samples at each training step based on the errors produced, bagging, on the other hand, selects each sample uniformly to create a training dataset (Zhang & Haghani, 2015). Semanjski and Gautama (2015) employed Gradient Boosting Trees (GBT) to model user decision-making while selecting transportation modes. They identified that boosting improved the stability of their developed machine-learning model by assigning greater weights to misclassified training instances, thus facilitating the generation of a new tree.

This paper focuses on several machine learning classifiers, namely SVM, Logistic Regression, Naïve Bayes, and Gradient Boosting. These classifiers are widely researched in sentiment analysis classification and have demonstrated strong performance in recent studies related to text classification and NLP tasks.

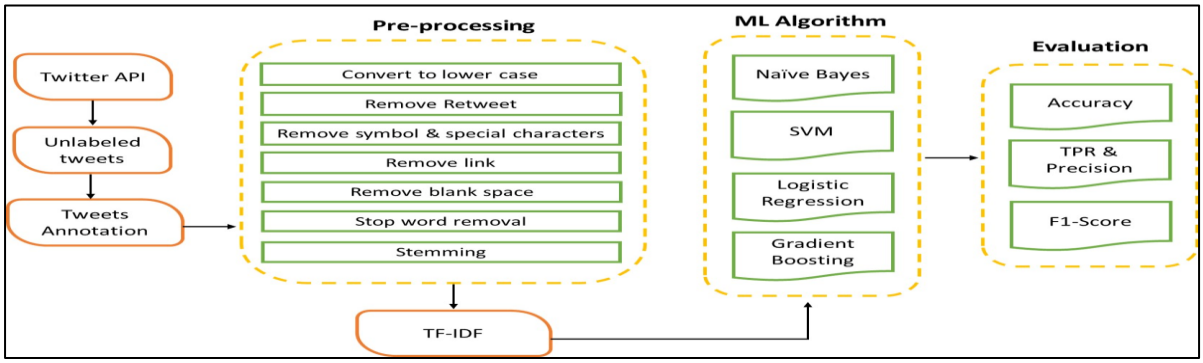
Methods

The steps in the research methodology are shown in Figure 1. Tweet annotations are part of the pre-processing step before they are vectorised using Term Frequency-Inverse Document Frequency (TF-IDF). TF-IDF prioritises terms with high relevance to specific individual documents by giving them more weight than uncommon terms throughout the dataset. This study used several classifiers, including Naïve Bayes, Support Vector Machines (SVM), Logistic Regression, and Gradient Boosting, to classify tweets to identify potential narcissistic traits. The performance of these classifier models was assessed using performance criteria such as accuracy, precision, true positive rate, and F1-score to determine which classifier was best suited for the study.

By normalising words, stemming can enhance the predictive power of text analytics models. Simplifying vocabulary ensures that different forms of the same word are treated similarly, improving the accuracy of models such as Naïve Bayes and Logistic Regression. Stemming can lead to faster computation by reducing the size of the dataset. By condensing multiple word forms into a single base form, the number of features the algorithm needs to process is decreased, leading to quicker training times.

Figure 1

Research Overview Diagram



Data Collection

In this research, common words, phrases, or language used by narcissists act as keywords to extract tweets. The keywords related to the tendency of a Twitter user to be a narcissist are based on DoctorRamani (2019), a clinical psychologist in Santa Monica. In her video, she identified phrases that a narcissist commonly utters. The example of keywords are: *'kau ni sensitif'; 'kau ni cepat emosi'; 'aku lebih mengetahui'; 'aku pandai'; 'aku cantik', 'jangan terasa tapi'; 'tiada yang lebih baik dari aku'; 'aku bergurau sahaja'; 'aku tak pernah kata pun'; 'tiada siapa suka kau'; 'aku seorang sahaja yang sayang kau'; 'dia ni kawan sahaja'; 'aku nak perhatian'; 'kau perlukan pertolongan'; 'kau ni bodoh'; 'kau ni gila'.* All the keywords were used as search filters to collect tweets. Since Malaysians usually use Manglish words (Malaysian English) to convey their thoughts and feelings, the phrases are translated to Bahasa Melayu and used as the search filter in this research. Table 1 shows the keywords used as search filters in Bahasa Melayu and English.

