

***SENTIMENT ANALYSIS PADA REVIEW BARANG BERBAHASA  
INDONESIA DENGAN METODE K-NEAREST NEIGHBOR (K-NN)***

**SKRIPSI**

Diajukan untuk memenuhi sebagian persyaratan memperoleh gelar Sarjana  
Komputer



Disusun Oleh :

**YUGO YUDANSHA LUHULIMA  
NIM. 0910683100**

**KEMENTERIAN PENDIDIKAN DAN KEBUDAYAAN NASIONAL  
UNIVERSITAS BRAWIJAYA  
PROGRAM TEKNOLOGI INFORMASI DAN ILMU KOMPUTER**

**MALANG**

**2013**

## LEMBAR PERSETUJUAN

***SENTIMENT ANALYSIS PADA REVIEW BARANG BERBAHASA  
INDONESIA DENGAN METODE K-NEAREST NEIGHBOR (K-NN)***

### SKRIPSI

Diajukan untuk memenuhi persyaratan memperoleh gelar Sarjana Komputer

Disusun Oleh:

**YUGO YUDANSHA LUHULIMA**

**NIM. 0910683100**

Menyetujui :

**Pembimbing I**

**Pembimbing II**

**Drs.Marji., M. T.**

**NIP. 196708011992031001**

**Lailil Mufflikhah, S.Kom., M.Sc.**

**NIP. 83101306120035**

## LEMBAR PENGESAHAN

***SENTIMENT ANALYSIS PADA REVIEW BARANG BERBAHASA  
INDONESIA DENGAN METODE K-NEAREST NEIGHBOR (K-NN)***

### SKRIPSI

Laboratorium Komputasi Cerdas dan Visualisasi

Diajukan untuk memenuhi persyaratan memperoleh gelar Sarjana Komputer

Disusun oleh:

**Yugo Yudansha Luhulima**

**NIM. 0910683100**

Skripsi ini telah diuji dan dinyatakan lulus tanggal 18 Juli 2013

**Penguji I**

**Penguji II**

**Dian Eka Ratnawati, S.Si, M.Kom**

**Suprpto, ST., MT.**

**NIP. 19730619 200212 2 001**

**NIP. 19710727 199603 1 001**

**Penguji III**

**Rekvan Regasari MP, S.T., M.T.**

**NIP. 77041406120253**

**Mengetahui  
Ketua Program Studi Informatika**

**Drs. Marji., M.T.**

**NIP. 19670801 199203 1 001**



## PERNYATAAN ORISINALITAS SKRIPSI

Saya menyatakan dengan sebenar-benarnya bahwa sepanjang pengetahuan saya, di dalam naskah SKRIPSI ini tidak terdapat karya ilmiah yang pernah diajukan oleh orang lain untuk memperoleh gelar akademik di suatu perguruan tinggi, dan tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis dikutip dalam naskah ini dan disebutkan dalam sumber kutipan dan daftar pustaka.

Apabila ternyata didalam naskah SKRIPSI ini dapat dibuktikan terdapat unsur-unsur PLAGIASI, saya bersedia SKRIPSI ini digugurkan dan gelar akademik yang telah saya peroleh (SARJANA) dibatalkan, serta diproses sesuai dengan peraturan perundang-undangan yang berlaku. (UU No. 20 Tahun 2003, Pasal 25 ayat 2 dan Pasal 70).

Malang, 18 Juli 2013  
Mahasiswa,

**Yugo Yudansha Luhulima**  
**NIM. 0910683100**

## KATA PENGANTAR

*Alhamdulillah* rabbil 'alamin. Puji syukur penulis panjatkan kehadirat Allah SWT, karena atas segala rahmat dan limpahan hidayah-Nya, penulis dapat menyelesaikan skripsi yang berjudul “*Sentiment Analysis* pada *Review* Barang Berbahasa Indonesia dengan Metode *K-Nearest Neighbor* (K-NN)”.

Dalam pelaksanaan dan penulisan skripsi ini penulis mendapatkan banyak bantuan dari berbagai pihak baik secara moral maupun material. Dalam kesempatan ini penulis ingin mengucapkan terima kasih yang sebesar – besarnya kepada :

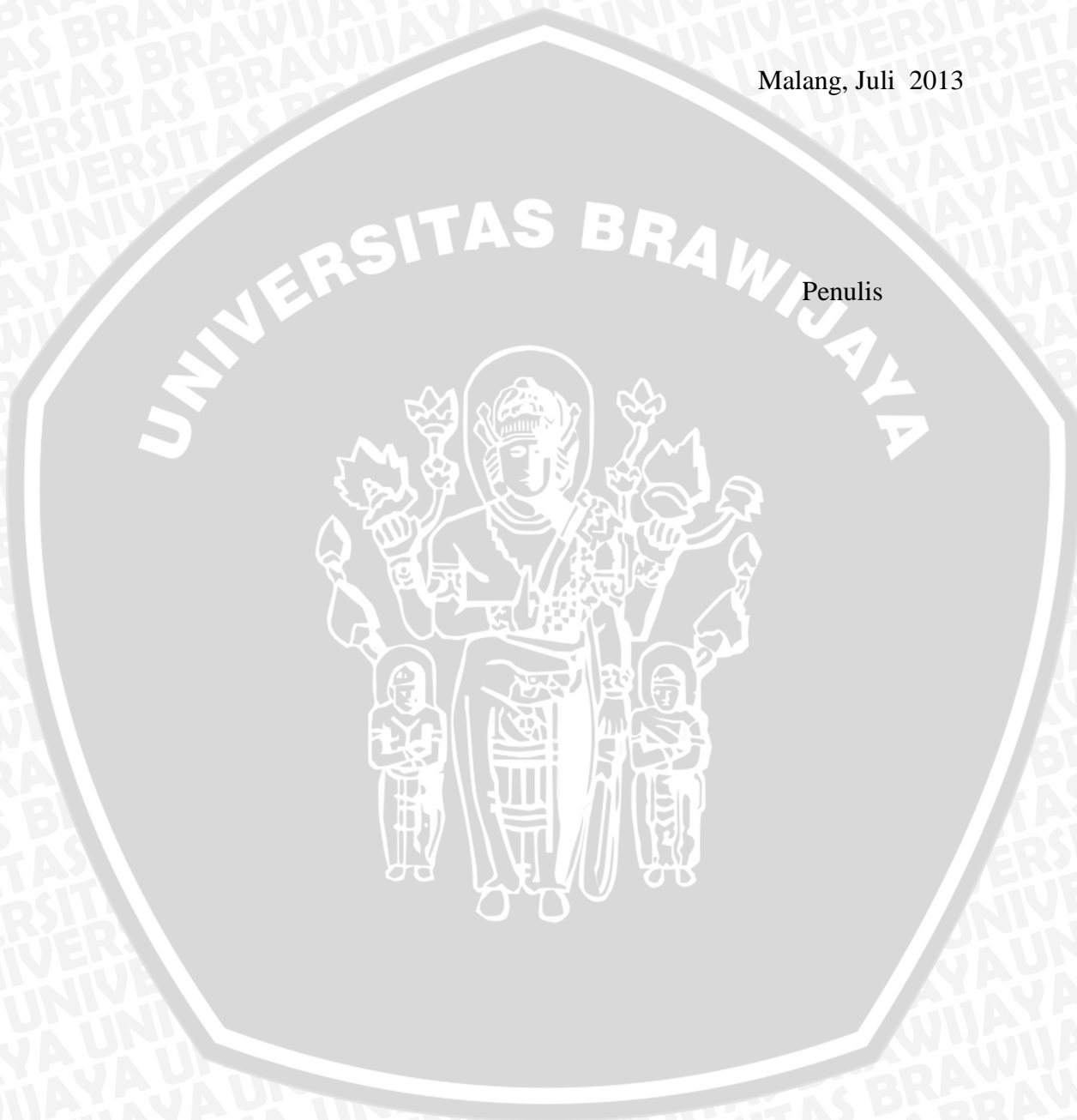
1. Bapak Drs. Marji., M. T . dan Ibu Lailil Muflikhah, S. Kom., M.Sc. selaku dosen pembimbing selama pelaksanaan skripsi.
2. Bapak Ir. Sutrisno, M.T, Bapak Ir. Heru Nurwasito, M.Kom, Bapak Himawat Aryadita, S.T, M.Sc, dan Bapak Eddy Santoso, S.Kom selaku Ketua, Wakil Ketua 1, Wakil Ketua 2 dan Wakil Ketua 3 Program Teknologi Informasi dan Ilmu Komputer Universitas Brawijaya.
3. Bapak Drs. Marji, M.T dan Bapak Issa Arwani, S.Kom, M.Sc selaku Ketua dan Sekretaris Program Studi Teknik Informatika Universitas Brawijaya.
4. Almarhum Ayahanda Lutfi Luhulima, Ibunda Yun Aryani, dan seluruh keluarga atas segenap dukungan dan kasih sayang yang telah diberikan.
5. Seluruh Dosen Teknik Informatika Universitas Brawijaya atas kesediaan membagi ilmunya kepada penulis.
6. Seluruh Civitas Akademika Teknik Informatika Universitas Brawijaya yang telah banyak memberi bantuan dan dukungan selama penulis menempuh studi di Teknik Informatika Universitas Brawijaya dan selama penyelesaian skripsi ini.
7. Seluruh pihak yang telah membantu kelancaran penulisan skripsi yang tidak dapat penulis sebutkan satu persatu.

