

ABSTRAK

Dalam dunia karir, analisis kepribadian merupakan hal yang penting dalam melakukan seleksi pemilihan minat dan karir. Oleh karena itu, dalam menganalisis kepribadian, klasifikasi kepribadian diikutsertakan, yang bertujuan agar kepribadian sesuai dengan minat dan karir yang dipilih. Hanya saja, proses melakukan analisis kepribadian memerlukan waktu yang cukup lama.

Oleh sebab itu, penulis merumuskan, bagaimana NLP (*Natural Language Processing*) dan CNN (*Convolutional Neural Network*) dapat digunakan untuk membantu dalam mengklasifikasikan kepribadian manusia. Tujuan penulis dari penelitian ini adalah untuk membuat sebuah program yang dapat memprediksi kepribadian manusia, dengan pendekatan NLP dan CNN, melalui forum komunitas PersonalityCafe.

Data yang digunakan diperoleh dari Kaggle, yaitu dataset kepribadian MBTI sebanyak 8675 baris data. Skenario penelitiannya adalah melakukan *preprocessing* serta *feature extraction* dataset berupa teks menggunakan pendekatan NLP serta TF-IDF (*term frequency – inverse document frequency*). Lalu melakukan *data balancing* menggunakan SMOTE untuk menyeimbangkan jumlah data yang tidak seimbang. Kemudian, membangun model klasifikasi menggunakan pendekatan CNN. Langkah terakhir adalah melakukan prediksi kepribadian dengan mengambil sampel teks seseorang untuk diprediksi. Hasil yang diperoleh dari metode *Convolutional Neural Network* menggunakan *9-layer* dengan jumlah *epoch* sebanyak 15, menghasilkan akurasi yang cukup optimal, dengan rata-rata akurasi setiap kategori kepribadian mencapai 0.7914.

Kata kunci: *Convolution Neural Network, term frequency, inverse document frequency, Natural Language Processing, klasifikasi*

ABSTRACT

In the career world, personality analysis is important when selecting interests and careers. Therefore, in analyzing personality, personality classification is included, which aims to make personality according to the chosen interests and career. However, the process of conducting a personality analysis takes quite a long time.

Therefore, the authors propose how NLP (Natural Language Processing) and CNN (Convolutional Neural Network) can be used to help predict the human personality. The author's aim of this study is to create a program that can predict human personality, using NLP and CNN approaches, through PersonalityCafe community forum.

The data used is obtained from Kaggle, which is the MBTI personality dataset of 8675 rows of data. The research scenario is to do preprocessing and feature extraction of datasets in the form of text using the NLP approach and TF-IDF (term frequency – inverse document frequency). Then, do a data balancing using SMOTE to balance the unbalanced amount of data. Afterwards, build a classification model using the CNN approach. The last step is to make a personality prediction by taking a sample of someone's text to predict. The results obtained from the Convolutional Neural Network method using 9-layers with a total of 15 epochs, resulting in quite optimal accuracy, with an average accuracy of each personality category reaching 0.7914.

Keywords: Convolution Neural Network, term frequency, inverse document frequency, Natural Language Processing, classification.