Table 1

Narcissist's Common Phrases in Bahasa Melayu and English

Bahasa Melayu phrases	**English phrases**
Kau ni sensitif	You're sensitive
Kau ni cepat emosi	You're too emotional.
Aku lebih mengetahui	I know better
Aku pandai	I'm smart/ I'm clever.
Aku cantik	I'm beautiful/I'm pretty.
Jangan terasa tapi	No offence but
Tiada yang lebih baik dari aku	No one is better than me.
Aku bergurau sahaja	I am just joking
Aku tak pernah kata pun	I never said that
Tiada siapa suka kau	No one likes you
Aku seorang sahaja yang sayang kau	I'm the only one who loves you.
Dia ni kawan sahaja	He/She is just a friend.
Aku nak perhatian	I want attention
Kau perlukan pertolongan	You need help
Kau ni bodoh	You're stupid
Kau ni gila	You're crazy

The tweets are filtered by location using the geocode. The longitude, latitude, and kilometres specified in geocode parameters refer to building a geofence within the preferred location. The purpose of filtering user location is to make sure all the tweets collected are within Malaysia.

This research, conducted using Malaysian tweets (now the X platform) collected from public accounts between September 10, 2019, and November 19, 2019, along with the analysis and findings, was completed by late 2020. The data privacy policy has anonymised user accounts; therefore, the username and user ID columns are excluded from the dataset. The dataset consists of tweets from Malaysian users

across various backgrounds, including students, working adults, public figures, and individuals who spend a significant amount of time socialising on Twitter. The dataset comprised a total of 2,129 tweets, organised into eight columns. These columns include the tweet text, user status count, number of user followers, user location, user verification status, favourite count, retweet count, and tweet date. Each column captures specific attributes relevant to the analysis, as outlined briefly below:

i. Tweets: A message on Twitter that allows a user to post 280 characters or less (except the link is counted as 23 characters no matter how long).
ii. User status count: Represents the total number of tweets posted by the user.
iii. User followers: The number of users' followers.
iv. User location: The location of the user.
v. User verification: Indicates the presence of the blue verified badge on Twitter, which signifies whether the account is authenticated as a genuine profile of public interest.
vi. Favorite count: The number of favorites or likes for the tweets.
vii. Retweet count: The total number of retweets.
viii. Tweet date: The user's tweet's time and date stamp.

Annotation

In machine learning, data labelling, also known as data annotation, is an essential step. It trains the model to learn from labelled data intelligently, allowing it to predict patterns of narcissistic traits in unlabeled data. Several techniques of data labelling are applied in machine learning training. This covers semi-supervised labelling and manual labelling. Based on the question "Does the user of this tweet exhibit tendencies of narcissism?", this study classified tweets into two groups: "yes" (narcissist) or "no" (not narcissist)". A counselling psychologist from the Department of Educational Psychology and Counseling at Universiti Malaya was consulted to provide an answer to this question and guarantee the reliability of the analysis. Her knowledge provided a comprehensive understanding of narcissistic traits and assisted in annotating the tweets. The American Psychiatric Association (APA, 2013) was also used to identify important traits associated with narcissism. Demeaning behaviour towards others, a lack of empathy, an excessive need for admiration, and a sense of entitlement were some of these characteristics. These criteria were a helpful guide for annotating the 2,129 rows, or tuples, in this study dataset. Nine hundred forty tweets were classified as potentially exhibiting narcissistic tendencies, while 1,189 tweets were labelled as not displaying such traits.

Text Pre-processing

As shown in Figure 1, the text pre-processing stage involves several steps that transform the raw dataset for use before classification. These steps include removing retweets, dropping empty rows, converting text to lowercase for consistency, cleaning the text (i.e., removing usernames, punctuation, URLs, symbols, special characters, numbers, and extra spaces), and expanding abbreviations, slang, and short forms (e.g., "you're," "weh," "power," "u").

A dictionary and re (regex) functions were used for English words to expand abbreviations and correct spelling. At the same time, the Malaya normaliser (a Bahasa Melayu NLP toolkit) was employed to correct Bahasa Melayu words. Additionally, due to mixed Bahasa Melayu and English in the tweets, English words were translated using the Python translate library.

