

## DAFTAR ISI

<b>Halaman Judul</b>	<b>ii</b>
<b>Halaman Pengesahan</b>	<b>iii</b>
<b>Halaman Pernyataan</b>	<b>iv</b>
<b>Halaman Persembahan</b>	<b>v</b>
<b>Halaman Motto</b>	<b>vi</b>
<b>PRAKATA</b>	<b>vii</b>
<b>DAFTAR ISI</b>	<b>viii</b>
<b>DAFTAR TABEL</b>	<b>xi</b>
<b>DAFTAR GAMBAR</b>	<b>xii</b>
<b>INTISARI</b>	<b>xv</b>
<b>ABSTRACT</b>	<b>xvii</b>
<b>I PENDAHULUAN</b>	<b>1</b>
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah	2
1.4 Tujuan Penelitian	2
1.5 Manfaat Penelitian	3
1.6 Metodologi Penelitian	3
1.7 Sistematika Penulisan	4
<b>II TINJAUAN PUSTAKA</b>	<b>6</b>
<b>III DASAR TEORI</b>	<b>13</b>
3.1 Plagiarisme	13
3.2 <i>Text Pre-Processing</i>	15

3.2.1	<i>Sentence Segmentation</i>	15
3.2.2	<i>Stop Word Removal</i>	15
3.2.3	<i>Stemming</i>	15
3.3	Metode Deteksi Similaritas	16
3.3.1	<i>Analisis Semantic</i>	16
3.3.2	<i>Analisis Syntactic</i>	17
3.4	<i>Word Embeddings</i>	17
3.4.1	<i>Arsitektur Skip-Gram</i>	18
3.5	Algoritma Nazief dan Adriani	20
3.6	Similaritas Jaccard	21
3.7	<i>Binary Search</i>	22
3.8	Faktor Pengujian	23
3.8.1	<i>Precision</i>	23
3.8.2	<i>Recall</i>	23
3.8.3	<i>F-Measure</i>	24
<b>IV</b>	<b>ANALISIS DAN RANCANGAN PENELITIAN</b>	<b>25</b>
4.1	Alat dan Bahan	25
4.2	Analisis	25
4.3	Perancangan Perangkat Lunak	25
4.4	Pengujian	31
<b>V</b>	<b>IMPLEMENTASI</b>	<b>33</b>
5.1	Pengolahan Input dan <i>Dataset</i>	33
5.2	<i>Model Training</i>	33
5.3	<i>Input Data dan Text Segmentation</i>	36
5.4	<i>Stop Words Removal, Word Stemming, dan Perhitungan Nilai Similaritas</i>	40
5.5	<i>Post Processing</i>	47
<b>VI</b>	<b>HASIL PENELITIAN DAN PEMBAHASAN</b>	<b>50</b>
6.1	Pendahuluan	50
6.2	Pengolahan Input dan <i>Dataset</i>	50
6.3	<i>Pre-Processing</i>	64
6.4	<i>Word Stemming dan Penghitungan Nilai Similaritas</i>	70
6.5	<i>Post Processing</i>	83

6.6 Perbandingan Performa Metode <i>Semantic</i> dan <i>Syntactic</i> dengan Metode Similaritas Jaccard	92
<b>VII PENUTUP</b>	<b>95</b>
7.1 Kesimpulan	95
7.2 Saran	95
<b>DAFTAR PUSTAKA</b>	<b>96</b>
<b>LAMPIRAN</b>	<b>100</b>
Hasil Performa Rata-Rata Pengujian Metode <i>Semantic</i> dan <i>Syntactic</i>	100

## DAFTAR TABEL

2.1	Perbandingan penelitian sebelumnya dengan yang akan dilakukan	11
6.1	Performa rata-rata metode <i>semantic</i> dan <i>syntactic</i> di setiap kasus uji	93
6.2	Performa rata-rata metode similaritas Jaccard di setiap kasus uji	94

