

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENGESAHAN.....	ii
HALAMAN PERNYATAAN.....	iii
HALAMAN PERSEMBAHAN	iv
HALAMAN MOTTO	v
PRAKATA	vi
DAFTAR ISTILAH	viii
DAFTAR ISI.....	ix
DAFTAR TABEL	xii
DAFTAR GAMBAR.....	xiii
DAFTAR LAMPIRAN	xiv
INTISARI	xv
ABSTRACT	xvi
1. PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	8
1.3 Tujuan Penelitian	9
1.4 Batasan Masalah	9
1.5 Manfaat Penelitian	10
1.6 Kontribusi Penelitian	10
2. TINJAUAN PUSTAKA.....	11
2.1 Penelitian Terkait Perbaikan Kata OOV	11
2.2 Penelitian Terkait Fitur Konteks	13
2.3 Penelitian Terkait Deteksi Sarkasme	15
2.4 Penelitian Terkait Sentence Encoder BERT	17
3. LANDASAN TEORI	20
3.1 Definisi Sarkasme	20
3.2 Kata-kata OOV	21
3.3 Algoritma Fonetik Soundex	23
3.3.1 Kode Soundex	24
3.3.2 Kode Soundex dengan aturan fonetik	25
3.3.3 Kode Soundex dengan dan tanpa karakter terakhir.....	26
3.4 Algoritma Menghitung Jarak dan Kemiripan Kata.....	27

3.4.1	Sequence Matcher	27
3.4.2	Levenshtein Distance	28
3.4.3	Jaro Similarity	29
3.4.4	Jaro-Winkler Similarity	29
3.5	Konteks Percakapan	29
3.6	Deep Learning	32
3.6.1	CNN	34
3.6.2	LSTM	35
3.6.3	GRU	36
3.6.4	BERT	36
3.7	Sentence Encoder	37
3.8	Pooling	39
3.9	Cosine Similarity	40
4.	METODOLOGI PENELITIAN	42
4.1	Kerangka Penelitian	42
4.2	Model Yang Dikembangkan	44
4.3	Pengumpulan Dataset	45
4.4	Pre-processing	46
4.5	Normalisasi OOV	47
4.6	Penggabungan Model Untuk Deteksi Kalimat Sarkasme	47
4.7	Klasifikasi	49
4.8	Metode Pengujian	49
4.8.1	Metode pengujian model normalisasi OOV	49
4.8.2	Metode pengujian sentence encoder dan cosine similarity	50
5.	PERBAIKAN KATA NON-STANDAR	52
5.1	Algoritma Perbaikan Kata OOV	52
5.1.1	Deteksi kata OOV	54
5.1.2	Input kata OOV	54
5.1.3	Pengkodean kata menggunakan kode fonetik Soundex	54
5.1.4	Penghitungan Sequence Matcher	56
5.1.5	Penghitungan Levenshtein Distance	57
5.1.6	Penghitungan Jaro Similarity	59
5.1.7	Penghitungan Jaro-Winkler Similarity	60
5.2	Pemilihan Kata Pengganti OOV	61
6.	MODEL DETEKSI SARKASME	63
6.1	Model Deteksi Kalimat Sarkasme	63
6.1.1	Lapisan input kalimat pada encoder BERT	65
6.1.2	Lapisan sentence encoding	66
6.1.3	Lapisan pooling	67
6.1.4	Lapisan cosine similarity	68
6.1.5	Lapisan fully connected	69
6.2	Pre-training dan fine-tuning	70
6.3	Model Pembandingan	71

7.	PENGUJIAN DAN PEMBAHASAN	77
7.1	Pembagian Dataset	77
7.2	Pengujian Perhitungan Jarak dan Kemiripan Kata	77
7.3	Analisis Kinerja Perhitungan Jarak dan Kemiripan Kata	79
7.4	Hasil Pengujian Normalisasi OOV	81
7.5	Analisis Kinerja Penggabungan Model	82
7.6	Hasil Pengujian Klasifikasi Sarkasme	85
7.7	Pembuktian Korelasi Metode Cosine, Hasil Prediksi Dan Hasil Label..	91
7.8	Performansi Waktu Proses	93
8.	PENUTUP.....	95
8.1	Kesimpulan	95
8.2	Saran	95
	DAFTAR PUSTAKA	96
	LAMPIRAN.....	107