The pre-processing stage also includes tokenising the text into individual words and applying stemming to convert words to their base or root forms (e.g., transforming "finds" to "find"). When working with Bahasa Melayu text, different forms of a word may have similar meanings (e.g., "tersangat," "sangatlah"). The Malaya deep learning stemming model helps reduce such words to their root form, replacing "tersangat" and "sangatlah" with "sangat."
Another critical step in text pre-processing involves eliminating stop words, which are frequently occurring words like "the," "a," "an," and "above" that typically contribute little to the overall meaning. The goal of removing stop words is to concentrate on the more significant terms within the text. Instead of using default stop words, this research utilised a customised set of stop words in Bahasa Melayu

(e.g., "ke," "yang," "sebab") to avoid removing words relevant to narcissism. After pre-processing, the dataset was reduced to 2,091 rows (930 labelled as "yes" and 1,161 labelled as "no").

Feature Extraction

Before building the machine learning classifiers, the dataset consisting of text must be converted into a numerical format. To achieve this, scikit-learn functions such as CountVectorizer and TfidfTransformer were used to measure word frequencies. CountVectorizer converts the text into a token or count matrix, which is then transformed into a normalised TF-IDF representation, assigning weights to each word.

Classification

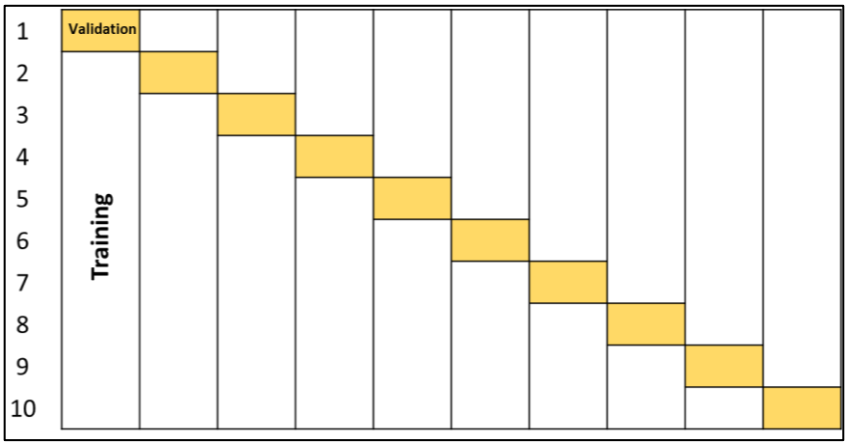
Four well-known machine learning classifiers—Naïve Bayes, Support Vector Machines (SVM), Logistic Regression, and Gradient Boosting—that are commonly employed in sentiment analysis are compared in this study. The performance of these classifier models was evaluated using cross-validation and 10-fold cross-validation techniques to identify the optimal model.

Using the train_test_split function to divide the dataset into training and test sets, cross-validation mitigates the bias-variance trade-off and offers a reliable method for evaluating predictive models. This approach ensures effective training with minimal computational overhead, given the dataset size in this study.

Each classifier is iteratively trained using 90% of the training set for 10-fold cross-validation, and the remaining 10% is tested over 10 iterations. Following the last iteration, the mean accuracy of all models is calculated to assess overall performance. The 10-fold cross-validation method employed in this study is depicted in Figure 2.

Figure 2

Illustration of the 10-fold Validation Approach Used in this Study

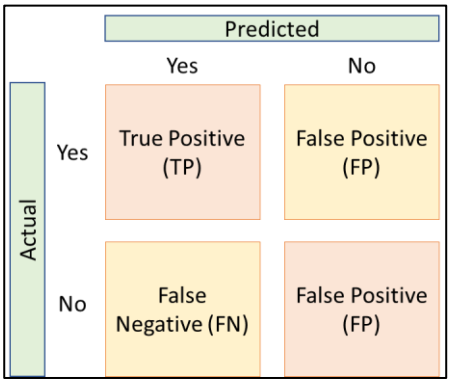


Parameter Performance

Metrics like accuracy, precision, F1-score, and true positive rate (recall) are used in this study to assess the performance of the machine learning classifier. By using established formulas, these measures are derived from the confusion matrix. The confusion matrix is a frequently used tool demonstrating a classifier's true performance. Figure 3 displays a confusion matrix diagram.

Figure 3

Diagram of the Confusion Matrix



The accuracy metric of a machine learning classifier calculates the proportion of correct predictions for every observation. Furthermore, the true positive rate (TPR), sometimes referred to as recall or sensitivity, is computed using a specific formula. The ratio of correctly identified positive cases to all positive cases is known as the TPR.

Precision is defined as he proportion of correctly predicted positive instances to all instances classified as positive. Lastly, the F1-score, sometimes called the F-measure, is the harmonic mean of precision and recall. A higher F1-score indicates better classifier performance. Table 2 displays the equations utilised in this investigation for these assessment metrics.

