

**KLASIFIKASI KATEGORI BERITA MENGGUNAKAN
METODE *SUPPORT VECTOR MACHINE* DAN REDUKSI
FITUR DENGAN *SINGULAR VALUE DECOMPOSITION***

SKRIPSI



Oleh :

MUHAMMAD SYAFRIANSYAH

NPM. 1534010032

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"
JAWA TIMUR
2019**

**LEMBAR PENGESAHAN
SKRIPSI**

**Judul : KLASIFIKASI KATEGORI BERITA MENGGUNAKAN
METODE SUPPORT VECTOR MACHINE DAN REDUKSI
FITUR DENGAN SINGULAR VALUE DECOMPOSITION**

**Oleh : MUHAMMAD SYAFRIANSYAH
NPM : 1534010032**

**Telah Diseminarkan Dalam Ujian Skripsi Pada
Pada Tanggal : 17 Mei 2019**

Menyetujui :

Dosen Pembimbing :

1.



**Intan Yuniar P., S.Kom, M.Sc
NPT. 3 8006 04 0198 1**

2.



**Wahyu S. J. Saputra, S.Kom, M.Kom
NPT : 3 8608 10 0295 1**

Dosen Penguji :

1.



**Intan Yuniar P., S.Kom, M.Sc
NPT. 3 8006 04 0198 1**

2.



**Eva Yulia P, S.Kom, M.Kom
NPT. 3 8907 13 0 346 1**

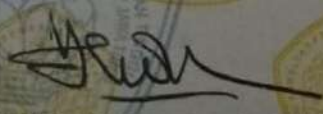
3.



**Firza Prima A., S.Kom, M.TI
NPT. 3 8605 13 0344 1**

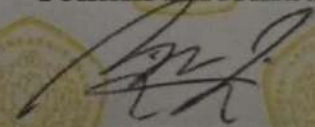
Mengetahui :

**Dekan
Fakultas Ilmu Komputer,**



**Dr. Ir. Ni Ketut Sari, MT
NIP. 19650731 199203 2 001**

**Koordinator Program Studi
Teknik Informatika**



**Budi Nugroho, S.Kom, M.Kom
NPT. 3 8009 05 0205 1**

SURAT PERNYATAAN ANTI PLAGIAT

Saya, mahasiswa Teknik Informatika UPN “Veteran” Jawa Timur, yang bertandatangan dibawah ini:

Nama : MUHAMMAD SYAFRIANSYAH

NPM : 1534010032

Menyatakan bahwa Judul Skripsi / Tugas Akhir yang Saya ajukan dan akan dikerjakan, yang berjudul:

“KLASIFIKASI KATEGORI BERITA MENGGUNAKAN METODE SUPPORT VECTOR MACHINE DAN REDUKSI FITUR DENGAN SINGULAR VALUE DECOMPOSITION”

Bukan merupakan plagiat dari Skripsi / Tugas Akhir / Penelitian orang lain dan juga bukan merupakan produk dan atau software yang saya beli dari pihak lain. Saya juga menyatakan bahwa Skripsi / Tugas Akhir ini adalah pekerjaan Saya sendiri, kecuali yang dinyatakan dalam Daftar Pustaka dan tidak pernah diajukan untuk syarat memperoleh gelar di UPN “Veteran” Jawa Timur maupun di institusi pendidikan lain.

Jika ternyata di kemudian hari pernyataan ini terbukti tidak benar, maka Saya siap menerima segala konsekuensinya.

Surabaya, 22 Mei 2019



MUHAMMAD SYAFRIANSYAH
NPM. 1534010032

KLASIFIKASI KATEGORI BERITA MENGGUNAKAN METODE SUPPORT VECTOR MACHINE DAN REDUKSI FITUR DENGAN SINGULAR VALUE DECOMPOSITION

Nama Mahasiswa : Muhammad Syafriansyah
NPM : 1534010032
Program Studi : Teknik Informatika
Dosen Pembimbing : Intan Yuniar Purbasari, S.kom., M.Sc.
Wahyu S. J. Saputra, S.Kom., M.Kom.

Abstrak

Berita yang dahulu banyak disebarakan melalui media cetak sekarang mulai memasuki atau bahkan berpindah pada media *online* untuk penyebarannya. Kelebihan dari media *online* adalah waktu distribusi yang lebih cepat. Pada saat ini pengkategorian berita masih menggunakan tenaga manusia atau manual. Kategori yang banyak beserta waktu yang cepat menyulitkan editor untuk mengkategorikan berita. Penelitian ini bertujuan untuk melakukan prediksi kategori berita menggunakan metode *Support Vector Machine (SVM)* dipadukan dengan *Singular Value Decomposition (SVD)*.

