

**TUGAS AKHIR**

**KLASIFIKASI DOKUMEN TEKS MENGGUNAKAN METODE  
SUPPORT VECTOR MACHINE DENGAN PEMILIHAN FITUR CIRI  
CHI-SQUARE**

Disusun Untuk Melengkapi dan Memenuhi Syarat Kelulusan  
Guna Meraih Gelar Sarjana Komputer  
Program Studi Teknik Informatika Universitas Muhammadiyah Jember



**MUHAMMAD ISA HIDAYATULLAH**

1210651073

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS TEKNIK  
UNIVERSITAS MUHAMMADIYAH JEMBER**

**2016**

**TUGAS AKHIR**

**KLASIFIKASI DOKUMEN TEKS MENGGUNAKAN METODE  
SUPPORT VECTOR MACHINE DENGAN PEMILIHAN FITUR CIRI  
CHI-SQUARE**



**MUHAMMAD ISA HIDAYATULLAH**

1210651073

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS TEKNIK  
UNIVERSITAS MUHAMMADIYAH JEMBER**

**2016**

**HALAMAN PENGESAHAN**

**KLASIFIKASI DOKUMEN TEKS MENGGUNAKAN METODE  
SUPPORT VECTOR MACHINE DENGAN PEMILIHAN FITUR CIRI  
CHI-SQUARE**

**MUHAMMAD ISA HIDAYATULLAH**

1210651073

Telah Mempertanggung Jawabkan Laporan Tugas Akhirnya Pada Sidang  
Tugas Akhir Tanggal 29 November 2016 Sebagai Salah Satu Syarat Kelulusan  
Guna Meraih Gelar Sarjana Komputer  
Program Studi Teknik Informatika Universitas Muhammadiyah Jember

Disetujui Oleh :

Dosen Pembimbing :  
Pembimbing I

Dosen Penguji :  
Penguji I

**Lutfi Ali Muharrom, S.Si., M.Si**  
NPK. 10 09 550

**Ari Eko Wardoyo, S.Kom., M.Kom**  
NIP. 19750214 200501 1 001

Pembimbing II

Penguji II

**Yeni Dwi Rahayu, S. ST., M.Kom**  
NPK. 11 03 590

**Wiwik Suharso, S.Kom., M.Kom**  
NPK. 19760906 200601 1 003

Mengesahkan,  
Dekan Fakultas Teknik

Mengetahui,  
Ketua Program Studi Teknik  
Informatika

**Ir. Suhartinah, MT.**  
NPK. 95 05 246

**Yeni Dwi Rahayu, S. ST., M.Kom**  
NPK. 11 03 590

## PERNYATAAN

Yang bertanda tangan di bawah ini :

NIM : 1210651073  
Nama : MUHAMMAD ISA HIDAYATULLAH  
Institusi : Teknik Informatika, Fakultas Teknik,  
Universitas Muhammadiyah Jember.

Menyatakan bahwa Tugas Akhir yang berjudul “**KLASIFIKASI DOKUMEN TEKS MENGGUNAKAN METODE SUPPORT VECTOR MACHINE DENGAN PEMILIHAN FITUR CIRI CHI-SQUARE**” Bukan merupakan karya orang lain kecuali dalam bentuk kutipan yang telah disebutkan sumbernya.

Demikian surat pernyataan ini dibuat dengan sebenar-benarnya dan apabila pernyataan ini tidak benar penulis bersedia mendapatkan sanksi dari akademik.

Jember, 1 Januari 2017

**Muhammad Isa Hidayatullah**  
NIM. 1210651073

## PERSEMBAHAN

Nikmat yang berlimpah dari Allah SWT sungguh menjadi semangat dan dasar bagi penulis dalam menyelesaikan tugas akhir yang berjudul **“KLASIFIKASI DOKUMEN TEKS MENGGUNAKAN METODE SUPPORT VECTOR MACHINE DENGAN PEMILIHAN FITUR CIRI CHI-SQUARE”**. Oleh karena itu, penulis mengucapkan syukur alhamdulillah karena penulis dapat menyelesaikan tugas akhir ini dengan baik. Sholawat serta salam yang selalu tercurahkan kepada nabi Muhammad SAW , keluarga , dan para sahabat beliau hingga pengikutnya sampai akhir zaman.

