

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENGESAHAN	ii
HALAMAN PERNYATAAN BEBAS PLAGIASI	iii
HALAMAN MOTTO	iv
PRAKATA	v
DAFTAR ISI	vii
DAFTAR TABEL	x
DAFTAR GAMBAR	xii
INTISARI	xv
ABSTRACT	xvi
I PENDAHULUAN	1
1.1. Latar Belakang Masalah	1
1.2. Rumusan Masalah	2
1.3. Tujuan Penelitian	3
1.4. Batasan Masalah	3
1.5. Manfaat Penelitian	3
1.6. Keaslian Penelitian	3
II TINJAUAN PUSTAKA	5
III LANDASAN TEORI	10
3.1. <i>Automated Essay Scoring</i> (AES)	10
3.2. Representasi Kata	10
3.2.1. Model <i>Skip-gram</i> dengan <i>Negative Sampling</i>	11
3.2.2. Model FastText dengan Informasi <i>Subword</i>	12
3.3. Klasifikasi <i>Cost-Sensitive XGBoost</i>	13
3.4. <i>Pseudo-Labeling</i> dengan <i>Self Training</i>	16
3.5. Evaluasi dan Validasi	17
3.5.1. Ukuran Evaluasi	18
3.5.2. <i>K-Fold Cross Validation</i>	19
IV ANALISIS DAN PERANCANGAN SISTEM	20
4.1. Analisis Sistem	20
4.1.1. Analisis Permasalahan	20
4.1.2. Deskripsi Umum Usulan Sistem	23
4.2. Deskripsi Dataset	25
4.2.1. Dataset Berlabel	25
4.2.2. Dataset Tidak Berlabel	29
4.3. Perancangan Sistem	32

4.3.1.	Pra-Pemrosesan Data	33
4.3.2.	Representasi Kata	34
4.3.3.	Perancangan Alur Pelatihan Model dengan menggunakan Metode <i>Cost-Sensitive XGBoost</i>	34
4.3.4.	Perancangan Alur Pelatihan Model dengan menggunakan Metode <i>Pseudo-Labeling</i> dengan <i>Self Training</i>	36
4.3.5.	Perancangan Alur Pelatihan Model dengan menggunakan Metode <i>Hybrid</i>	38
4.4.	Perancangan Pengujian	40
4.4.1.	Pengujian Model <i>Cost-Sensitive XGBoost</i>	41
4.4.2.	Pengujian <i>Pseudo-Labeling</i> dengan <i>Self Training</i>	41
4.4.3.	Perbandingan antara Model <i>Hybrid</i> dengan Model AdaBoost	42
V	IMPLEMENTASI	43
5.1.	Spesifikasi Perangkat Keras dan Perangkat Lunak	43
5.2.	Implementasi Pemilihan Data	44
5.3.	Implementasi Pra-Pemrosesan	44
5.4.	Implementasi Representasi Kata <i>fastText</i>	46
5.5.	Implementasi Metode <i>Cost-Sensitive XGBoost</i>	46
5.6.	Implementasi Metode <i>Pseudo-Labeling</i> dengan <i>Self Training</i>	47
5.7.	Implementasi Program Utama Metode <i>Hybrid</i>	47
5.8.	Implementasi Metode AdaBoost	48
VI	HASIL DAN PEMBAHASAN	51
6.1.	Hasil Pengujian Model <i>Cost-Sensitive XGBoost</i>	52
6.2.	Hasil Pengujian <i>Pseudo-Labeling</i> dengan <i>Self Training</i>	58
6.2.1.	Hasil Pengujian Banyaknya Data Awal	58
6.2.2.	Hasil Pengujian Banyaknya Data Tambahan	67
6.3.	Hasil Perbandingan antara Model <i>Hybrid</i> dengan Model AdaBoost	75
VII	KESIMPULAN DAN SARAN	85
7.1.	Kesimpulan	85
7.2.	Saran	85
	DAFTAR PUSTAKA	87
A	Pertanyaan dari Setiap Dataset	89
B	Source Code	92
B.1	Source Code Representasi Kata <i>fastText</i>	92
B.2	Source Code <i>Cost-Sensitive XGBoost</i>	94
B.3	Source Code <i>Pseudo-Labeling</i> dengan <i>Self Training</i>	96
C	Hasil Pengujian <i>Pseudo-Labeling</i> dengan <i>Self Training</i>	97

C.1	Hasil Pengujian <i>Pseudo-Labeling</i> dengan <i>Self Training</i> (Banyak Data Awal)	97
C.2	Hasil Pengujian <i>Pseudo-Labeling</i> dengan <i>Self Training</i> (Banyak Data Tambahan)	105