

Big five personality prediction based in Indonesian tweets using machine learning methods

Warih Maharani¹, Veronikha Effendy²

¹Data Science Department, School of Computing, Telkom University, Bandung, Indonesia

²Informatics Department, School of Computing, Telkom University, Bandung, Indonesia

Article Info

Article history:

Received Apr 3, 2021

Revised Sep 17, 2021

Accepted Oct 10, 2021

Keywords:

Big Five model

K-nearest neighbors

Naïve Bayes

Personality

Support vector machine

ABSTRACT

The popularity of social media has drawn the attention of researchers who have conducted cross-disciplinary studies examining the relationship between personality traits and behavior on social media. Most current work focuses on personality prediction analysis of English texts, but Indonesian has received scant attention. Therefore, this research aims to predict user's personalities based on Indonesian text from social media using machine learning techniques. This paper evaluates several machine learning techniques, including naive Bayes (NB), K-nearest neighbors (KNN), and support vector machine (SVM), based on semantic features including emotion, sentiment, and publicly available Twitter profile. We predict the personality based on the Big Five personality model, the most appropriate model for predicting user personality in social media. We examine the relationships between the semantic features and the Big Five personality dimensions. The experimental results indicate that the Big Five personality exhibit distinct emotional, sentimental, and social characteristics and that SVM outperformed NB and KNN for Indonesian. In addition, we observe several terms in Indonesian that specifically refer to each personality type, each of which has distinct emotional, sentimental, and social features.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Warih Maharani

School of Computing, Telkom University

Telekomunikasi Street, No.1, Bandung, Indonesia

Email: wmaharani@telkomuniversity.ac.id

1. INTRODUCTION

Nowadays, there is advanced progress in applying computing technologies [1]–[8]. Researchers are interested in approaches to shortening different issues. With the rapid growth of social media, approaches to solving psychological research problems, such as personality prediction and analysis of social behavior, are continually improved. Personality can be defined as patterns of behavior, manners, thinking, motives, and emotions that provide character to individuals all the time and in various situations. Identifying personality type is not a simple task, given that each person possesses a unique set of psychological characteristics. With hundreds of millions of users on social media sharing their content, social media presents enormous personality modeling opportunities.

A major issue in conventional personality assessments involving self-reported inventory costs a long time and many human resources. The recent focus of researchers on the development of automatic personality recognition systems demonstrates the critical nature of personality recognition in social networks. Generally, these applications have been based on the central philosophy of several well-known personality models. Many models have been used to characterize personality traits, but the Big Five model is the most extensively studied and widely accepted to describe personality traits [9]–[12]. The Big Five model consists

of openness (O), conscientiousness (C), extraversion (E), agreeableness (A), and neuroticism (N)-OCEAN. In contrast to conventional personality tests, there is no requirement for formal questionnaires. It takes time for businesses to collect sufficient user data to understand their personalities fully. Previous work has shown the relation between users' profiles in social media and their actual personalities [13], [14]. Zidan *et al.* [6] stated that the difficulty of identifying psychological types varies, i.e., some of them are easier to recognize than other types. Golbeck *et al.* [15] predicted users' personality traits using Facebook and Twitter datasets. They demonstrated that users do not want to make Facebook profiles that show only their best but rather want Facebook profiles that reflect their real lives. Unlike Facebook, users must reveal personal information like name and age to post anonymously; users on Twitter are free to project everything they think. Personality traits extracted from Twitter users' tweets are assumed to be accurate because everybody is on Twitter, so they do not worry about what words to use. Quercia *et al.* [16] were the first to investigate the relationship between personality and Twitter use in general; they also proposed a model for estimating users' personalities based on their followers, followers, and count numbers. The earliest attempts at personality prediction relied heavily on machine learning techniques such as support vector machine (SVM), which exploited syntactic and lexical features [17], [18]. Tausczik *et al.* [19] stated that daily words include thought patterns, social interests, and mood characteristics. Linguistic cues allow for a prediction of approximate mental well-being. Some variables cannot be standardized or measured because of cultural differences, genders, ages, and other variables. However, researchers have claimed that social networks are simple and easy to work with for both extroverted [16].

Numerous previous studies have successfully predicted the personality of social media users by analyzing social features in English tweets [20]–[23]. Majumder *et al.* [24], [25] implemented emotional detection features and sentiment analysis to predict personality. However, these features are only available in English Tweets, and research into the Indonesian language is still in its early stages. Pratama incorporated term frequency-inverse document frequency (TF-IDF) into machine learning modeling as a feature to predict personality based on Indonesian tweets. However, the TF-IDF feature is only effective at the lexical level and does not capture semantics [22]. Therefore, we attempt to address the aforementioned problems by incorporating semantics including emotion, sentiment, and social features into the predictive modeling of user personality on Twitter. In addition, we are interested in finding links between the different characteristics of the Indonesian user profiles and the type of personality.

