What You Say Define Who You Are? Word Exploration And Automatic Personality Detection

Ahmad Fikri Iskandar, Ema Utami, Agung Budi Prasetio

Abstract: The personality of people being have their individual differences affects every aspect of their life. In the psychology field, the concept of personality is considered a powerful but imprecisely defined construct. There are some popular personality assessments namely, Big Five, Dominance Influence Steadiness Compliance (DISC), and Myers-Briggs Type Indicator (MBTI). This works do exploration about the word used each dimension of Myers-Briggs Type Indicator (MBTI) personality trait and use machine learning technique to classify text into different personality traits such as Introversion-Extroversion (IE), Sensing-iNtuition (NS), Thinking- Feeling (FT) and Judging-Perceiving(JP). After doing some hypothesis tests, there is difference between each axis about people-related word for IE dimension, counterfactual word for NS dimension, objective word for TF dimension, and rigid word for JP dimension. The best accuracy user MBTI classification result for IE dimension is 75.80%, NS dimension is 55.52%, TF dimension is 95.02% and JP dimension is 88.26%.

Index Terms: Personality Detection, Myers Briggs Type Indicator, Classification

1 Introduction

Personality of people being have their individual differences affects every aspect of their life. In the psychology field, the concept of personality is considered a powerful but imprecisely defined construct. Personality not only describe an individual preferences but also what they thinking and feeling about life, so that it affects our life [1]. Automatic detection of a person's personality traits has many important practical applications. Personality used in marketing field to knowing people preferences regarding certain products to provide different promotions to each person. While human resource company needs, personality preferences are the initial process of recruiting applicants to knowing ability to work together and collaborate as a team [2]. Education field, personality can used by Susilawati [3] as learning outcome. 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 one alternative linguistic approach that can be used to see the user's mental illness. Social networking platforms like Facebook, Instagram or Twitter have become an increasingly popular for individuals to share their ideas and emotions with each other. The way that an individual present himself/herself online reflects their attitude, behavior and personality [5].

There are some popular personality assessments namely, Big Five, DISC and MBTI. Big Five is dominant approach for representing human trait today [6]. Big Five model of personality is the most prominent Goldberg, McCrae & John. It posits five continuous personality traits: openness to experience (open-minded vs. traditional), conscientiousness (disciplined vs. disorganized), extraversion (outgoing vs. reserved), agreeableness (compassionate vs. antagonistic), and neuroticism (emotionally unstable vs. stable) [7]. DISC is a simple personality assessment personality introduced in 1928 [8]. This personality assessment divide into four different dimensions: Dominance, Influence, Stability and Conscientious, [2]. DISC personality assessment will help human to understand why someone does what he does because Model DISC based on two ideas: 1) we consider our environment to be good or unfavorable, and 2) a person considers himself more or less strong than his environment [9]. MBTI personality traits that described by Isabel Briggs Myers Katherine Cook Briggs, MBTI as a test used in job selection [10]. MBTI is most popular assessment for the process of recruitment, placement, even mutation because MBTI easy to use and the accuracy of the results [11]. MBTI divided human psychological functions based on 4 dimensions namely, Extrovert (E) vs Introvert (I), Sensing (S) vs Intuition (N), Feeling (F) vs Thinking (T), Judging (J) vs Perceiving (P)[11].

This works intend to exploration about word used each dimension of Myers-Briggs Type Indicator (MBTI) personality trait and use machine learning technique to classify tweet and user to MBTI personality type. Some preprocessing steps were performed to clean the noisy data and finally is mapped on MBTI personality. Some machine learning model was applied [5] as the initial benchmark personality recognition dataset to classify the text in this work into different personality traits such as Introversion-Extroversion (IE), Sensing-iNtuition (NS), Thinking- Feeling (FT) and Judging-Perceiving(JP) . This works consist of 5 sections. We start the introduction on section 1. Section 2 discuss about executive summary of previous work related with automated personality trait prediction. Section 3 provides data preprocessing used and methodology. Section 4 show result of performance and compare to others model also discussion. Finally, section 5 conclude with state of conclusion and future research.

