

## Personality Classification of Myers Briggs Type Indicators (MBTI) Using BERT and Machine Learning

Agus Sihabuddin<sup>\*1</sup>, Dian Ekapratwi<sup>2</sup>

<sup>1</sup>Department of Computer Science and Electronics, FMIPA UGM, Yogyakarta, Indonesia

<sup>2</sup>Bachelor Program of Computer Science, FMIPA UGM, Yogyakarta, Indonesia

e-mail: <sup>\*1</sup>a\_sihabudin@ugm.ac.id, <sup>2</sup>dianekapratwi72@mail.ugm.ac.id

### Abstrak

Klasifikasi kepribadian menggunakan data tekstual yang bersumber dari media sosial ataupun forum daring menimbulkan berbagai tantangan yang disebabkan oleh sifat teks yang tidak terstruktur dan kompleksitas penilaian kepribadian. Walaupun, model kepribadian Myers-Briggs Type Indicator (MBTI) menyediakan kerangka kerja yang komprehensif, mengadaptasinya ke data media dan menangani pola linguistik yang beragam memerlukan algoritma yang efektif. Selain itu, dasar psikologis model MBTI yang rumit diklasifikasi, terutama saat menggunakan metode yang kompleks seperti pembelajaran mendalam.

Penelitian ini mengklasifikasikan tipe kepribadian berdasarkan perilaku masing-masing individu pada forum online dengan memerhatikan pola linguistik dari data tekstual yang diunggah menggunakan algoritma SVM, Random Forest, BERT, dan Word2Vec. Algoritma SVM dan Random Forest merupakan algoritma pembelajaran mesin tradisional yang dikenal dengan kemampuan dan efektivitasnya dalam peran klasifikasi teks. Sedangkan, BERT dan Word2Vec digunakan untuk mengidentifikasi hubungan semantik dan informasi kontekstual dari data tekstual. Selain itu, untuk model BERT akan digunakan model IndoBERT karena penelitian ini berfokus pada klasifikasi teks berbahasa Indonesia.

Pengujian dilakukan dengan menggunakan data tekstual dari unggahan yang terdapat pada forum PersonalityCafe. Hasil pengujian menunjukkan bahwa kombinasi model SVM dan IndoBERT mengungguli model lainnya dengan tingkat akurasi sebesar 82% dan skor F1 sebesar 75%.

**Kata kunci**—BERT, MBTI, Klasifikasi Kepribadian, Random Forest, SVM

### Abstract

Personality classification using textual data from social media or online forums is a complex task due to the unstructured text and the multifaceted nature of personality. While the Myers-Briggs Type Indicator (MBTI) provides a comprehensive framework, adapting it to media data and handling diverse linguistic patterns requires effective algorithms. The psychological basis of MBTI is intricate, especially when using complex methods like deep learning, which can be challenging.

This study classifies personality types based on each individual's behavior on an online forum by observing the linguistic patterns of posted textual data using the SVM, Random Forest, BERT, and Word2Vec algorithms. The SVM and Random Forest algorithms are traditional machine learning algorithms known for their capabilities and effectiveness in text classification. Meanwhile, BERT and Word2Vec identify semantic relationships and contextual information from textual data. In addition, the IndoBERT model will be used for the BERT model because this study focuses on the classification of Indonesian language texts.

Testing was carried out using textual data from posts on the PersonalityCafe forum. The test results showed that the combination of the SVM and IndoBERT models outperformed other models with an accuracy rate of 82% and an F1 score of 75%.

**Keywords**—BERT, MBTI, Personality Classification, Random Forest, SVM

## 1. INTRODUCTION

In this digital era, social media or online forums have become the primary platform for users to share their emotions, opinions, or feelings in real time. An online forum is a web-based platform that brings together individuals with similar thoughts and interests, where forum members can submit their opinions in a post, interact, and receive feedback from other members [1]. Currently, many online forums are being developed, including the PersonalityCafe forum. Based on data listed on the PersonalityCafe forum website as of August 2024, there were 168.4 thousand members who joined the forum, with a total of 11 million posts. This forum focuses on various theories and personality types that are appropriate for analysis or study related to personality because they tend to have a variety of proper linguistic behaviors that can be further analyzed compared to posts on social media.

Currently, user activity on social media can be used to find out the personality of the account user. This is usually done to study the behavior, thoughts, or feelings of individuals that can be used in the recruitment process, choosing the right career, and others [2]. Generally, the personality assessment process is carried out by reflecting on existing personality models. One of them is the MBTI model, which uses a questionnaire as its assessment system. However, this assessment system can also be done by utilizing existing technology, namely by using computer language to observe the linguistic patterns in someone's writing.