The main contribution of this paper consists of two distinct components first, personality prediction based on semantic features, including emotion, sentiment, and social features. We evaluate and compare several machine learning techniques on Indonesia tweets, including naive Bayes (NB), K-nearest neighbors (KNN), and SVM. The second is a detailed discussion of personality prediction in Indonesia and its associated models based on these features. The rest of this paper is organized as follows. Section 2 describes the research method, section 3 discusses the result, and section 4 provides the conclusion and future works.

2. RESEARCH METHOD

This paper proposes a personality classification based on emotion, sentiment, and social features to capture the semantics in Indonesian tweets. We classify the personality of a Twitter user into Big Five personality classes (OCEAN) using NB, KNN, and SVM. Tweets weighted by the number of times the word appears in the document. Based on the weighted results, the emotions of each word is detected by using the NRC emotion lexicon, which categorizes words into eight categories, namely anger, anticipation, disgust, fear, joy, sadness, surprise, and trust [26]. Likewise, the analysis of sentiment using lexicon sentiment categorizes both positive and negative polarity. Finally, we retrieved social features based on statistical data, including following, follower, retweet, mentions, replies, and favorites from each account. Figure 1 shows the overall architecture of our personality prediction system that we propose in this study.

2.1. Data collection

The paper used the Twitter's application programming interface (API) to collect tweets from 800 Twitter users who have completed a personality test questionnaire consisting of 44 questions based on the Big Five inventory (BFI) [9]. We collected information on their profiles and posts for each user, including the number of followers, number of followings, number of mentions, replies, hashtags, favorites, and several links. Previous research has demonstrated that linguistic features can determine personality traits [13], [14], [25], [26].

2.2. Preprocessing data

Preprocess is the critical and first stage in the process of sentiment analysis and personality prediction. It converts raw data into an analyzable format. This process is fundamentally based on cleaning

and transforming required data. We implemented several pre-processing stages in this research, including tokenization, stopwords removal, punctuation removal, case folding, URL links removal, hashtag removal, and user mentions (tokens started with “@”) removal. The example can be shown in Table 1.

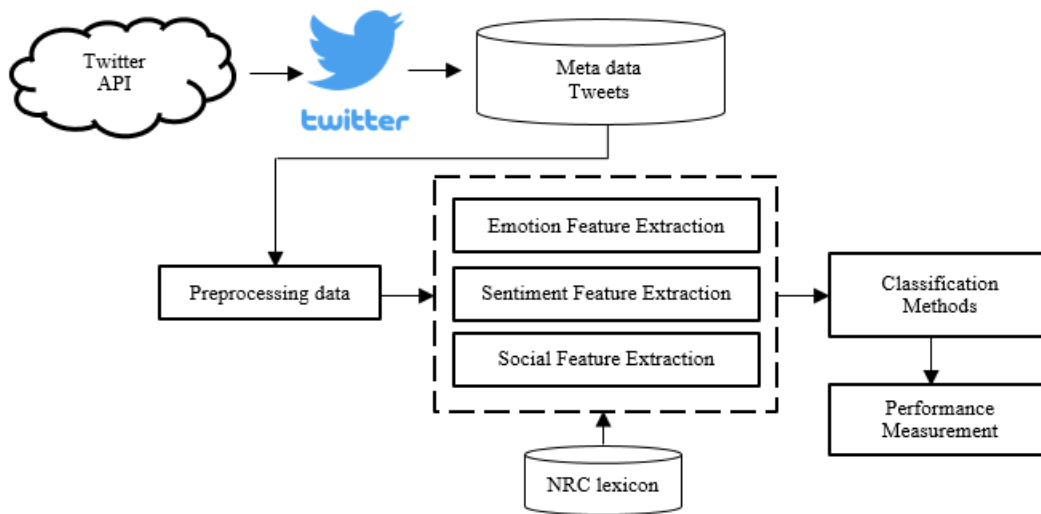


Figure 1. The architecture of personality prediction system

Table 1. Example of preprocessing data

Tweet	Tokenization Result	Filtering Result
@A_ID Covid ini bikin gue stress. ada yang sakit tapi teteeeeeppp jalan, tapi senang juga karena bisa rebahan terussss kesenangan diatas kesedihan :D.	'@', 'A_ID', 'Covid', 'ini', 'bikin', 'gue', 'stres', 'ada', 'yang', 'sakit', 'tapi', 'teteeeeppp', 'jalan', 'tapi', 'senang', 'juga', 'karena', 'bisa', 'rebahan', 'terussss', 'kesenangan', 'diatas', 'kesedihan', ':D', '.'	'Covid', 'ini', 'bikin', 'gue', 'stres', 'ada', 'yang', 'sakit', 'tapi', 'tetep', 'jalan', 'tapi', 'senang', 'juga', 'karena', 'bisa', 'rebahan', 'terus', 'kesenangan', 'diatas', 'kesedihan'