Data yang digunakan merupakan data yang diambil dengan teknik *crawling* dari situs <https://www.kompas.com/> berjumlah 280 data yang terdiri dari empat kategori yaitu ekonomi, olahraga, teknologi, dan entertainment. Pertama-tama dilakukan *text preprocessing*, dalam proses ini nantinya terdapat beberapa proses yaitu *case folding*, tokenisasi, *filtering*, dan *stemming*. Selanjutnya proses seleksi fitur menggunakan penghitungan bobot *TF-IDF* serta proses *SVD* yang berfungsi untuk mereduksi fitur. Kemudian proses klasifikasi dengan metode *multiclass SVM one-against-one* menggunakan *kernel polynomial*.

Dari hasil penelitian ini diketahui bahwa penggunaan *SVD* dapat meningkatkan nilai akurasi, presisi, dan recall yang dihasilkan oleh *SVM*. Rata-rata nilai yang dihasilkan oleh *TF-IDF + SVD + SVM* adalah akurasi = 0.98, presisi = 0.95, dan recall = 0.96.

Kata Kunci: *Klasifikasi, Berita, SVD, SVM*

KATA PENGANTAR

Puji syukur penulis panjatkan pada Tuhan Yang Maha Esa yang telah menganugerahkan rahmat serta karunia-Nya kepada kita, sehingga penulis diberikan kekuatan dan kesabaran untuk menyelesaikan laporan skripsi dengan judul **“KLASIFIKASI KATEGORI BERITA MENGGUNAKAN METODE SUPPORT VECTOR MACHINE DAN REDUKSI FITUR DENGAN SINGULAR VALUE DECOMPOSITION”** tepat pada waktunya. Laporan ini dibuat untuk memenuhi persyaratan untuk menyelesaikan tugas akhir pada program studi Teknik Informatika di Fakultas Ilmu Komputer UPN “Veteran” Jawa Timur.

Penulis menyadari bahwa laporan skripsi ini masih jauh dari kesempurnaan. Atas segala kekurangan penulis sangat mengharapkan masukan, kritik, dan saran yang bersifat membangun kearah perbaikan dan penyempurnaan laporan skripsi ini.

Akhir kata penulis mengucapkan terimakasih kepada semua pihak yang berperan dalam penyusunan laporan skripsi ini. Penulis berharap semoga laporan skripsi ini dapat bermanfaat bagi semua pihak.

Surabaya, Mei 2019

Penulis

UCAPAN TERIMA KASIH

Segala puji hanyalah milik Allah, sumber segala hikmah dan ilmu pengetahuan karena berkat rahmat dan berkah-Nya penulis dapat menyusun dan menyelesaikan laporan Skripsi ini hingga selesai tepat pada waktunya. Dengan segala hormat, Ucapan terima kasih saya sampaikan ke berbagai pihak yang turut membantu memperlancar penyelesaian laporan Skripsi ini, yaitu kepada:

1. Kedua orang tua, Ruswandi dan Suwaibah yang selalu mendoakan keberhasilan penulis serta selalu memberi motivasi untuk menyelesaikan semua tugas, dan selalu memberikan fasilitas yang mendukung untuk mempercepat penyelesaian laporan skripsi ini.
2. Pakde dan Bude, yang telah memperbolehkan menempati tempat tinggal selama kuliah sehingga penulis dapat menyelesaikan laporan skripsi ini tepat pada waktunya.
3. Ibu Dr. Ir. Ni Ketut Sari, MT selaku dekan Fakultas Ilmu Komputer UPN “Veteran” Jawa Timur.
4. Bapak Budi Nugroho, S.Kom. M.Kom selaku kepala program studi Teknik Informatika Fakultas Ilmu Komputer UPN “Veteran” Jawa Timur.
5. Ibu Intan Yuniar Purbasari, S.Kom., M.Sc selaku dosen pembimbing 1 yang dengan sabar membimbing, mengarahkan serta memberikan masukan sejak awal penelitian ini berlangsung hingga akhir.

