

ABSTRAK

Prediksi Tipe Kepribadian *Big five* Pada Pengguna Twitter Dengan Algoritma *Long Short-Term Memory*

Zhafran Nouvaldi – 1187050103
Jurusan Teknik Informatika

Penelitian prediksi kepribadian *big five* pengguna twitter ini melanjutkan dari penelitian sebelumnya yang menggunakan algoritma *random forest* yang merupakan salah satu algoritma *traditional machine learning* dan pada penelitian ini menggunakan algoritma LSTM yang merupakan salah satu algoritma *deep learning*. Penelitian ini bertujuan untuk mengetahui cara penerapan dan kinerja dari algoritma LSTM dalam melakukan klasifikasi teks pada prediksi kepribadian *big five* pengguna twitter. Data yang digunakan pada penelitian ini merupakan *public dataset* yang dipublikasikan oleh Rendy Putra Pratama pada penelitian sebelumnya yang berisi 269 data pengguna akun twitter dengan 240.000 data tweet. Metode penelitian yang digunakan adalah metode CRISP-DM. Penelitian ini menggunakan variasi data latih dan data uji dan juga algoritma CNN sebagai perbandingan dari hasil penelitian. Hasil dari penelitian ini adalah algoritma LSTM menghasilkan akurasi terbaik sebesar 77,72% pada perbandingan data latih dan data uji 80:20, dan itu lebih besar dibandingkan dengan menggunakan algoritma CNN dan juga algoritma *random forest* pada penelitian sebelumnya.

UNIVERSITAS ISLAM Negeri
SUNAN GUNUNG DJATI
BANDUNG

Kata Kunci: *Big Five*, LSTM, *Random Forest*, CRISP-DM, CNN

ABSTRACT

Big Five Personality Prediction on Twitter Users With Long Short-Term Memory Algorithm

Zhafran Nouvaldi – 1187050103
Jurusan Teknik Informatika

Prediction of the big five personality of Twitter users research continues from previous research using the random forest algorithm which is one of the traditional machine learning algorithms and in this study uses the LSTM algorithm which is one of the deep learning algorithms. This study aims to find out how to implement and perform the LSTM algorithm in classifying text on the big five personality predictions of Twitter users. The data used in this study is a public dataset that was published by Rendy Putra Pratama in a previous study which contained 269 Twitter account user data with 240,000 tweet data. The research method used is the CRISP-DM method. This study uses a variety of training data and test data and CNN algorithm too as a comparison of the research results. The results of this study are that the LSTM algorithm produces the best accuracy of 77.72% on a comparison of training data and test data of 80:20, and that is greater than using the CNN algorithm and also the random forest algorithm in previous studies.

Keywords: Big Five, LSTM, Random Forest, CRISP-DM, CNN