2.3. Feature extraction

We extract three types of semantic features: emotional, sentimental, and social. Emotion features will recognize a word and classify it into one of eight emotion categories based on the NRC emotion lexicon: anger, anticipation, disgust, fear, joy, sadness, surprise, and trust [27]. The sentiment feature does the same thing, but it divides it into positive and negative sentiment polarity. Finally, the social features take into account each user account's social behavior data.

2.3.1. Emotion feature

Farnadi *et al.* [27] stated that at least one emotion could be derived from a tweet. In addition, they showed that the emotions and sentimental expressions of each person's traits are different. This study inspired us to produce an expressive NRC emotion lexicon, which contains words translated from English into many languages, including Indonesian [19]. The purpose of this analysis is to provide additional information on the frequency of emotions calculated for all tweets of a particular user. The rationale for including emotional characteristics is that individuals with varying personality traits will express themselves differently and employ various words (phrases) and emotions. Previous research has also found a link between emotions and personality traits [23]. Although the annotations of emotions and feelings in this lexicon were made in English and then translated into Indonesian, Mohammad *et al.* [26] stated that the majority of affective norms are stable across languages, so we also expect the quality of results for all languages to remain relatively similar to English. Table 2 shows the example of emotion features.

Table 2. Emotion features

Term	Anger	Anticipation	Disgust	Fear	Joy	Sadness	Surprise	Trust
sakit	1	0	1	1	0	1	0	0
senang	0	1	0	0	1	0	0	1
stres	1	1	0	1	0	1	0	0

2.3.2. Sentiment feature and social features

In addition to eight categories of emotions, there are also two categories of sentiment, positive and negative, in the NRC Word-Emotion Association Lexicon. So, in addition to detecting emotions, we also detect polarity. Based on Table 2, we categorized “*senang*” as positive polarity and “*sakit*”, “*stress*” as negative polarity. According to Quercia *et al.* [16], the personality prediction can be made by knowing the publicly available number of followers, the following number, and the correlation values between them and the Big Five personality traits. Attributes for social features, including the number of followers, followers, retweets, and favorites for each user, can be found in Table 3.

Table 3. Social features

Username	Following	Follower	Retweet	Mention	Replies	Favorite
X	375	10430	12	15	23	43
Y	893	3279	2	5	4	12
Z	268	56	1	2	2	4

2.4. Classification methods

The first step is data collection, which involves gathering tweets from several users who have agreed to complete the respondent form. Simultaneously, they were asked to complete the BFI44 questionnaire to obtain their gold standard personality label. The next stage is data preprocessing, which includes tokenization, stop-word removal, and filtering. To measure the performance of the classification model, we use a confusion matrix and mean square error (MAE) to calculate the performance of personality prediction.

The naive Bayes method, KNN, and SVM classify the data into five personality types (OCEAN). The algorithm of MNB can be shown in Figure 2(a), where D is a document, C is the set of classes $C = \{c_1, c_2, \dots, c_j\}$, and V denotes the collection of all words (w) that occurs in the training corpus. Figure 2(b) shows the KNN algorithm where ϕ is set $\{c_1, c_2, \dots, c_j\}$ of all classes, D is Set $\{<d_1, c_1>, \dots, <d_N, c_N>\}$ of all labeled documents, $S_k(d)$ is the set of d 's K -nearest neighbors, and p_j is an estimate for $P(c_j|S_k) = P(c_j|d)$; c_j denotes the set of all documents in the class c_j . The Naïve Bayes uses a multinomial distribution that estimates the number of distinct words that have occurred in each sequence as a function. KNN is a classification algorithm that uses the distance between training data and the number of closest neighbors to determine the classification results [8]. Cosine similarity is a function widely used in the classification of documents to determine the similarity between documents. Close distance shows similarities between the two documents in such a way that they have the same category. SVM is a supervised learning method that analyzes data and recognizes patterns. The SVM model represents data as space dots, mapped into categories separated by hyperplane/dividing lines [28].