## DAFTAR GAMBAR

3.1	Arsitektur <i>Skip-gram</i> (Mikolov et al., 2013)	18
3.2	Ilustrasi <i>skip-gram</i> (Alammar, J., 2019)	19
3.3	Ilustrasi <i>skip-gram</i> ke- 2 (Alammar, J., 2019)	19
3.4	Ilustrasi <i>skip-gram</i> ke- 3 (Alammar, J., 2019)	20
4.1	Flowchart Algoritma (A. Abdi et al. 2015)	30
4.2	Flowchart <i>Detailed Comparison</i> (A. Abdi et al. 2015)	31
5.1	Kode sumber untuk <i>model training</i>	34
5.2	Kode sumber untuk menerima masukan data dan <i>text segmentation</i> (dilanjutkan)	37
5.3	Kode sumber untuk menerima masukan data dan <i>text segmentation</i> (lanjutan)	39
5.4	Kode sumber untuk <i>Stop Word Removal</i> dan pembuatan <i>constructor stemmer</i> untuk <i>Word Stemming</i>	41
5.5	Implementasi kode untuk <i>Word Stemming</i> dan Perhitungan Nilai Similaritas (dilanjutkan)	44
5.6	Kode sumber untuk <i>Word Stemming</i> dan Perhitungan Nilai Similaritas (lanjutan)	46
5.7	Kode sumber untuk <i>Post Processing</i>	48
6.1	Abstrak <i>suspected</i> kasus uji A dengan nilai similaritas 0%	50
6.2	Abstrak sumber kasus uji A dengan nilai similaritas 0%	51
6.3	Abstrak <i>suspected</i> kasus uji B dengan nilai similaritas 0%	51
6.4	Abstrak sumber kasus uji B dengan nilai similaritas 0%	52
6.5	Abstrak <i>suspected</i> kasus uji C dengan nilai similaritas 0%	52
6.6	Abstrak sumber kasus uji C dengan nilai similaritas 0%	53
6.7	Abstrak <i>suspected</i> kasus uji D dengan nilai similaritas 0%	53
6.8	Abstrak sumber kasus uji D dengan nilai similaritas 0%	54
6.9	Abstrak <i>suspected</i> kasus uji A dengan nilai similaritas 50%	55
6.10	Abstrak sumber kasus uji A dengan nilai similaritas 50%	55
6.11	Abstrak <i>suspected</i> kasus uji B dengan nilai similaritas 50%	56
6.12	Abstrak sumber kasus uji B dengan nilai similaritas 50%	56
6.13	Abstrak <i>suspected</i> kasus uji C dengan nilai similaritas 50%	57