The supervised learning algorithm can classify personality models from textual data. Supervised learning is a learning model that utilizes labeled datasets to train algorithms so that the algorithm can classify or predict data accurately. There are two algorithm models in supervised learning: classification and regression. In the classification algorithm, the output results are expected to be discrete. Meanwhile, the regression algorithm's output results are continuous or real-valued. Researchers often use classification algorithms such as SVM and Random Forest (RF) when identifying personality. For example, [3] used SVM to develop a classifier to determine a person's personality based on Facebook account activity without requiring a questionnaire. In developing a classifier model, algorithm performance can be optimized with various strategies, including feature extraction, cross-validation, ensemble learning, and handling class imbalance. Several researchers have carried out this optimization strategy, one of which is [4], who developed a deep learning-based approach model to automatically detect personality from text by applying Bagged-SVM combined with Bidirectional Encoder Representations from Transformer (BERT). Applying the BERT model for feature extraction yields higher accuracy in identifying personality traits from text compared to models using non-contextualized word representations. In addition, handling class imbalance in the dataset can be overcome by using oversampling techniques, one of which is the Synthetic Minority Oversampling Technique (SMOTE), which was applied by [2].

### *1.1 MBTI Personality Type*

In the early 1940s, a mother and daughter, Katherine Briggs and Isabel Briggs Myers embarked on a journey to develop the Myers-Briggs Type Indicator (MBTI) personality model. The journey began because it was motivated by Katherine's learning about personality type theory after focusing on Carl Jung's work, "Psychological Type." The MBTI, the result of Briggs and Myers' innovation, describes the personality model through four main aspects that are dichotomous or opposing [5]. Each of these dichotomies displays contrasting characteristics that indicate how individuals interact with the environment around them. Here are the four main aspects or characters that make up the MBTI model:

1. Introvert (I) / Extrovert (E). This dichotomy refers to where individuals reflect their energy. Introverts prefer solitude and minimal social interaction, while extroverts thrive in social settings and enjoy engaging with others.
2. Sensing (S)/Intuition (I). This dichotomy reflects on how individuals process information. Sensing focuses on concrete, practical, and realistic details, while intuition emphasizes abstract concepts, patterns, and future possibilities. Sensing types prioritize present, actionable solutions, whereas intuitive types are innovative and see future outcomes.
3. Thinking (T)/Feeling (F). This dichotomy reflects on decision-making approaches. Thinkers prioritize logic and analysis, making firm, principle-based decisions. Feelers rely on empathy and emotions to create harmony and balance in relationships.
4. Judging (J) / Perceiving (P). This dichotomy reflects how individuals manage daily life. Judging types prefer systematic planning and dislike deviations from their plans, while Perceiving types are more flexible and spontaneous and adapt easily to unexpected changes.

Based on the explanation above, the MBTI personality type is composed of four characteristics found in the four MBTI classes. From the characters found in the MBTI class, 16 MBTI personality types will be formed, as shown in Figure 1.

<b>MBTI PERSONALITY TYPE</b>			
<b>Extroverts</b> <b>Introverts</b>	<b>Sensors</b> <b>Intuitives</b>	<b>Thinkers</b> <b>Feelers</b>	<b>Judgers</b> <b>Perceives</b>
ENTP	ENFP	INTP	INFP
ENTJ	ENFJ	INTJ	INFJ
ESTP	ESFP	ISTP	ISFP
ESTJ	ESFJ	ISTJ	ISFJ

Figure 1 MBTI Personality Types

### 1.2 Literature Review

Personality classification is a field of combined technology and psychology that categorizes individuals based on their psychological characteristics. The Myers-Briggs Type Indicator (MBTI) is the most popular model in personality classification. MBTI categorizes individuals into 16 different personality types based on four classification classes, namely Introvert (I) and Extrovert (E), Sensing (S) and Intuition (N), Thinkers (T) and Feelers (F), and Judgers (J) and Perceivers (P). By utilizing the personality model, several researchers have focused on the application of supervised learning methods to text-based data, especially from social media platforms, to classify personality types [3], [4], [7], [8], [9], [10], [11]. This study was conducted based on the MBTI personality type of each social media user that had previously been collected through a questionnaire or based on information shared by the account owner [12]. Furthermore, [12] uses the same dataset as the dataset used in this study, namely Mitchell J's MBTI dataset obtained from Kaggle.

The supervised learning method is often used in the field of text classification. This is because the way this method works is to train the model by training a labeled dataset to learn the mapping function between each variable [10]. In addition to using supervised learning methods, both machine learning and deep learning, researchers usually combine these methods with other

models to get more optimal results with various strategies, such as feature extraction, cross-validation, ensemble learning, and handling class imbalance. These researchers modify the existing model architecture to obtain a more effective and optimal model.

