

BAB 5 SIMPULAN DAN SARAN

5.1 Simpulan

Implementasi algoritma BERT untuk melakukan klasifikasi topik penelitian telah berhasil dilakukan dan telah dilakukan demonstrasi terhadap pihak LPPM UMN. Model BERT diterapkan untuk mengklasifikasikan teks judul dan abstrak ke dalam 17 kategori UN SDG. UN SDG merupakan parameter utama yang digunakan untuk meningkatkan akreditasi fakultas dan program studi oleh Universitas Multimedia Nusantara. Penelitian ini menggunakan dataset yang didapatkan dari *website* Huggingface yang menyediakan *model repository*, *transformers library*, *Application Programming Interface* (API) dan dataset. Data yang digunakan selama penelitian berlangsung merupakan data dengan artikel berbahasa Inggris yang berjumlah 76.958 data. Selama pembuatan model berlangsung parameter yang digunakan adalah 4 *epochs*, *learning rate* $2e-5$, dan *batch size* 32. Terdapat 16 model yang dilakukan uji coba untuk mencari model terbaik dalam klasifikasi teks dan mendapatkan hasil terbaik pada *classification report* dan hasil akurasi. Model ketujuh merupakan model terbaik karena model ini memperoleh nilai *classification report* dan akurasi terbesar berupa nilai *precision* sebesar 0.99, nilai *recall* sebesar 0.82, nilai *f1-score* sebesar 0.87 dan memperoleh hasil akurasi sebesar 90.68% mengungguli model lainnya. Pembuatan model ketujuh ini menggunakan tahapan *preprocessing* dengan *library NLTK* untuk melatih dataset sehingga memperoleh hasil yang memuaskan dan model ini akan digunakan untuk melakukan pengklasifikasian teks karena mendapatkan hasil terbaik dibandingkan model lainnya.

5.2 Saran

Terdapat beberapa saran yang dapat digunakan untuk meningkatkan kualitas penelitian ini di masa depan, yaitu:

1. Menerapkan implementasi model terhadap tampilan *interface*, dimana dengan adanya implementasi ini model yang telah dirancang akan dikembangkan menjadi suatu aplikasi *website*, sehingga kualitas dari penelitian ini dapat mengalami peningkatan dengan model yang dapat diakses banyak orang.

2. Menambahkan jenis dataset yang digunakan terutama dataset dalam bahasa Indonesia. Dengan menambahkan dataset bahasa Indonesia, pengklasifikasian teks akan semakin akurat dalam melakukan klasifikasi pada artikel berbahasa Indonesia.



UMMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA