

**PERBANDINGAN PERFORMA KERNEL PADA ALGORITMA
SUPPORT VECTOR MACHINE DAN *TERM FREQUENCY*
TERHADAP ANALISIS SENTIMEN**

SKRIPSI



disusun oleh

Martha Presina

17.11.1200

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2021**

**PERBANDINGAN PERFORMA KERNEL PADA ALGORITMA
SUPPORT VECTOR MACHINE DAN *TERM FREQUENCY*
TERHADAP ANALISIS SENTIMEN**

SKRIPSI

untuk memenuhi sebagian persyaratan
mencapai gelar Sarjana
pada Program Studi Informatika



disusun oleh

Martha Presina

17.11.1200

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2021**

PERSETUJUAN

SKRIPSI

**PERBANDINGAN PERFORMA KERNEL PADA ALGORITMA
SUPPORT VECTOR MACHINE DAN *TERM FREQUENCY* TERHADAP
ANALISIS SENTIMEN**

yang dipersiapkan dan disusun oleh

Martha Presina

17.11.1200

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 30 Oktober 2020

Dosen Pembimbing,

Anggit Dwi Hartanto, M.Kom.
NIK. 190302163

PENGESAHAN

SKRIPSI

PERBANDINGAN PERFORMA KERNEL PADA ALGORITMA *SUPPORT VECTOR MACHINE* DAN *TERM FREQUENCY* TERHADAP ANALISIS SENTIMEN

yang dipersiapkan dan disusun oleh

Martha Presina

17.11.1200

telah dipertahankan di depan Dewan Penguji
pada tanggal 17 Juni 2021

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Bambang Sudaryatno, Drs., M.M.
NIK. 190302029

Lilis Dwi Farida, S.Kom, M.Eng
NIK. 190302288

Anggit Dwi Hartanto, M.Kom
NIK. 190302163

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 22 Juni 2021

DEKAN FAKULTAS ILMU KOMPUTER

Hanif Al Fatta, M.Kom
NIK. 190302096

PERNYATAAN

Saya yang bertandatangan dibawah ini menyatakan bahwa, skripsi ini merupakan karya saya sendiri (ASLI), dan isi dalam skripsi ini tidak terdapat karya yang pernah diajukan oleh orang lain untuk memperoleh gelar akademis di suatu institusi pendidikan tinggi manapun, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis dan/atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Segala sesuatu yang terkait dengan naskah dan karya yang telah dibuat adalah menjadi tanggungjawab saya pribadi.