Penulis menyadari bahwa dalam penyusunan skripsi ini masih banyak kekurangan baik format penulisan maupun isinya. Oleh karena itu, saran dan

kritik membangun senantiasa diharapkan guna perbaikan bagi penelitian selanjutnya. Semoga skripsi ini dapat memberikan manfaat bagi semua pihak, Amin.

Malang, Juli 2013

Penulis





## ABSTRAK

Yugo Yudansha Luhulima. 2009. *Sentiment Analysis* pada *Review* Barang Berbahasa Indonesia dengan Metode *K-Nearest Neighbor* (K-NN). Skripsi Program Studi Teknik Informatika, Program Teknologi Informasi dan Ilmu Komputer, Universitas Brawijaya. Pembimbing: Drs. Marji., M. T. dan Ibu Lailil Muflikhah, S. Kom., M.Sc. .

Analisis sentimen merupakan domain *Text Mining* atau penggalian data berupa teks, yang diantaranya terdapat proses mengolah dan mengekstrak data tekstual secara otomatis untuk mendapatkan informasi. Manfaat analisis sentimen dalam dunia usaha antara lain untuk melakukan pemantauan terhadap sebuah produk. Analisis sentimen dapat digunakan sebagai alat bantu untuk melihat respon konsumen atau masyarakat terhadap suatu produk tertentu, sehingga dapat segera diambil langkah-langkah strategis berikutnya. Proses pada analisis sentimen diawali dengan *preprocessing*, dilanjutkan dengan pembobotan kata, kemudian penghitungan *cosine similarity*, dan klasifikasi.

*Preprocessing* terdiri dari beberapa tahap yaitu *Cleaning*, *Casefolding*, *stopword removal*, dan *stemming*. Metode pembobotan kata yang digunakan pada skripsi ini adalah *Term Frequecnny – Inverse Document Frequency* (TF-IDF) dan menggunakan *K-Nearest Neighbor* (KNN) sebagai metode klasifikasinya. *K-Nearest Neighbor* (KNN) adalah suatu metode pengklasifikasian berdasarkan mayoritas dari kategori. Metode ini bertujuan untuk mengklasifikasi objek baru berdasarkan atribut dan *training sample*.

Pengujian akurasi dari analisis sentimen pada *review* barang berbahasa Indonesia dengan metode *K-Nearest Neighbor* (KNN) menghasilkan rata-rata akurasi sebesar 59,160 % dan nilai tertinggi akurasi sebesar 76,19048 %.