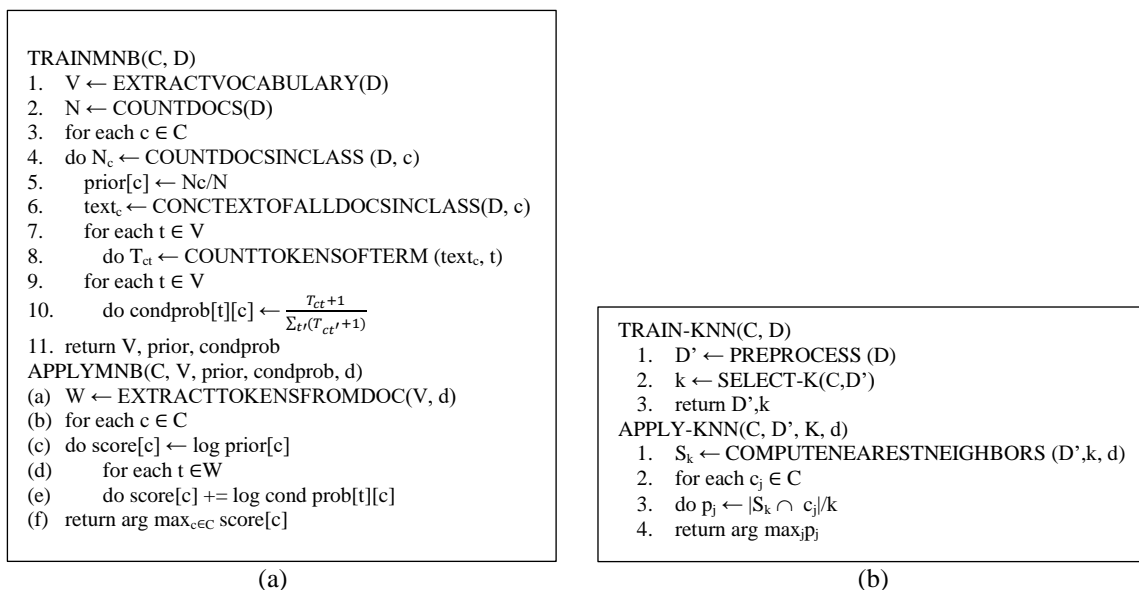


Figure 2. The algorithms (a) multinomial naive Bayes (MNB) algorithm and (b) KNN [28]

3. RESULTS AND DISCUSSION

Data was collected from 800 Twitter users as respondents, where 200 tweets will be collected from each respondent to have a total of 160,000 tweets. These respondents have completed the Big Five personality test by the provisions of the BFI, which will become the gold standard. Each respondent will have a value for each type of personality. The most dominant value will be the type of personality of the respondent. In order to predict the personality trait score, we performed three classification methods in Weka. Table 4 shows the MAE of our classification methods.

Table 4. MAE of personality classification methods

Methods	O	C	E	A	N
Naïve Bayes	0.113	0.148	0.177	0.131	0.192
KNN	0.116	0.140	0.160	0.128	0.172
SVM	0.101	0.122	0.144	0.111	0.148

We discovered that openness was the most uncomplicated trait to quantify and neuroticism was the most challenging. Based on the performance of classification methods, SVM was able to predict all personality traits within 59.45%. We believe that larger sample size and a variety of features will produce much better results. We used 10 cross-fold validation in the NB determined by the occurrence of TF-IDF, emotion, sentiment, and social features. In the SVM, we used radial base function (RBF) kernels with $C=1$, $\gamma=0.1$, max iteration=100, and degree=1. Moreover, for the KNN, we used the Euclidian distance with a $k=10$. Table 5 shows the detail of the result.

To determine the correlation of each feature category to the Big Five personality, we used the Pearson correlation coefficient. Table 6 shows the Pearson correlation values between the features and personality scores. Significant correlations are shown in bold for $p<0.05$. Some interesting correlations between the features and personality traits were discovered through the study.

Table 5. The performance of personality classification methods

Methods	Accuracy	Precision	Recall	F-Measure
Naïve Bayes	45.92%	0.55	0.72	0.62
KNN	48.02%	0.58	0.68	0.63
SVM	59.45%	0.72	0.88	0.79

Table 6. Pearson Correlation values between feature and personality scores

Features	Personality				
	O	C	E	A	N
Sentiment feature					
Positive	0,015	0,051	0,071	0,045	-0,072
Negative	0,055	-0,008	0,043	-0,045	0,010
Emotion feature					
Anger	0,027	0,016	0,066	-0,056	0,035
Anticipation	-0,004	0,105	0,071	0,087	-0,069
Disgust	0,026	-0,024	0,056	-0,080	0,048
Fear	0,035	0,003	0,055	-0,044	-0,023
Joy	-0,030	0,062	0,075	0,079	-0,059
Sadness	0,067	0,003	0,013	-0,041	0,040
Surprise	-0,096	0,021	0,067	0,058	-0,070
Trust	-0,068	0,076	0,110	0,045	-0,144
Social feature					
Following	0,010	-0,009	0,113	-0,017	-0,036
Follower	-0,064	0,057	0,073	-0,032	-0,102
Retweet	0,044	-0,009	-0,078	-0,016	-0,025
Mentions	0,020	-0,010	-0,014	0,020	-0,021
Replies	0,071	-0,017	-0,044	0,019	-0,015
Favorite	0,107	-0,026	-0,071	-0,065	-0,011