Yogyakarta, 26 Juni 2021



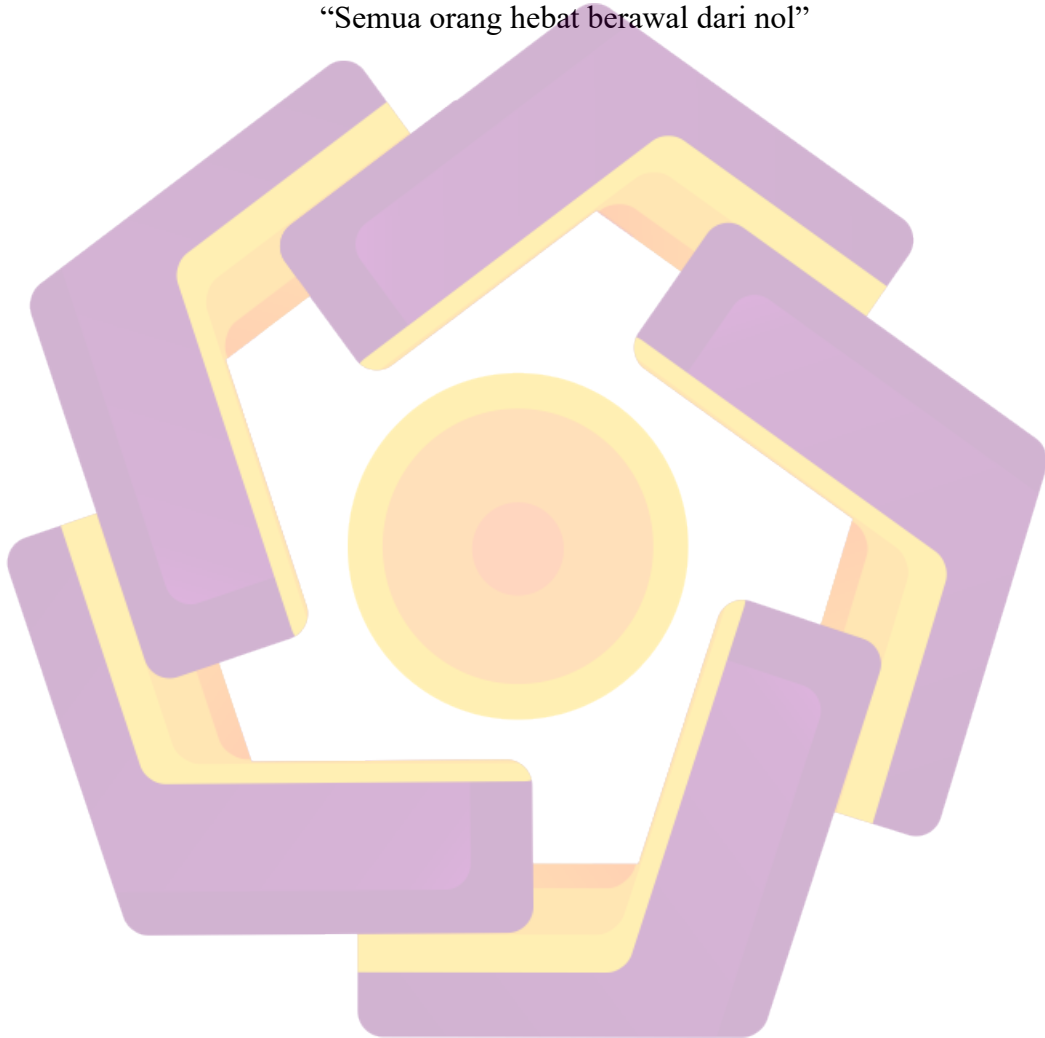
Martha Presina

NIM. 17.11.1200

MOTTO

“Apabila menginginkan hasil yang terbaik, maka lakukanlah yang terbaik pula
sedini mungkin”

“Semua orang hebat berawal dari nol”



PERSEMBAHAN

Dengan penuh rasa syukur, saya persembahkan skripsi ini kepada semua pihak yang terlibat secara langsung maupun tidak langsung dalam proses pembuatan skripsi.

- Allah SWT yang sentiasa memberi nikmat dan kasih sayang kepada seluruh umat-Nya.
- Kedua orang tua beserta keluarga saya yang selalu mendoakan, memberi motivasi dan dukungan kepada saya.
- Bapak Anggit Dwi Hartanto, M.Kom selaku dosen pembimbing yang telah membimbing saya dari awal hingga akhir pembuatan skripsi ini.
- Dosen-dosen Universitas Amikom Yogyakarta yang telah memberikan banyak ilmu pengetahuan kepada saya selama perkuliahan.
- Teman-teman seperjuangan dari awal perkuliahan hingga skripsi, Youse Nur fauzi, Renalda Aji, Valoryan, Yan Gurin Ivanda, dan Bill gani yang selalu berbagi keluh kesah, memberikan solusi di masa sulit selama pembuatan skripsi.
- Teman-teman kelas 17 IF 05 yang selalu menemani perkuliahan dan memberi semangat sampai saat ini.
- Muttafiah yang selalu menemani, memberikan semangat, menjadi motivasi bagi saya untuk menyelesaikan skripsi ini.

KATA PENGANTAR

Puji dan syukur kehadiran Allah SWT yang telah melimpahkan rahmat, taufik, dan hidayah-Nya kepada kita semua. Tak lupa pula shalawat serta salam penulis curahkan kepada junjungan kita Nabi Muhammad SAW sehingga penulis dapat menyelesaikan skripsi yang berjudul “PERBANDINGAN PERFORMA KERNEL PADA ALGORITMA *SUPPORT VECTOR MACHINE* DAN *TERM FREQUENCY* TERHADAP ANALISIS SENTIMEN”.

Skripsi ini disusun guna menyelesaikan studi jenjang Strata Satu (S1) pada program studi Informatika fakultas Ilmu Komputer Universitas Amikom Yogyakarta. Selain itu juga sebagai bukti bahwa mahasiswa telah memenuhi salah satu syarat untuk memperoleh gelar Sarjana Komputer.