Kata kunci: Analisis Sentimen, *Text Mining*, *K-Nearest Neighbor* (KNN)

## ABSTRACT

*Sentiment analysis is a branch of research in text mining or data mining domain, which contains process and extract textual data automatically to obtain information. One of the benefit of sentiment analysis in the business world is for monitoring a product. Sentiment analysis can be used as a tool to see customer response or a particular community to a product, so the next strategic steps can be taken. The process of sentiment analysis begins with the preprocessing, followed by a weighting word, then cosine similarity calculation and classification.*

*Preprocessing consist of several steps, Cleaning, Casefolding, stopword removal, and stemming. Term weighting method used in this research is Term Frequency – Inverse Document Frequency (TF-IDF) and using K-Nearest Neighbor (KNN) as the classification method. K-Nearest Neighbor (KNN) is a classification method based on the category majority. This method aims to classify new object based on attribute and training sample.*

*The accuracy testing of sentiment analysis of product review in Indonesian using K-Nearst Neighbor (KNN) obtained 59,160% as the average of accuracy and the highest accuracy is 76,19048%.*

**Keyword:** *Sentiment Analysis, Text Mining, K-Nearest Neighbor (KNN)*

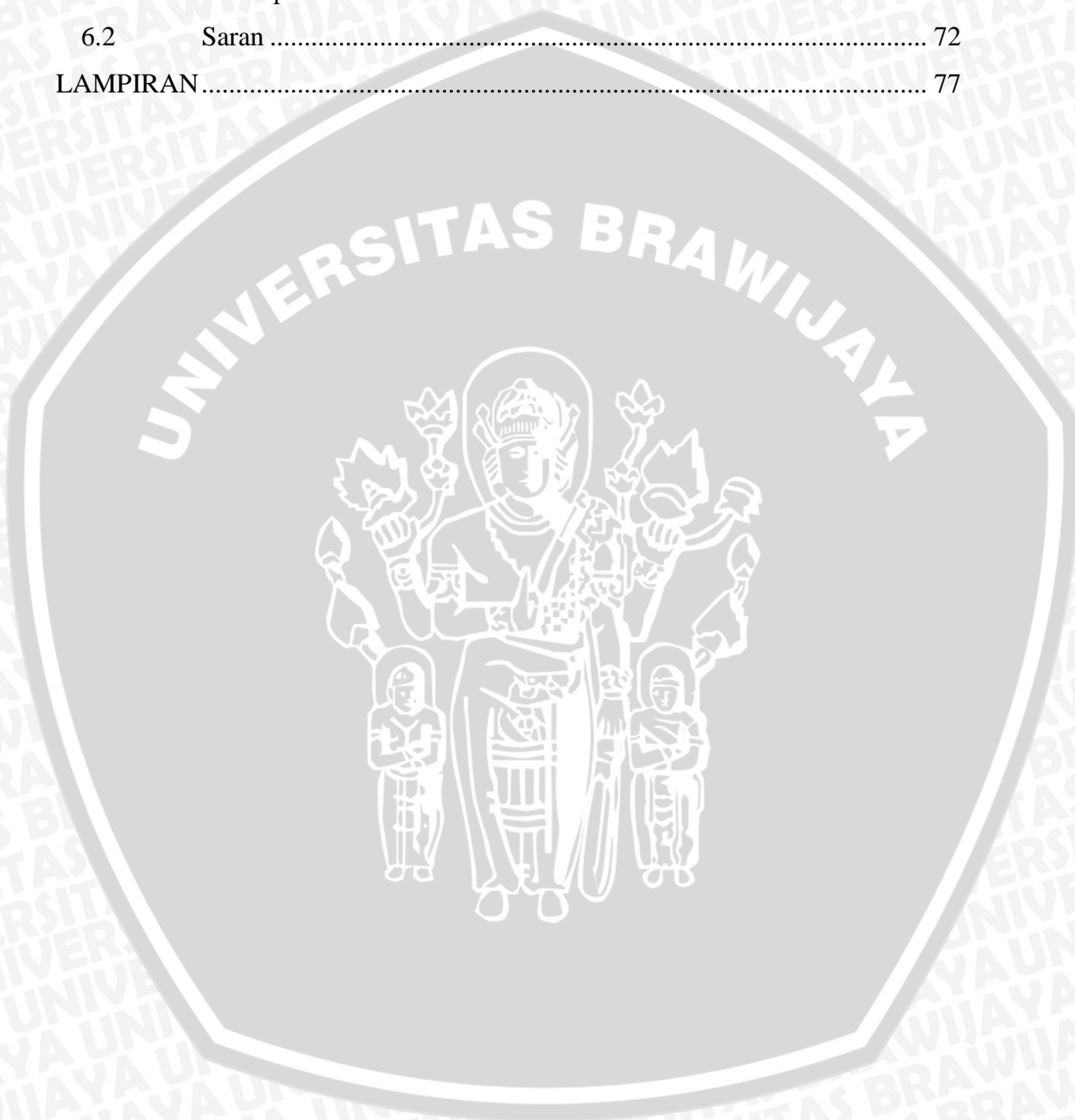


## DAFTAR ISI

LEMBAR PERSETUJUAN.....	ii
LEMBAR PENGESAHAN .....	iii
PERNYATAAN ORISINALITAS SKRIPSI.....	iv
KATA PENGANTAR .....	v
ABSTRAK.....	vii
<i>ABSTRACT</i> .....	viii
DAFTAR ISI.....	ix
DAFTAR GAMBAR .....	xii
DAFTAR TABEL.....	xiii
DAFTAR SOURCE CODE .....	xv
DAFTAR LAMPIRAN .....	xvi
<b>BAB I PENDAHULUAN.....</b>	<b>1</b>
1.1    Latar Belakang.....	1
1.2    Rumusan Masalah.....	3