6. Bapak Wahyu S. J. Saputra, S.Kom, M.Kom selaku dosen pembimbing 2 yang dengan sabar membimbing, mengarahkan serta memberikan masukan sejak awal penelitian ini berlangsung hingga akhir.
7. Bapak dan Ibu dosen program studi Teknik Informatika UPN “Veteran” Jawa Timur yang telah mendidik dan memberikan ilmunya sehingga penulis memiliki bekal untuk dapat melakukan penelitian ini.
8. Teman-teman UK Persma UPN “Veteran” Jawa Timur yang selalu memberi semangat dan motivasi dalam pengerjaan laporan skripsi ini.
9. Teman-teman jurusan Teknik Informatika, serta semua pihak yang telah membantu proses perkuliahan di Teknik Informatika yang tidak bisa penulis sebutkan satu persatu.

Semoga Tuhan Yang Maha Esa memberikan balasan yang setimpal kepada semua pihak yang telah memberikan bantuan, bimbingan maupun motivasi yang bermanfaat bagi penulis. Penulis juga menyadari bahwa masih banyak kekurangan pada penulisan Laporan Skripsi ini. Oleh karena itu, saran dan kritik yang membangun akan penulis terima dengan senang hati dan penulis berharap semoga Laporan Skripsi ini dapat memberikan manfaat dalam perkembangan ilmu pengetahuan bagi pembaca.

DAFTAR ISI

LEMBAR PENGESAHAN SKRIPSI	i
SURAT PERNYATAAN ANTI PLAGIAT	ii
Abstrak	iii
KATA PENGANTAR	iv
UCAPAN TERIMA KASIH.....	v
DAFTAR ISI.....	vii
DAFTAR GAMBAR	x
DAFTAR TABEL.....	xi
DAFTAR KODE.....	xiii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang Masalah.....	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah.....	4
1.4 Tujuan Penelitian.....	4
1.5 Manfaat Penelitian.....	4
BAB II TINJAUAN PUSTAKA.....	5
2.1 Penelitian Terdahulu.....	5
2.2 Berita	7
2.3 <i>Text Mining</i>	7
2.4 <i>Text Preprocessing</i>	8
2.5 <i>Term Frequency-Inverse Document Frequency</i>	11
2.6 <i>Singular Value Decomposition</i>	12

2.7	<i>Support Vector Machine</i>	13
BAB III METODOLOGI.....		18
3.1	Jenis dan Sumber Data	18
3.2	Metode Penelitian.....	18
3.2.1	Alur Kerja Sistem.....	19
3.2.2	<i>Text Preprocessing</i>	20
3.2.3	<i>TF-IDF</i>	23
3.2.4	<i>SVD</i>	28
3.2.5	Proses Hitung Kernel	31
3.2.6	Proses <i>Training SVM</i>	32
3.2.7	Proses <i>Testing SVM</i>	34
3.2.8	<i>SVM one-against-one</i>	35
3.2.9	Pengukuran Performa.....	36
3.2.10	Skenario Pengujian.....	37
3.2.11	Contoh Perhitungan Kernel.....	38
3.2.12	Contoh Perhitungan Training SVM	39
3.2.13	Contoh Perhitungan Testing SVM.....	42
3.3	<i>Class Diagram</i>	43
BAB IV HASIL DAN PEMBAHASAN		46
4.1	Pengumpulan <i>Dataset</i>	46
4.2	Implementasi Sistem	49
4.2.1.	<i>Text Preprocessing</i>	49
4.2.2.	<i>TF-IDF</i>	52
4.2.3.	<i>SVD</i>	54

4.2.4.	Proses <i>Training SVM</i>	55
4.2.5.	Proses <i>Testing SVM</i>	57
4.2.6.	<i>Multiclass SVM</i>	58
4.2.7.	<i>Confusion Matrix</i>	59
4.2.8.	Implementasi Antarmuka	60
4.3	Uji Coba	62
4.3.1.	Percobaan I.....	62
4.3.2.	Percobaan II	64
4.3.3.	Percobaan III	66
4.3.4.	Percobaan IV	68
4.3.5.	Analisa Hasil Uji Coba.....	70
4.3.6.	Pengujian nilai k-rank	72
4.4	Validasi.....	72
BAB V KESIMPULAN DAN SARAN.....		76
5.1	Kesimpulan.....	76
5.2	Saran	77
DAFTAR PUSTAKA		78
BIODATA PENULIS		80