Dalam penyusunan skripsi ini, penulis menyadari bahwa masih banyak kekurangan dan jauh dari kata sempurna. Selain itu dengan terselesaikannya skripsi ini, penulis ingin menyampaikan rasa terima kasih kepada:

1. Allah SWT atas segala berkah dan karunia serta ridho-Nya sehingga penulis dapat menyelesaikan skripsi ini dengan baik.
2. Nabi Muhammad SAW yang selalu menjadi panutan dan suri tauladan bagi umat manusia.
3. Kedua orang tua dan keluarga penulis yang selalu memberikan dukungan dalam bentuk apapun.
4. Bapak Prof. Dr. M. Suyanto, M.M., selaku Rektor Universitas Amikom Yogyakarta.

5. Bapak Anggit Dwi Hartanto, M.Kom, selaku dosen pembimbing yang telah memberikan bimbingan, arahan, dan masukan kepada penulis.
6. Seluruh staf pengajar dan karyawan Universitas Amikom Yogyakarta yang telah memberikan ilmu yang bermanfaat.
7. Dia yang selalu menemani, memberi semangat dan bantuan kepada saya.
8. Teman-temanku semuanya yang selalu memberikan bantuannya.
9. Serta semua pihak yang telah membantu penulis baik secara langsung maupun tidak langsung.

Semoga Allah SWT memberikan balasan lebih kepada semua yang telah ikut membantu penulis hingga terselesainya skripsi ini. Semoga skripsi ini dapat bermanfaat bagi penulis maupun pembaca.

Yogyakarta, 26 Juni 2021



Martha Presina

DAFTAR ISI

HALAMAN JUDUL.....	i
PERSETUJUAN	ii
PENGESAHAN	iii
PERNYATAAN.....	iv
MOTTO	v
PERSEMBAHAN.....	vi
KATA PENGANTAR	vii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xii
DAFTAR GAMBAR	xiii
INTISARI.....	xv
<i>ABSTRACT</i>	xvi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah.....	4
1.4 Tujuan Penelitian.....	4
1.5 Manfaat Penelitian.....	5
1.6 Metode Penelitian.....	5
1.6.1 Pengumpulan Data	5
1.6.2 Metode Analisis	6
1.6.3 Metode Pengujian	7
1.7 Sistematika Penulisan.....	8

BAB I	PENDAHULUAN.....	8
BAB II	LANDASAN TEORI.....	8
BAB III	METODE PENELITIAN.....	8
BAB IV	HASIL DAN PEMBAHASAN.....	8
BAB V	PENUTUP.....	9
BAB II	LANDASAN TEORI.....	10
2.1	Kajian Pustaka.....	10
2.2	Dasar Teori.....	18
2.2.1	Analisis Sentimen.....	18
2.2.2	Data Mining.....	19
2.2.3	Text Preprocessing.....	21
2.2.4	Feature Extraction.....	22
2.2.5	Support Vector Machine (SVM).....	24
2.2.6	Confusion Matrix.....	33
BAB III	METODE PENELITIAN.....	36
3.1	Alur Penelitian.....	36
3.2	Alat dan Bahan Penelitian.....	37
3.2.1	Alat.....	37
3.2.2	Bahan.....	38
3.3	Metode Analisis.....	38
3.3.1	Pengumpulan Data.....	38
3.3.2	Pengolahan Awal Data.....	40
3.3.3	Klasifikasi SVM.....	51
3.4	Perancangan Sistem.....	58
3.4.1	Flowchart.....	58

3.5	Perancangan Antarmuka Pengguna.....	59
3.5.1	Halaman Prediksi Data Crawling.....	59
3.5.2	Halaman Perbandingan Kombinasi Kernel SVM dan <i>Term-Frequency</i>	60
BAB IV HASIL DAN PEMBAHASAN		63
4.1	Pengujian.....	63
4.1.1	Confusion Matrix	63
4.1.2	K-Fold Cross Validation	68
4.1.3	Hasil Perbandingan	70
4.2	Antarmuka.....	71
4.2.1	Halaman Utama.....	71
4.2.2	Halaman Prediksi Sentimen	72
4.2.3	Halaman Komparasi Model	74
BAB V PENUTUP.....		78
5.1	Kesimpulan.....	78
5.2	Saran.....	78
DAFTAR PUSTAKA		80
LAMPIRAN.....		83

