ICIC Express Letters ICIC International ©2021 ISSN 1881-803X
Volume **15**, Number **12**, December **2021** pp. **1243–1251**

PERSONALITY PREDICTION FROM TEXT ON SOCIAL MEDIA WITH MACHINE LEARNING

Syaiful Huda[1] and Andry Chowanda[2]

[1]Computer Science Department, BINUS Graduate Program – Master of Computer Science
[2]Computer Science Department, School of Computer Science
Bina Nusantara University
Jl. K. H. Syahdan No. 9, Kemanggisan, Palmerah, Jakarta 11480, Indonesia
syaiful.huda@binus.ac.id; achowanda@binus.edu

Received February 2021; accepted May 2021

Abstract. *Twitter is a social media that is widely used by people today and contains various kinds of information that can be used to detect a person's personality. The purpose of this thesis is to simplify and shorten the detection of a person's personality without using traditional methods such as using a questionnaire which requires a lot of time. This study also creates a dataset based on tweets of Indonesian-speaking users. The data collected is labeled based on the Big Five Personality method which consists of openness, cautiousness, extraversion, agreeableness and neuroticism for further personality classification. The proposed method uses four machine learning algorithms, namely Support Vector Machine (SVM), Naïve Bayes, Decision Tree and K-Nearest Neighbor (K-NN). The results of the four models were then analyzed using cross validation and calculating the F1-score to find the most suitable model for each personality trait. Based on the results of the evaluation, it shows that the Naïve Bayes model slightly outperforms other models in predicting each personality by excelling at three personality traits.*
Keywords: Natural language processing, Text mining, Machine learning, Big Five Personality

1. **Introduction.** Social media has become the most used communication and interaction tool by people over the last few years. Direct interaction between humans is reduced because people tend to communicate indirectly via smartphones. Thus, it was very difficult to remember someone. However, what is written on social media can help us get the neccessary information. Most people spend their time accessing social media and expressing their feelings and thoughts through status, comments and updates. Social media produces a collection of accumulated written language and a rich source of psychological data with unrealized scientific potential [1]. Knowledge of a person's personality allows us to make predictions about preferences through context and environment, and improve recommendation systems [2]. A person's personality can influence the decision-making process and has been shown to influence preferences for websites, products, brands and services [3], as well as to influence one's preferences in selecting content such as films, TV shows, and books [4].

The current dominant technique to solve problems in Natural Language Processing (NLP) is to use the supervised learning method. Several machine learning techniques can be used as a text-based classification model, namely Support Vector Machine (SVM), Decision Tree and Naïve Bayes. In [11] by translating the MyPersonality dataset into Indonesian and techniques such as Naïve Bayes, K-Nearest Neighbor (K-NN) and SVM as classification algorithms, it was obtained 72.29% as the highest accuracy. [5] used Term Frequency-Inverse Document Frequency (TF-IDF) for n-gram as a feature and SVM,

DOI: 10.24507/icicel.15.12.1243

Multinomial Naïve Bayes (MNB), Decision Tree or K-NN as a supervised learning algorithm. The best results for each personality trait were 0.59 and 0.33 in binary and multiclass representations according to the F1-score. [6] studied the use of machine learning techniques (SVM, K-NN, Naïve Bayes) for automatic recognition of personality traits of Facebook users based on their social network properties, status update text and the frequency with time of posting. From the results of this study, it was found that adding training examples from other social media could improve personality recognition results.

The social media that will be used as research material here is Twitter. Personality datasets based on tweets that use Indonesian are still rare, so this research will create a personality dataset for Indonesian Twitter users. Then based on previous research for the classification process through text, we will use the supervised learning method with four machine learning algorithms, namely SVM, Naïve Bayes, Decision Tree and K-NN to predict personality traits from tweet data. Then from the four classification algorithms, we will look for which algorithm can provide the best results from the five existing personality traits. There are several personality models that can be used to predict personality, including the Big Five Personality, MBTI (Myers Briggs Type Indicator) or DISC (Dominance Influence Steadiness Conscientiousness). In this study, the personality model to be used is the Big Five Personality [19-21]. The Big Five Model was first studied in the 1990s and is still in use today. Moreover, the Big Five Personality is the most widely accepted model for describing the basic structure of human personality [12].

