

DAFTAR ISI

HALAMAN PENGESAHAN	ii
PERNYATAAN BEBAS PLAGIASI	iii
PRAKATA	iv
DAFTAR ISI	v
DAFTAR GAMBAR	viii
DAFTAR TABEL	x
INTISARI	xii
ABSTRACT	xiii
BAB I PENDAHULUAN	1
1.1 Rumusan Masalah	5
1.2 Batasan Masalah	5
1.3 Tujuan	6
1.4 Manfaat	6
BAB II TINJAUAN PUSTAKA	7
BAB III LANDASAN TEORI	18
3.1 Deteksi Ujaran Kebencian Tersirat	18
3.2 Arsitektur <i>Triplet Network</i>	19
3.3 Proyeksi Representasi Laten Pada Himpunan Vektor <i>Embedding</i>	20
3.4 <i>Supervised Contrastive Loss (SCL)</i> Dan <i>ImpCon</i>	21
3.4.1 Representasi Metrik Jarak Berdasarkan Proxi Sudut dan Jarak <i>Angular</i> Dari Fungsi Kesamaan Kosinus	21
3.5 <i>Triplet Loss</i> Dan Metode Pengambilan Sampel <i>Semi-Hard Negative Mining</i>	23
3.6 Fungsi Penghimpun Nilai <i>Loss (Reducer)</i>	25
3.7 <i>Backpropagation</i> Pada Representasi <i>Angular</i> Dengan Margin Aditif Dalam Domain Kosinus Dan <i>Angular</i>	26
3.8 <i>Multi Layer Perceptron</i>	28
3.9 <i>Loss Cross Entropy</i>	28
3.10 <i>Loss</i> Gabungan Klasifikasi Dan <i>Metric</i>	29

3.11	Arsitektur Pralatih Semantik Bahasa Berbasis BERT	29
3.12	Evaluasi Pada Identifikasi Ujaran Kebencian Tersirat	32
BAB IV METODOLOGI PENELITIAN		34
4.1	Deskripsi Umum Penelitian	34
4.2	Akuisisi Data	35
4.2.1	<i>Dataset Implicit Hate Corpus</i>	35
4.2.2	<i>Dataset Social Bias Frame Corpus</i>	36
4.2.3	Dataset DynaHate	37
4.3	Rancangan Model Dan Metode Yang Diajukan	38
4.4	Rancangan Pengujian	41
BAB V IMPLEMENTASI SISTEM		42
5.1	Alat Dan Bahan Kerja	42
5.2	Pembagian Data Latih Dan Data Uji	43
5.3	Prapemrosesan Data	43
5.3.1	Agregasi dan Splitting data Implicit hate corpus (IHC)	43
5.3.2	Agregasi Data Dan Pembangunan Fitur <i>implied statement</i> Pada Dataset SBIC	46
5.3.3	Tokenisasi Untuk Seluruh Data	54
5.3.4	Pembangunan Model <i>Triplet network</i> dengan <i>Multi Layer Perceptron Head</i>	56
5.3.5	Pembangunan Algoritma <i>Triplet Loss, Mining, Representasi Distance</i> dan <i>Reducer</i>	58
5.3.6	Pelatihan, Pengujian Sistem	64
BAB VI HASIL DAN PEMBAHASAN		69
6.1	Performa F1-Score Pada Metode Yang Diajukan Dengan Metode <i>ImpCon</i>	69
6.2	Grafik <i>Loss</i> Dan <i>Confusion Matrix</i> Pada Model Terbaik Untuk Setiap <i>Dataset Latih</i>	76
6.3	Akurasi Pada Metode Terbaik Yang Diajukan	78
6.4	Pengaruh Metode <i>Mining</i> Pada <i>Triplet Loss</i>	79