1.3    Tujuan .....	3
1.4    Batasan Masalah .....	3
1.5    Manfaat .....	3
1.6    Sistematika Penulisan .....	4
<b>BAB II KAJIAN PUSTAKA DAN DASAR TEORI.....</b>	<b>6</b>
2.1    Kajian Pustaka .....	6
2.2    Text Mining .....	7
2.3    Preprosesing Dokumen .....	8
2.3.1    Cleaning .....	8
2.3.2    Parsing .....	8
2.3.3    Casefolding .....	8
2.3.4    Filtering/Stopword Removal.....	9
2.3.5    Stemming .....	9
2.4    Analisis Sentimen .....	13
2.5 <i>K-Nearest Neighbor</i> .....	13
2.6 <i>KNN Decision Rule</i> .....	14

2.7	Pengukuran Kemiripan ( <i>Cosine Similarity</i> ).....	15
2.8	Pembobotan ( <i>Term Weighting</i> ).....	16
<b>BAB III METODE PENELITIAN DAN PERANCANGAN.....</b>		<b>18</b>
3.1	Studi Literatur .....	19
3.2	Penyusunan Dasar Teori .....	19
3.3	Analisa dan Perancangan .....	19
3.3.1	Kebutuhan Antar Muka .....	19
3.3.2	Kebutuhan Data .....	20
3.3.3	Kebutuhan Fungsional .....	20
3.3.4	Arsitektur Program .....	20
3.3.5	Diagram Alir Sistem .....	21
3.4	Desain Antar Muka .....	31
3.5	Manualisasi Analisis Sentimen .....	33
3.6	Klasifikasi .....	39
3.7	Implementasi.....	40
3.8	Pengujian .....	41
3.8.1.	Skenario Pengujian .....	41
3.9	Penulisan Laporan.....	42
<b>BAB IV IMPLEMENTASI .....</b>		<b>43</b>
4.1	Spesifikasi Sistem .....	43
4.1.1	Spesifikasi Perangkat Keras.....	43
4.1.2	Spesifikasi Perangkat Lunak.....	44
4.2	Batasan-Batasan Implementasi .....	44
4.3	Implementasi Algoritma .....	45
4.3.1	Implementasi Algoritma Proses <i>Input</i> Dokumen.....	45
4.4	Implementasi Antar Muka .....	52
4.4.1	Tampilan Halaman Data Training .....	52
4.4.2	Tampilan Halaman Pengujian.....	53
4.4.3	Tampilan Halaman Akurasi .....	53
4.4.4	Tampilan Halaman <i>Preprocessing</i> .....	54
<b>BAB V PENGUJIAN DAN ANALISIS.....</b>		<b>55</b>
5.1	Pengujian Sistem.....	55