DAFTAR TABEL

Tabel 2.1	Perbedaan penelitian	14
Tabel 2.2	<i>Confusion Matrix</i>	34
Tabel 3.1	Hasil <i>crawling</i> data <i>twitter</i>	39
Tabel 3.2	Hasil Pelabelan Data	42
Tabel 3.3	Data Setelah <i>Cleaning</i>	44
Tabel 3.4	Data Setelah <i>Case Folding</i>	46
Tabel 3.5	Hasil Akhir <i>Text Preprocessing</i>	48
Tabel 3.6	Perbandingan Hasil TF-IDF dan TF-RF	51
Tabel 4.1	<i>Confusion Matrix Polynomial</i> – TF-IDF	63
Tabel 4.2	<i>Confusion Matrix RBF</i> – TF-IDF	65
Tabel 4.3	<i>Confusion Matrix Polynomial</i> – TF-RF	66
Tabel 4.4	<i>Confusion Matrix RBF</i> – TF-RF	67
Tabel 4.5	Tabel Perbandingan Model	70

DAFTAR GAMBAR

Gambar 2.1	Kerangka proses <i>text mining</i>	21
Gambar 2.2	<i>Hard Margin SVM</i>	25
Gambar 2.3	<i>Hyperplane</i> terbaik	26
Gambar 2.4	<i>Miss</i> klasifikasi pada <i>soft margin SVM</i>	29
Gambar 2.5	Kernel SVM untuk memisahkan data secara linear	30
Gambar 3.1	Diagram alur penelitian	36
Gambar 3.2	<i>Script Crawling Data Twitter</i>	39
Gambar 3.3	<i>Script labelling data</i>	41
Gambar 3.4	<i>Script cleaning data</i>	44
Gambar 3.5	<i>Script Case Folding</i>	45
Gambar 3.6	<i>Script Tahap Tokenizing, Stopword Removal, dan Stemming</i>	47
Gambar 3.7	<i>Script Pembagian Data Latih dan Data Uji</i>	49
Gambar 3.8	<i>Script TF-IDF</i>	50
Gambar 3.9	<i>Script TF-RF</i>	50
Gambar 3.10	Tuning Parameter	52
Gambar 3.11	Hasil Tuning Skema 1	53
Gambar 3.12	Hasil Tuning Skema 2	53
Gambar 3.13	Hasil Tuning Skema 3	53
Gambar 3.14	Hasil Tuning Skema 4	54
Gambar 3.15	Pengaplikasian hasil tuning	55
Gambar 3.16	Klasifikasi dengan Kernel Poly dan TF-IDF	55
Gambar 3.17	Klasifikasi dengan Kernel RBF dan TF-IDF	56
Gambar 3.18	Klasifikasi dengan Kernel Poly dan TF-RF	56
Gambar 3.19	Klasifikasi dengan Kernel RBF dan TF-RF	57
Gambar 3.20	<i>Script</i> untuk menyimpan model	58
Gambar 3.21	<i>Flowchart Sistem</i>	59
Gambar 3.22	Rancangan Halaman Prediksi Data Crawling	60
Gambar 3.23	Rancangan Halaman Hasil Prediksi Data Crawling	60
Gambar 3.24	Rancangan Halaman Utama Perbandingan Kombinasi	61

Gambar 3.25	Rancangan Halaman untuk Menampilkan Data Upload	61
Gambar 3.26	Rancangan Halaman Hasil Preprocessing Data	62
Gambar 3.27	Rancangan Halaman Hasil Perbandingan Kombinasi	62
Gambar 4.1	Performa Kernel <i>Polynomial</i> – TF-IDF	64
Gambar 4.2	<i>Classification Report Polynomial</i> – TF-IDF	64
Gambar 4.3	Performa Kernel RBF – TF-IDF	65
Gambar 4.4	<i>Classification Report RBF</i> – TF-IDF	66
Gambar 4.5	Performa Kernel <i>Polynomial</i> – TF-RF	67
Gambar 4.6	<i>Classification Report Polynomial</i> – TF-RF	67
Gambar 4.7	Performa Kernel RBF – TF-RF	68
Gambar 4.8	<i>Classification Report RBF</i> – TF-RF	68
Gambar 4.9	Hasil <i>K-Fold Cross Validation</i> Kernel <i>Polynomial</i> – TF-IDF	69
Gambar 4.10	Hasil <i>K-Fold Cross Validation</i> Kernel RBF – TF-IDF	69
Gambar 4.11	Hasil <i>K-Fold Cross Validation</i> Kernel <i>Polynomial</i> – TF-RF	69
Gambar 4.12	Hasil <i>K-Fold Cross Validation</i> Kernel RBF – TF-RF	70
Gambar 4.13	Antarmuka Halaman Utama	72
Gambar 4.14	Antarmuka Halaman Prediksi Sentimen	73
Gambar 4.15	Antarmuka Halaman Hasil Prediksi Sentimen	74
Gambar 4.16	Antarmuka Halaman Komparasi Model	75
Gambar 4.17	Antarmuka Halaman Sampel File Terupload	76
Gambar 4.18	Antarmuka Halaman Hasil <i>Preprocessing Data</i>	76
Gambar 4.19	Antarmuka Halaman Hasil Komparasi Model	77