6.5	Analisa Representasi <i>Embedding</i> Pada Metode <i>ImpCon</i> Dengan Metode Yang Diajukan	80
6.5.1	Kemampuan Pemisahan Antar Kelas Dan Target Pada Data Uji Internal	81
6.5.2	Kemampuan Pemisahan Antar Kelas Dan Target Pada Data Uji Silang Pada <i>Triplet Network</i>	84
6.6	Kemampuan Generalisasi Pada Ujaran Kebencian Tersirat Serta Kebencian Eksplisit	85
6.6.1	<i>False Negative</i> Pada Data Ujaran Kebencian Pada Metode Yang Diajukan	85
6.6.2	Konteks Spesifik dan Ambiguitas Linguistik	90
6.6.3	Kemampuan Model Dalam Mengidentifikasi Ujaran Kebencian Tersirat Dengan Benar	91
6.6.4	Kemampuan Generalisasi Pada Data Ujaran Kebencian Eksplisit Pada <i>DynaHate</i>	94
BAB VII KESIMPULAN DAN SARAN		96
7.1	Kesimpulan	96
7.2	Saran	96
DAFTAR PUSTAKA		104

DAFTAR GAMBAR

3.1	Arsitektur <i>Siamese Network</i>	20
3.2	Model <i>HateBERT</i>	30
3.3	<i>Encoder</i> pada <i>Transformer</i>	31
3.4	BERT untuk <i>Word Embeddings</i>	31
4.1	Diagram alur kerja	34
4.2	Perancangan model triplet network dengan Semi hard negative mining	39
5.1	Agregasi data dan data <i>Splitting</i> dataset IHC(bagian 1))	44
5.2	Agregasi Data dan Data <i>Splitting</i> dataset IHC (bagian 2)	45
5.3	Agregasi data SBIC (bagian 1)	46
5.4	Agregasi data SBIC (bagian 2)	47
5.5	Pembangunan fitur <i>implied_statement</i>	51
5.6	Pembangunan fitur <i>implied_statement</i> (bagian 2)	52
5.7	Tokenisasi (Bagian 1)	54
5.8	Tokenisasi (Bagian 2)	55
5.9	Pembangunan model <i>triplet network</i> dengan <i>MLP head</i> (bagian 1) . .	56
5.10	Pembangunan model <i>triplet network</i> dengan <i>MLP head</i> (bagian 2) . .	57
5.11	<i>triplet loss</i>	58
5.12	<i>cosine similarity distance</i>	58
5.13	Fungsi Representasi Jarak Untuk <i>Embedding</i>	59
5.14	Pembangunan Fungsi <i>Reducer SmoothMax, AttentiveReducer dan Me-</i> <i>an</i>	60
5.15	<i>Semi-Hard Negative Mining</i> (bagian 1)	62
5.16	<i>Semi-Hard Negative Mining</i> (bagian 2)	63
5.17	Pelatihan Model Triplet-Network Dengan Loss Yang Diajukan	66
5.18	Pengujian Model Yang Telah Dilatih	67
6.1	Grafik <i>loss</i> (kiri) <i>baseline (ImpCon)</i> dan (kanan) <i>loss angular smoo-</i> <i>thmax</i> $\beta = 5 \alpha = 0.45$ pada dataset IHC	76
6.2	Grafik <i>loss</i> (kiri) <i>baseline (ImpCon)</i> dan (kanan) <i>loss angular aditif</i> dengan <i>attentive reducer</i> $\alpha = 0.45$ pada dataset SBIC	76
6.3	<i>Confusion matrix</i> (kiri) <i>baseline (ImpCon)</i> dan (kanan) <i>loss angular</i> <i>smoothmax</i> $\beta = 5 \alpha = 0.45$ pada dataset IHC	77
6.4	<i>Confusion matrix</i> (kiri) <i>baseline (ImpCon)</i> dan (kanan) <i>loss angular</i> <i>aditif</i> dengan <i>attentive reducer</i> $\alpha = 0.45$ pada dataset SBIC	77

