IJCCS (Indonesian Journal of Computing and Cybernetics Systems)
Vol.14, No.4, October 2020, pp. 365~376
ISSN (print): 1978-1520, ISSN (online): 2460-7258
DOI: https://doi.org/10.22146/ijccs.58595

Word Analysis of Indonesian Keirsey Temperament

Ahmad Fikri Iskandar*[1], Ema Utami[2], Agung Budi Prasetio[3]
[1,2,3]AMIKOM University Yogyakarta, Indonesia
e-mail: ***[1]andfikri@gmail.com**, [2]ema.u@amikom.ac.id, [3]agung.bp@excelindo.co.id

Abstrak

Kepribadian secara unik menghubungkan perasaan dan pola tindakan. Perilaku ini akan berubah melalui pengalaman, pendidikan formal, dan lingkungan sekitar. Penelitian ini fokus berdasarkan Keirsey Temperament Sorter, kuesioner kepribadian yang dikembangkan oleh David Keirsey. Model temperamen ini membagi kepribadian menjadi empat kategori sebagai Idealists, Rationals, Guardians, dan Artisans. Konsep ini umumnya diakui untuk interpretasi tren spesialis, berpotensi berkontribusi pada proses rekrutmen atau seleksi dan bidang potensial untuk analisis data media sosial. Kata-kata dipilih dengan menggunakan Chi-Square dengan kesalahan 5%. Akurasi pendekatan leksikon adalah 34%, sedangkan pendekatan machine learning terbaik dengan algoritma Random Forest dengan 69.59%

Kata kunci*— Keirsey, Temperamen, Kepribadian, Chi-Square*

Abstract

Personality uniquely relates to our feeling and pattern to the aspect of actions. This behavior will change through the experience, formal education, and the surrounding environment. This works based on the Keirsey Temperament Sorter, a personality questionnaire developed by David Keirsey. This model divides the personality into four categories as Idealists, Rationals, Guardians, and Artisans. This concept is commonly recognized for the interpretation of specialist trends, potentially contributes to the process of recruitment or selection, and potential fields for analysis of social media data. Words selected by using Chi-Square with an error of 5%. Accuracy of the lexicon approach is 34%, while the best machine learning approach is Random Forest algorithm with 69.59%

Keywords— *Keirsey, Temperament, Personality, Chi-Square*

1. INTRODUCTION

1.1 Background

Personality may differentiate a person from others. Personality explains the combination of characteristics and qualities which create an individual's character. Personality will uniquely relate our feeling and pattern to the aspect of actions. This behavior will change through learning, experience, formal education, and the environment. There is some application personality useful for our daily life. Type of personality can be found in the application for

Received August 14th,2020; Revised October 27th, 2020; Accepted October 30th, 2020

marketing, education, health, talent career, recruitment process or selection applicant, and other application.

In the marketing aspect, personality can also be used to determine the marketing strategy. Personality is used as a tool for knowing people's preferences regarding certain products to provide different promotions to each person [1]. For Human Resource Development Department needs, personality preferences are the initial process of recruiting applicants, where the goal is to detect an applicant's psychological problems. Then, applicant's and employee's personalities can also show the ability to work together and collaborate as a team [2]. Furthermore in education, Susilawati [3] explain that personality and good character are part of learning outcome on civic education. The research about user behavior in social media related to psychological illnesses has been done by Preotiuc-Pietro [4], which analyzed the language used by social media users. The results obtained that the language on social media can be alternative linguistic approach that can be used to know user mental illness.

Since this world now relies much more on text-based communication than on face-to-face interactions, it is becoming highly essential to develop text-based predictive behavior models. It is also believed that the underlying patterns of personality can be obtained from the text. Unstructured data was extracted from social media, most of which relate to human interaction and behavior. As a result, social media can be seen as repositories of behaviors that can be modeled to psychological characteristics.

The following research is based on the Keirsey Temperament Sorter, a personality questionnaire developed by David Keirsey, which group individuals into four temperaments. Keirsey's work is based mainly on the Myers-Briggs Type. Keirsey's research divide into four categories: Idealists, Rationals, Guardians, and Artisans. In this context, as regards progress in human behavior research, this article presents a temperament prediction classification system based on the temperament model developed by David Keirsey [5]. Keirsey's model leads us to identify how person corresponds to their world, either by making choices or absorbing information.