5.2	Pengujian Akurasi dengan Stemming.....	56
5.3	Pengujian Akurasi tanpa Stemming.....	64
BAB VI PENUTUP .....		72
6.1	Kesimpulan .....	72
6.2	Saran .....	72
LAMPIRAN .....		77





## DAFTAR GAMBAR

Gambar 3. 1 Desain Skripsi .....	18
Gambar 3. 2 Arsitektur Perancangan Aplikasi.....	21
Gambar 3. 3 Diagram Alir Sistem Keseluruhan .....	22
Gambar 3. 4 Diagram Alir <i>Preprocessing</i> .....	23
Gambar 3. 5 Proses <i>Cleaning</i> .....	24
Gambar 3. 6 Proses Parsing .....	25
Gambar 3. 7 Proses Casefolding .....	26
Gambar 3. 8 Proses Stopword Removal .....	27
Gambar 3. 9 Diagram Alir Proses Stemming Arifin-Setiono .....	29
Gambar 3. 10 Proses Pembobotan .....	30
Gambar 3. 11 Proses K-Nearest Neighbor (KNN).....	31
Gambar 3. 12 Perancangan Desain Antar Muka Halaman Awal Pengujian.....	32
Gambar 3. 13 Perancangan Desain Antar Muka Halaman Hasil.....	33
Gambar 4. 1 Tampilan Halaman dan <i>Form</i> Tambah Data Training .....	52
Gambar 4. 2 <i>Form</i> Pengujian.....	53
Gambar 4. 3 <i>Form</i> Halaman Akurasi.....	54
Gambar 4. 4 <i>Form</i> Halaman <i>Preprocessing</i> .....	54
Gambar 5. 1 Grafik Hasil Pengujian Skenario 1 .....	58
Gambar 5. 2 Grafik Hasil Pengujian Skenario 2.....	59
Gambar 5. 3 Grafik Hasil Pengujian Skenario 3.....	60
Gambar 5. 4 Grafik Hasil Pengujian Skenario 4.....	61
Gambar 5. 5 Grafik Hasil Pengujian 4 Skenario.....	62
Gambar 5. 6 Grafik Hasil Pengujian 4 Skenario.....	63
Gambar 5. 7 Grafik Hasil Pengujian Skenario 1 Tanpa Stemming .....	65
Gambar 5. 8 Grafik Hasil Pengujian Skenario 2 Tanpa Stemming .....	66
Gambar 5. 9 Grafik Hasil Pengujian Skenario 3 Tanpa Stemming .....	67
Gambar 5. 10 Grafik Hasil Pengujian Skenario 4 Tanpa Stemming .....	68
Gambar 5. 11 Grafik Hasil Pengujian 4 Skenario Tanpa Stemming .....	69
Gambar 5. 12 Grafik Hasil Pengujian 4 Skenario.....	70