3.1. Relationship between emotion feature and personality traits

Based on our findings, emotional features are correlated with all personality types. Table 7 shows that openness personality (O) has high correlations with emotional features: sadness, fear, anger, and disgust. Tweets by openness users convey emotions more frequently than posts by other personalities, whereas neurotic users are less emotional. Extraversion users post the most emotional tweets. Surprisingly,

agreeableness users on Twitter express emotions that are very similar to conscientious users. Conscientious, extraverted, or agreeable users are expressing anticipation, joy, surprise, and positive emotions. The more open and neurotic users express less happiness than other individuals, and their posts tend to be more sad, disgusted, scared, fearful, and negative emotions. Sadness is expressed more than other emotions by neurotic and openness users. Whereas extraversion, conscientiousness, and agreeableness users express the most joy.

Table 7. Relationship between Indonesian emotions and all personality traits

Personality-Emotion Features	Terms
C, E, A Anticipation, joy, surprise, trust	<i>bebas, cinta, doa, hiburan, lengkap, manis, menang, merdeka, mulia, optimis, pesta, semangat, sempurna, suka cita, uang, tertawa, selamat, puji, cakap, awet, aspirasi, penuh harap, rangkul, aksi, moral, seru, untung, wajib, suci</i>
O, N Sadness, disgust, anger, fear	<i>aib, akibat, ancam, aniaya, antisosial, asusila, bahaya, bangkrut, bencana, benci, binasa, bodoh, bunuh, cabut, celaka, gila, jahat, mati, muak, pecat, parah, pelit, cemburu, cacat, cekik, cela, kolusi, mutilasi, malapetaka, mesum, dosa, sabotase, sekarat, selingkuh, suram, wabah, terorisme</i>

Additionally, openness users frequently tweet about their fear and anger. The relationship between Facebook status updates and the user's age, gender, and individuality was investigated by Farnadi *et al.* [27]. They conducted a study on Facebook status updates and age, gender, and users' individuality to examine the relationship between Facebook statuses and user demographics. They found that open users tend to be more emotional than users with neuroticism in their status posts. Extraversion is significantly correlated with emotional expression, but openness has a stronger relationship.

In Table 7, we have identified some relevant terms to Indonesian emotions and personality traits based on our dataset. Our results are similar to Sumner *et al.* [29], who examined the correlation between users' personalities and their use of Facebook, content, and emotions. Their result showed an affinity with words expressing negative emotions, anger, taboo, money, religion, and death. This new knowledge can be used to identify the personality types of Indonesian Twitter users. Although most tweets in the dataset represent anger and sadness based on our observations, some also contain much joy.

3.2. Relationship between sentiment feature and personality traits

Based on Table 8, we can conclude that conscientiousness, extraversion, and agreeableness users have a positive correlation to positive sentiment. In contrast, openness and neuroticism are related to negative sentiment. Table 8 shows some sentiment words that are frequently tweeted by all personality traits. Similarly, as in the previous sections, positive feelings are mostly expressed by conscientiousness, extraversion, and agreeableness users. In comparison, openness and neuroticism show no relation with positive feelings.

Table 8. Frequent sentiment words

Personality-Emotion Features	Terms
C, E, A Positive feelings	<i>asa, aspirasi, awet, bebas, berkat, cinta, doa, fajar, hiburan, megah, moral, mulia, optimis, penuh harap, piknik, rangkul, raya, suci, tertawa, akrab, andai,antisipasi, awas, bintang, cahaya, fokus, gembira, gentar, goyah, hadiah, hibah, ilham, karier, karunia, kilau, klimaks, kompetisi</i>
O, N Negative feelings	<i>adu, kotor, akibat, amuk, anarkis, angkuh, anonim, bahaya, bakar, bandel, banting, benci, biadab, binasa, blokade, bohong, boikot, bual, bubar, bunuh diri, cabut, dendam derita, dusta, jatuh, jelek, kejam, konyol, kumuh, provokatif</i>