2 Previous Work

A review of literature about previous personality detection from text is presented in this section. Studies about personality recognition using different social media platforms like Instagram, Facebook or Twitter. There are many models available for personality assessment, like MBTI [12][13], DISC Assessment, and Big Five Personality Traits etc. The literature studies of this work is categorized based personality type can show as TABLE 1:

- *Ahmad Fikri Iskandar is currently pursuing magister of informatics engineering of Amikom University Yogyakarta, Indonesia, E-mail: andfikri@gmail.com*
- *Ema Utami is currently lecture on magister of informatics engineering of Amikom University Yogyakarta, Indonesia. E-mail: ema.u@amikom.ac.id*
- *Agung Budi Prasetio is currently lecture on magister of informatics engineering of Amikom University Yogyakarta, Indonesia. E-mail: agungbp@amikom.ac.id*

TABLE 1
PREVIOUS WORK

Author	Model/ Approach	Result	Limitation or Future Work
MBTI Personality			
Bharadwaj et al [12]	SVM, Neural Network, Naïve Bayes	Best Accuracy: IE: 84.9% NS: 88.4% TF: 87.0% JP: 78.8%	There are no balancing method to solve unbalance data
Verhoeven et al [13]	SVM, Logistic Regression	Best F1-score: IE: 77.78 % NS: 79.21 % TF: 52.13 % JP: 47.01 %	In future, the model can be trained enough to predict all four dimensions of MBTI efficiently.
Lukito, P.H , et all [14]	Naïve Bayes	Best Accuracy: IE: 80 % NS: 60 % TF: 60 % JP: 60 %	Naive Bayes statistical model but needed compare result using other machine learning model
Fikry & Yusra [15]	SVM	Accuracy of IE 88%	This work provide increased from previous model NBC with 83.33 %, but we add more data other dimension of MBTI to enrich the data
Iskandar, et al[5]	Naïve Bayes KNN	Best Accuracy: IE: 81.25% NS: 84.62% TF: 84.55% JP: 75.00%	This data not balance, needed some method to solve it.
DISC Personality			
Hartanto, et al [16]	Naïve Bayes	Accuracy: 36.67% Precision: D:36.67% I:20% S:50% C=41.67%	Data looks like it's not balancing, and also Naive Bayes prediction focuses more on the D label resulting in an accuracy of 36.67
Utami, et al [2]	SVM	Accuracy: 37.41%,	Compare to other machine learning to better result performance classification and can use chi square as feature selection
Temperament Personality			
Sarwani & Mahmudy [17]	Naïve Bayes	Accuracy: 100%	data used only 20 user only show the result of the model NBC
Claudy, Y.I, et al [18]	KNN	Accuracy: 66%	comparing some word weighting and other machine learning models and Correcting non-standard words by taking a certain approach (Word Normalization)
Big Five Personality			
Ong, et all [19]	SVM XGBoost	Average Accuracy: SVM: 76.23% XGBoost=:97. 99%	compare Frequency and word weighting (TF-IDF)
Adi, et al [20]	Logistic Regression XGBoost SVD	Best perfomance: SGD: 99% XGB: 84.60% SL: 99.20%	Update and also use the stopword

Based on limitation on TABLE 1, this work conduct to using scenario to improve of classification of MBTI personality 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 that will discuss on Section 3.

3 METHODOLOGY

3.1 Data
This works uses data from [5], this data taken from social media Twitter and consist of three columns, namely text as tweet, MBTI label, and code user. For exploration we used data aggregation to process faster, and also we remove some outlier of data before classification.