1.2 Temperament Model

Temperament is a configuration of observable personality traits, such as communications, action, attitudes, values, and talents. Temperament denotes a set of innate and particular characteristics of an individual, closely connected with biological or physiological determinants. Carl Gustav Jung introduced one of the essential concepts in 1920. Jung explains how the mind works of every person consist of an interaction between attitudes and functions. Attitudes can be the factor of psychic energy and maybe Extraversion (E) and Introversion (I). The functions are defined by how people view the world, so we will have two ways to receive knowledge (Sensing (S) and iNtuition (N)) and two ways to make decisions (Thinking (T) and Feeling (F)). Afterward, Isabel Briggs Myers and Katharine Cook Briggs introduced a new combination of functions to Jung's proposed typology: Judgment (J) and Perception (P) [6]. This pair determine if an entity's approach in reaction against the outside world derives from logical (Judging) or illogical (Perceiving) function.

The temperament model suggested by David Keirsey [5] divides the personality into four categories namely, Idealists, Rationals, Guardians, and Artisans. This concept is commonly recognized for the interpretation of specialist trends, potentially contribute to the process of recruitment and selection and potential fields for analysis of social media data. D. Keirsey [5] focus on his research on the connection between the taxonomy of Myers-Briggs and the evaluation of personality in practice at the time of choosing, behavior patterns, reasoning, and consistency. He believed that the character-associated temperament determines the individual's personality that inherent and arises from the experience of the temperament with the environment. Hence, the categories are directed by aspirations and interests that motivate us to survive, behave, move, and play a part in society[5]. He stated that expectations are more linked to perception (S-N), completely instinctive, than decision-making (T-F), which is entirely logical. Sensing (S) can be combined with judgment (J) or perception (P), whereas intuition (N) can be combined with feeling (F) or thinking (T). This identification did result in four categories of personality: the Guardian (SJ), the Artisan (SP), Idealist (NF), and Rational (NT).

1.3 Previous Works

There has been some automatic prediction of personality, initially taken by Lukito [1] in trying to develop Indonesian MBTI personality classification using three approach, namely machine learning based, lexicon based, and grammatical rule with 97 users data. Train data and test data is 84.5%:15.5%. Naive Bayes model performs better than the others with Introvert-Extrovert (IE) accuracy is 72.5%. Next, Adi [7] developed the classifier model with 286 data for classifying the Indonesian Big-5 personality traits. There are 12 extraction features, namely the number of tweets, retweets, replies, followers, retweeted, hashtags, following, quotes, URLs, favorites, mentions and tweet content. Each label of features is labeled as 1 for high and 0 for low. The selection of features that used in this works is the Decision Tree with four scenarios, combination of hyper parameter tuning, selection of features, and sampling with 80:20 train test ratio. Meanwhile, temperament prediction framework was done by Lima [8]. There are scenarios done, combination of models, Linguistic Inquiry and Word Count (LIWC), Medical Research Council (MRC), Psycholinguistic Database, oNLP. This works not only focus on temperament but also MBTI prediction.

Relatively similar work has been done by Fikry [9] for the classification of extroverted and introverted characters that use feature extraction from posts on Twitter. Extraction of the feature is the number of tweets, URL, hashtag, retweet, liked, mention, follow, active ratio, mention without retweet, reply, word on profile, average word per tweet, tweet character, emoticon/emoji, and media. The training process that uses three proportions of training data and test data is 70:30, 80:20, and 90:10. It seems good accuracy, but this works a tiny scope which is only 60 users. Ong [10] also developed an Indonesian-language of Big Five personality classification system. There are 12 feature selections, namely, the number of tweets, followers, following, favorites, retweets, tweet retweets, quote tweets, mentions, replies, hashtags, extracted tweet URL form, and the time difference between each tweet. This works compared 12 scenarios with the parameters of word weighting, topic modeling, stop word, and n-gram. The proportion of data used for training data only 329 and 30 for testing data.