3.3. Relationship between social feature and personality traits

As for social features, openness personality has the strongest positive correlation in the favorite category and the strongest negative correlation in the follower category. In contrast, conscientiousness has the strongest positive correlation in the follower category and the strongest negative correlation in the favorite category. Our finding is consistent with Golbeck *et al.* [18], with reported correlation coefficients having the same polarity as openness and neuroticism in terms of user characteristics. This finding is consistent with Farnadi *et al.* [27], who proposed that extraversion, agreeableness, and openness to new experiences are all associated with interpersonal selection. Based on the description of personality traits [23], imagination, creativity, curiosity, tolerance, and spontaneity are all associated with openness. Individuals with a high openness score enjoy change, are receptive to new and unusual ideas, and have a strong sense of aesthetics. There may be a high degree of openness, a desire to broaden and deepen one's range of ideas,

perspectives, and experiences. In general, a lack of openness may indicate a more conservative attitude. Extraversion users, by nature, have a large number of friends and engage in social interaction via likes, retweets, comments, and replies. Based on our results, extraversion users have a high number of followers and following. They place a high value on maintaining close personal connections, having good social networks, communicating with others, and engaging with them. This is supported by Sumner *et al.* [29], who describe extroverts as people who feel alone when they spend much time alone and therefore tend to spend the rest of their time with other people. This is also confirmed by Farnadi *et al.* [27], who discovered that extroverts engage in more social interaction than most people because they love it.

In contrast, agreeableness users participate in very little interaction through likes, retweets, and sharing. They have fewer followers and are followed by fewer individuals. Agreeableness, which has a small connection with social characteristics, has a weaker association with social features. Individuals with a high level of neuroticism are highly cautious when disclosing too many personal details. They are more likely to post material that elicits negative emotions. Individuals with lower Neuroticism values have a stronger sense of self-worth. They have less depressive behaviors than those with higher Neuroticism values. Due to their decreased sense of isolation and psychological distress, emotionally healthy people with lower Neuroticism beliefs are less likely to use social media at all. Additionally, we find a positive correlation between usage intensity and neuroticism. Individuals with low neuroticism values expend less time on social media, change their status less often, are members of fewer communities, and are less reliant on social media. They are more likely to retweet and more likely to reply to a tweet. Based on our observations, they are more likely to like and retweet posts that express anger and other negative emotions.

3.4. Theoretical and practical contributions

This work will have significant contributions to our society. Our research explores personality traits on Twitter in Indonesia. This experiment clearly illustrates the dynamics of the personal expression of Indonesian users. Numerous studies have identified a correlation between personality and language use. Our research has significant implications for practice, especially concerning Indonesian tweets. Understanding the relationship between microblogging and personality is extremely promising for evaluating personality without resorting to lengthy questionnaire surveys, as microblogging becomes more common and widely accessible. However, it should be noted that observer decisions are more closely correlated with linguistic cues than with self-reported personality. As a result, assessing an individual's personality solely based on their profile and linguistic clues in tweets is more likely to represent the personality as viewed by others than the actual personality of that person. Individuals can view a Twitter profile on another's page based on their familiarity with users, not their true personalities. Numerous studies have analyzed tweets and created personality profiles based on the user's emotional and behavioral characteristics. Although our research does not aim to develop a new algorithm for automatically detecting personality traits in Indonesian tweets, it does provide empirical evidence for the expression of personality in tweets. It demonstrates that personality can be predicted from tweets, especially in Indonesian. On the other hand, there is a reasonably strong correlation between term categories and personality traits.

3.5. Limitations and future directions

Currently, we are limited to exploring only emotions, sentiments, and social features by using the NRC Word-Emotion lexicon [23]. In addition to these features, we believe that it is critical to consider the context when determining personality traits. Additional semantic analysis, such as detecting writing styles, may be incorporated in future research better to understand the relationship between personality and linguistic characteristics. The majority of tweets are likely to be highly unstructured and noisy, with numerous typos and abbreviations. In this paper, we eliminate typos and abbreviations. Adding preprocessing to handle noisy data is one of the future development opportunities. Twitter has more interaction than websites or blogs. Emotion, sentiment, and social features may only play a minor role in personality traits. Accurate personality traits through social media may necessitate based on these features and additional behavioral cues such as interactions with other users and users' profiles. While this study focuses on the relationship between personality traits and these three features, future research may focus on personality expression using additional behavioral features. Additionally, the frequency of words, their connection, and their similarities and patterns of use can also vary over time. Since the number of participants in our experiment is equivalent to other studies of personality traits in social media, future research should involve many participants with a range of user profiles to validate our findings.