2. Method.

Previous research has shown that linguistic features can be used to predict personality traits [8,13,14]. So this research will develop a personality prediction model based on text taken from tweets on Twitter that focus on users in Indonesia. Due to the limited dataset of Indonesian-language related to personality predictions on Twitter, this study will also create a personality dataset for Twitter users in Indonesia. The first step is to collect tweet data that will be used in the research. After data collection process, data will be preprocessed before feature extraction and classification process is carried out by four machine learning algorithms.

2.1. Data collection.

Data to be used is sourced from Twitter and the language to be used is Indonesian. User tweet data retrieval is taken through the Twitter application programmed using a Twitter API called twint. The number of tweet data taken was 10,000 tweets in a period of six months starting from January 1, 2020 to June 30, 2020. Indonesia's geolocation coordinates are used to filter tweet streaming. We do not include tweets from news portals and government offices. Commercial promotional tweets are also removed from collected data.

There are two annotators that are used to annotate data. The collected tweet data is labeled based on the Big Five Model [7], namely openness, cautiousness, extraversion, agreeableness and neuroticism. After annotation, the tweet data obtained were 1989 tweets. Then the dataset that has been labeled with the five personality traits will then go through a preprocessing process consisting of data normalization, tokenizing, stopword removal and stemming.

Table 1. Class distribution of dataset

Traits	**Total tweet**
Openness	190
Cautiousness	134
Extraversion	723
Agreeableness	281
Neuroticism	661
	1989

2.2. **Preprocessing.** The collected and labeled dataset from Twitter will be preprocessed before proceeding to the feature extraction and classification process. In the data preprocessing stage, the first step is data normalization to clean data from noise and uninformative data. The steps taken in data normalization are case folding, which is to change all letters in the document to lowercase. Then remove the numeric characters and punctuation marks because they are not related to what will be analyzed. The emoticon character is not deleted because it can represent one of the five personality traits. Tokenizing is done to change sentences into a collection of single words into pieces called tokens for later analysis. To normalize abbreviations, slang words and misspellings in tweets, we use a manually created Indonesian typography dictionary. Stopword removal is done to retrieve important words from the token results using a stoplist algorithm (removing less important words) or a wordlist (storing important words). Stopword is a common word that usually appears in large numbers and is considered meaningless. Examples of stopwords in Indonesian are "yang", "dan", "di", "dari", etc. The meaning behind using stopwords is that by removing words that have low information from a text, we can focus on the important words instead. Furthermore, stemming is done to return the word to its root form by removing the existing affixes, but the basic form does not mean the same as the root word. For example, the words "mendengarkan", "dengarkan", "didengarkan" will be transformed into the word "dengar". The stopword removal and stemming algorithm for Indonesian that will be used is the Sastrawi algorithm [15].

2.3. **Feature extraction.** Previous research [8,9] showed that the use of the Term Frequency-Inverse Document Frequency (TF-IDF) as a feature extraction method can be relied on in statistical measurements to measure how important a word is in a document. In addition, TF-IDF is one of the most versatile statistical methods showing the relative importance of a word or phrase in a document or set of documents compared to other corpuses and is one of the most widely used ways representing documents in text mining [10].

2.4. **Classification process.** At the learning stage, labeled training features are used as input for the machine learning classifier. This learning process produces a model that will be used for feature testing for unlabeled data. 10-fold K-fold cross-validation is used to evaluate the performance of a model or algorithm where data is separated into two subsets, namely learning process data and validation or evaluation data. The model or algorithm is trained by the learning subset and validated by the validation subset. 10-fold cross validation is one of the K-fold cross validation recommended for selecting the best model because it tends to provide less biased accuracy estimates compared to ordinary cross validation. Then for the classification process use four machine learning algorithms, namely Naïve Bayes, Decision Tree, SVM and K-NN. To measure the performance of the classification model and the features used, the macro F1-score was selected as the evaluation method.

2.5. **Evaluation method.** The evaluation method that will be used to measure the performance of the personality classification model and the features used is the macro F1-score. The F1-score is a weighted comparison of the average precision and recall. In addition, for the case of imbalance class, such as our dataset, we can use the F1-score [16].