Table 2

Equation Formula for the Evaluation Metric

Evaluation Metric	Equation
Accuracy	$\frac{TP+TN}{(TP+TN+FP+FN)}$
True Positive Rate	$\frac{TP}{(TP+FN)}$
Precision	$\frac{TP}{(TP+FP)}$
F1-score	$2*\frac{1}{\frac{1}{Precision}+\frac{1}{Recall}}$

Analysis of Results

This section outlines and analyses the performance of classifiers in identifying potential narcissistic tendencies among Malaysians using text data from Twitter. The assessment was carried out using both cross-validation and 10-fold cross-validation techniques.

During cross-validation, the accuracy scores for Naïve Bayes, Support Vector Machine (SVM), Logistic Regression, and Gradient Boosting were 64%, 66%, 57%, and 57%, respectively. The true positive rate (recall) values closely aligned with the accuracy scores. Precision scores were 65% for Naïve Bayes, 66% for SVM, and 57% for both Logistic Regression and Gradient Boosting. The F1-

scores were 62%, 65%, 57%, and 57%, respectively. A summary of the cross-validation results is provided in Table 3.

Table 3

Result of Evaluation Metric for Cross-validation

Algorithm	Accuracy	True Positive Rate	Precision	F1-Measure
Naïve Bayes	0.64	0.64	0.65	0.62
SVM	0.66	0.66	0.66	0.65
Logistic Regression	0.57	0.57	0.57	0.57
Gradient Boosting	0.57	0.57	0.57	0.57

The outcomes of the 10-fold cross-validation reveal a notable enhancement over standard cross-validation. Both the accuracy and the true positive rate for Naïve Bayes, Support Vector Machine (SVM), Logistic Regression, and Gradient Boosting were consistent, reaching 76%, 80%, 60%, and 74%, respectively. Table 4 displays each classifier's accuracy, true positive rate, precision, and F1-score under the 10-fold cross-validation method.

Table 4

Result of Evaluation Metric for Cross-validation

Algorithm	Accuracy	True Positive Rate	Precision	F1-Measure
Naïve Bayes	0.76	0.76	0.74	0.74
SVM	0.80	0.80	0.80	0.78
Logistic Regression	0.60	0.60	0.68	0.62
Gradient Boosting	0.74	0.74	0.70	0.71

The objective of 10-fold cross-validation is to improve the performance of classifier models. The findings reveal a significant enhancement compared to the standard cross-validation method. Support Vector Machine (SVM) achieved the highest accuracy at 80%. Moreover, the accuracy of Gradient Boosting rose significantly from 57% to 74%, highlighting its effectiveness with 10-fold cross-validation. In contrast, Logistic Regression exhibited only marginal improvement and recorded the lowest accuracy in both cross-validation approaches. Consequently, SVM is the most appropriate classifier for this research dataset.

Conclusion

Linguistic indicators of narcissistic tendencies were successfully used to predict narcissistic traits in Malaysians through social media content. TF-IDF feature extraction was combined with Naïve Bayes, SVM, Logistic Regression, and Gradient Boosting to classify the tweets. The effectiveness of each

classifier in detecting the propensity for narcissistic traits was assessed using both cross-validation and 10-fold cross-validation techniques.

The SVM demonstrated the best classifier performance in this investigation, achieving an accuracy of 80% in 10-fold cross-validation. Additionally, with 10-fold cross-validation, Gradient Boosting's accuracy rose sharply from 57% to 74%, demonstrating its efficacy. In contrast, Logistic Regression showed little improvement and received the lowest accuracy scores in 10-fold and cross-validation. All things considered, 10-fold cross-validation is a tried-and-true technique for assessing models on this research dataset since it efficiently uses the data at hand while preserving computational overhead and lowering the possibility of overfitting.

Additional words, phrases, or language frequently used by people with narcissistic traits can be added to the dataset to improve classification accuracy. To improve the dataset, more advice on narcissism from clinical psychologists and psychiatrists would be very helpful. Moreover, future research could explore the use of state-of-the-art (SOTA) word embedding methods, like Generative Pre-trained Transformers (GPT) and Bidirectional Encoder Representations from Transformers (BERT), which have demonstrated remarkable effectiveness in capturing contextual relationships in text and can achieve coherent paragraphs of text (Devlin et al., 2018; Radford et al., 2019). By better capturing the linguistic subtleties linked to narcissistic traits, these sophisticated models have been shown to improve classification accuracy.

