

INTISARI

AUTHOR VERIFICATION MENGGUNAKAN SIAMESE CONVOLUTIONAL NEURAL NETWORK

Oleh

FAQIH ETHANA PRABANDARU

17/409433/PA/17740

Author verification adalah metode untuk menentukan apakah dua dokumen berasal dari penulis yang sama atau bukan. Model tersebut dapat memecahkan masalah anonimitas plagiarisme. Tantangan dalam memecah permasalahan tersebut adalah gaya penulisan seseorang sangat beranekaragam serta jumlah data yang tersedia pada masalah nyata sangat sedikit.

Salah satu solusi dalam menyelesaikan masalah *author verification* adalah melalui pendekatan *deep learning*. *Siamese network* merupakan arsitektur yang digunakan untuk melakukan perbandingan dua hal. Dengan menggunakan 2 jaringan yang sama, dilakukan proses komputasi yang menghasilkan output berbeda yang kemudian dilakukan perhitungan jarak kedua output tersebut. Jarak yang dihasilkan akan dijadikan acuan sebagai prediksi dari model.

Model tersebut dievaluasi berdasarkan parameter akurasi, presisi, recall, dan F1 skor. Model tersebut memiliki rata-rata parameter evaluasi sebesar 68.25%. Penelitian ini juga melakukan perbandingan akurasi yang dihasilkan dari *word embedding FastText* dengan *Glove*. Jenis arsitektur dari *layer* yang digunakan juga dilakukan perbandingan antara banyaknya *hidden layer* yang digunakan serta penggunaan *CNN*, *LSTM*, serta *GRU*. Analisis juga dilakukan pada penggunaan data besar dan data kecil.

Kata-kata kunci : *author verification*, *siamese network*, *word embedding*, *convolutional neural network*

ABSTRACT

AUTHOR VERIFICATION USING SIAMESE CONVOLUTIONAL NEURAL NETWORK

By

FAQIH ETHANA PRABANDARU

17/409433/PA/17740

Author verification is a method to determine whether two documents are from the same author or not. The model can solve the problem of anonymity plagiarism. The challenge in solving these problems is that a person's writing style is very diverse and the amount of data available on real problems is very small.

One solution in solving the author verification problem is through a deep learning approach. Siamese network is an architecture that is used to compare two things. By using the same 2 networks, a computational process is carried out that produces different outputs which then calculates the distance between the two outputs. The resulting distance will be used as a reference as a prediction of the model.

The model was evaluated based on the parameters of accuracy, precision, recall, and F1 score. The model has an average evaluation parameter of 68.25%. This study also compares the accuracy resulting from word embedding FastText with Glove. The type of architecture of the layer used is also made a comparison between the number of hidden layers used and the use of CNN, LSTM, and GRU. Analysis was also carried out on the use of big data and small data.

Keywords : *author verification, siamese network, word embedding, convolutional neural network*