Many researchers apply optimization algorithms as a strategy for optimizing classifier models. The first optimization strategy is to perform feature extraction, which can be done with the BERT model as done by [4] and [2], or with the Word2Vec model as done by [8] and [10], as well as several other feature extraction models applied by [9], [10], [11]. The second optimization strategy is to perform cross-validation on the classifier model to be tested. This cross-validation process can be done using the K-Fold Cross Validation model by [2] and [10]. Furthermore, ensemble learning stages can be carried out to optimize the model, one of which is the application of Bagged-SVM [4]. The last strategy is to handle unbalanced classes or imbalanced data, one of which is using MLSMOTE as carried out by [2]. A summary or outline of each study can be seen in Table 1.

Table 1 Previous Research on Personality Type Prediction

Study	Method	Personality Model
[3]	SVM	Big-Five
[4]	SVM, BERT	Big-Five
[2]	XGBoost, BERT, MLSMOTE, K-Fold	Big-Five
[7]	LDA, SVM, LR, MLP	MBTI
[8]	RF, LR, KNN, SVM, Word2Vec	MBTI
[9]	RF, XGBoost, SGD, LR, KNN, SVM, CountVectorizer	MBTI
[10]	CNN, LSTM, Word2Vec, fastText, GridSearchCV, K-fold	Big-Five
[12]	XGBoost	MBTI
[11]	TF-IDF, BOW, KNN, MNB, LR, DT, SVM	Big-Five

## 2. METHODS

In this study, model development was carried out using two types of machine learning algorithms: SVM and Random Forest. Both algorithms were tested using two word embedding models: Word2Vec and BERT. The Word2Vec model uses the Skip-Gram architecture in its application. The personality classification to be carried out is included in the multi-label classification type of text data, where the labels consist of 16 personality types. Each personality type of the MBTI model combines four opposing aspects, as mentioned in Section 1.1.

### 2.1 Implementation Tools

This study will use several tools, including an Intel(R) Core(TM) i5-8250U CPU @ 1.60GHz (8 CPUs) with a speed of ~1.80GHz, 8 GB of RAM, and a Windows 11 operating system (64-bit). For software, the study will be conducted through Google Colab and utilize several Python libraries, such as TensorFlow, PyTorch, Seaborn, NLTK, Scikit-learn, Sastrawi, Transformers, Gensim, and Googletrans.

### 2.2 Research Data

This study will use two datasets: embedding training data and classification model training/testing data. The Indonesian Wikipedia database, known as the Indonesia Wikimedia Dump, is utilized for word embedding. This database contains regularly updated structured text, including articles and metadata, which helps improve the word embedding model by introducing new words over time, enhancing contextual accuracy.

For MBTI personality classification, the Mitchell J. Myers-Briggs dataset, sourced from the PersonalityCafe forum, is used. It includes data from 8,600 individuals, covering their MBTI

type and last 50 posts. This dataset is translated into Indonesian to focus on classifying Indonesian text.

### 2.3 Research Stage

This section outlines the stages of researching MBTI model classification based on textual data. First, datasets are collected from the Kaggle platform. Next, data preprocessing prepares the text for classification through tokenization, filtering, stemming, standardization, and normalization. Additionally, word embeddings are generated using Word2Vec and BERT to convert text into vector representations for machine learning algorithms.

After the preprocessing stage, the dataset is labeled, and classification models using SVM and Random Forest are designed. This stage includes 5-fold cross-validation for model validation and training with the processed data. After training, the model testing phase uses prepared test data. Finally, each model is evaluated using a confusion matrix to assess accuracy, error rate, precision, recall, and F-measure, identifying the best-performing model.

The stages of the research are illustrated in a diagram, which can be seen in Figure 2.

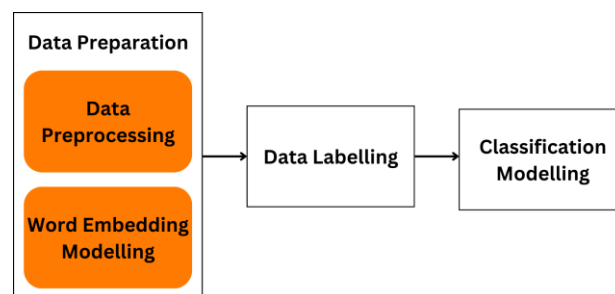


Figure 2 Research Stage

### 2.4 Data Preprocessing

Data preprocessing is an important stage in the research process in text mining, where raw data is processed into more structured data. This stage aims to improve model performance and accuracy and speed up the training process. In this study, the data preprocessing stages that will be carried out consist of case folding, tokenization, filtering, stemming, standardization, and normalization. These stages can be seen in Figure 3.