In the classification of the Big Five Personality, which was done by Jeremy [11], there is an addition of 4 feature extraction approaches. This research-based on metadata approaches such as the number of followers, following, tweets, favorites, retweets, mentions, quotes, replies, and hashtags. Compared to the approach, the approach is not getting significant results without adding extraction of the feature. In computing the Big Five personality, the Naive Bayes and K-NN models get quite good results, and the Sequential Minimal Optimization (SMO) model is the best in the classification process. This work did not use a reduction dimension or selection of features. Utami [2] used an open-vocabulary approach to classify the personality Dominant, Influence, Steadiness, and Compliant (DISC). An exciting part of analytics is the synonyms of every word. The word weighting for first synonym is 0.85, while for the second synonym is 0.35.

Table 1 Related Works of Personality Detection

Author	Model/ Approach	Result	Limitation or Future Work
MBTI Personality			
Lukito, P.H [1]	Naïve Bayes	Best Accuracy: IE: 80 % NS: 60 % TF: 60 % JP: 60 %	The classification needed comparison result using other machine learning model
Fikry & Yusra [9]	SVM	Accuracy : 88%	This work provide increased from previous model NBC with 83.33 %, but we add more data about other dimension of MBTI.
Iskandar, A.F. [12]	Naïve Bayes KNN	Best Accuracy: IE: 81.25% NS: 84.62% TF: 84.55% JP: 75.00%	This data is not balance, so it is needed a method to solve it.
DISC Personality			
Utami, E. [2]	SVM	Accuracy: 37.41%,	Compare the SVM model with other models to produce a better performance classification and select features using chi-square
Temperament Personality			
Claudy [13]	KNN	Accuracy: 66%	Compare feature TF-IDF and other machine learning models and normalize non-standard words
Lima [8]	Random Forests LIWC	Best Accuracy: Artisan: 96.46% Guardian: 92.19%, Idealist: 78.68%, Rational: 83.82%.	Balancing data method is needed to better accuracy for idealist and rational
Big Five Personality			
Ong [10]	SVM XGBoost	Average Accuracy: SVM: 76.23% XGBoost: 97.99%	Compare frequency and word weighting (TF-IDF)
Adi [7]	Logistic Regression XGBoost SVD	Best performance: SGD: 99% XGB: 84.60% SL: 99.20%	Compare result using feature selection

Based on the limitation in Table 1, this work conduct using scenario to classify personality Keirsey framework using some model machine learning like logistic regression, Naïve Bayes, KNN, SVM, etc. and also this work use balancing method SMOTE and Chi-square feature selection. The research focuses on words on each dimension of the temperament. There are several discussions, namely (1) explore the words of each dimension of the temperament, (2) the relationship between each dimension based on words, and (3) classification based on these words.

In summary, contributions of this work, the processed text data are used to explore and classify user personality based on the Keirsey Temperament framework two-approach, namely based on the lexicon and machine learning approach. We applied different pre-processing techniques for the extraction feature to combine Categorical Proportional Difference (CPD).

Performance of classification model using Naïve Bayes (NB), Random Forest (RF), Logistic Regression (LR), and Support Vector Machines (SVM).

This works is organized as follows, section 1 discuss the background, Keirsey Temperament concepts, and recent research about automated personality prediction. Section 2 includes a description of the methodology exploration and classification. Section 3 presents and analyzes performance. Section 4 concludes this work and future research.

2. METHODS

In this part will be introducted the data to be used, the process of preprocessing data into a lexicon, and rules so that the words can be categorized into one of classes namely Idealists, Rationals, Guardians, and Artisans. More detail of this works are as follows:

2.1 Data

Data used in this works is Twitter social media personality data by Iskandar [12]. The data consists of 2 columns, namely text and their label MBTI. The detail type MBTI from this data shown in Figure 1 below:

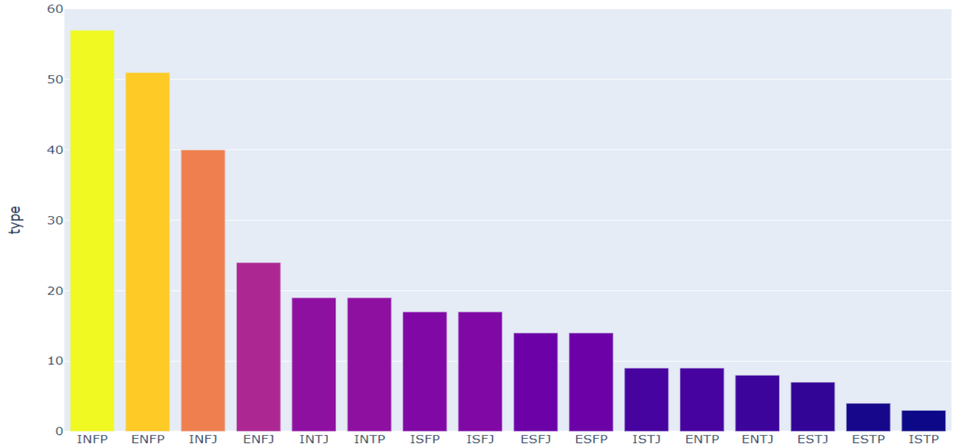

Figure 1 Type MBTI Data

Type of MBTI personality preference will be broken down into 4 classes based on Role Temperaments namely Idealists, Rationals, Guardians, and Artisans. The rules of MBTI classes into role Temperaments shown in Table 2 below:

Table 2 Comparison Tempraments

	Temperament			
	Idealists	**Rationals**	**Guardians**	**Artisans**
MBTI	Champion (ENFP)	Architect (INTP)	Inspector (ISTJ)	Composer (ISFP)
	Counselor (INFJ)	Fieldmarshal (ENTJ)	Protector (ISFJ)	Crafter (ISTP)
	Healer (INFP)	Inventor (ENTP)	Provider (ESFJ)	Performer (ESFP)
	Teacher (ENFJ)	Mastermind (INTJ)	Supervisor (ESTJ)	Promoter (ESTP)

Source: D. Keirsey [5]

Summary of user based on role temperament shown in Table 3.

Table 3 Comparison Tempraments

No	Temperaments	Number of user
1	Idealists	172
2	Rationals	55
3	Guardians	47
4	Artisans	38

Table 3 shows that the Temperaments class data are not balanced. User data are more dominated by users with Temperaments Idealists type as many as 172 while the class with other types is almost 1/3 of the Temperaments Idealists type class. So, it is necessary to do a data balancing of the Idealists class.

2.2 Preprocessing

After collecting the data, the information on the behavioral category was extracted from each user account, while the grammatical information was obtained from each user label. Its behavioral and grammatical information represents each user. Some steps must do on natural language processing research which is preprocessing. Step of preprocessing namely case folding, remove stop word, non-numeric, stemming, normalize word, translate to Indonesia language.

2.3. TF-IDF and CPD

Feature extraction on this work consists of TF and also TF-IDF. Term Frequency (TF) explains the number of times the word appears within the document. Similarly, Inverse Document Frequency (IDF) a measure of the final importance of the term the number of documents that contain the term t within the entire document[14]. While categorical proportional difference or called CPD is an easy selection method for multiclass classification problems. CPD estimates how much a word adds to separating a specific classification from different classes in a text corpus. CPD may be defined in equation (1):

$$CPD = \frac{|Positive\ document - negative\ document|}{Positive\ dociment + negative\ document} \quad (1)$$

CPD process positive document and negative document of 1 term exclusively, and next, it computes the relative distinction of 1 term in both positive and negative classes [15].

2.4 Analysis

Words were selected by using Chi-Square with an error of 5%. The lower the error will select words that have no correlate with the label class. This works analyze the number of words in each user, the number of unique words before and after it so it will be categorized word to the label class. Similar to other research about automated personality, Eealuation of the classification models is Accuracy, Precision, Recall, and F1-Score. More details about the evaluation model are shown in Table 4.

Table 4 Evaluation Model

No.	Evaluation	Formula	Description
1.	Accuracy	$\frac{TN+TP}{(TP+FP+TN+FN)}$	Accuracy is used to evaluate the number of predictive labels that correspond to the actual label.
2.	Precision	$\frac{TP}{(TP+FP)}$	Precision is the level of accuracy between the information requested by the user and the answer given by the system.
3.	Recall	$\frac{TP}{(TP+FN)}$	Recall is the success rate of the system in rediscovering information.
4.	F1-Score	$\frac{2 \times precision \times recall}{precision + recall}$	F1 Score is the weighted average of Precision and Recall

Source: Willy [16]

Where TP is true positive, TN is true negative, FP is false positive, and FN is false negative.