Tugas akhir ini dapat menjadi alternatif untuk metode klasifikasi dokumen teks yang ada saat ini karena penelitian ini membahas mengenai metode klasifikasi dokumen teks menggunakan pemilihan fitur ciri.

Atas segala upaya, bimbingan dan arahan dari semua pihak, tanpa mengurangi rasa hormat penulis mengucapkan terima kasih yang sebesar-besarnya kepada :

1. Ibu Ir. Suhartinah, MT. selaku Dekan Fakultas Teknik, Universitas Muhammadiyah Jember.
2. Ibu Yeni Dwi Rahayu, S. ST., M. Kom. selaku ketua Program Studi Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Jember dan juga selaku dosen pembimbing 2 yang telah memberikan ilmu , bimbingan dan semangat dengan penuh kesabaran.
3. Bapak Lutfi Ali Muharrom , S.Si, M.Si selaku dosen pembimbing 1 yang telah memberikan ilmu , bimbingan dan semangat dengan penuh kesabaran.
4. Ari Eko Wardoyo, S.Kom, M.Kom selaku dosen penguji 1 dan bapak Wiwik Suharso, S.Kom, M.Kom selaku dosen penguji 2 yang telah memberikan saran dan kritik yang membangun dalam penyelesaian tugas akhir ini.

5. Bapak dan Ibu Dosen Fakultas Teknik Universitas Muhammadiyah Jember yang telah memberikan banyak ilmu dan pengalaman.
6. Kedua orang tua dan keluarga yang selalu mendukung penulis melalui do'a dan memotivasi penulis untuk menyelesaikan tugas akhir ini.
7. Sahabat – sahabat terbaik, Javar, Vanny, Ihwan, Reza, Lukman, Irfan, Adi, Silvi, Lia, terima kasih atas dukungan dan semangatnya.
8. Semua pihak yang telah membantu menyelesaikan tugas akhir ini dari awal hingga akhir. Penulis menyadari bahwa dalam penyusunan tugas akhir ini masih jauh dari sempurna, maka penulis menerima segala saran dan kritik yang sifatnya membangun dari semua pihak demi kesempurnaan tugas akhir ini di masa mendatang. Semoga tugas akhir ini dapat bermanfaat baik sebagai sumber informasi maupun sumber inspirasi bagi pembaca.

Jember, 1 Januari 2017

Penulis

## **KATA PENGANTAR**

Puji syukur kehadiran Allah SWT yang Maha Pengasih lagi Maha Penyayang, Yang hanya kepadaNya-lah segala sesuatu bergantung. Alhamdulillah tak lupa senantiasa saya panjatkan karena hanya dengan ridho, kemurahan dan kekuasaanNya-lah proyek akhir yang berjudul :

### **“KLASIFIKASI DOKUMEN TEKS MENGGUNAKAN METODE SUPPORT VECTOR MACHINE DENGAN PEMILIHAN FITUR CIRI CHI-SQUARE”**

dapat diselesaikan dengan segala kelebihan dan tak lepas dari kekurangan yang terdapat di dalamnya.

Tugas akhir ini menjelaskan klasifikasi dokumen teks menggunakan metode *support vector machine* dengan pemilihan fitur ciri *chi-square*.

Dengan segala kerendahan hati, penulis memohon maaf jika ternyata di kemudian hari diketahui bahwa hasil dari proyek akhir ini masih jauh dari kesempurnaan. Semoga bermanfaat bagi setiap insan yang mempergunakannya untuk kebaikan di jalan Allah SWT.