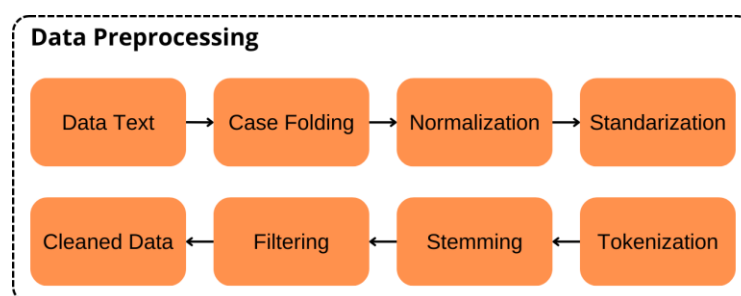


Figure 3 Data Preprocessing Stage

### 2.5 Data Labelling

After the preprocessing stage, labeling will be done on the training data, and the classification model will be tested. Data labeling will be carried out based on the labels of the 16 MBTI types contained in the data into four MBTI classes, where the class is dichotomous or

consists of two characters that conflict with each other, as explained in section 1.1. The dataset has 8600 rows, where each row represents one user, two columns containing the MBTI type of each user, and the last 50 uploads made by the user. Figure 4 visualizes the distribution of the sixteen MBTI types included in the dataset. Labeling will then be carried out on the data into four MBTI classes, where the division is based on the four letters contained in the MBTI type column.

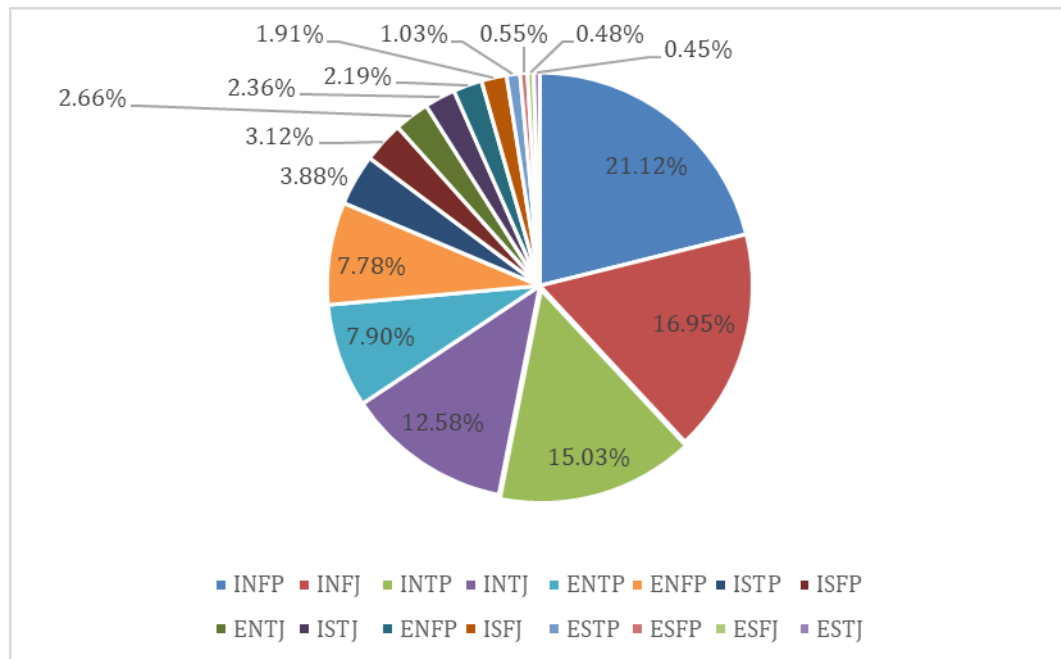


Figure 4 Visualization of MBTI Type Distribution in Classification Dataset

## 2.6 Word2Vec

Word2Vec is one of the word embedding models developed by Google that utilizes neural networks to learn the meaning of words from text data. This allows the model to create a numeric representation, or vector, for each word and capture the relationships between the words. In this study, the Skip-Gram architecture will be used, which can analyze the context around a particular word in a predetermined category to predict the meaning of the words around it. For example, Figure 5 illustrates the architecture of the Skip-Gram model in processing input data, namely "The wide road shimmered in the hot sun," into a vector. Furthermore, vector dimensions ranging from 50 to 300 will be utilized (50, 100, 200, 300).