3. RESULTS AND DISCUSSION

3.1 Word Exploration

The preprocessing words aim to eliminate words to reduce noise from the data. The number of words after preprocessing is 310 words. Furthermore, words will be analyzed the number of words, the number of words after and before, and the number of users who used these words. This analysis aims to know the context of the word and group the words into the Idealists, Rationals, Artisans, or Guardians classes. Results of generating the three features above to analyze the correlation of each word to the class. Pearson correlation results from these words shown in Figure 2 the following:

	b_Idealists	uk_Idealists	us_Idealists
b_Idealists	1.0	0.9725	0.8518
uk_Idealists	0.9725	1.0	0.9082
us_Idealists	0.8518	0.9082	1.0

	b_Rationals	uk_Rationals	us_Rationals
b_Rationals	1.0	0.9828	0.8549
uk_Rationals	0.9828	1.0	0.9055
us_Rationals	0.8549	0.9055	1.0

	b_Artisans	uk_Artisans	us_Artisans
b_Artisans	1.0	0.9389	0.806
uk_Artisans	0.9389	1.0	0.8684
us_Artisans	0.806	0.8684	1.0

	b_Guardians	uk_Guardians	us_Guardians
b_Guardians	1.0	0.9697	0.846
uk_Guardians	0.9697	1.0	0.8913
us_Guardians	0.846	0.8913	1.0

Figure 2 Correlation Feature

Based on Figure 2 above, there are three features with prefixes b_, uk_, and us_. Prefix b_means total of words, Prefix uk_ means the number of unique words, and Prefix us_ means the number of users that use it. Lexical Diversity (LD) refers to the variety of words used in a text. LD indices generally measure the number of types (i.e., unique words occurring in the text) by tokens. Average LD for Idealists words is 1.2, and Rationals words are 0.9, Artisan words are 1.1, and Guardians words is 1.04. The average lexical diversity in all classes is 1. It means the number of word types is equal to the total number of tokens; all of the words are different. Lexical diversity measures relate to the number of words a user knows. The proportion weight for prefix b_ is 0.26, prefix us_ is 0.63, and uk_ is 0.11. The weight of this feature is based on the researchers' judgment of these words. From the calculation of the weights to each class, the number of the word for Idealists is 116 words, Rationals are 47 words, Guardians are 59 words, and Artisans are 88 words. More explore, this works also mapped the word into the scatter plot with two variables generated using PCA shown in Figure 3 below:

Figure 3 Word Temperament

Scatter of words in Figure 3 shows four colors, namely blue show words on class Idealists, red show words on class Rationals, green show words on Artisan class, and purple show words on class Guardians. Furthermore, the words will be grouped to become keywords that can get described these words in general. Based on these words can generate keyword each class is shown in Table 5 below:

Table 5 Top 5 Keyword Temperament

Idealists	**Rationals**	**Guardians**	**Artisans**
Character	Character	Character	Character
Valentine's	Math	Zodiac	Weddings
Earth	Literature	Earth	Photography
Food	World	Person	Money
World	Life	Life	Fashion

Table 5 explains the top 5 keyword words that are often used by users in each class. The keyword “Character” dominates each label. Keywords for Idealist related to Earth and World. Rational user related to Math and Literature. Guardian users related to Earth and Life, and Artisan users related to Money, Fashion, and Photography. These are unique keyword for each user of the Keirsey Temperament Model.

3.2 Classification

The words that have been categorized then will be tested against those words using classification. Classification is done using two approaches, namely lexicon based and machine learning based.

1. Lexicon
   This approach will do the classification based on the words that have been filtered using chi-square. Then, each of these words will be counted the number of words that appear then presented to the total words so that the percentage of Idealists, Rationals, Guardians, and Artisans will be obtained. Based on the highest percentage of those words, the sentence will be classified into the class.
2. Machine Learning
   The words on which have been cleaned from noise. In this part, we classify with three scenario-based on feature extraction. Machine learning model used is Multinomial Naïve Bayes, Random Forest, Logistic Regression, and SVM.