DAFTAR TABEL

Tabel 2.1 Penelitian terkait kata OOV	13
Tabel 2.2 Penelitian terkait fitur konteks	14
Tabel 2.3 Perkembangan metode pada penelitian deteksi sarkasme	15
Tabel 2.4 Penelitian terkait sentence encoder	18
Tabel 3.1 Contoh kalimat <i>cuitan</i>	22
Tabel 3.2 Aturan kode fonetik	24
Tabel 4.1 Statistik dataset kata OOV	45
Tabel 5.1 Penghitungan jarak dua kata menggunakan sequence matcher	56
Tabel 5.2 Nilai jarak kata “aagr” menggunakan sequence matcher	57
Tabel 5.3 Penghitungan nilai Levenshtein distance	58
Tabel 5.4 Nilai jarak kata “aagr” menggunakan Levenshtein distance	58
Tabel 5.5 Nilai jarak kata “aagr” menggunakan Jaro similarity	59
Tabel 5.6 Nilai jarak kata “aagr” menggunakan Jaro-Winkler similarity	61
Tabel 5.7 Contoh nilai jarak dan kemiripan	62
Tabel 7.1 Statistik dataset sarkasme	77
Tabel 7.2 Nilai korelasi distribusi antara semua algoritma	78
Tabel 7.3 Performansi Sequence Matcher dan Jaro-Winkler Similarity	80
Tabel 7.4 Perolehan nilai pengujian	82
Tabel 7.5 Contoh kata OOV dengan probabilitas kemunculan kecil	84
Tabel 7.6 Hasil Confusion Matrix Semua Model	85
Tabel 7.7 Hasil Pengujian Spearman	92
Tabel 7.8 Contoh pengujian pada kalimat <i>cuitan</i>	93
Tabel 7.9 Waktu proses	94

DAFTAR GAMBAR

Gambar 2.1 Perkembangan penelitian deteksi sarkasme	16
Gambar 3.1 Kode Soundex dengan satu karakter terakhir.....	26
Gambar 3.2 Kode Soundex tanpa dan dengan satu karakter terakhir	27
Gambar 3.3 Contoh tampilan media sosial Twitter.....	31
Gambar 3.4 Model CNN.....	35
Gambar 3.5 Model LSTM	36
Gambar 3.6 Model GRU	36
Gambar 3.7 Model BERT transformer.....	37
Gambar 3.8 Arsitektur sentence-BERT.....	40
Gambar 3.9 Nilai kemiripan dua kalimat.....	41
Gambar 4.1 Kerangka penelitian	43
Gambar 4.2 Alur model yang dikembangkan.....	44
Gambar 4.3 Tampilan website labelling dataset	46
Gambar 4.4 Diagram penggabungan model deteksi kalimat sarkasme	48
Gambar 5.1 Diagram normalisasi kata OOV	53
Gambar 5.2 Pseudocode kode Soundex	55
Gambar 6.1 Pipeline fine-tuning BERT untuk deteksi sarkasme	63
Gambar 6.2 Arsitektur BERT dan cosine similarity	64
Gambar 6.3 Representasi input pada lapisan segmen embedding	65
Gambar 6.4 Implementasi representasi input.....	66
Gambar 6.5 Struktur input kalimat sentence encoder BERT	67
Gambar 6.6 Proses pre-training dan fine-tuning BERT	71
Gambar 6.7 Arsitektur model BERT	72
Gambar 6.8 Arsitektur model BERT-CNN.....	73
Gambar 6.9 Arsitektur model BERT-LSTM	74
Gambar 6.10 Arsitektur model BERT-GRU.....	75
Gambar 7.1 Nilai korelasi distribusi empat algoritma yang diajukan	79
Gambar 7.2 Confusion matrix BERTweet	86
Gambar 7.3 Confusion matrix BERTweet dan NormOOV	87
Gambar 7.4 Confusion matrix BERTweet-CS	87
Gambar 7.5 Confusion matrix BERTweet-CS dan NormOOV	88
Gambar 7.6 Confusion matrix BERTweet-CNN	88
Gambar 7.7 Confusion matrix BERTweet-CNN dan NormOOV	89
Gambar 7.8 Confusion matrix BERTweet-LSTM.....	89
Gambar 7.9 Confusion matrix BERTweet-LSTM dan NormOOV	90
Gambar 7.10 Confusion matrix BERTweet-GRU	90
Gambar 7.11 Confusion matrix BERTweet-GRU dan NormOOV.....	91