

BAB 5

SIMPULAN DAN SARAN

5.1 Simpulan

Berdasarkan hasil penelitian dari Implementasi *Word Embedding* untuk Klasifikasi Berita Penyandang Disabilitas Menggunakan Metode Logistic Regression dan uji coba aplikasi, simpulan dari penelitian adalah sebagai berikut.

1. Penelitian *Word Embedding* dengan FastText untuk Klasifikasi Berita Penyandang Disabilitas Menggunakan Metode Logistic Regression telah berhasil diimplementasikan dalam bentuk aplikasi *website* dengan fitur klasifikasi teks berita dan *file* berita. Penelitian ini menggunakan data hasil *scraping* yang berjumlah 378 data dan telah melalui proses pelabelan oleh pakar dengan sembilan kategori berita.
2. Hasil uji coba yang telah dilakukan membuktikan bahwa teknik *Full augmented data* membantu meningkatkan nilai *precision*, *recall*, *f1-score*, dan *accuracy* dari Model Logistic Regression. Teknik *Augmentation* yang digunakan adalah dengan cara *back-translation* semua data *training* yang digunakan. Pada penelitian ini dapat disimpulkan pembagian dengan teori pareto dengan rasio 80:20 mendapatkan hasil yang lebih baik daripada pembagian rasio yang lainnya. Dengan menggunakan *pre-trained* Model Skip-gram FastText dengan dimensi sebesar 300 dan Logistic Regression dengan parameter Solvers: *lbfgs*, C: 100.0, *max_iter*: 10, dan *Penalty*: 12

mempunyai hasil terbaik dengan nilai *Precision: 77,78%*, *Recall: 77%*, *f1-score: 76,24%*, dan *Accuracy: 76,31%*. Berdasarkan data berita yang dipakai ditemukan bahwa data berita kategori nasional mempunyai kemiripan data dengan kategori hukum, internasional, regional, edukasi, dan tekno.

5.2 Saran

Berdasarkan penelitian yang telah dilakukan, masih ada beberapa kekurangan dan keterbatasan, terdapat beberapa saran yang dapat digunakan untuk pengembangan penelitian yang lebih baik, yaitu sebagai berikut.

1. Menggunakan *pre-trained* model langsung dari FastText, karena penelitian ini hanya menggunakan model FastText yang dibuat dari data yang digunakan, sehingga memiliki keterbatasan kosakata. Penelitian ini tidak menggunakan *pre-trained* model dari FastText karena keterbatasan perangkat keras yang digunakan.
2. Menggunakan *library feature extraction* yang lain seperti tf-idf, karena di beberapa penelitian tf-idf lebih baik dari pada FastText (Tussenbroek, Makrodimitris, Viering, Jahfari, & Krijthe, 2020).