3.2 Preprocessing Text
As we know the main problem on text mining is there are some not useful character like special character (#,@,!,$, etc) or link. We need to do some text processing before. Text preprocessing is the stage where select data so that the data becomes more structured. This work using Python to do text preprocessing. Generally, preprocessing almost using on text mining are case folding, tokenization, stemming and stop word removal. Case Folding: aims to convert all letters in a text document into a lower case. Tokenization: breaks the document into word parts called tokens. Stemming: aims to transform a word into a root word by removing all word affixes. Stop word removal: removes a set of words that are irrelevant to the main text, although these words often appear in the data used. and other steps, like slang, typo and word normalization.

3.3 Text Classification
Text classification is an important task in supervised learning. Naive Bayes classifier is the baseline and powerful on solve text classification [5]. It is very useful when there is limited memory. It is very fast to train and sometimes out performs other classifiers. Support Vector Machine (SVM) is similar to Logistic Regression. But it is very useful when the data is not linearly separable. It has been reported that SVM works best for text classification when steemed and weight the word [2]. Neural Network [12] uses the processing of the brain concept as a basic to develop algorithms that can be used to mathematically model to classification problems. The training module operates separately on the four classes of MBTI. Each class consists of two traits or labels. These two traits are mapped to binary 0 or 1. Count vector feature vector is generated for the data. The parameters of classifier are adjusted to obtain probabilities of MBTI traits in the test data. The classifier that use on this work is Decision Tree Classifier, K Neighbors Classifier, Linear Discriminant Analysis, Logistic Regression, Naive Bayes, Random Forest Classifier, SVM - Linear Kernel. The parameter scenario are 80:20 train test split, feature selection chi-square and balancing method using SMOTE.

3.4 Evaluation
Similarity with other research about automated personality recognition, evaluation of the classification models this work are accuracy, precision, recall, and f1-score. More details about the evaluation model can be shown in TABLE 2:

TABLE 2
EVALUATION MODEL

Evaluation	Formula	Description
Accuracy	(TN+TP)÷ (TP+FP+TN+FN)	Accuracy is used to evaluate the number of predictive labels that correspond to the actual label.
Precision	(TP)÷ (FP+FP)	Precision is the level of accuracy between the information requested by the user and the answer given by the system.
Recall	(TP)÷ (TP+FN)	Recall is the success rate of the system in rediscovering information.
F1-Score	2 x (precision x recall) ÷ (precision + recall)	F1 Score is the weighted average of Precision and Recall

Source: Willy [21]

Where:
True Positive (TP) is the true amount of label 0. True Negative (TN) is the true amount of label 1. False Positive (FP) is the sum of all column values for each class, except for the TP value. False Negative (FN) is the sum of all row values for each class, except the TP value [16].

4 RESULT & DISCUSSION

This section explain about result of exploration and classification. Exploration start from unique of word, amount of word, lexicon diversity, word cloud, some hypothesis testing each dimension. While classification of MBTI personality trait using feature selection, and balancing method.