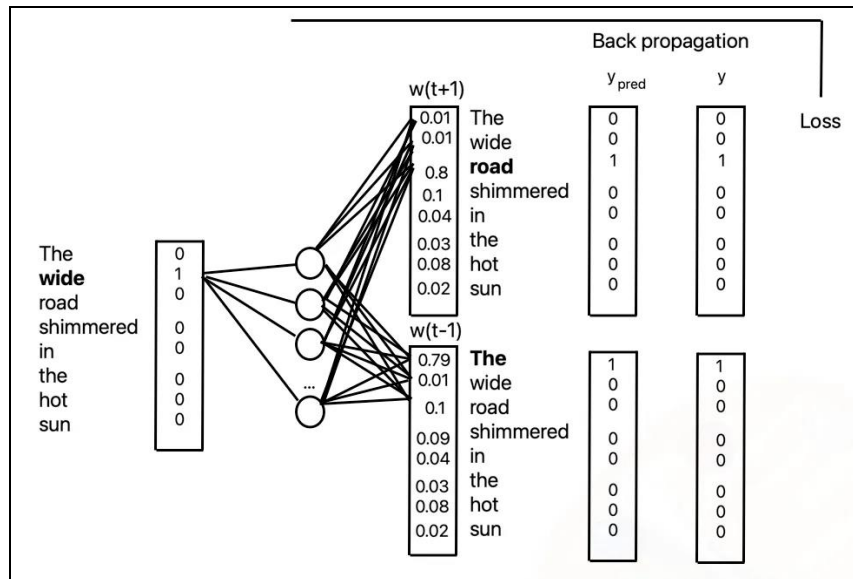


Figure 5 Illustration of the Word2Vec Process (Marklin, C., 2022)

### 2.7 BERT

BERT is a word embedding model that utilizes transformers. In its application, BERT uses a bidirectional approach, considering the entire sentence simultaneously and analyzing features before and after a particular word to understand its meaning. For example, in the sentence "He wore a bright yellow scarf and a gloomy expression," a unidirectional model might focus on the words before ("He wore a bright yellow scarf"). Based on this context, it might predict a positive word like "smile" or "happy." However, considering the entire sentence bidirectionally, BERT can analyze the contrast between "bright yellow scarf" and "gloomy expression." This allows it to understand that the expression is most likely the opposite of the brightness of the scarf. The process of how BERT works is illustrated in Figure 6, with the final output being a series of vectors of size 768.

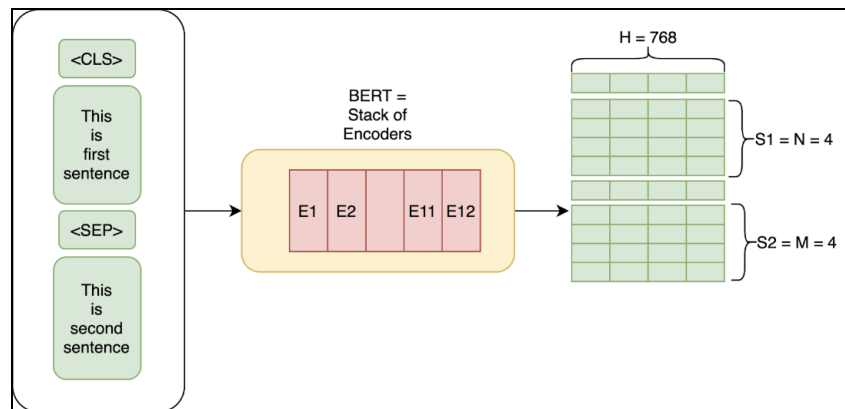


Figure 6 Illustration of the BERT Process (Vu, K., 2021)

### 2.8 SVM

The SVM architecture applied in this study consists of three parts, there are input layer, hyperparameter tuning, and output layer. The input layer contains feature vectors from data converted into numeric format. This process produces vectors that represent semantic and syntactic information from text data. Hyperparameter tuning includes kernel functions, such as

RBF, which calculate the similarity between two data points based on the Gaussian function of the Euclidean distance, as written in Equation (1). The output layer is the stage where the SVM predicts characteristics or personality types based on text data, referring to characters from the four MBTI classes. This determines the personality class that the system can expect from a given text sample. The architecture of the SVM classification model can be seen in Figure 7.

$$K(x, x') = \exp\left(-\frac{|x-x'|^2}{2\sigma^2}\right) \quad (1)$$

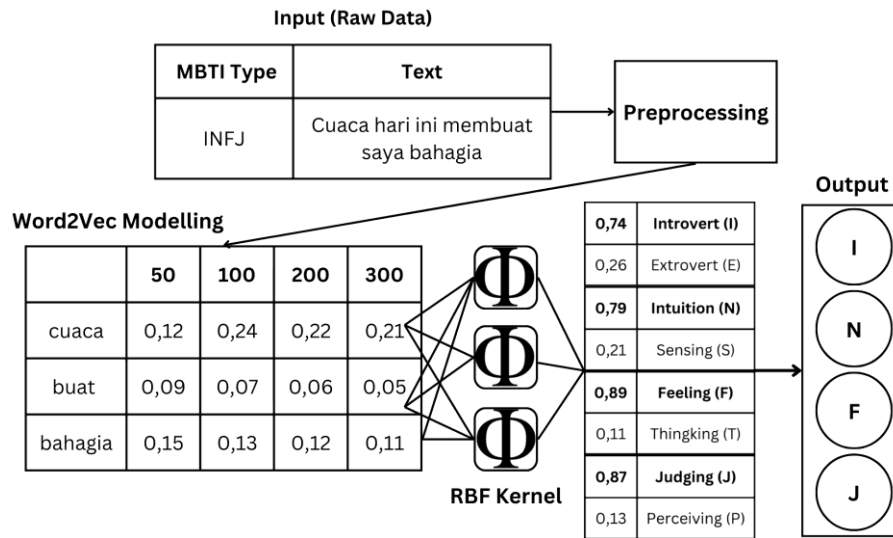


Figure 7 SVM Architecture Overview in Personality Classification (with Word2Vec)

