

ABSTRAK

Nama : Khairul Sholeh
NIM : 41519120068
Program Studi : Teknik Informatika
Judul Laporan Skripsi : Analisis Sentimen Twitter untuk Menentukan Kepribadian Seseorang Berdasarkan Model *The Big Five Personality Traits* dengan Menggunakan Algoritma *Naïve Bayes*
Dosen Pembimbing : Saruni Dwiasnati, ST.MM., M.Kom

Penelitian ini mengimplementasikan algoritma *Naïve Bayes* untuk menganalisis kepribadian seseorang berdasarkan data Twitter sesuai dengan *The Big Five Personality Traits* dengan pengguna Twitter yang berpengaruh di Indonesia dengan masing-masing 500 tweet per-pengguna. Melalui proses Crawling data menghasilkan 50.024 tweets dan setelah melalui Pre-processing data menjadi 46.028 tweet. Langkah selanjutnya dengan menerapkan teknik pembobotan Term Frequency-Inverse Document Frequency (TF-IDF) dan melakukan analisis NRC-Lexicon, penelitian ini berhasil menggali informasi sentimen dan emosi yang terkandung dalam setiap tweet. Kombinasi kedua metode ini menjadi dasar penentuan label pada dimensi kepribadian yang mengacu pada model OCEAN (Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism). Algoritma *Naïve Bayes* digunakan untuk klasifikasi, mencapai akurasi tertinggi 75.56% pada dataset uji dengan perbandingan 90:10. Evaluasi model menunjukkan peningkatan performa seiring dengan pertumbuhan ukuran dataset uji, menekankan pentingnya dataset yang lebih besar untuk meningkatkan ketepatan klasifikasi. Penelitian ini memberikan wawasan mendalam untuk riset lebih lanjut dan aplikasi praktis dalam memahami dinamika kepribadian yang diperoleh dari media sosial.

Kata Kunci : Algoritma *Naïve Bayes*, *The Big Five Personality Traits*, NRC-Lexicon, Analisis Sentimen, TF-IDF.

ABSTRACT

Name : Khairul Sholeh
NIM : 41519120068
Study Program : Informatics Engineering
Title Thesis : Twitter Sentiment Analysis to Determine Personality Based on The Big Five Personality Traits Model Using Naïve Bayes algorithm
Counsellor : Saruni Dwiasnati, ST.MM., M.Kom

This research implements the Naïve Bayes algorithm to analyze a person's personality based on Twitter data according to The Big Five Personality Traits with influential Twitter users in Indonesia with a total data of 500 tweets per user. Crawling data produces 50.024 tweets and after data becomes 46.028 tweets. The next step by applying the Term Frequency-Inverse Document Frequency (TF-IDF) weighting technique and conducting NRC-Lexicon analysis, this research succeeded in extracting sentiment and emotion information contained in each tweet. The combination of the two methods is the basis for determining the label on the personality dimension that refers to the OCEAN (Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism). The Naïve Bayes algorithm was used for classification, achieving the highest accuracy of 75.56% on a test dataset with a 90:10 ratio. Model evaluation showed improved performance as the size of the test dataset increased, emphasizing the importance of larger dataset to improve classification accuracy. This research provides insights for further research and practical applications in understanding personality dynamics derived from social media.

Keywords : Naïve Bayes algorithm, The Big Five Personality Traits, NRC-Lexicon, Sentiment Analysis, TF-IDF.