

Prediksi Kepribadian *Big Five* Pengguna Twitter Menggunakan Metode *Decision Tree* dengan Pendekatan Semantik *BERT*

Jammie Reyhan Widyanto, Dr. Erwin Budi Setiawan S.Si., M.T.

Fakultas Informatika, Universitas Telkom, Bandung

jammiereyhan@students.telkomuniversity.ac.id
erwinbudisetiawan@telkomuniversity.ac.id

Abstrak

Kepribadian individu dapat dilihat dengan mudah pada zaman ini. Ada beberapa pendekatan dalam mengklasifikasi kepribadian, salah satunya kepribadian *big five*. Kepribadian *big five* terdiri dari 5 dimensi, yaitu *Openness*, *Conscientiousness*, *Extraversion*, *Agreeableness*, dan *Neuroticism*. Salah satu cara mengetahui kepribadian individu dapat dilihat dari media sosialnya, karena zaman ini hampir semua individu mempunyai media sosial. Salah satu media sosial yang masih ramai digunakan adalah Twitter. Twitter adalah media sosial yang berisi cuitan-cuitan atau *tweet* dari tiap individu yang maksimal 280 karakter per-*tweet*. Sudah ada beberapa penelitian terkait kepribadian *big five* dari pengguna twitter. Berdasarkan permasalahan penelitian kepribadian *big five* sebelumnya, maka pada penelitian ini dilakukan prediksi kepribadian *big five* pengguna twitter dengan menggunakan metode *Decision Tree Classification And Regression Tree (CART)*, *Term Frequency – Invers Document Frequency (TF-IDF)*, *Synthetic Minority Oversampling Technique (SMOTE)*, *Linguistic Inquiry Word Count (LIWC)*, dan *Bidirectional Encoder Representations from Transformers (BERT)*. Penelitian bertujuan untuk mengetahui penerapan metode-metode yang digunakan pada penelitian ini terhadap prediksi kepribadian *big five* dan untuk mendapatkan hasil akurasi yang lebih baik dari penelitian sebelumnya. Data yang didapatkan pengguna twitter berjumlah 315 pengguna twitter dan 672.866 *tweet* yang diperoleh dari survey dan telah dilabeli kepribadian *big five*, dihasilkan akurasi sebesar 97,62% dari baseline dengan kenaikan 23.1%, dengan menerapkan metode *CART+TF-IDF+SMOTE+LIWC+BERT*. Kata kunci: *CART*, Kepribadian Big Five, *LIWC*, *SMOTE*, *BERT*, Twitter.

Abstract

Individual personality can be seen easily in this day. There are several approaches in classifying personality, one of which is the big five personality. The big five personality consists of 5 dimensions, namely *Openness*, *Conscientiousness*, *Extraversion*, *Agreeableness*, and *Neuroticism*. One way of knowing an individual's personality can be seen from their social media, because today almost all individuals have social media. One of the social media that is still widely used is Twitter. Twitter is a social media that contains tweets from each individual with a maximum of 280 characters per tweet. There have been several studies related to the big five personalities of Twitter users. Based on previous big five personality research problems, this study carried out predictions of the big five personalities of Twitter users using the *Decision Tree Classification And Regression Tree (CART)*, *Term Frequency – Inverse Document Frequency (TF-IDF)*, *Synthetic Minority Oversampling Technique (SMOTE)*, *Linguistic Inquiry Word Count (LIWC)*, and *Bidirectional Encoder Representations from Transformers (BERT)* methods. The study aims to determine the application of the methods used in this study to the prediction of big five personalities and to get better accuracy results than previous studies. Data obtained from 315 twitter users and 672,866 tweets obtained from surveys and have been labeled with big five personalities, resulting in an accuracy of 97.62% from the baseline with an increase of 23.1%, by applying the *CART+TF-IDF+SMOTE+LIWC+BERT* method.

Keywords: *CART*, *Kepribadian Big Five*, *LIWC*, *SMOTE*, *BERT*, Twitter.

1. Pendahuluan

Kepribadian merupakan hal yang melekat pada tiap individu. Kepribadian juga mempengaruhi setiap individu pada berbagai macam kegiatan [1]. Salah satunya yaitu ada kepribadian *big five*. Kepribadian *big five* merupakan suatu pendekatan dalam psikologi untuk menentukan kepribadian dari individu dalam beberapa dimensi. Menurut Lewis R. Goldberg ada 5 dimensi yaitu, *Openness*, *Conscientiousness*, *Extraversion*,