

## DAFTAR ISI

<b>PRAKATA</b> .....	iv
<b>INTISARI</b> .....	x
<b>ABSTRACT</b> .....	xi
<b>BAB I PENDAHULUAN</b> .....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	3
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian.....	3
1.5 Manfaat Penelitian.....	4
1.6 Keaslian Penelitian.....	4
1.7 Metode Penelitian.....	4
1.8 Sistematika Penulisan.....	5
<b>BAB II TINJAUAN PUSTAKA</b> .....	7
<b>BAB III LANDASAN TEORI</b> .....	12
3.1 Text Mining.....	12
3.2 Ujaran Kebencian.....	15
3.3 Natural Language Processing.....	16
3.4 Word Embeddings.....	19
3.5 Twitter.....	23
3.6 Gated Recurrent Unit.....	24
<b>BAB IV METODOLOGI PENELITIAN</b> .....	26
4.1 Analisis Kebutuhan.....	26
4.2 Gambaran Umum Penelitian.....	26
4.3 Deskripsi Data.....	28
4.4 Preprocessing.....	28
4.5 Ekstraksi Fitur ( <i>Word2vec</i> ).....	34
4.6 Analisis Membangun Model Klasifikasi.....	40
4.7 Rancangan Pengujian.....	42
<b>BAB V IMPLEMENTASI</b> .....	45
5.1 Implementasi Pelatihan <i>Word2vec</i> Model.....	45
5.2 Implementasi Preprocessing.....	46
5.3 Implementasi Proses Klasifikasi.....	52
5.4 Implementasi Pengujian.....	56
<b>BAB VI HASIL DAN PEMBAHASAN</b> .....	58
6.1 Hasil Pengujian Ekstraksi Fitur.....	58
6.2 Penentuan Parameter GRU.....	61
6.3 Menguji Kinerja Metode GRU.....	65
<b>BAB VII KESIMPULAN DAN SARAN</b> .....	67



UNIVERSITAS  
GADJAH MADA

**Deteksi Hate Speech Bahasa Indonesia Menggunakan Word Embedding dan Gated Recurrent Unit**  
JUNANDA PATIHULLAH, Drs. Edi Winarko, M.Sc., Ph.D.

Universitas Gadjah Mada, 2018 | Diunduh dari <http://etd.repository.ugm.ac.id/>

7.1 Kesimpulan.....	67
7.2 Saran.....	67

## DAFTAR TABEL

Tabel 2.1 Tinjauan Pustaka.....	10
Tabel 3.1 representasi vektor satu kata.....	20
Tabel 4.1 Contoh tweet ujaran kebencian.....	28
Tabel 4.2 Escapping html character pada data tweet.....	29
Tabel 4.3 Menghapus simbol dan tanda baca pada data tweet.....	30
Tabel 4.4 Pemisah kata yang digabung pada data tweet.....	31
Tabel 4.5 Penyeragaman huruf pada data tweet.....	31
Tabel 4.6 Tokenisasi pada data tweet.....	32
Tabel 4.7 Convert slangword pada data tweet.....	33
Tabel 4.8 Stopword pada data tweet.....	34
Tabel 4.9 Contoh konten window.....	37
Tabel 4.10 Nilai representasi vektor pada kata dengan word2vec.....	39
Tabel 4.11 Confusion Matrix prediksi 2 kelas.....	43
Tabel 4.12 Rancangan pengujian dan evaluasi.....	44
Tabel 6.1 hasil akurasi model dengan ekstraksi fitur word2vec, tf dan tf-idf.....	59
Tabel 6.2 Akurasi dari jumlah neuron pada lapisan tersembunyi.....	64
Tabel 6.3 Hasil akurasi model untuk mendeteksi ujaran kebencian.....	66

## DAFTAR GAMBAR

Gambar 3.1 Tahapan Proses Text mining (Fieldman dan Sanger, 2006).....	14
Gambar 3.2 Arsitektur model Word2Vec (Brownlee, 2017).....	22
Gambar 3.3 Arsitektur Gated Recurrent Neural Network (Gupta, 2017).....	24
Gambar 4.1 Arsitektur sistem deteksi ujaran kebencian.....	27
Gambar 4.2 Arsitektur utama Word2vec.....	35
Gambar 4.3 Proses pembangun kosakata.....	36
Gambar 4.4 struktur representasi masukan dan keluaran pada jaringan syaraf tiruan word2vec.....	38
Gambar 4.5 Proses representasi kata pada model word2vec.....	39
Gambar 4.6 Proses perancangan klasifikasi.....	40
Gambar 4.7 Rancangan arsitektur model GRU.....	41
Gambar 5.1 Implementasi pembuatan model word2vec.....	46
Gambar 5.2 Implementasi pre-processing.....	47
Gambar 5.3 Kode program hapus karakter html.....	48
Gambar 5.4 Penghapusan simbol dan tanda baca.....	48
Gambar 5.5 Implementasi pemisah dua kalimat.....	49
Gambar 5.6 Implementasi split attached.....	49
Gambar 5.7 fungsi tokenize dan case folding.....	50
Gambar 5.8 implementasi proses slangword.....	51
Gambar 5.9 Proses pengecekan stopword.....	51
Gambar 5.10 program inisialisasi variabel yang digunakan.....	52
Gambar 5.11 Implementasi arsitektur GRU.....	53
Gambar 5.12 Implementasi SGD.....	54
Gambar 5.13 Implementasi pelatihan GRU.....	55
Gambar 5.14 Implementasi pelatihan GRU.....	56
Gambar 5.15 Implementasi proses ekstraksi fitur.....	56
Gambar 5.16 Pengujian dengan K-Fold.....	57
Gambar 6.1 perbandingan akurasi model dengan ekstraksi fitur word2vec, tf dan tf-idf.....	59
Gambar 6.2 perbandingan akurasi ekstraksi fitur word2vec.....	60
Gambar 6.3 Penentuan learning rate.....	62
Gambar 6.4 Perbandingan akurasi dari jumlah neuron hidden layer.....	65