

BAB 5 SIMPULAN DAN SARAN

5.1 Simpulan

Implementasi algoritma *logistic regression* untuk dapat melakukan klasifikasi teks pada judul penelitian telah berhasil dilakukan dan telah sesuai dengan tahap-tahap pembangunan model. Model pada penelitian ini menggunakan *dataset* yang berasal dari sebuah situs yang menyediakan sumber daya dan alat, seperti *dataset* untuk *natural language processing* (NLP) bernama Hugging Face. *Dataset* yang digunakan ini adalah sekumpulan data judul artikel, jurnal atau deskripsi singkat berbahasa Inggris yang terkait dan telah dikategorikan ke salah satu dari 17 tujuan *Sustainable Development Goals* (SDG) dari Perserikatan Bangsa-Bangsa (PBB). *Dataset* ini memiliki data sebesar 76.958 data. Pada proses implementasi dan pembangunan model ini, terdapat 2 skenario dengan 9 model masing-masing pada skenario yang dibangun untuk melakukan uji coba untuk menentukan model terbaik untuk melakukan klasifikasi teks. Model ini juga menggunakan parameter penalti "L1" dan *solver* "SAGA". Model keenam yang menggunakan *library* TF-IDF tanpa melakukan *data cleaning* dengan menggunakan 80% *dataset* untuk melakukan *training* dan 20% untuk melakukan *testing* merupakan model terbaik atau dengan tingkat akurasi, presisi, *recall*, dan *f1-score* tertinggi dari model lainnya. Model ini memiliki tingkat akurasi sebesar 90.01% dengan rata-rata tingkat presisi sebesar 0.86, *recall* sebesar 0.82, dan *f1-score* sebesar 0.84. Model ini juga telah didemonstrasikan kepada pihak LPPM UMN melalui wawancara yang dilakukan secara *online* bersama Bapak Ifan Bagus Haryanto, S.Si selaku staf *Research Center Officer* dari LPPM UMN. Berdasarkan respon dari Bapak Ifan, model yang telah dibangun sudah cukup bagus dan membantu dalam meningkatkan efektifitas dalam melakukan pengkategorian judul penelitian dosen ke 17 tujuan SDG dari PBB.

5.2 Saran

Setelah melakukan pembangunan model, terdapat beberapa saran yang berkaitan untuk meningkatkan hasil dari penelitian di waktu yang mendatang, yaitu sebagai berikut.

1. Menambahkan tampilan *interface* untuk mengimplementasikan model yang

telah dibangun ini menjadi sebuah aplikasi *website* agar dapat digunakan oleh banyak orang.

2. Menggunakan *dataset* yang serupa yang telah digunakan pada penelitian ini tetapi yang telah digunakan pada penelitian yang telah terpublikasi.
3. Menambahkan *dataset* yang serupa yang telah digunakan pada penelitian ini dalam berbahasa Indonesia, ini berguna dalam meningkatkan akurasi dalam melakukan klasifikasi teks pada judul penelitian dengan bahasa Indonesia.
4. Mengembangkan model yang telah dibangun lebih lanjut dengan metode *data cleaning* yang berbeda seperti menggunakan *lemmatization* daripada menggunakan *stemming* ataupun menggunakan *library data preprocessing* yang berbeda selain dari *Count Vectorizer*, TF-IDF, dan *Word2Vec* sehingga dapat meningkatkan kualitas *dataset*.