Despite being a successful and efficient technique for locating important terms in textual data, TF-IDF has some significant drawbacks, such as sensitivity to document length, difficulty managing synonyms and polysemy, and a lack of semantic understanding. These limitations result from TF-IDF’s exclusive focus on term and document frequency, which ignores the semantic relationships between words. As a result, it may overemphasise words in longer documents or fail to recognise the nuanced meanings of words that vary depending on the context. To improve the efficacy of text analytics techniques, recent research avenues are still looking into and addressing these issues (Ushio et al., 2021).

Lemmatisation preserves the semantic integrity of words by distilling them to their most basic form according to context and part of speech (Analytics Vidhya, 2025). To accurately interpret the emotional and psychological subtleties that may indicate narcissism, lemmatisation ensures that words retain their meaning (e.g., distinguishing between "I," "me," and "my" in context). On the other hand, stemming can result in forms that are too generalised, which could change the meaning of words and make it more difficult to identify subtle linguistic elements. For example, stemming could make it more difficult for the model to discern between various forms of narcissism in the text by reducing "narcissist" and "narcissistic" to a single root. Based on these arguments, we suggest that both lemmatisation and stemming should be carefully examined for tasks where meaning and context are crucial, like identifying narcissistic traits in text.

According to Kacewicz et al. (2014), future research should look at how their findings affect the use of pronouns, which indicate people's places in social hierarchies in various social and organisational contexts. According to Holtzman et al. (2019), certain linguistic characteristics can serve as indicators of narcissistic traits in both spoken and written communication. According to Mariconti et al. (2019), automated systems can reduce the negative impact of harmful behaviours on social media and moderate online content. In addition, according to Casale and Banchi (2020), more long-term research is required to assess how their findings, i.e., trends, evolve and to look into the psychological mechanisms and causal relationships underlying the link between narcissism and PSMU.

Our study contributes significantly to computational psychology and social media analytics by offering insights into the psychological dimensions of digital self-presentation. The results advance our understanding of the manifestation of narcissistic behaviours within online communities and emphasise

their behavioural implications for examining the engagement of individuals with narcissistic traits with social media users on online platforms.

Acknowledgement

The authors wish to acknowledge the support given by Universiti Malaya, Malaysia, for the facilities provided to complete this research.

Conflict of Interest

No conflicts of interest were disclosed.

Author Contribution Statement

SSAR: Conceptualisation, Research Supervision, Writing of the Initial Draft. AANR: Dataset Acquisition, Text Pre-processing and Annotation, Analysis of Results and Findings. RAMJ: Dataset Annotation

Funding

No funding was received for this research.

Ethics Statement

This research did not require IRB approval because it used publicly available secondary data (self-scraped from Twitter, now X) with no identifying personal information.

Data Access Statement

Research data supporting this publication are available upon request to the corresponding author.

Author Biography

Dr. Siti Soraya Abdul Rahman holds a PhD in Cognitive Science from the University of Sussex, UK (2012). She is a Senior Lecturer in the Department of Artificial Intelligence at Universiti Malaya. Her expertise includes Machine Learning, Large Language Models, and AI applications. She also researches Cognitive Science, AI in Education, and Data Science.

Amira An-Nur Rusli holds a Master of Data Science from the Universiti Malaya. Her master's dissertation focused on the application of linguistic indicators to predict narcissistic traits in Malaysians through the analysis of social media content.

Dr. Rafidah Aga Mohd Jaladin holds a PhD from Monash University, Australia. She is a counselling psychologist at Universiti Malaya's Faculty of Education. With over 20 years of experience, she specialises in multicultural counselling, counsellor education, and mental health literacy.

References

Al-Garadi, M. A., Varathan, K. D., & Ravana, S. D. (2016). Cybercrime detection in online communications: The experimental case of cyberbullying detection in the Twitter network. *Computers in Human Behavior*, *63*, 433–443. https://doi.org/10.1016/j.chb.2016.05.051

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (5th ed.). https://doi.org/10.1176/appi.books.9780890425596

Analytics Vidhya. (2025, May 1). Stemming vs. lemmatization in NLP: Must-know differences. https://www.analyticsvidhya.com/blog/2022/06/stemming-vs-lemmatization-in-nlp-must-know-differences/

Anandarajan, M., Hill, C., & Nolan, T. (2019). Introduction to Text Analytics. In *Practical Text Analytics, Advances in Analytics and Data Science* (Vol. 2, pp. 1–10). Switzerland: Springer Nature Switzerland AG. https://doi.org/10.1007/978-3-319-95663-3

Anggarawati, S., Armelly, A., Salim, M., Odilawati, W., Saputra, F. E., & Atmaja, F. T. (2023). The mediating role of narcissistic behavior in the relationship between materialistic orientation and conspicuous online consumption behavior on social media. Cogent Business & Management, 10(3), DOI: 10.1080/23311975.2023.2285768

Brindha, S., Sukumaran, D. S., & Prabha, D. K. (2016). A SURVEY ON CLASSIFICATION TECHNIQUES FOR TEXT MINING. In *3rd International Conference on Advanced Computing and Communication Systems (ICACCS -2016)* (pp. 1–5). Coimbatore, India

Burnap, P., Scourfield, J., & Colombo, G. (2015). Machine Classification and Analysis of Suicide-Related Communication on Twitter. *ACM*, 75–84. https://doi.org/10.1145/2700171.2791023

Carey, A. L., Brucks, M.S., Küfner, A.C., Holtzman, N., große Deters, F., Back, M.D., Donnellan, M., Pennebaker, J., & Mehl, M.R. (2015). Narcissism and the use of personal pronouns revisited. Journal of personality and social psychology, 109 3, e1-e15.

Casale, S., & Banchi, V. (2020). Narcissism and problematic social media use: A systematic literature review. Addictive Behaviors Reports, 11, 100252. https://doi.org/10.1016/j.abrep.2020.100252

Czarna, A. Z., & Zajenkowski, M. (2018). How Does It Feel to Be a Narcissist? Narcissism and Emotions. In *Handbook of Trait Narcissism* (pp. 255–263). Springer International Publishing AG. https://doi.org/10.1007/978-3-319-92171-6

Davenport, S. W., Bergman, S. M., Bergman, J. Z., & Fearrington, M. E. (2014). Computers in Human Behavior Twitter versus Facebook: Exploring the role of narcissism in the motives and usage of different social media platforms. *Computers in Human Behavior*, *32*, 212–220. https://doi.org/10.1016/j.chb.2013.12.011

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

DeWall, C. N., Buffardi, L. E., Bonser, I., & Campbell, W. K. (2011). Narcissism and implicit attention seeking: Evidence from linguistic analyses of social networking and online presentation. *Personality and Individual Differences*, *51*, 57–62. https://doi.org/10.1016/j.paid.2011.03.011

DoctorRamani. (2019, April 22). What's gaslighting? (Individual, tribe, and societal gaslighting) [Video]. YouTube. https://youtu.be/UTS5XsZe9Jg

DoctorRamani. (2019, August 9). What does it mean when a narcissist tells you you're being too sensitive or too dramatic? [Video]. YouTube. https://youtu.be/IBV_RXRShcE

Hart, W., Adams, J., Burton, K. A., & Tortoriello, G. K. (2017). Narcissism and self-presentation: Profiling grandiose and vulnerable Narcissists' self-presentation tactic use. *Personality and Individual Differences*, *104*, 48–57. https://doi.org/10.1016/j.paid.2016.06.062

Holtzman, N. S., & Strube, M. J. (2013). People with dark personalities tend to create a physically attractive veneer. Social Psychological and Personality Science, 4(4), 461–467. https://doi.org/10.1177/1948550612461284

Holtzman, N. S., Tackman, A. M., Carey, A. L., Brucks, M. S., Küfner, A. C. P., Deters, F. G., Back, M. D., Donnellan, M. B., Pennebaker, J. W., & Mehl, M. R. (2019). Linguistic markers of grandiose narcissism: A LIWC analysis of 15 samples. Journal of Language and Social Psychology, 38(5-6), 773–786. https://doi.org/10.1177/0261927X19871092

Holtzman, N. S., Vazire, S., & Mehl, M. R. (2010). Sounds like a narcissist: Behavioral manifestations of narcissism in everyday life. Journal of Research in Personality, 44(4), 478–484. https://doi.org/10.1016/j.jrp.2010.06.001

Ismail, I. R. (2018). Dark Triad Personality Traits: Evaluation of Self versus Others among Employees in Malaysia. In *Proceeding of the 5th International Conference on Management and Muamalah* (pp. 105–115). UKM Graduate School of Business.