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}$$

$$\mathrm{TNR} = \frac{\mathrm{TN}}{\mathrm{FP} + \mathrm{TN}}$$

$$\text{F-Score} = 2 \times \frac{\mathrm{TPR} \times \mathrm{TNR}}{\mathrm{TPR} + \mathrm{TNR}}$$

where
TPR = True Positive Rate/Sensitivity
TNR = True Negative Rate/Specificity
TP = True Positive
TN = True Negative
FP = False Positive

2.6. **Big Five Personality.** In psychology, a theory based on the Big 5 factors is the most widely recognized model for describing the basic structure of human personality. A theory based on these factors is called the Big 5 factor model and is the most widely recognized model of personality. The five-factor model provides terms and a conceptual framework that brings together many research findings in the field of psychology, namely differences in individuals and personalities. The five-factor model also reduces most personal adjectives to the five major personality traits that make up the acronym OCEAN [17,18]. Five factors or personality traits were studied and first discovered in the 1990s, and have been used until now. In Table 2, the individuals in the Big Five Model vary based on the concept of OCEAN, namely openness, cautiousness, extraversion, agreeableness and neuroticism. This model represents a complete set of traits that can capture personality differences [7].

Table 2. Big Five Personality traits

Personality traits	**Characteristics**
Openness (O)	From cautious/consistent to curious/inventive, intellectual, polished, creative, independent, open-minded, imaginative, creative, curious, tolerant
Cautiousness (C)	From careless/easy-going to organized/efficient reliable, consistent, self-disciplined, organized, hardworking, has long-term goals, planner
Extroversion (E)	From solitary/reserved to outgoing/energetic, express positive emotions, excited, satisfied, friendly, seeks stimulation in the company of others, talkative
Agreeableness (A)	From cold/unkind to friendly/compassionate, kind, concerned, truthful, good natured, trustful, cooperative, helpful, nurturing, optimistic
Neuroticism (N)	From secure/calm to unconfident/nervous, angry, anxious, neurotic, upset, depressed, sensitive, moody

3. Result and Discussion.

3.1. **Preprocessing result.** The following is an example of text information taken from the personality dataset before the preprocessing process:

"Sedih ya rasanya kalo media internasional udh nge lirik indo because a bad thing. @nytimes pic.twitter.com/mxLeJpUXdx".

The following is the preprocessing process that will be carried out.

1) Case folding by changing all letters to lowercase then removing non-letter characters, URLs and punctuation marks.

"sedih ya rasanya kalo media internasional udh nge lirik indo because a bad thing".

2) The tokenizing process based on space characters.

Token 1	Token 2	Token 3	Token 4	Token 5
sedih	ya	rasanya	kalo	media
Token 6	**Token 7**	**Token 8**	**Token 9**	**Token 10**
internasional	udh	nge	lirik	indo
Token 11	**Token 12**	**Token 13**	**Token 14**	
because	a	bad	thing	

3) Spell checking to correct spelling, abbreviations and slang words.

Token 1	Token 2	Token 3	Token 4	Token 5
sedih	iya	rasanya	kalau	media
Token 6	**Token 7**	**Token 8**	**Token 9**	**Token 10**
internasional	sudah	nge	lirik	indonesia
Token 11	**Token 12**	**Token 13**		
karena	hal	buruk		

4) Stopword removal to retrieve important words.

Token 1	Token 2	Token 3	Token 4	Token 5
sedih	iya	rasanya	kalau	media
Token 6	**Token 7**	**Token 8**	**Token 9**	**Token 10**
internasional	lirik	indonesia	hal	buruk

5) Stemming to return the word to its basic form (root word).

Token 1	Token 2	Token 3	Token 4	Token 5	Token 6
sedih	rasa	kalau	lirik	hal	buruk

3.2. **Classification result.** In this step, after we run process on Rapid Miner, the results are analyzed to find out which classifier for each personality trait is more accurate than the others. The four classification techniques were run on the same test data; therefore, a total of 20 models were run. The reason for this repetition is to find the most appropriate classifier for each personality factor because it is possible that classifiers will respond better to one personality factor than other classifiers on other factors.

accuracy: 64.91% +/- 2.25% (micro average: 64.91%)

	true openness	true cautiousness	true extraversion	true agreeableness	true neuroticism	class precision
pred. openness	132	1	10	1	32	75.00%
pred. cautiousness	1	88	4	9	26	68.75%
pred. extraversion	51	33	677	100	345	56.14%
pred. agreeableness	2	11	18	165	29	73.33%
pred. neuroticism	4	2	13	6	229	90.16%
class recall	69.47%	65.19%	93.77%	58.72%	34.64%	