Classification is done on 198 users. This data divided 160 users to training data and 38 to testing data. The result's lexicon approach is 34%. For details is shown in Table 6 below:

Table 6 Performance Lexicon Approach

Label	**Precision**	**Recall**	**F1-score**
Idealists	33%	67%	44%
Artisans	38%	60%	46%
Rationals	36%	29%	32%
Guardians	30%	23%	26%

Table 6 shows average precision is 34.25%, average recall is 44.75% and f1-score is 37%. While best accuracy using machine learning model is 69.59% with random forest model. Detail of performance precision, recall, and f1-score is shown in Table 7 below:

Table 7 Performance Machine Learning Approach

Model	**Precision**	**Recall**	**F1-Score**
	TF		
Naïve Bayes	58.47%	46.56%	43.38%
Random Forest	69.05%	67.54%	67.51%
Logistic Regression	52.01%	51.07%	50.54%
Support Vector Machine	62.54%	46.72%	42.45%
	TF-IDF		
Naïve Bayes	63.36%	41.75%	38.49%
Random Forest	71.78%	66.84%	66.72%
Logistic Regression	52.60%	48.84%	48.28%
Support Vector Machine	54.24%	47.42%	46.27%
	TF-IDF + CPD		
Naïve Bayes	45.26%	31.54%	28.72%
Random Forest	**75.72%**	**69.88%**	**69.98%**
Logistic Regression	41.41%	37.63%	37.30%
Support Vector Machine	37.31%	32.14%	23.53%

Based on Table 7 above, the best machine learning approach obtained by Random Forest with precision 75.72%, recall 69.88%, and f1-score 69.98%. This best performance is obtained by using feature TF-IDF with CPD and balancing method SMOTE.

3.3 Ethic and Privacy

This study only focuses on analyzing words in social media based on the Keirsey temperament model. So this research only takes general topics, not focus on the user's private information. Kosinski et all [17] explained social media research to use publicly available private user information without agreement with the provisions assuming that the data was intentionally made public, user data anonymized after collection and no attempt was made to define it and no interaction or communication with individuals in the sample.

During data collection, exploration until classification, research remains focused on maintaining the privacy of Twitter users who have taken their tweets and ethics in researching social media data. Even we know, Twitter is one or part accessible data, the researcher also keeps Twitter users who have taken data by doing a rename with sample code to disappear judging from researchers. This work was done so that the focus on the words they use is not the focus of the Twitter user [2].

4. CONCLUSIONS

This research was done to understanding the behavior of users on social media using word that what they said. Here, we did an exploratory study aimed at understanding the potential of machine learning techniques for Keirsey Temperament prediction. We used data from 16 types of Myers-Brigss typology and mapped them into the Keirsey temperament model. This is based on the lexical hypothesis, which shows that the majority of individual differences is encoded in the language. Accuracy of Lexicon approach is 34%, while best perfomance approach to classify using machine learning with Random Forest algorithm is 69.59%.

The understanding of Keirsey temperament framework can be used in various fields, such as professional guidance, leadership training, pedagogical approaches, group dynamics, sales training and customer service, profile audiences, self-understanding, educational aptitude and professional achievement, conflict resolution and stress management, understand decision making, among others. We would like to expand this research to new databases both from Twitter and other social media, do some hypothesis toward each user temperament, and utilize feature extraction and deep learning models to get better results.

REFERENCES

[1] L. C. Lukito, A. Erwin, J. Purnama, and W. Danoekoesoemo, “Social media user personality classification using computational linguistic,” *Proc. 2016 8th Int. Conf. Inf. Technol. Electr. Eng. Empower. Technol. Better Futur. ICITEE 2016*, no. October 2016, 2017, doi: 10.1109/ICITEED.2016.7863313.

[2] E. Utami, A. D. Hartanto, S. Adi, I. Oyong, and S. Raharjo, “Profiling analysis of DISC personality traits based on Twitter posts in Bahasa Indonesia,” *J. King Saud Univ. - Comput. Inf. Sci.*, no. xxxx, 2019, doi: 10.1016/j.jksuci.2019.10.008.

[3] E. Susilawati, H. Sitompul, and J. Situmorang, “The Differences in Using Direct Instruction (DI) Learning Strategy Based on Competitive Behavior to Civic Education

Learning Achivement," no. January 2018, 2018, doi: 10.2991/aisteel-18.2018.47.

[4] D. Preoţiuc-Pietro *et al.*, "The role of personality, age, and gender in tweeting about mental illness," pp. 21–30, 2015, doi: 10.3115/v1/w15-1203.

[5] D. Keirsey, *Please Understand Me II: Temperament, Character, Intelligence*. 1998.

[6] A. L. Hammer, "Myers-Briggs type indicator. comparison report : Work styles," pp. 1–13, 2015, [Online]. Available: https://www.cpp.com/pdfs/smp261182.pdf.

[7] G. Y. N. N. Adi, M. H. Tandio, V. Ong, and D. Suhartono, "Optimization for Automatic Personality Recognition on Twitter in Bahasa Indonesia," *Procedia Comput. Sci.*, vol. 135, pp. 473–480, 2018, doi: 10.1016/j.procs.2018.08.199.

[8] A. C. E. S. Lima and L. N. De Castro, "Tecla: A temperament and psychological type prediction framework from Twitter data," *PLoS One*, vol. 14, no. 3, pp. 1–18, 2019, doi: 10.1371/journal.pone.0212844.

[9] M. Fikry, "Ekstrover atau Introver : Klasifikasi Kepribadian Pengguna Twitter dengan Menggunakan Metode Support Vector Machine," *J. Sains dan Teknol. Ind.*, vol. 16, no. 1, p. 72, 2018, doi: 10.24014/sitekin.v16i1.5326.

[10] V. Ong *et al.*, "Personality prediction based on Twitter information in Bahasa Indonesia," *Proc. 2017 Fed. Conf. Comput. Sci. Inf. Syst. FedCSIS 2017*, vol. 11, pp. 367–372, 2017, doi: 10.15439/2017F359.

[11] N. H. Jeremy, C. Prasetyo, and D. Suhartono, "Identifying personality traits for Indonesian user from twitter dataset," *Int. J. Fuzzy Log. Intell. Syst.*, vol. 19, no. 4, pp. 283–289, 2019, doi: 10.5391/IJFIS.2019.19.4.283.

[12] A. F. Iskandar and E. Utami, "Impact of Feature Extraction and Feature Selection Using Naïve Bayes on Indonesian Personality Trait,". ICOIACT. Pending Publication.

[13] Y. I. Claudy, R. S. Perdana, and M. A. Fauzi, "Klasifikasi Dokumen Twitter Untuk Mengetahui Karakter Calon Karyawan Menggunakan Algoritme K-Nearest Neighbor (KNN)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 8, pp. 2761–2765, 2018, [Online]. Available: https://www.researchgate.net/publication/322959490.

[14] R. K. Roul, J. K. Sahoo, and K. Arora, "Modified TF-IDF Term Weighting Strategies for Text Categorization," *2017 14th IEEE India Counc. Int. Conf. INDICON 2017*, no. October, 2018, doi: 10.1109/INDICON.2017.8487593.

[15] J. R. Chang, H. Y. Liang, L. S. Chen, and C. W. Chang, "Novel feature selection approaches for improving the performance of sentiment classification," *J. Ambient Intell. Humaniz. Comput.*, no. 0123456789, 2020, doi: 10.1007/s12652-020-02468-z.

[16] Willy, E. B. Setiawan, and F. N. Nugraha, "Implementation of Decision Tree C4.5 for Big Five Personality Predictions with TF-RF and TF-CHI2 on Social Media Twitter," *2019 Int. Conf. Comput. Control. Informatics its Appl. Emerg. Trends Big Data Artif. Intell. IC3INA 2019*, pp. 114–119, 2019, doi: 10.1109/IC3INA48034.2019.8949601.

[17] M. Kosinski, D. Stillwell, and T. Graepel, "Private traits and attributes are predictable from digital records of human behavior," Proceedings of the National Academy of Sciences, 110(15), 5802–5805, doi:10.1073/pnas.1218772110