

ABSTRAK

Kesalahan penulisan dokumen jurnal menjadi masalah yang dapat mempengaruhi proses lamanya *publish* pada suatu jurnal, karena proses pembuatan hingga rilisnya sebuah jurnal membutuhkan proses yang panjang terutama pada proses *journal review*. Untuk mengatasi hal ini, diimplementasikan model *Generative Pre-Trained Transformers 2* (GPT-2) yang telah dilatih khusus untuk bahasa Indonesia, yaitu model Cahya/GPT-2-small-Indonesian-522M dari huggingface.co, untuk digunakan pada tugas koreksi kesalahan penulisan pada dokumen jurnal 6-8 halaman. Dalam penelitian ini, dilakukan metode *fine tuning* pada model Cahya/GPT-2-small-Indonesian-522M menggunakan *dataset* dengan format teks (.txt) yang terdiri dari pasangan teks asli dan teks yang telah dikoreksi dengan benar. Tujuan dari *fine tuning* adalah membuat model dapat melakukan tugas spesifik (koreksi penulisan) dari tugas awalnya dengan mempelajari pola kesalahan penulisan bahasa Indonesia agar dapat meningkatkan kemampuan model dalam memperbaiki kesalahan tersebut. Hasil implementasi menunjukkan bahwa penggunaan model Cahya/GPT-2-small-Indonesian-522M yang telah melalui proses *fine tuning* memberikan skor BLEU sebesar 0,482. Skor BLEU tersebut mengindikasikan hasil yang cukup baik pada sebuah model dalam mengukur tingkat kesamaan teks asli dengan teks keluaran dari model. Penelitian juga menunjukkan bahwa GPT-2 lebih unggul dan cocok digunakan untuk tugas *paraphrase* sebuah teks.

Kata Kunci: BLEU, *fine tuning*, GPT-2, jurnal, koreksi teks.

ABSTRACT

The errors in journal document writing have become a problem that can affect the length of the publishing process in a journal, as the entire process from creation to release of a journal requires a considerable amount of time, especially during the journal review process. To address this issue, an implementation involving the usage of the Generative Pre-Trained Transformers 2 (GPT-2) model, specifically trained for the Indonesian language, known as the Cahya/GPT-2-small-Indonesian-522M model from huggingface.co, was employed for the purpose of rectifying writing errors within journal documents spanning 6 to 8 pages in length. In this research, the fine-tuning method was employed on the Cahya/GPT-2-small-Indonesian-522M model using a dataset in text format (.txt) consisting of pairs of original texts and their corresponding corrected versions. The goal of fine-tuning was to make the model capable of performing a specific task (writing correction) derived from its original task by learning the patterns of Indonesian writing errors to improve its ability to rectify such errors. The implementation results indicated that the use of the fine-tuned Cahya/GPT-2-small-Indonesian-522M model achieved a BLEU score of 0,482. This BLEU score suggests a relatively good performance of the model in measuring the similarity between the original text and the model's output. The research also revealed that GPT-2 excelled and was suitable for the task of paraphrasing a text.

Keywords: BLEU, fine tuning, GPT-2, journal, text correction.