4. CONCLUSION

This paper predicts the personality of Indonesian Twitter users into a Big Five model consisting of five personality categories based on emotional, sentiment, and social features. We observe several terms in

Indonesian that specifically refer to each personality type, each of which has distinct emotional, sentimental, and social features. This research supports previous research, which has demonstrated the relationship between user characteristics and emotions. SVM performed better than naïve Bayes and KNN in personality classification. We have shown that the Big Five personality can be predicted using public information data and Indonesian tweets they share on Twitter. Due to the nature of this study, using Twitter has its unique problems. The future application of personality identification applications is a challenging issue. There are numerous opportunities for talent management with the ability to identify a user's personality traits.

ACKNOWLEDGEMENTS

The research was sponsored by Telkom University, Bandung, Indonesia.




REFERENCES

- [1] W. Liu, Q. Wu, J. Shen, J. Zhao, M. Zidan, and L. Tong, "An optimized quantum minimum searching algorithm with sure-success probability and its experiment simulation with Cirq," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 11, pp. 10425–10434, Nov. 2021, doi: 10.1007/s12652-020-02840-z.
- [2] M. Zidan, "A novel quantum computing model based on entanglement degree," *Modern Physics Letters B*, vol. 34, no. 35, Dec. 2020, doi: 10.1142/S0217984920504011.
- [3] M. Zidan, H. Eleuch, and M. Abdel-Aty, "Non-classical computing problems: toward novel type of quantum computing problems," *Results in Physics*, vol. 21, Feb. 2021, doi: 10.1016/j.rinp.2020.103536.
- [4] M. Zidan *et al.*, "A quantum algorithm based on entanglement measure for classifying boolean multivariate function into novel hidden classes," *Results in Physics*, vol. 15, Dec. 2019, doi: 10.1016/j.rinp.2019.102549.
- [5] A. Sagheer, M. Zidan, and M. M. Abdelsamea, "A novel autonomous perceptron model for pattern classification applications," *Entropy*, vol. 21, no. 8, Aug. 2019, doi: 10.3390/e21080763.
- [6] M. Zidan, A.-H. Abdel-Aty, A. El-Sadek, E. A. Zanaty, and M. Abdel-Aty, "Low-cost autonomous perceptron neural network inspired by quantum computation," in *AIP Conference Proceedings*, 2017, doi: 10.1063/1.5012145.
- [7] M. Zidan *et al.*, "Quantum classification algorithm based on competitive learning neural network and entanglement measure," *Applied Sciences*, vol. 9, no. 7, Mar. 2019, doi: 10.3390/app9071277.
- [8] A.-H. Abdel-Aty, H. Kadry, M. Zidan, Y. Al-Sbou, E. A. Zanaty, and M. Abdel-Aty, "A quantum classification algorithm for classification incomplete patterns based on entanglement measure," *Journal of Intelligent & Fuzzy Systems*, vol. 38, no. 3, pp. 2809–2816, Mar. 2020, doi: 10.3233/JIFS-179566.
- [9] T. Ryan and S. Xenos, "Who uses Facebook? An investigation into the relationship between the Big Five, shyness, narcissism, loneliness, and Facebook usage," *Computers in Human Behavior*, vol. 27, no. 5, pp. 1658–1664, Sep. 2011, doi: 10.1016/j.chb.2011.02.004.
- [10] L. R. Goldberg *et al.*, "The international personality item pool and the future of public-domain personality measures," *Journal of Research in Personality*, vol. 40, no. 1, pp. 84–96, Feb. 2006, doi: 10.1016/j.jrp.2005.08.007.
- [11] S. Srivastava and O. P. John, "The big-five trait taxonomy: history, measurement, and theoretical perspectives," in *Handbook of personality: Theory and research*, 1999, pp. 102–138.
- [12] L. R. Goldberg, "A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models," *Personality psychology in Europe*, vol. 7, pp. 7–28, 1999.
- [13] S. D. Gosling, P. J. Rentfrow, and W. B. Swann, "A very brief measure of the big-five personality domains," *Journal of Research in Personality*, vol. 37, no. 6, pp. 504–528, Dec. 2003, doi: 10.1016/S0092-6566(03)00046-1.
- [14] S. D. Gosling, A. A. Augustine, S. Vazire, N. Holtzman, and S. Gaddis, "Manifestations of personality in online social networks: self-reported Facebook-related behaviors and observable profile information," *Cyberpsychology, Behavior, and Social Networking*, vol. 14, no. 9, pp. 483–488, Sep. 2011, doi: 10.1089/cyber.2010.0087.
- [15] J. Golbeck, C. Robles, and K. Turner, "Predicting personality with social media," in *Proceedings of the 2011 annual conference extended abstracts on Human factors in computing systems - CHI EA '11*, 2011, pp. 253–262, doi: 10.1145/1979742.1979614.
- [16] D. Quercia, M. Kosinski, D. Stillwell, and J. Croccroft, "Our Twitter profiles, our selves: predicting personality with Twitter," in *2011 IEEE Third Int'l Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third Int'l Conference on Social Computing*, 2011, pp. 180–185, doi: 10.1109/PASSAT/SocialCom.2011.26.
- [17] K. Sharma and A. Kaur, "Personality prediction of Twitter users with logistic regression classifier learned using stochastic gradient descent," *IOSR Journal of Computer Engineering*, vol. 17, no. 4, pp. 39–47, 2015, doi: 10.9790/0661-17433947.
- [18] J. Golbeck, C. Robles, M. Edmondson, and K. Turner, "Predicting personality from Twitter," in *2011 IEEE Third Int'l Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third Int'l Conference on Social Computing*, 2011, pp. 149–156, doi: 10.1109/PASSAT/SocialCom.2011.33.
- [19] Y. R. Tausczik and J. W. Pennebaker, "The psychological meaning of words: LIWC and computerized text analysis methods," *Journal of Language and Social Psychology*, vol. 29, no. 1, pp. 24–54, Mar. 2010, doi: 10.1177/0261927X09351676.
- [20] M. M. Tadesse, H. Lin, B. Xu, and L. Yang, "Personality predictions based on user behavior on the facebook social media platform," *IEEE Access*, vol. 6, pp. 61959–61969, 2018, doi: 10.1109/ACCESS.2018.2876502.
- [21] S. Vosoughi, "Automatic detection and verification of rumors on Twitter," Massachusetts Institute of Technology, 2015.
- [22] B. Y. Pratama and R. Sarno, "Personality classification based on Twitter text using Naive Bayes, KNN and SVM," in *2015 International Conference on Data and Software Engineering (ICoDSE)*, 2015, pp. 170–174, doi: 10.1109/ICoDSE.2015.7436992.
- [23] D. Preoțiuc-Pietro, S. Volkova, V. Lampos, Y. Bachrach, and N. Aletras, "Studying user income through language, behaviour and affect in social media," *PLOS ONE*, vol. 10, no. 9, Sep. 2015, doi: 10.1371/journal.pone.0138717.
- [24] N. Majumder, S. Poria, A. Gelbukh, and E. Cambria, "Deep learning-based document modeling for personality detection from text," *IEEE Intelligent Systems*, vol. 32, no. 2, pp. 74–79, Mar. 2017, doi: 10.1109/MIS.2017.23.
- [25] Y. Mehta, N. Majumder, A. Gelbukh, and E. Cambria, "Recent trends in deep learning based personality detection," *Artificial Intelligence Review*, vol. 53, no. 4, pp. 2313–2339, Apr. 2020, doi: 10.1007/s10462-019-09770-z.