## 2.9 Random Forest

Random Forest is an ensemble learning algorithm that combines multiple Decision Trees to produce more accurate predictions. The architecture of this model consists of three layers: input layer, decision tree layer, and output layer, as seen in Figure 8. In the first layer, text data is collected and prepared to be processed into features such as word frequency, sentiment scores, or syntactic patterns. The second layer, the decision tree, contains several decision trees that separate the data based on these features until leaf nodes are formed that represent data with similar characteristics. Each tree then provides a prediction related to a particular personality type, such as introvert (I) or extrovert (E). In the last layer, the predictions from each tree are combined using a majority voting system, where the majority prediction of the trees determines the final result. For example, if most trees predict introversion, the classification result is introversion.



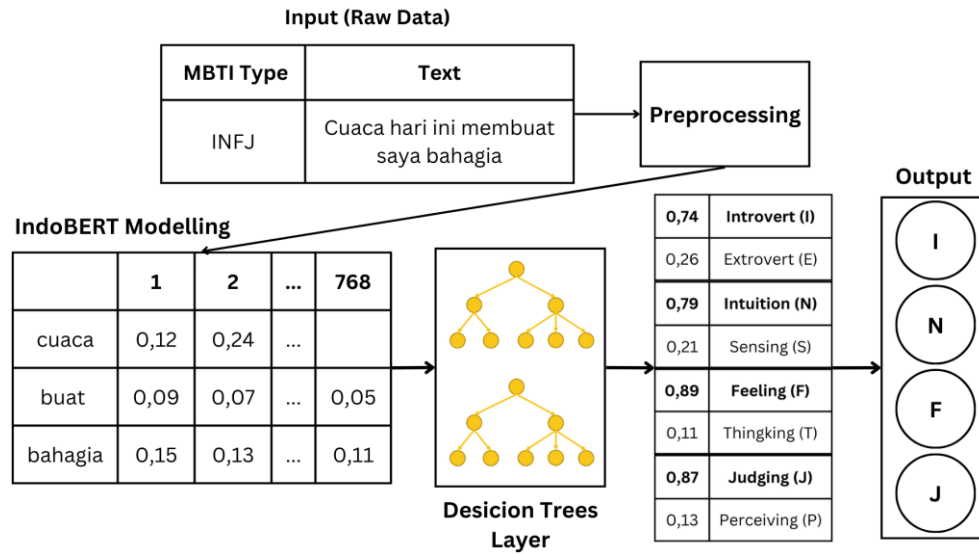


Figure 8 Random Forest Architecture Overview in Personality Classification (with BERT)

### 2.10 Confusion Matrix

Confusion matrix analysis explains the relationship between different features and data objects by revealing the inherent structure of the data set itself [13]. The confusion matrix consists of predicted and actual values, allowing the identification of True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). It helps calculate key metrics like accuracy, error rate, precision, recall, and F-measure, providing a detailed assessment of model performance.

## 3. RESULTS AND DISCUSSION

The analysis of the performance of the classification model was carried out by comparing the evaluation results of the models used in this study, namely the SVM model with Word2Vec, the SVM model with IndoBERT, the Random Forest model with Word2Vec, and the Random Forest model with IndoBERT, as well as with previous studies. From the research, overall, the SVM model with IndoBERT produced a better accuracy level than other models. Table 2 shows the average performance of each classification model based on the four MBTI classes in the form of an average of the evaluation metrics.

Table 2 Results of Classification Model Performance

Model	Acc	Error	Prec	Recall	F1
SVM + Word2Vec50	0.75	0.24	0.48	0.57	0.52
SVM + Word2Vec100	0.76	0.22	0.58	0.60	0.56
SVM + Word2Vec200	0.79	0.21	0.68	0.61	0.59
SVM + Word2Vec300	0.80	0.20	0.71	0.62	0.61
SVM + IndoBERT	<b>0.81</b>	0.19	0.79	0.70	0.72
Random Forest + Word2Vec50	0.76	0.23	0.74	0.59	0.58
Random Forest + Word2Vec100	0.76	0.24	0.74	0.59	0.58
Random Forest + Word2Vec200	0.77	0.23	0.74	0.60	0.58
Random Forest + Word2Vec300	0.77	0.23	0.73	0.60	0.59
Random Forest + IndoBERT	<b>0.77</b>	0.23	0.65	0.60	0.59

Based on the testing process, the error convergence graph of the SVM and Random Forest models is visualized, which can be seen in Figure 9 and Figure 10. The graph is used to compare

the error rate of the SVM and Random Forest models with Word2Vec (50, 100, 200, and 300) and IndoBERT based on four MBTI classes. In both graphs, it can be seen that the higher the MBTI class, the higher the error rate also increases. The Word2Vec model with smaller vector dimensions, namely 50 and 100, shows the highest error rate compared to more significant vectors, namely 200 and 300. However, the performance of Word2Vec is still less effective when compared to the IndoBERT model, which has a lower error rate in almost every class except class 3. The test results in the Random Forest model show a similar pattern, where models with smaller vector sizes perform worse than those with larger vector sizes. Although IndoBERT outperforms Word2Vec in classes 1, 2, and 4, the model suffers from increased errors in class 3.