6.5	Plot kelompok TSNE per kelas dengan 15 target terbanyak pada data uji dan latih IHC metode ImpCon	81
6.6	Plot kelompok TSNE per kelas dengan 15 target terbanyak pada data uji dan latih SBIC metode ImpCon	82
6.7	Plot kelompok TSNE per kelas dengan 15 target terbanyak pada data uji dan latih IHC metode <i>triplet loss</i> yang diajukan	83
6.8	Plot kelompok TSNE per kelas dengan 15 target terbanyak pada data uji dan latih SBIC pada metode <i>triplet loss</i> yang diajukan	83
6.9	Plot kelompok TSNE kemampuan generalisasi data pada data latih IHC dengan data uji SBIC (Kiri) dan <i>DynaHate</i> (Kanan) pada metode yang diajukan	84
6.10	Plot kelompok TSNE kemampuan generalisasi data pada data latih IHC dengan data uji <i>DynaHate</i> dan SBIC pada metode yang diajukan	85

DAFTAR TABEL

2.1	Ringkasan tinjauan penelitian	14
2.2	Ringkasan tinjauan penelitian (Lanjutan)	15
2.3	Ringkasan tinjauan penelitian (Lanjutan)	16
2.4	Ringkasan tinjauan penelitian (Lanjutan)	17
3.1	Tabel Contoh Tipologi Makna Perbedaan Ujaran Kebencian Tersirat dan Eksplisit	18
3.2	Tabel Contoh Tipologi Makna Perbedaan Ujaran Kebencian Tersirat dan Eksplisit (Lanjutan)	19
4.1	Ringkasan Distribusi Data pada Setiap Tahap IHC	35
4.2	Ringkasan Distribusi Data SBIC	36
4.3	Deskripsi Fitur Dataset SBIC	37
4.4	Distribusi label per split dataset	38
4.5	<i>Hyperparameter</i> yang digunakan dalam proses pelatihan	40
5.1	Daftar Alat dan Bahan Kerja	42
5.2	Distribusi Data Setelah Agregasi	46
5.3	Atribut Token dan Deskripsinya	48
5.4	Penjelasan Part-of-Speech (Universal POS)	49
5.5	Penjelasan Tag Penn Treebank	49
5.6	Aturan Pemilihan Template Berdasarkan Token Pertama	50
5.7	Distribusi data SBIC setelah agregasi	54
6.1	Hasil pengujian model pada <i>dataset</i> utama IHC dengan nilai terbaik pada uji internal data terbaik	70
6.2	Hasil pengujian model pada <i>dataset</i> utama SBIC dengan nilai terbaik pada uji internal data terbaik	72
6.3	Hasil pengujian model pada <i>dataset</i> utama IHC	74
6.4	Hasil pengujian model pada <i>dataset</i> utama SBIC	75
6.5	Akurasi dataset IHC dengan model terbaik dibandingkan dengan baseline ImpCon	79
6.6	Akurasi dataset SBIC dengan model terbaik dibandingkan dengan baseline ImpCon	79
6.7	Studi ablasi F1-Score Pengaruh metode mining	80

6.8	Tabel perbandingan ujaran kebencian tersirat yang merupakan <i>false negative</i> pada data latih dan uji IHC dengan metode terbaik yang diajukan	86
6.9	Tabel perbandingan ujaran kebencian tersirat yang merupakan <i>false negative</i> pada data latih dan uji SBIC dengan metode terbaik yang diajukan	88
6.10	Daftar Kalimat yang <i>false negative</i> dengan permasalahan <i>multi-lingual</i> atau konteks	91
6.11	Contoh Kasus Ujaran Kebencian tersirat yang teridentifikasi dengan benar pada data latih dan uji IHC	92
6.12	Contoh Kasus Ujaran Kebencian Tersirat yang teridentifikasi dengan benar pada data latih dan uji SBIC	93
6.13	Contoh Kasus Ujaran Kebencian Tersirat <i>false negative</i> pada data latih IHC uji <i>DynaHate</i>	94
6.14	Contoh Kasus Ujaran Kebencian Tersirat <i>false negative</i> pada data latih SBIC uji <i>DynaHate</i>	95