INTISARI

Konsep dasar dalam analisis sentimen adalah *text mining*, dimana data teks yang tidak terstruktur harus diubah menjadi data semi terstruktur agar dapat ditemukan pola pada suatu dokumen. Proses pengubahan data teks menjadi data semi terstruktur tersebut merupakan tahap *text preprocessing*. Tahapan yang penting dalam *text preprocessing* yaitu *term weighting* (pembobotan kata). Metode yang sering digunakan untuk melakukan pembobotan kata terhadap suatu dokumen yaitu *Term Frequency-Inverse Document Frequency* (TF-IDF) dan *Term Frequency-Relevance Frequency* (TF-RF). Selain itu pada tahap klasifikasi terdapat metode yang populer digunakan untuk analisis sentimen yaitu *Support Vector Machine* (SVM) dengan beberapa jenis kernelnya. Kernel SVM yang mampu menghasilkan performa terbaik diantaranya *Polynomial* dan *Radial Basis Function* (RBF). Sehingga dapat dianalisis untuk mendapatkan kombinasi terbaik antara metode pembobotan kata dan kernel SVM.

Langkah awal yang dilakukan yaitu *crawling* data menggunakan twitter API, data yang berhasil dikumpulkan diberikan label terlebih dahulu menggunakan *library* dari *python*. Setelah itu melakukan *preprocessing* terhadap data, data yang telah bersih akan dilakukan pembobotan kata menggunakan metode TF-IDF dan TF-RF sehingga terdapat 2 dokumen yang berbeda untuk diklasifikasi. Langkah terakhir yaitu melakukan klasifikasi menggunakan kernel SVM yaitu *Polynomial* dan RBF, dimana tiap kernel tersebut diaplikasikan ke dalam tiap dokumen TF-IDF dan TF-RF, sehingga akan terbentuk 4 skema berbeda.

Untuk pengujian performa digunakan metode *confusion matrix* dan *K-Fold Cross Validation*. Dari hasil pengujian yang dilakukan, skema TF-RF & RBF menghasilkan performa terbaik dibanding dengan skema lainnya yaitu akurasi rata-rata sebesar 73%. Sedangkan skema yang memperoleh performa paling rendah yaitu TF-IDF & *Polynomial* sebesar 64%.

Kata-kunci: Kernel Polynomial SVM, Kernel Radial Basis Function SVM, TF-IDF, TF-RF, Analisis sentimen

ABSTRACT

The basic concept in sentiment analysis is text mining, where unstructured text data must be converted into semi-structured data so that patterns can be found in a document. The process of converting text data into semi-structured data is a text preprocessing stage. An important step in text preprocessing is term weighting. The methods that are often used to weight words on a document are Term Frequency-Inverse Document Frequency (TF-IDF) and Term Frequency-Relevance Frequency (TF-RF). In addition, at the classification stage, there is a popular method used for sentiment analysis, namely the Support Vector Machine (SVM) with several types of kernels. SVM kernels that are able to produce the best performance include Polynomial and Radial Basis Function (RBF). So it can be analyzed to get the best combination of word weighting method and SVM kernel.

The first step is data crawling using the Twitter API, the data that has been collected is labeled first using the python library. After that, preprocessing the data, the clean data will be weighted using the TF-IDF and TF-RF methods so that there are 2 different documents to classify. The final step is to classify using the SVM kernel, namely Polynomial and RBF, where each kernel is applied to each TF-IDF and TF-RF document, so that 4 different schemes will be formed.

For performance testing, confusion matrix and K-Fold Cross Validation methods are used. From the results of the tests carried out, the TF-RF & RBF scheme produced the best performance compared to other schemes, namely an average accuracy of 73%. Meanwhile, the scheme with the lowest performance was TF-IDF & Polynomial by 64%.

Keyword: SVM Polynomial Kernel, SVM Radial Basis Function Kernel, TF-IDF, TF-RF, Sentiment Analysis