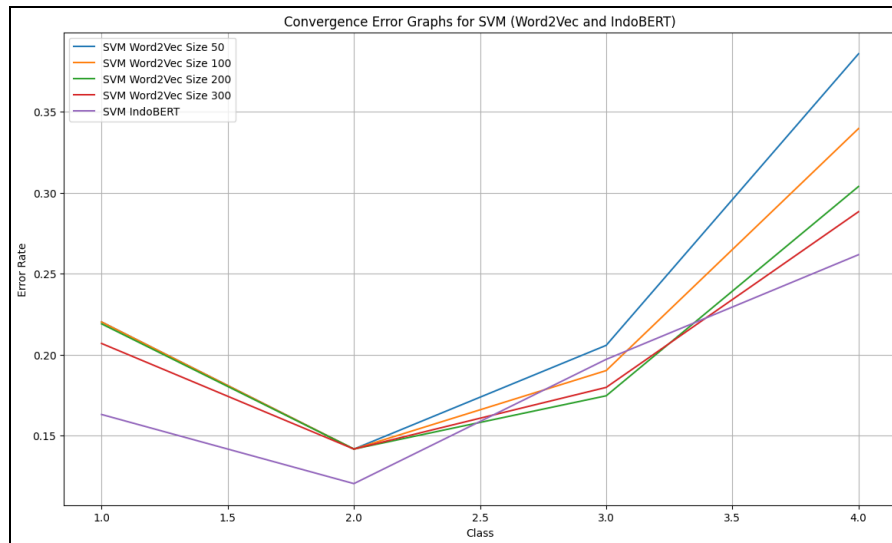


Figure 9 SVM Error Convergence Graph

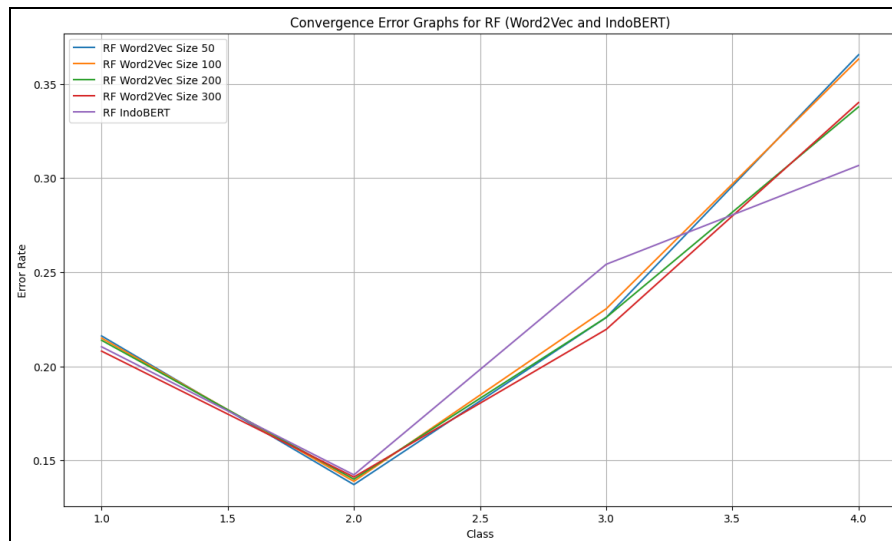


Figure 10 Random Forest Error Convergence Graph

[12] conducted a study using the same dataset as this study, namely Mitchell J.'s Myers-Briggs Personality Type dataset. In their study, [12] also conducted a study based on the four MBTI classes, namely IE (1<sup>st</sup> class), NS (2<sup>nd</sup> class), FT (3<sup>rd</sup> class), and JP (4<sup>th</sup> class), using the

XGBoost classification model. Based on the results obtained, the classification model in this study received a better level of accuracy overall than [12]. The difference between the accuracy levels of the classification models in the two studies can be observed in Table 3, which shows that the SVM model with IndoBERT is superior in three MBTI classes (IE, NS, JP) and the SVM model with Word2Vec300 is superior in one class (FT).

