

## INTISARI

### **KLASIFIKASI *MYERS-BRIGGS TYPE INDICATOR* BERDASARKAN DATA TEKSTUAL BERBAHASA INDONESIA MENGGUNAKAN *BERT* DAN *MACHINE LEARNING***

Oleh

Dian Ekapratiwi  
20/459262/PA/19923

Klasifikasi kepribadian menggunakan data tekstual yang bersumber dari media sosial ataupun forum daring menimbulkan tantangan yang perlu diperhatikan. Tantangan tersebut disebabkan oleh sifat teks yang tidak terstruktur dan kompleksitas penilaian kepribadian. Meskipun model *Myers-Briggs Type Indicator* (MBTI) memberikan kerangka komprehensif untuk klasifikasi kepribadian, mengadaptasinya ke data media sosial memerlukan algoritma efektif yang mampu menangani pola linguistik yang beragam.

Penelitian ini menganalisis dan mengklasifikasikan tipe kepribadian berdasarkan perilaku masing-masing individu pada forum *online* dengan memerhatikan pola linguistik dari data tekstual yang diunggah menggunakan algoritma *support vector machines* (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 8600 individu pada forum PersonalityCafe, yang mana akan diambil 50 unggahan terakhir dari masing-masing individu dan menerjemahkannya ke bahasa Indonesia. Hasil pengujian pada kedua model menunjukkan bahwa kombinasi model SVM dan IndoBERT mengungguli Random Forest dan Word2Vec maupun IndoBERT. Model SVM mendapat hasil dengan tingkat akurasi sebesar 82% dan skor F1 sebesar 75%.

**Kata Kunci:** BERT, Klasifikasi kepribadian, MBTI, *Random Forest*, SVM

## ABSTRACT

### CLASSIFICATION OF MYERS-BRIGGS TYPE INDICATOR BASED ON TEXTUAL DATA IN INDONESIAN LANGUAGE USING BERT AND MACHINE LEARNING

By

Dian Ekapratwi  
20/459262/PA/19923

Personality classification using textual data from social media or online forums poses challenges that need to be considered. These challenges are caused by the unstructured nature of text and the complexity of personality assessment. Although the Myers-Briggs Type Indicator (MBTI) model provides a comprehensive framework for personality classification, adapting it to social media data requires an effective algorithm that can handle diverse linguistic patterns.

This study analyzes and classifies personality types based on the behavior of each individual on an online forum by observing the linguistic patterns of uploaded textual data using the support vector machines (SVM), Random Forest, BERT, and Word2Vec algorithms. The SVM and Random Forest algorithms are traditional machine learning algorithms known for their robustness 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.

The test was carried out using textual data from 8600 individuals on the PersonalityCafe forum, where the last 50 posts from each individual will be taken and translated into Indonesian. The test results on both models show that the combination of the SVM and IndoBERT models outperforms Random Forest and Word2Vec as well as IndoBERT. The SVM model produces results with an accuracy rate of 82% and an F1 score of 75%.

**Keywords:** BERT, MBTI, Personality classification, Random Forest, SVM