4.1 Exploration

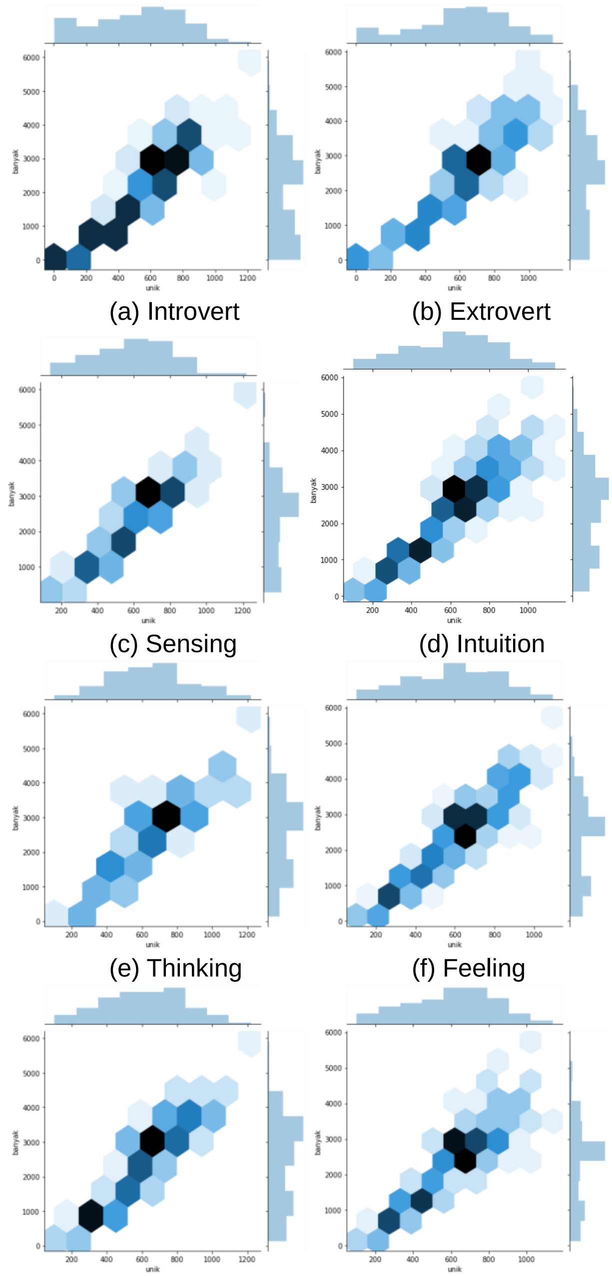
Total of sample after preprocessing is 281 users. Average lexicon diversity this sample is 3.6, means each user use one unique word 3 until 4 times. It also showed from average of unique words is 608 with standard deviation is ± 227 and number of words is ±2358 with standard deviation 1181. Detail of each dimension MBTI personality trait can be shown in TABLE 3:

TABLE 3
DESCRIPTION USER & WORD MBTI

Axis MBTI	NU	AUW	AAW	LD
Introvert	161	571	2202	3.6
Extrovert	120	657	2566	3.7
Sensing	75	594	2331	3.7
Intuition	206	612	2368	3.6
Thinking	68	649	2488	3.6
Feeling	213	594	2316	3.6
Judging	122	610	2407	3.7
Perceiving	159	606	2320	3.6

Note:
NU : Number of User
AUW : Average Unique Words
AAW : Average Amount Words
LD : Lexicon Diversity

Based TABLE 3, lexicon diversity around between 3.6 – 3.7, means even each dimension of MBTI also used one word 3-4 times. It is not be general for each word, but it is only simple summary of each MBTI. Comparison unique words and number of words used on each label dimension shown on Fig.1. below:



(a) Introvert (b) Extrovert
(c) Sensing (d) Intuition
(e) Thinking (f) Feeling
(g) Judging (h) Perceiving

***Fig. 1**. Unique words versus Number of Words*

Data visualization in this works using word clouds for concepts most prevalently words. In the case of words shared between both extractions for a given dimension, the extraction with the smaller number of instances of that word had all its instances of that word removed and the other extraction had the same number of instances of that word removed (in order to better capture the disproportionate use of certain words by certain types). The data visualization can be shown in Fig.2.:

(a) Introvert (b) Extrovert
(c) Intuition (d) Sensing
(e) Thinking (f) Feeling
(g) Judging (h) Perceiving

***Fig. 2**. Word cloud MBTI*

Based on Fig.2., we can see that only then are these word clouds created, where the size of each word is proportional to its appearance frequency in the respective extractions. We consider these word clouds to be illustrative of some unique ways different MBTIs use language.

4.2 Hypothesis

This works also do some hypothesis testing. This testing used for knowing different word between each axis on dimension. First, we do for normality test, and t-test. If the sample of data is not normal, so the hypothesis testing use Mann Whitney test.

Hypothesis 1:

H_0: There is no different people-related words between Extroverts and Introverts.

H_a: There is different people-related words between Extroverts and Introverts.

P-value of hypothesis 1 is 0.00224 or less than the significance level of alpha is 0.05. We reject the null hypothesis, so that we conclude that there is different people-related words between Extroverts and Introverts. We know using of people related word on this dimension for axis Introvert and Extrovert is different. People related word rarely using for introvert people than extrovert people because their introvert's activities most writing, reading and drawing than in activities which require them to act in an outgoing way like speaking [22].

Hypothesis 2:

H_0: There is no different counterfactual words between Intuition and Sensing.

H_a: There is different counterfactual words between Intuition and Sensing.

P-value of hypothesis 2 is 0.0434 or less than the significance level of alpha is 0.05. We reject the null hypothesis, so that we conclude that there is different counterfactual words between Sensing and Intuition. This second pair of MBTI, sensing more attention to information that comes in through their five sense. While intuition more attention to the patterns and possibilities that you see in the information you receive.

Hypothesis 3:

H_0: There is no different objective words between Thinking and Feeling.

H_a: There is different objective words between Thinking and Feeling.

P-value of hypothesis 3 is 0.02624 or less than the significance level of alpha is 0.05. We reject the null hypothesis, so that we conclude that there is different objective words between Thinking and Feeling. This third preference pair describes how you like to make decisions. Thinking put more weight on objective principles and impersonal facts. While feeling put more weight on personal concerns and the people involved.

Hypothesis 4:

H_0: There is no different rigid words between Judging and Perceiving.

H_a: There is different rigid words between Judging and Perceiving.

P-value of hypothesis 1 is 0.02624 or less than the significance level of alpha is 0.05.We reject the null hypothesis, so that we conclude that there is different rigid words between Judging and Perceiving. This fourth preference pair describes how you like to live your outer life what are the behaviors others tend to see. Judging axis prefer a more structured and decided lifestyle. While perceiving more flexible and adaptable lifestyle. This preference may also be thought of as their orientation to the outer world.

4.3 Classification

This section discusses test results of classification that has been built. This works do some machine learning model. To make different with previous work, we add some parameter namely, feature selection Chi-Square, and balancing method with SMOTE. This classification also using training-testing split 80:20 with k-fold validation 5. Evaluation of performance model using accuracy, recall, precision and F1-score.

DT : Decision Tree Classifier
KNN : K Neighbors Classifier
LDA : Linear Discriminant Analysis
LR : Logistic Regression
NB : Naive Bayes
RF : Random Forest Classifier
SVM : SVM - Linear Kernel

TABLE 4
PERFORMANCE IE DIMENSION

Model	IE Dimension			
	Accuracy	Recall	Precision	F1-Score
DT	64.40%	62.55%	74.31%	67.92%
KNN	64.88%	75.68%	69.20%	72.10%
LDA	68.57%	62.41%	81.08%	70.53%

LR	68.78%	62.36%	81.48%	70.64%
NB	56.22%	29.69%	92.72%	44.97%
RF	65.74%	65.26%	74.69%	69.66%
SVM	63.66%	57.49%	82.67%	62.72%

TABLE 4 show performance of IE dimension, best accuracy is Logistic Regression with 68.78%, best recall is KNN with 75.68%, best precision is Naïve Bayes with 92.72% and f1-score is KNN with 72.10%
.

TABLE 5
PERFORMANCE NS DIMENSION

Model	NS Dimension			
	Accuracy	Recall	Precision	F1-Score
DT	65.01%	63.74%	43.68%	51.84%
KNN	70.67%	36.56%	50.50%	42.40%
LDA	67.00%	75.91%	46.42%	57.60%
LR	67.88%	76.23%	47.30%	58.37%
NB	47.64%	96.03%	35.65%	52.00%
RF	66.09%	65.45%	44.92%	53.27%
SVM	71.78%	54.71%	60.44%	51.62%

TABLE 5 show performance of NS dimension, best accuracy is SVM - Linear Kernel with 71.78%, best recall is Naïve Bayes with 96.03%, best precision is 60.44% with SVM - Linear Kernel and f1-score is 58.37% with Logistic Regression.