Table 3 Comparison of Model Accuracy Level with Previous Research

	IE	NS	FT	JP
XGBoost	0.78	0.86	0.71	0.65
SVM + Word2Vec50	0.78	0.86	0.80	0.61
SVM + Word2Vec100	0.78	0.86	0.81	0.66
SVM + Word2Vec200	0.78	0.86	0.82	0.70
SVM + Word2Vec300	0.79	0.86	<b>0.83</b>	0.71
SVM + IndoBERT	<b>0.84</b>	<b>0.88</b>	0.80	<b>0.74</b>
Random Forest + Word2Vec50	0.78	0.86	0.76	0.64
Random Forest + Word2Vec100	0.79	0.86	0.78	0.65
Random Forest + Word2Vec200	0.79	0.86	0.77	0.67
Random Forest + Word2Vec300	0.79	0.86	0.77	0.66
Random Forest + IndoBERT	0.79	0.86	0.76	0.68

#### 4. CONCLUSIONS

Based on a series of research stages that have been carried out, several conclusions were obtained. First, overall, the combination of the SVM classification model and IndoBERT word embedding showed superior performance in classifying MBTI personalities, with an accuracy rate of 82% and an F1 score of 75%. Second, based on the results obtained, it can be concluded that in applying the word embedding model, BERT has advantages over Word2Vec. This advantage is due to BERT's ability to manage contextual information through the transformer architecture better.

#### REFERENCES

- [1] Biriya, A.H. dan Thomas, E.V., 2014, Online Discussion Forum: A Tool for Effective Student-Teacher Interaction, *International Journal of Applied Sciences*, 1, 3, 111-116.
- [2] Lucky, H., Roslynia, dan Suhartono, D., 2021, Towards Classification of Personality Prediction Model: A Combination of BERT Word Embedding and MLSMOTE, *2021 1st International Conference on Computer Science and Artificial Intelligence (ICCSAI)*, 1, 346-350.
- [3] Utami, N. A., Maharani, W., dan Atastina, I., 2021, Personality Classification of Facebook Users According to Big Five Personality Using SVM (Support Vector Machine) Method, *Procedia Computer Science*, 1, 179, 177-184.
- [4] Kazameini, A., Fatehi, S., Mehta, Y., Eetemadi, S. dan Cambria, E., 2020, Personality Trait Detection Using Bagged SVM over BERT Word Embedding Ensembles. Available: <https://arxiv.org/abs/2010.01309>. [Accessed: 18-Jan-2024]
- [5] Murdrika, N., 2014, MBTI (Myer Briggs Type Indicator). Available: <http://dewihardiningtyas.lecture.ub.ac.id/files/2012/04/mbti.pdf>. [Accessed: 14-Mar-2024]
- [6] Gjurković, M. dan Šnajder, J., 2018, Reddit: A Gold Mine for Personality Prediction, *Proceedings of the Second Workshop on Computational Modeling of People's Opinions, Personality, and Emotions in Social Media*, New Orleans.

- [7] Abidin, N.H.Z., Akmal, M., Mohd, N., Nincarean, D., Yusoff, N., Karimah, H., dan H, A., 2020, Improving Intelligent Personality Prediction using Myers-Briggs Type Indicator and Random Forest Classifier, *International Journal of Advanced Computer Science and Applications*, 11, 11, 192-199.
- [8] Zumma, Md. T., Munia, J. A., Halder, D., dan Rahman, Md. S., 2022, Personality Prediction from Twitter Dataset using Machine Learning, *2022 13th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, 1-5.
- [9] Apriansyah, R., 2021, KLASIFIKASI KEPERIBADIAN PENGGUNA TWITTER BERDASARKAN DATA TWEET MENGGUNAKAN DEEP LEARNING, *Tesis*, Jurusan Ilmu Komputer FMIPA UGM, Yogyakarta.
- [10] Alsubhi, S.M., Alhothali, A.M., dan AlMansour, A.A., 2023, AraBig5: The Big Five Personality Traits Prediction Using Machine Learning Algorithm on Arabic Tweets, *IEEE Access*, 11, 112526–112534.
- [11] Amirhosseini, M.H. dan Kazemian, H., 2020, Machine Learning Approach to Personality Type Prediction Based on the Myers–Briggs Type Indicator®, *Multimodal Technologies and Interaction*, 4, 9.
- [12] Marklin, C., 2022, Word2Vec — Skip-Gram, <https://medium.com/@corymaklin/word2vec-skip-gram-904775613b4c>, [Accessed: 5-Feb-2024]
- [13] Vu, K., 2021, BERT Transformers: How Do They Work?, <https://dzone.com/articles/bert-transformers-how-do-they-work>, [Accessed: 3-Feb-2024]
- [14] Kristinic, D., Braović, M., Šerić, L., Božić-Štulić, D., 2020, Multi-Label Classifier Performace Evaluation with Confusion Matrix, *International Conference on Soft Computing, Artificial Intelligence, and Machine Learning (SAIM 2020)*, 1-14.