Figure 1. Confusion matrix for the Decision Tree model

accuracy: 74.96% +/- 4.09% (micro average: 74.96%)

	true openness	true cautiousness	true extraversion	true agreeableness	true neuroticism	class precision
pred. openness	134	1	5	0	21	83.23%
pred. cautiousness	3	77	2	2	9	82.80%
pred. extraversion	41	31	611	62	139	69.12%
pred. agreeableness	2	11	23	189	12	79.75%
pred. neuroticism	10	15	81	28	480	78.18%
class recall	70.53%	57.04%	84.63%	67.26%	72.62%	

FIGURE 2. Confusion matrix for the SVM model

accuracy: 83.47% +/- 8.65% (micro average: 84.09%)

	true openness	true cautiousness	true extraversion	true agreeableness	true neuroticism	class precision
pred. openness	16	0	4	2	0	72.73%
pred. cautiousness	0	8	1	0	0	88.89%
pred. extraversion	2	1	22	3	0	78.57%
pred. agreeableness	1	0	3	37	2	86.05%
pred. neuroticism	2	3	3	1	65	87.84%
class recall	76.19%	66.67%	66.67%	86.05%	97.01%	

FIGURE 3. Confusion matrix for the Naïve Bayes model

accuracy: 68.58% +/- 2.83% (micro average: 68.58%)

	true openness	true cautiousness	true extraversion	true agreeableness	true neuroticism	class precision
pred. openness	141	5	24	6	24	70.50%
pred. cautiousness	2	29	9	2	7	59.18%
pred. extraversion	33	55	541	49	133	66.71%
pred. agreeableness	4	15	41	189	33	67.02%
pred. neuroticism	10	31	107	35	464	71.72%
class recall	74.21%	21.48%	74.93%	67.26%	70.20%	

FIGURE 4. Confusion matrix for the K-NN model

For example, if the Decision Tree is chosen as a suitable classifier for extraversion, it may not necessarily mean that the model is suitable for cautiousness. The experimental results were evaluated and analyzed using the F1-score according to Table 3 to Table 7.

By sending a tweet from a particular user, personality classifiers will have the ability to predict which class the tweet made by that user belongs to the five existing personality traits. Based on Table 3, SVM with an accuracy of 83.23% and an F1-score of 76.36% can be selected as the recommended model to predict the openness and Decision Tree models with an accuracy of 75.00% and an F1-score of 72.13% is not recommended as a suitable model. Based on Table 4, Naïve Bayesian with an accuracy of 88.89% and F1-score reaching 76.19% can be selected as the recommended model to predict cautiousness and the K-NN model with an accuracy of 59.18% and F1-score of 31.52% is not recommended

TABLE 3. Precision, Recall and F1-score in openness factor

Model	Precision (%)	Recall (%)	F1-score (%)
Naïve Bayesian	72.73	76.19	74.42
Decision Tree	75.00	69.47	72.13
SVM	83.23	70.53	76.36
K-NN	70.50	74.21	72.31

TABLE 4. Precision, Recall and F1-score in cautiousness factor

Model	Precision (%)	Recall (%)	F1-score (%)
Naïve Bayesian	88.89	66.67	76.19
Decision Tree	68.75	65.19	66.92
SVM	82.80	57.04	67.55
K-NN	59.18	21.48	31.52

TABLE 5. Precision, Recall and F1-score in extraversion factor

Model	Precision (%)	Recall (%)	F1-score (%)
Naïve Bayesian	78.57	66.67	72.13
Decision Tree	56.14	93.77	70.23
SVM	69.12	84.63	76.09
K-NN	66.71	74.93	70.58

TABLE 6. Precision, Recall and F1-score in agreeableness factor

Model	Precision (%)	Recall (%)	F1-score (%)
Naïve Bayesian	86.05	86.05	86.05
Decision Tree	73.33	58.72	65.22
SVM	79.75	67.26	72.97
K-NN	67.02	67.26	67.14

TABLE 7. Precision, Recall and F1-score in neuroticism factor

Model	Precision (%)	Recall (%)	F1-score (%)
Naïve Bayesian	87.84	97.01	92.20
Decision Tree	90.16	34.64	50.05
SVM	78.18	72.62	75.30
K-NN	71.72	70.20	70.95