TABLE 6
PERFORMANCE TF DIMENSION

Model	TF Dimension			
	Accuracy	Recall	Precision	F1-Score
DT	72.39%	45.70%	63.26%	53.05%
KNN	71.19%	40.19%	62.05%	48.76%
LDA	72.89%	60.53%	60.29%	60.40%
LR	71.71%	64.25%	58.25%	60.82%
NB	51.24%	93.03%	40.65%	56.58%
RF	73.12%	47.43%	64.46%	54.65%
SVM	63.43%	55.29%	68.19%	46.50%

TABLE 6 show performance of TF dimension, best accuracy is Random Forest Classifier with 73.12%, best recall is Naïve Bayes with 93.03%, best precision is SVM - Linear Kernel with 68.19% and f1-score is 60.82% with Logistic Regression.

TABLE 7
PERFORMANCE JP DIMENSION

Model	JP Dimension			
	Accuracy	Recall	Precision	F1-Score
DT	67.13%	78.24%	71.11%	74.50%
KNN	66.65%	81.89%	69.34%	75.09%
LDA	70.25%	76.92%	75.19%	76.04%
LR	70.13%	76.68%	75.17%	75.91%
NB	52.91%	26.09%	90.35%	40.46%
RF	68.50%	80.67%	71.62%	75.87%
SVM	68.32%	78.57%	74.88%	72.42%

TABLE 7 show performance of JP dimension, best accuracy is Linear Discriminant Analysis with 70.25%, best recall is Random Forest Classifier with 80.67%, best precision is Naïve Bayes with 90.35% and f1-score is 76.04% with Linear Discriminant Analysis. From the classification results in TABLE 4, TABLE 5, TABLE 6, and TABLE 7 above classification carried out on raw data. Furthermore, to find out personality type of MBTI personality, predict results must aggregate the raw data and find mode each user. The classification to predict raw data, we use best accuracy each dimension namely, IE dimensions use LRC, NS dimensions use SVM-Linear, TF dimensions use RF and JP dimensions use LDA. From the results of aggregation and find mode from label, the results of user classification with confusion matrix are obtained as follows Fig.3., Fig.4., Fig.5., and Fig.6. below:

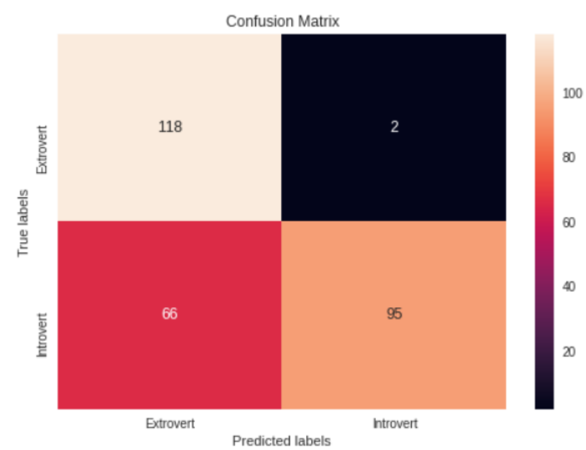


***Fig. 3.** Confusion Matrix User Classification IE Dimension*

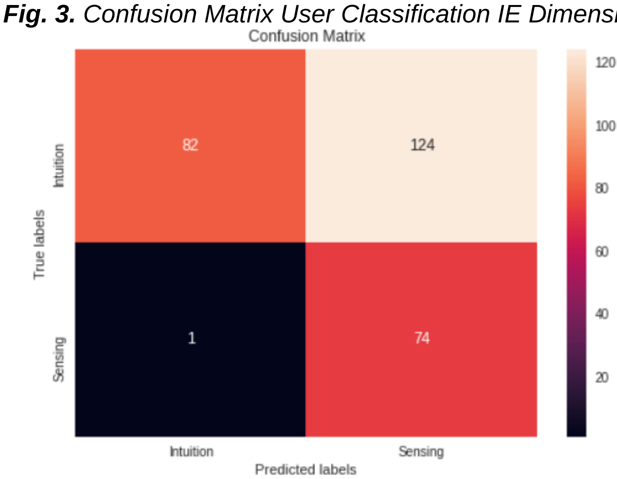


***Fig. 4**. Confusion Matrix User Classification NS Dimension*

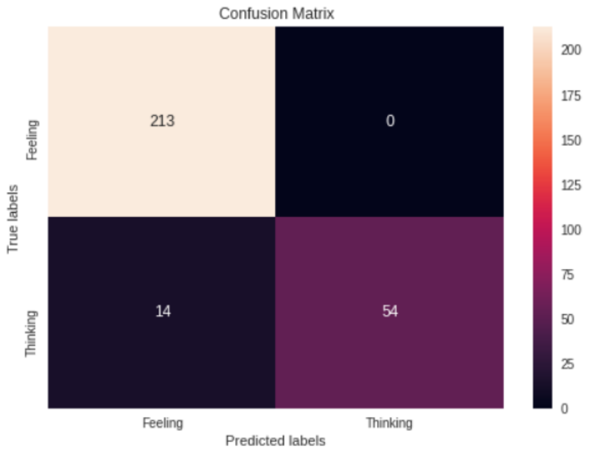


***Fig. 5**. Confusion Matrix User Classification TF Dimension*

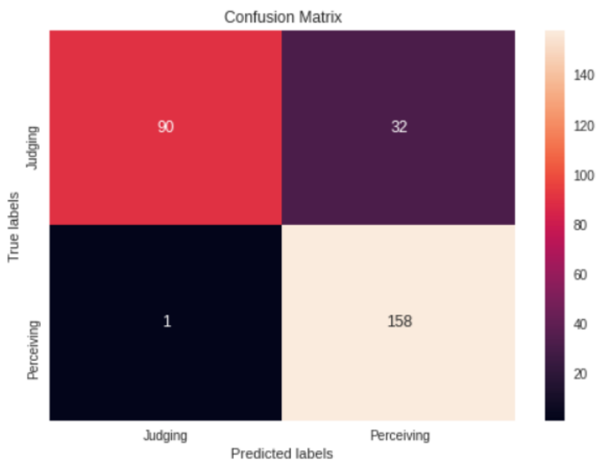


***Fig. 6**. Confusion Matrix User Classification JP Dimension*

From the Confusion matrix in Fig.3., Fig.4., Fig.5., and Fig.6. above we can see the accuracy of IE dimension is 75.80%, for NS dimension is 55.52%, for TF Dimension is 95.02% and last JP dimension is 88.26% user MBTI classification. As initial problem statement, we compare the performance classification with previous work of MBTI personality [5]. Comparison of performance classification can be shows in table 8:

TABLE 8
COMPARISON PERFORMANCE

Dimension	Best Accuracy	
	This Work	Iskandar [5]
Introvert-Extrovert	75.80%	81.25%
Intuition-Sensing	55.52%	84.62%
Thinking-Feeling	95.02%	84.55%
Judging-Perceiving	88.26%	75.00%

The comparison in TABLE 8 is a comparison of the user classification results obtained this work contributed increased accuracy of third and fourth dimension MBTI, Thinking-Feeling with 10.47% and Judging-Perceiving with 13.26%. This scenario using count vector, train-split 80:20, k-fold validation is 5 and balancing method using SMOTE.