DAFTAR GAMBAR

Gambar 2. 1 Bidang Pemisah Terbaik <i>Support Vector Machine</i>	14
Gambar 3. 1 Rancangan Tabel Berita	18
Gambar 3. 2 <i>Flowchart</i> Alur Sistem	19
Gambar 3. 3 <i>Flowchart Text Preprocessing</i>	20
Gambar 3. 4 <i>Flowchart</i> Proses Perhitungan <i>TF-IDF</i>	24
Gambar 3. 5 <i>Flowchart</i> Proses Perhitungan <i>SVD</i>	28
Gambar 3. 6 <i>Flowchart</i> Proses Hitung Kernel Data <i>Training</i>	31
Gambar 3. 7 <i>Flowchart</i> Proses <i>Training SVM</i>	32
Gambar 3. 8 <i>Flowchart</i> Proses <i>Testing SVM</i>	34
Gambar 3. 9 Rancangan <i>Class Diagram</i>	44
Gambar 4. 1 Halaman <i>Crawl</i> Berita	46
Gambar 4. 2 <i>Database</i> Tempat Menyimpan Berita	48
Gambar 4. 3 Implementasi Antarmuka Halaman Utama	60
Gambar 4. 4 Tampilan Antarmuka <i>Form</i> Validasi	61

DAFTAR TABEL

Tabel 3. 1 Perhitungan <i>TF</i>	25
Tabel 3. 2 Perhitungan <i>IDF</i>	26
Tabel 3. 3 Perhitungan <i>TF-IDF</i>	27
Tabel 3. 4 Gambaran Matriks <i>A</i>	29
Tabel 3. 5 Gambaran Klasifikasi <i>one-against-one</i>	35
Tabel 3. 6 Gambaran <i>Multiclass Confusion Matrix</i>	36
Tabel 3. 7 Hasil Perhitungan <i>Kernel</i>	38
Tabel 3. 8 Hasil Perhitungan Matriks <i>Hessian</i>	39
Tabel 4. 1 Hasil Pengujian Percobaan I <i>TF-IDF + SVM</i>	63
Tabel 4. 2 Hasil Pengujian Percobaan I <i>TF-IDF + SVD + SVM</i>	63
Tabel 4. 3 Hasil Percobaan I.....	64
Tabel 4. 4 Hasil Pengujian Percobaan II <i>TF-IDF + SVM</i>	64
Tabel 4. 5 Hasil Pengujian Percobaan II <i>TF-IDF + SVD + SVM</i>	65
Tabel 4. 6 Hasil Percobaan II.....	66
Tabel 4. 7 Hasil Pengujian Percobaan III <i>TF-IDF + SVM</i>	66
Tabel 4. 8 Hasil Pengujian Percobaan III <i>TF-IDF + SVD + SVM</i>	67
Tabel 4. 9 Hasil Percobaan III.....	67
Tabel 4. 10 Hasil Pengujian Percobaan IV <i>TF-IDF + SVM</i>	68
Tabel 4. 11 Hasil Pengujian Percobaan IV <i>TF-IDF + SVD + SVM</i>	69
Tabel 4. 12 Hasil Percobaan IV.....	69
Tabel 4. 13 Hasil Pengujian Akurasi.....	70
Tabel 4. 14 Hasil Pengujian Presisi.....	70

Tabel 4. 15 Hasil Pengujian Recall	70
Tabel 4. 16 Hasil Pengujian Waktu.....	71
Tabel 4. 17 Hasil Terbaik Setiap Aspek	71
Tabel 4. 18 Hasil Pengujian k-rank.....	72
Tabel 4. 19 Hasil Validasi.....	73
Tabel 4. 20 <i>Confusion Matrix</i> Hasil Validasi.....	74
Tabel 4. 21 Hasil Pengujian Validasi.....	74

DAFTAR KODE

Kode 4. 1 <i>Script Proses Crawling</i>	47
Kode 4. 2 <i>Script Proses Case Folding</i>	49
Kode 4. 3 <i>Script Proses Tokenisasi</i>	50
Kode 4. 4 <i>Script Proses Filtering</i>	50
Kode 4. 5 <i>Script Proses Stemming</i>	51
Kode 4. 6 <i>Script Perhitungan TF-IDF</i>	53
Kode 4. 7 <i>Script Proses SVD</i>	54
Kode 4. 8 <i>Script Perhitungan Vektor Dokumen SVD</i>	55
Kode 4. 9 <i>Script Proses Training SVM</i>	56
Kode 4. 10 <i>Script Proses Testing SVM</i>	57
Kode 4. 11 <i>Script Proses Training Multiclass SVM</i>	58
Kode 4. 12 <i>Script Proses Testing Multiclass SVM</i>	59
Kode 4. 13 <i>Script Pembentukan Confusion Matrix</i>	59