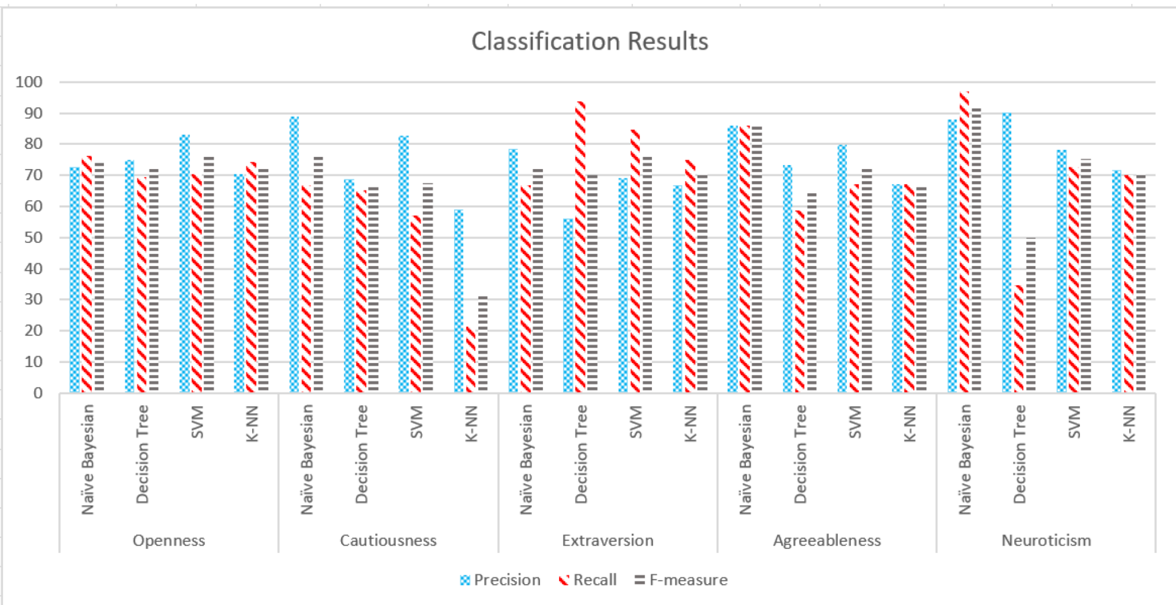

FIGURE 5. The results of the classification of the four algorithm models

as a suitable model. Based on Table 5, SVM with an accuracy of 69.12% and an F1-score of 76.09% can be selected as the recommended model to predict people with extraversion and Decision Tree characteristics with an accuracy of 56.14% and F1-score of 70.23% not recommended as a suitable model. Based on Table 6, Naïve Bayesian with an accuracy of 86.05% and an F1-score of 86.05% can be chosen to predict people with agreeableness and the Decision Tree model with an accuracy of 73.33% and F1-score of 65.22% is not recommended as a suitable model. Based on Table 7, Naïve Bayesian with an accuracy of 87.84% and F1-score reaching 92.20% can be chosen to predict people with neurotic properties and the Decision Tree model with an accuracy of 90.16% and F1-score of 50.05% is not recommended as the appropriate model.

From the results of the overall model, Naïve Bayes slightly outperformed other models with the highest F1-score calculation for the neuroticism class, namely 92.20%. Meanwhile, the lowest prediction result is the K-NN model with an F1-score calculation result of 31.52% for the cautiousness class. Then for the cautiousness class analysis, the overall results obtained were less than satisfactory because the number of tweets for that class was the least compared to the number of other classes. In general, the performance of the classification model in our new dataset can be said to be moderate and can be used as a benchmark model to detect personality traits in tweets of Twitter users in Indonesia.

4. **Conclusion.** In this study, we made a publicly available dataset for the classification of personality traits of Twitter users in Indonesia. This dataset consists of five personality classes (openness, cautiousness, extraversion, agreeableness and neuroticism). We try to identify the user's personality indirectly without using traditional methods so that the user's personality is not only recognized based on a questionnaire by providing a statement, whereas we can predict the user's personality through the tweets they make. In the results achieved using our proposed model, Naïve Bayes outperforms other models in three personality classes: cautiousness, agreeableness and neuroticism.

For future research, the proposed method can be changed in several aspects so that we plan to predict personality by other techniques such as text mining through Twitter user profiles to predict personality. In addition, research can also be carried out on what photos Twitter users will use based on the five personality types as their profile photos. Then, we could consider building a larger data set and a more balanced distribution of data for each class. With a larger dataset, we would like to use deep learning models to predict personality traits.

REFERENCES

[1] G. Park, H. A. Schwartz, J. C. Eichstaedt, M. L. Kern, M. Kosinski, D. J. Stillwell, L. H. Ungar and M. E. P. Seligman, Automatic personality assessment through social media language, *Journal of Personality and Social Psychology*, vol.108, pp.934-952, 2015.