Jember, 1 Januari 2017

Penulis

## DAFTAR ISI

Halaman Judul.....	i
Halaman Pengesahan .....	ii
Halaman Pernyataan Keaslian.....	iii
Abstrak .....	iv
Abstract .....	v
Lembar Persembahan .....	vi
Kata Pengantar .....	viii
Daftar Isi.....	ix
Daftar Gambar.....	xi
Daftar Tabel .....	xii
Daftar Rumus .....	xiii
<b>BAB I PENDAHULUAN .....</b>	<b>1</b>
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	2
1.3 Batasan Masalah .....	3
1.4 Tujuan Penelitian .....	3
1.5 Manfaat Penelitian .....	3
<b>BAB II TINJAUAN PUSTAKA.....</b>	<b>4</b>
2.1 Klasifikasi Dokumen .....	4
2.2 Preprosesing Dokumen .....	4
2.2.1 Tokenisasi .....	5
2.2.2 Stopwords .....	5
2.2.3 Pembobotan .....	5
2.2.4 Pemilihan Fitur Ciri .....	6
2.3 SVM.....	8
<b>BAB III METODELOGI .....</b>	<b>10</b>
3.1 Metode Penelitian .....	10
3.2 Arsitektur Sistem .....	11
3.3 Tahapan Uji Coba .....	13



3.4	Skenario Penelitian .....	21
BAB IV HASIL DAN PEMBAHASAN .....		23
4.1	Uji Coba.....	23
4.2	Hasil Pengujian.....	23
4.3	Penggunaan Aplikasi .....	31
4.3.1	Data Latih.....	31
4.3.2	Data Uji .....	34
BAB V KESIMPULAN DAN SARAN.....		36
5.1	Kesimpulan.....	36
5.2	Saran .....	36
Daftar Pustaka .....		37
Lampiran .....		38
Biografi Penulis.....		50

## DAFTAR GAMBAR

Gambar 3.1 Flowchart Data Latih.....	11
Gambar 3.2 Flowchart Data Uji .....	12
Gambar 4.1 Halaman Utama Data Latih.....	32
Gambar 4.2 Proses Buka Folder Dokumen.....	32
Gambar 4.3 Hasil Buka Folder Dokumen.....	33
Gambar 4.4 Proses Perhitungan Data Latih .....	33
Gambar 4.5 Proses Simpan Data.....	34
Gambar 4.6 Halaman Utama Data Uji .....	34
Gambar 4.7 Proses Perhitungan Data Uji .....	35
Gambar 4.8 Hasil Klasifikasi .....	35

## DAFTAR TABEL

Tabel 2.1 Kontigensi .....	7
Tabel 3.1 Hasil tokenisasi .....	15
Tabel 3.2 Hasil pembuangan stopwords .....	16
Tabel 3.3 Hasil perhitungan idf.....	17
Tabel 3.4 Hasil perhitungan tfidf .....	18
Tabel 3.5 Hasil chi-square.....	19
Tabel 3.6 Hasil svm.....	20
Tabel 4.1 Deksripsi data.....	23
Tabel 4.2 Hasil tokenisasi dan pembuangan stopwords dokumen 1.....	25
Tabel 4.3 Hasil perhitungan dokumen 1 data latih .....	27
Tabel 4.4 Pelabelan .....	28
Tabel 4.5 Hasil perhitungan dokuemn 1 data uji .....	29
Tabel 4.6 Hasil pengujian skenario 1 .....	30
Tabel 4.7 Hasil akurasi.....	31
Tabel 4.8 Waktu klasifikasi.....	31

## DAFTAR RUMUS

Rumus 2.1 Perhitungan idf.....	5
Rumus 2.2 Perhitungan tfidf .....	5
Rumus 2.3 Perhitungan chi-square .....	6
Rumus 2.4 Perhitungan frekuensi harapan.....	6
Rumus 2.5 Penyederhanaan chi-square.....	7
Rumus 2.6 Perhitungan chi-sqiare dengan tfidf.....	7
Rumus 2.7 Perhitungan svm .....	8
Rumus 2.8 Keluaran svm .....	8
Rumus 2.9 Akurasi.....	8