Kacewicz, E., Pennebaker, J. W., Davis, M., Jeon, M., & Graesser, A. C. (2014). "Pronoun use reflects standings in social hierarchies." Journal of Language and Social Psychology, 33(2), 125–143. https://doi.org/10.1177/0261927X13502654

Ksinan, A. J., & Vazsonyi, A. T. (2016). Narcissism, Internet, and social relations: A study of two tales. *Personality and Individual Differences*, *94*, 118–123. https://doi.org/10.1016/j.paid.2016.01.016

Macenczak, L. A., Campbell, S., Henley, A. B., & Campbell, W. K. (2016). Direct and interactive effects of narcissism and power on overconfidence. *PAID*, *91*, 113–122. https://doi.org/10.1016/j.paid.2015.11.053

Mariconti, E., Suarez-Tangil, G., Blackburn, J., De Cristofaro, E., Kourtellis, N., Leontiadis, I., Luque Serrano, J., & Stringhini, G. (2019). "You know what to do": Proactive detection of YouTube videos targeted by coordinated hate attacks. *Proceedings of the 22nd ACM Conference on Computer-Supported Cooperative Work and Social Computing (CSCW '19)*, 205–217. https://doi.org/10.1145/3359246

McGregor, S. A. (2010). The Analysis of Personality through Language: Narcissism Predicts Use of Shame-Related Words in Narratives. University of Michigan.

Milosevic, N., Dehghantanha, A., & Choo, K. R. (2017). Machine learning aided Android malware classification. *Computers and Electrical Engineering*, 1–9. https://doi.org/10.1016/j.compeleceng.2017.02.013

Narcissistic Personality Disorder Facts. (2018). Retrieved May 13, 2019, from https://barendspsychology.com/narcissism/

Pennebaker, J. W. (n.d.). Linguistic inquiry and word count (LIWC). Retrieved from https://www.liwc.app/

Pratama, B. Y., & Sarno, R. (2016). Personality classification based on Twitter text using Naive Bayes, KNN and SVM. In *Proceedings of 2015 International Conference on Data and Software Engineering, ICODSE 2015* (pp. 170–174). https://doi.org/10.1109/ICODSE.2015.7436992

Qiu, L., Lu, J., Yang, S., Qu, W., & Zhu, T. (2015). What does your selfie say about you?. *Computers in Human Behavior*, *52*, 443-449.

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language Models are Unsupervised Multitask Learners.

Rathner, E.-M., Djamali, J., Terhorst, Y., Schuller, B., Cummins, N., Salamon, G., Hunger-Schoppe, C., & Baumeister, H. (2018). How did you like 2017? Detection of language markers of depression and narcissism in personal narratives. In Proceedings of Interspeech 2018 (pp. 3388–3392). ISCA. https://doi.org/10.21437/Interspeech.2018-2040

Semanjski, I., & Gautama, S. (2015). Smart City Mobility Application—Gradient Boosting Trees for Mobility Prediction and Analysis Based on Crowdsourced Data. *Sensors*, *15*, 15974–15987. https://doi.org/10.3390/s150715974

Social media and narcissism. (n.d.). https://doi.org/10.1017/CBO9781107415324.004

Staff, M. C. (2017). Narcissistic Personality Disorder. Retrieved May 22, 2019, from https://www.mayoclinic.org/diseases-conditions/narcissistic-personality-disorder/symptoms-causes/syc-20366662

Sumner, C., Byers, A., Boochever, R., & Park, G. J. (2012). Predicting Dark Triad Personality Traits from Twitter usage and a linguistic analysis of Tweets, 386–393. https://doi.org/10.1109/ICMLA.2012.218

Thorpe, J. (2016). The 2 Main Types Of Narcissism - And How To Spot The Difference. Retrieved May 20, 2019, from https://www.bustle.com/articles/198007-the-2-main-types-of-narcissism-and-how-to-spot-the-difference

Ushio, A., Liberatore, F., & Camacho-Collados, J. (2021). Back to the basics: A quantitative analysis of statistical and graph-based term weighting schemes for keyword extraction. arXiv. https://arxiv.org/abs/2104.08028

Zhang, Y., & Haghani, A. (2015). A gradient boosting method to improve travel time prediction. *Transportation Research Part C Journal*, 2015. https://doi.org/10.1016/j.trc.2015.02.019