[2] R. Lambiotte and M. Kosinski, Tracking the digital footprints of personality, *Proc. of the Institute of Electrical and Electronics Engineers (PIEEE)*, pp.1935-1939, 2014.

[3] M. Kosinski, Y. Bachrach, P. Kohli, D. Stillwell and T. Graepel, Manifestations of user personality in website choice and behaviour on online social networks, *Mach. Learn.*, vol.95, no.3, pp.1-24, 2013.

[4] I. Cantador, I. Fernández-Tobías, A. Bellogín, M. Kosinski and D. Stillwell, Relating personality types with user preferences in multiple entertainment domains, *Proc. of the 1st Workshop on Emotion and Personality in Personalized Services (EMPIRE)*, 2013.

[5] M. S. Salem, S. S. Ismail and M. Aref, Personality traits for Egyptian Twitter users dataset, *Proc. of 2019 8th International Conference on Software and Information Engineering*, pp.206-211, 2019.

[6] G. Farnadi, S. Zoghbi, M.-F. Moens and M. De Cock, Recognising personality traits using Facebook status updates, *Proc. of the Workshop on Computational Personality Recognition (WCPR13) at the 7th International AAAI Conference on Weblogs and Social Media (ICWSM13)*, 2013.

[7] O. P. John and S. Srivastava, The big five trait taxonomy: History, measurement, and theoretical perspectives, *Handbook of Personality: Theory and Research*, vol.2, pp.102-138, 1999.

[8] D. R. J. Moreno, J. C. Gomez, D. L. Almanza-Ojeda and M. A. Ibarra-Manzano, Prediction of personality traits in Twitter users with latent features, *International Conference on Electronics, Communications and Computers (CONIELECOMP)*, pp.176-181, 2019.
[9] H. Zheng and C. Wu, Predicting personality using Facebook status based on semi-supervised learning, *Proc. of 2019 11th International Conference on Machine Learning and Computing*, pp.59-64, 2019.
[10] W. Zhang and T. Yoshida, A comparative study of TF-IDF, LSI and multi-words for text classification, *Expert Systems with Applications*, vol.38, pp.2758-2765, 2011.
[11] B. Y. Pratama and R. Sarno, Personality classification based on Twitter text using Naïve Bayes, KNN and SVM, *International Conference on Data and Software Engineering (ICoDSE)*, pp.170-174, 2015.
[12] 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.
[13] E. Bassignana, M. Nissim and V. Patti, Matching theory and data with personal-ITY: What a corpus of Italian YouTube comments reveals about personality, *arXiv Preprint*, arXiv: 2011.07009, 2020.
[14] V. Balakrishnan, S. Khan and H. R. Arabnia, Improving cyberbullying detection using Twitter users' psychological features and machine learning, *Computers & Security*, vol.90, DOI: 10.1016/j.cose.2019.101710, 2020.
[15] M. A. Rosid, A. S. Fitrani, I. R. I. Astutik, N. I. Mulloh and H. A. Gozali, Improving text preprocessing for student complaint document classification using sastrawi, *IOP Conference Series: Materials Science and Engineering*, vol.874, no.1, DOI: 10.1088/1757-899X/874/1/012017, 2020.
[16] J. Han, K. Micheline and P. Jian, *Data Mining: Concepts and Techniques*, 3rd Edition, Elsevier Inc., Massachusetts, 2012.
[17] L. R. Goldberg, The structure of phenotypic personality traits, *Amer. Psychol.*, vol.48, no.1, pp.26-34, 1993.
[18] E. C. Tupes and R. E. Christal, Recurrent personality factors based on trait ratings, *J. Pers.*, vol.60, no.2, pp.225-251, 1992.
[19] D. Azucar, D. Marengo and M. Settanni, Predicting the big 5 personality traits from digital footprints on social media: A meta-analysis, *Personality and Individual Differences*, vol.124, pp.150-159, 2018.
[20] V. Kaushal and M. Patwardhan, Emerging trends in personality identification using online social networks – A literature survey, *ACM Trans. Knowledge Discovery from Data (TKDD)*, vol.12, no.2, 2018.
[21] P. Adamopoulos, A. Ghose and V. Todri, The impact of user personality traits on word of mouth: Text-mining social media platforms, *Information Systems Research*, vol.29, no.3, DOI: 10.1287/isre.2017.0768, 2018.