6.14	Abstrak sumber kasus uji C dengan nilai similaritas 50%	57
6.15	Abstrak <i>suspected</i> kasus uji D dengan nilai similaritas 50%	58
6.16	Abstrak sumber kasus uji D dengan nilai similaritas 50%	58
6.17	Abstrak <i>suspected</i> kasus uji A dengan nilai similaritas 100%	59
6.18	Abstrak sumber kasus uji A dengan nilai similaritas 100%	60
6.19	Abstrak <i>suspected</i> kasus uji B dengan nilai similaritas 100%	61
6.20	Abstrak sumber kasus uji B dengan nilai similaritas 100%	62
6.21	Abstrak <i>suspected</i> kasus uji C dengan nilai similaritas 100%	62
6.22	Abstrak sumber kasus uji C dengan nilai similaritas 100%	63
6.23	Abstrak <i>suspected</i> kasus uji D dengan nilai similaritas 100%	63
6.24	Abstrak sumber kasus uji D dengan nilai similaritas 100%	64
6.25	<i>Pre-processing</i> kasus uji A dengan nilai similaritas 0%	65
6.26	<i>Pre-processing</i> kasus uji B dengan nilai similaritas 0%	65
6.27	<i>Pre-processing</i> kasus uji C dengan nilai similaritas 0%	65
6.28	<i>Pre-processing</i> kasus uji D dengan nilai similaritas 0%	66
6.29	<i>Pre-processing</i> kasus uji A dengan nilai similaritas 50%	66
6.30	<i>Pre-processing</i> kasus uji B dengan nilai similaritas 50%	67
6.31	<i>Pre-processing</i> kasus uji C dengan nilai similaritas 50%	67
6.32	<i>Pre-processing</i> kasus uji D dengan nilai similaritas 50%	68
6.33	<i>Pre-processing</i> kasus uji A dengan nilai similaritas 100%	68
6.34	<i>Pre-processing</i> kasus uji B dengan nilai similaritas 100%	69
6.35	<i>Pre-processing</i> kasus uji C dengan nilai similaritas 100%	69
6.36	<i>Pre-processing</i> kasus uji D dengan nilai similaritas 100%	70
6.37	Hasil <i>word stemming</i> dan penghitungan nilai similaritas kasus uji A dengan nilai similaritas aktual 0%	71
6.38	Hasil <i>word stemming</i> dan penghitungan nilai similaritas kasus uji B dengan nilai similaritas aktual 0%	72
6.39	Hasil <i>word stemming</i> dan penghitungan nilai similaritas kasus uji C dengan nilai similaritas aktual 0%	73
6.40	Hasil <i>word stemming</i> dan penghitungan nilai similaritas kasus uji D dengan nilai similaritas aktual 0%	74
6.41	Hasil <i>word stemming</i> dan penghitungan nilai similaritas kasus uji A dengan nilai similaritas aktual 50%	75
6.42	Hasil <i>word stemming</i> dan penghitungan nilai similaritas kasus uji B dengan nilai similaritas aktual 50%	76

6.43 Hasil <i>word stemming</i> dan penghitungan nilai similaritas kasus uji C dengan nilai similaritas aktual 50%	77
6.44 Hasil <i>word stemming</i> dan penghitungan nilai similaritas kasus uji D dengan nilai similaritas aktual 50%	78
6.45 Hasil <i>word stemming</i> dan penghitungan nilai similaritas kasus uji A dengan nilai similaritas aktual 100%	79
6.46 Hasil <i>word stemming</i> dan penghitungan nilai similaritas kasus uji B dengan nilai similaritas aktual 100%	80
6.47 Hasil <i>word stemming</i> dan penghitungan nilai similaritas kasus uji C dengan nilai similaritas aktual 100%	81
6.48 Hasil <i>word stemming</i> dan penghitungan nilai similaritas kasus uji D dengan nilai similaritas aktual 100%	82
6.49 Hasil <i>post-processing</i> kasus uji A dengan nilai similaritas 0%	83
6.50 Hasil <i>post-processing</i> kasus uji B dengan nilai similaritas 0%	84
6.51 Hasil <i>post-processing</i> kasus uji C dengan nilai similaritas 0%	84
6.52 Hasil <i>post-processing</i> kasus uji D dengan nilai similaritas 0%	85
6.53 Hasil <i>post-processing</i> kasus uji A dengan nilai similaritas 50%	85
6.54 Hasil <i>post-processing</i> kasus uji B dengan nilai similaritas 50%	86
6.55 Hasil <i>post-processing</i> kasus uji C dengan nilai similaritas 50%	87
6.56 Hasil <i>post-processing</i> kasus uji D dengan nilai similaritas 50%	87
6.57 Hasil <i>post-processing</i> kasus uji A dengan nilai similaritas 100%	88
6.58 Hasil <i>post-processing</i> kasus uji B dengan nilai similaritas 100%	89
6.59 Hasil <i>post-processing</i> kasus uji C dengan nilai similaritas 100%	90
6.60 Hasil <i>post-processing</i> kasus uji D dengan nilai similaritas 100%	91