- [26] S. Kiritchenko, X. Zhu, C. Cherry, and S. Mohammad, "NRC-Canada-2014: detecting aspects and sentiment in customer reviews," in *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, 2014, pp. 437–442, doi: 10.3115/v1/S14-2076.
- [27] G. Farnadi *et al.*, "Computational personality recognition in social media," *User Modeling and User-Adapted Interaction*, vol. 26, no. 2–3, pp. 109–142, Jun. 2016, doi: 10.1007/s11257-016-9171-0.
- [28] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to information retrieval*, vol. 12, no. 5. Cambridge: Cambridge University Press, 2008.
- [29] C. Sumner, A. Byers, R. Boochever, and G. J. Park, "Predicting dark triad personality traits from Twitter usage and a linguistic analysis of tweets," in *2012 11th International Conference on Machine Learning and Applications*, 2012, pp. 386–393, doi: 10.1109/ICMLA.2012.218.

BIOGRAPHIES OF AUTHORS



Warih Maharani    received her bachelor's degree in Informatics Engineering from Telkom Institute of Technology (STT Telkom) in 2001, the Master's Degree in the School of Electrical Engineering and Informatics from the Institute of Technology Bandung (ITB), Bandung, Indonesia, in 2006, and the Ph.D. on Electrical Engineering and Informatics from Institut Teknologi Bandung (ITB) in 2019. She is a lecturer and researcher in School of Computing Telkom University. Her current research interests include sentiment analysis and social network analysis. She can be contacted at email: wmaharani@telkomuniversity.ac.id.



Veronikha Effendy    received her bachelor's degree in Informatics Engineering from Telkom Institute of Technology (STT Telkom) in 2003, the Master's Degree in Informatics from Telkom University, Bandung, Indonesia, in 2013. She has worked in IT industry for 8 years, and now she is a lecturer and researcher in School of Computing Telkom University. Her current research interests include human computer interaction and user experience. She can be contacted at email: veffendy@telkomuniversity.ac.id.