**DAFTAR TABEL**

Tabel 3. 1 Tabel Manualisasi Dokumen <i>Review</i> .....	34
Tabel 3. 2 Tabel Manualisasi Dokumen <i>Cleaning</i> .....	34
Tabel 3. 3 Tabel Manualisasi Dokumen <i>Casefolding</i> .....	35
Tabel 3. 4 Tabel Manualisasi Dokumen <i>Parsing</i> .....	35
Tabel 3. 5 Tabel Manualisasi Dokumen <i>Stopword Removal</i> .....	36
Tabel 3. 6 Tabel Manualisasi Dokumen <i>Stemming</i> .....	36
Tabel 3. 7 Tabel Informasi Dokumen .....	37
Tabel 3. 8 Tabel Informasi TF .....	38
Tabel 3. 9 Tabel Informasi DF .....	38
Tabel 3. 10 Tabel Informasi IDF .....	39
Tabel 3. 11 Tabel Hasil Perhitungan <i>Cossine Similarity</i> .....	39
Tabel 3. 12 Tabel Hasil Pengurutan <i>Cossine Similarity</i> .....	39
Tabel 3. 13 Tabel Hasil Akhir Sistem .....	40
Tabel 3. 14 Contoh Tabel Pengujian .....	41
Tabel 3. 15 Contoh Tabel Skenario Pengujian .....	42
Tabel 4. 1 Spesifikasi Perangkat Keras Komputer .....	43
Tabel 4. 2 Spesifikasi Perangkat Lunak Komputer .....	44
Tabel 4. 3 Daftar Fungsi Pada Sistem .....	45
Tabel 5. 1 Hasil Pengujian Uji Model .....	55
Tabel 5. 2 Skenario Pengujian .....	57
Tabel 5. 3 Akurasi Pada Skenario 1 .....	57
Tabel 5. 4 Akurasi Pada Skenario 2 .....	58
Tabel 5. 5 Akurasi Pada Skenario 3 .....	59
Tabel 5. 6 Akurasi Pada Skenario 4 .....	60
Tabel 5. 7 Akurasi Rata-Rata .....	61
Tabel 5. 8 Nilai k Terhadap Skenario .....	62
Tabel 5. 9 Skenario Pengujian Tanpa Stemming .....	64
Tabel 5. 10 Akurasi Pada Skenario 1 Tanpa Stemming .....	65
Tabel 5. 11 Akurasi Pada Skenario 2 Tanpa Stemming .....	66
Tabel 5. 12 Akurasi Pada Skenario 3 Tanpa Stemming .....	67



Tabel 5. 13 Akurasi Pada Skenario 4 Tanpa Stemming ..... 68  
Tabel 5. 14 Akurasi Rata-Rata Tanpa Stemming ..... 69  
Tabel 5. 15 Nilai k Terhadap Skenario (Tanpa Stemming) ..... 70





## DAFTAR SOURCE CODE

<i>Sourcecode 4. 1 Implementasi Algoritma Cleaning Pada Dokumen</i> .....	47
<i>Sourcecode 4. 2 Implementasi Algoritma Casefolding</i> .....	47
<i>Sourcecode 4. 3 Implementasi Algoritma Parsing</i> .....	47
<i>Sourcecode 4. 4 Implementasi Algoritma Stopword Removal</i> .....	47
<i>Sourcecode 4. 5 Implementasi Algoritma Stemming</i> .....	48
<i>Sourcecode 4. 6 Implementasi Algoritma TF-IDF, Nilai TF Dan DF Pada Dokumen</i> .....	49
<i>Sourcecode 4. 7 Implementasi Algoritma Pembobotan TF-IDF Pada Dokumen</i> .	50
<i>Sourcecode 4. 8 Implementasi Algoritma Kemiripan Dokumen Dengan Cosine Similarity</i> .....	51



**DAFTAR LAMPIRAN**

LAMPIRAN 1 Daftar *Stopword/stoplist*..... 77

LAMPIRAN 2 Tabel Hasil Percobaan 2 Skenario 1 ..... 78

LAMPIRAN 3 Tabel Hasil Percobaan 2 Skenario 2 ..... 78

LAMPIRAN 4 Tabel Hasil Percobaan 2 Skenario 3 ..... 79

LAMPIRAN 5 Tabel Hasil Percobaan 2 Skenario 4 ..... 79

LAMPIRAN 6 Tabel Hasil Percobaan 3 Skenario 1 ..... 79

LAMPIRAN 7 Tabel Hasil Percobaan 3 Skenario 2 ..... 80

LAMPIRAN 8 Tabel Hasil Percobaan 3 Skenario 3 ..... 80

LAMPIRAN 9 Tabel Hasil Percobaan 3 Skenario 4 ..... 80

