

DAFTAR ISI

LEMBAR PENGESAHAN	ii
LEMBAR PERNYATAAN ORISINALITAS	iii
ABSTRAK	iv
KATA PENGANTAR	vi
UCAPAN TERIMA KASIH	vii
DAFTAR ISI	viii
DAFTAR GAMBAR	xii
DAFTAR TABEL	xxii
DAFTAR KODE SUMBER	xxiv
DAFTAR SINGKATAN	xxvii
1 USULAN GAGASAN	1
1.1 Deskripsi Umum Masalah	1
1.2 Analisis Masalah	5
1.3 Analisis Solusi yang Sudah Ada	8
2 SPESIFIKASI DAN BATASAN SOLUSI	12
2.1 Dasar Penentuan Spesifikasi	12
2.1.1 Aturan Pemerintah untuk <i>Bullying</i>	12
2.1.2 Pemahaman Semantik	13
2.1.3 Komputasi dan Efisiensi Model	13
2.2 Batasan dan Spesifikasi	15
2.2.1 Kemampuan Memahami Bahasa Indonesia	15
2.2.2 Pemilihan <i>Keyword</i>	15
2.2.3 Dataset	15
2.2.4 Jumlah Dataset	16
2.2.5 Model	16

2.2.6	<i>Fine-tuning dan Hyperparameter Tuning</i>	17
2.2.6.1	<i>Fine-tuning</i>	17
2.2.6.2	<i>Hyperparameter Tuning</i>	18
2.3	Pengukuran atau Verifikasi Spesifikasi	21
2.3.1	<i>Overfitting dan Underfitting</i>	21
2.3.2	<i>Loss Training dan Loss Validation</i>	22
2.3.3	<i>Accurate Training dan Accurate Validation</i>	24
2.3.4	Metode Evaluasi	24
2.3.5	Pengaturan Parameter	26
2.3.6	Analisis Kesalahan	27
3	DESAIN RANCANGAN SOLUSI	28
3.1	Alternatif Usulan Solusi	28
3.1.1	<i>Recurrent Neural Network (RNN)</i>	28
3.1.2	<i>Support Vector Machine (SVM)</i>	33
3.1.3	Naive Bayes	36
3.1.4	<i>Term Frequency-Inverse Document Frequency (TF-IDF)</i>	38
3.1.5	<i>Bidirectional Encoder Representations from Transformers (BERT)</i>	41
3.1.6	<i>Indonesia Bidirectional Encoder Representations from Transformers (IndoBERT)</i>	44
3.1.7	Cendol	46
3.2	Analisis dan Pemilihan Solusi	48
3.3	Desain Solusi Terpilih	51
3.3.1	Model IndoBERT	52
3.3.2	Model Cendol	52
3.3.3	Integrasi dan Penerapan	53
3.4	Jadwal Pengerjaan	53
4	IMPLEMENTASI	55
4.1	Deskripsi Umum Implementasi	55
4.1.1	<i>PyTorch</i>	55
4.1.2	<i>TensorFlow</i>	56
4.1.3	<i>Sklearn</i>	56
4.1.4	<i>Transformers</i>	57
4.1.5	<i>Google Collaboratory</i>	57
4.1.6	<i>Natural Language Processing (NLP)</i>	58
4.1.7	X (Twitter)	59

4.2	Detil Implementasi	59
4.3	Prosedur Pengoperasian	60
4.3.1	Survei	61
4.3.2	<i>Data Crawling</i>	65
4.3.3	<i>Data Cleaning dan Preprocessing</i>	67
4.3.3.1	Menggunggah Data CSV	67
4.3.3.2	Menghapus Duplikasi Data CSV	68
4.3.3.3	Menghapus Nilai Kosong pada Data CSV	69
4.3.3.4	Membersihkan Data	69
4.3.3.5	Menyaring Data Lebih Dari Dua Kata	70
4.3.3.6	Normalisasi Teks	71
4.3.3.7	Menghapus Data Lebih dari Sepuluh Kata Asing	72
4.3.3.8	Menghapus <i>Stopwords</i>	73
4.3.3.9	Tokenisasi	74
4.3.3.10	<i>Stemming</i>	75
4.3.4	Pelabelan Data	77
4.3.5	Penyeimbangan Label	78
4.3.6	Pembagian Dataset	80
4.3.7	<i>Modeling Data</i>	82
4.3.7.1	Pemodelan Menggunakan IndoBERT	82
4.3.7.2	Pemodelan Menggunakan Cendol	83
4.3.8	<i>Tokenizer</i>	84
4.3.8.1	<i>Tokenizer</i> Menggunakan INDOBERT	84
4.3.8.2	<i>Tokenizer</i> Menggunakan Cendol	85
4.3.9	Pemilihan Perangkat	88
4.3.10	Pelatihan dan Evaluasi Model	89
4.3.10.1	Pelatihan Model	89
4.3.10.2	Evaluasi Model	90
4.3.11	Analisis dan Interpretasi Hasil	91
4.3.12	Menyimpan Model <i>Trainer</i>	97
4.3.13	Pengoptimalan dan <i>Tuning</i> Model	98
5	Pengujian	101
5.1	Skema Umum Pengujian	101
5.2	Detil Pengujian	102
5.2.1	<i>Cross-Entropy Loss</i>	103
5.2.1.1	IndoBERT	104
5.2.1.2	Cendol	104

5.2.2	Evaluasi Model	105
5.2.2.1	IndoBERT	106
5.2.2.2	Cendol	115
5.2.3	Analisis Hasil Pengujian	125
5.3	Perbandingan Model IndoBERT dan Cendol	127
5.3.1	Karakteristik Model	127
5.3.2	Evaluasi Performansi	127
5.3.3	Analisis	128
5.3.4	Percobaan Training Model Cendol Menggunakan Dataset Terbaru	128
6	KESIMPULAN DAN SARAN	135
6.1	Kesimpulan	135
6.2	Saran	135
	DAFTAR PUSTAKA	137
	LAMPIRAN I	147
	LAMPIRAN II	152
	LAMPIRAN III	156
	LAMPIRAN IV	158
	LAMPIRAN V	223