5 CONCLUSION

In this paper, we present an exploration word of a personality assessment namely MBTI. From data, we first extract exploration words using text-mining methods. We got average lexicon diversity between 3.6 – 3.7 each axis. Next, there is different between each dimension about people related word for IE dimension, counterfactual word for NS dimension, objective word for TF dimension, and rigid word for JP dimension. The best accuracy user MBTI classification result for IE dimension is 75.80% with Logistic Regression model, NS dimension is 55.52% with SVM-Linear, TF dimension is 95.02% with RF and JP dimension is 88.26% with LDA. In future, we will utilize distributional contextual representations of the keywords each dimension and also determine the meaning of the word. This will provide psychologists with a vital tool to deeply study and interpret personality traits. Furthermore, to get better result of the classification to use deep learning.

REFERENCES

[1] T. Yılmaz, A. Ergil, and B. İlgen, “Deep Learning-Based Document Modeling for Personality Detection from Turkish Texts,” Adv. Intell. Syst. Comput., vol. 1069, pp. 729–736, 2020, doi: 10.1007/978-3-030-32520-6_53.

[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] A. F. Iskandar and E. Utami, “Impact of Feature Extraction and Feature Selection Using Naïve Bayes on Indonesian Personality Trait,” 2020 3rd International Conference On Information And Communications Technology (ICOIACT), Pending Publication.

[6] S. Roccas, L. Sagiv, S. H. Schwartz, and A. Knafo, “The Big Five personality factors and personal values,” Personal. Soc. Psychol. Bull., vol. 28, no. 6, pp. 789–801, 2002, doi: 10.1177/0146167202289008.

[7] S. C. Matz and G. M. Harari, “Personality-Place Transactions: Mapping the Relationships Between Big Five Personality Traits, States, and Daily Places,” J. Pers. Soc. Psychol., vol. 2, no. 999, 2020, doi: 10.1037/pspp0000297.

[8] N. Ahmad and J. Siddique, “Personality Assessment using Twitter Tweets,” Procedia Comput. Sci., vol. 112, pp. 1964–1973, 2017, doi: 10.1016/j.procs.2017.08.067.

[9] “The DiSC model - Theory and background - The story behind DiSC – it is older than you think The DiSC model : Theory and background The traditional theories,” pp. 1–16.

[10] D. Pittenger, “Measuring the MBTI and coming up short,” Consult. Psychol. J. Pract. Res., 2005, doi: citeulike-article-id:3171710.

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

[12] S. Bharadwaj, S. Sridhar, R. Choudhary, and R. Srinath, “Persona Traits Identification based on Myers-Briggs Type Indicator(MBTI) - A Text Classification Approach,” 2018 Int. Conf. Adv. Comput. Commun. Informatics, ICACCI 2018, pp. 1076–1082, 2018, doi:

10.1109/ICACCI.2018.8554828.

[13] B. Verhoeven, W. Daelemans, and B. Plank, "TwiSty: A multilingual Twitter stylometry corpus for gender and personality profiling," Proc. 10th Int. Conf. Lang. Resour. Eval. Lr. 2016, pp. 1632–1637, 2016.

[14] 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.

[15] 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.

[16] A. D. Hartanto, E. Utami, S. Adi, and H. S. Hudnanto, "Job seeker profile classification of twitter data using the naïve bayes classifier algorithm based on the DISC method," 2019 4th Int. Conf. Inf. Technol. Inf. Syst. Electr. Eng. ICITISEE 2019, pp. 533–536, 2019, doi: 10.1109/ICITISEE48480.2019.9003963.

[17] M. Z. Sarwani and W. F. Mahmudy, "Analisis Twitter Untuk Mengetahui Karakter Seseorang Menggunakan Algoritma Naïve Bayess Classifier," Semin. Nas. Sist. Inf. Indones., no. November, pp. 2–3, 2015.

[18] 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.

[19] 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.

[20] 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.

[21] 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.

[22] Z. Zainuddin, "The Impact of Personality: Extrovert vs. Introvert on the Ability in Syntax in Essay Writing," Stud. English Lang. Educ., vol. 3, no. 2, p. 162, 2016, doi: 10.24815/siele.v3i2.4963.