

**IMPLEMENTASI SUPPORT VECTOR MACHINE DENGAN KERNEL  
TRICK UNTUK FILTRASI KOMENTAR  
CYBERBULLYING YOUTUBE**

**SKRIPSI**



Disusun oleh

**Ahmad Farid Mansur**

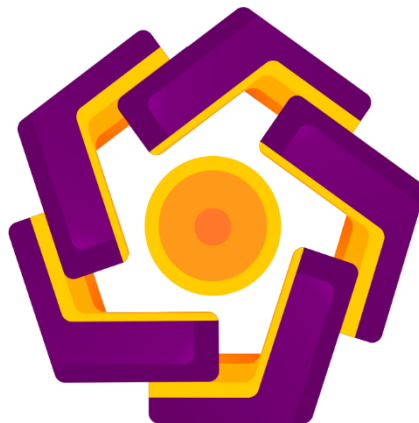
**16.11.0786**

**PROGRAM STUDI INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA  
2020**

**IMPLEMENTASI SUPPORT VECTOR MACHINE DENGAN KERNEL  
TRICK UNTUK FILTRASI KOMENTAR  
CYBERBULLYING YOUTUBE**

**SKRIPSI**

untuk memenuhi sebagian persyaratan  
mencapai gelar Sarjana  
pada Program Studi Informatika



Disusun oleh

**Ahmad Farid Mansur**

**16.11.0786**

**PROGRAM STUDI INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA  
2020**

**PERSETUJUAN**

**SKRIPSI**

**IMPLEMENTASI SUPPORT VECTOR MACHINE DENGAN KERNEL  
TRICK UNTUK FILTRASI KOMENTAR  
CYBERBULLYING YOUTUBE**

yang dipersiapkan dan disusun oleh

**Ahmad Farid Mansur**

**16.11.0786**

telah disetujui oleh Dosen Pembimbing Skripsi  
pada tanggal 19 Maret 2020

**Dosen Pembimbing,**

**Dina Maulina, M.Kom.**

**NIK. 190302xx**

**PENGESAHAN**  
**SKRIPSI**  
**IMPLEMENTASI SUPPORT VECTOR MACHINE DENGAN KERNEL**  
**TRICK UNTUK FILTRASI KOMENTAR**  
**CYBERBULLYING YOUTUBE**

yang dipersiapkan dan disusun oleh  
**Ahmad Farid Mansur**

**16.11.0786**

telah dipertahankan di depan Dewan Penguji  
pada tanggal 19 Maret 2020

**Susunan Dewan Penguji**

**Nama Penguji**

**Tanda Tangan**

**Dina Maulina, M.Kom**  
**NIK. 190302xxx**

**Banu Santoso, S.T., M.Eng**  
**NIK. 190302xxx**

**Haryoko, S.Kom, M.Cs**  
**NIK. 190302xxx**

Skripsi ini telah diterima sebagai salah satu persyaratan  
untuk memperoleh gelar Sarjana Komputer  
Tanggal 2020

**KETUA UNIVERSITAS AMIKOM YOGYAKARTA**

**Prof. Dr. M. Suyanto, M.M.**

**NIK. 190302001**

## 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, .....

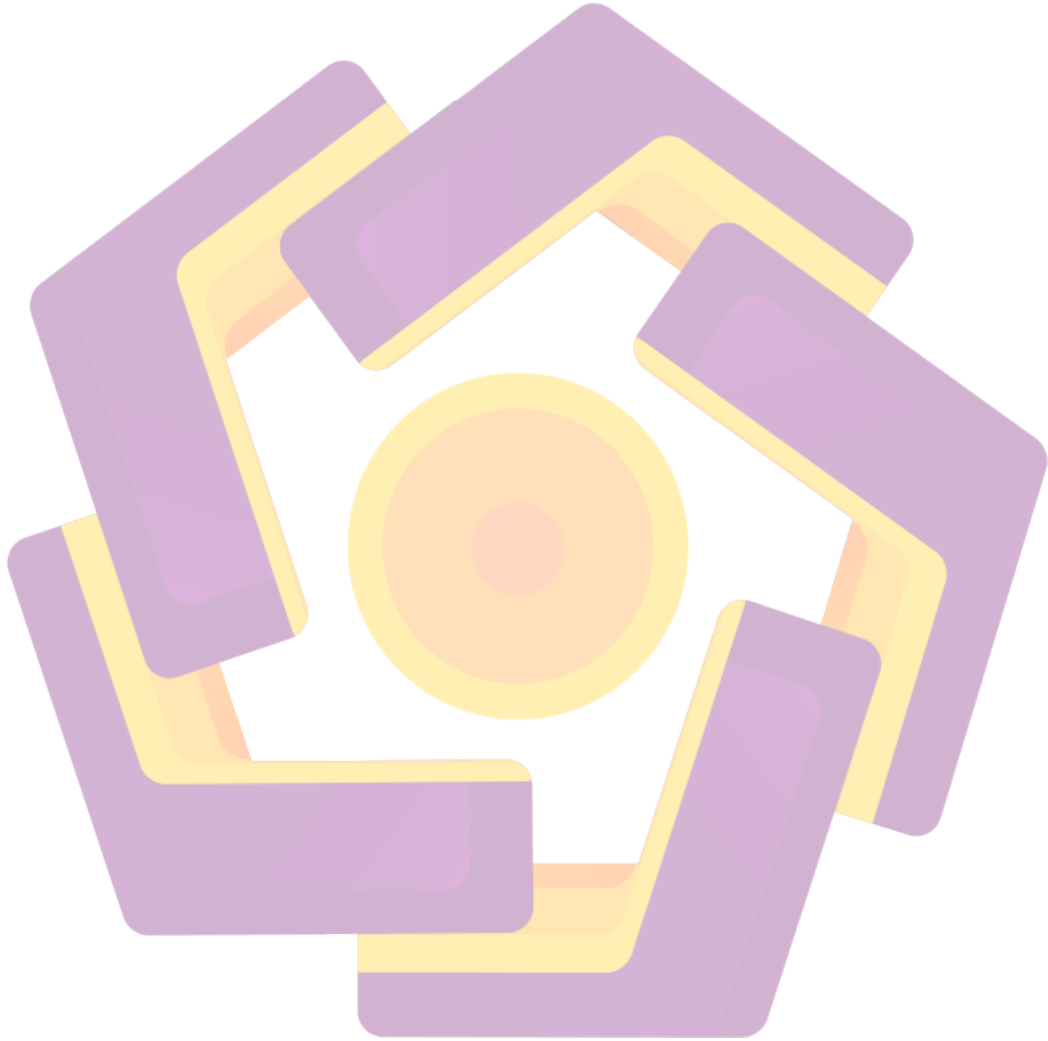
Meterai

Ahmad Farid Mansur

NIM. 16.11.0786

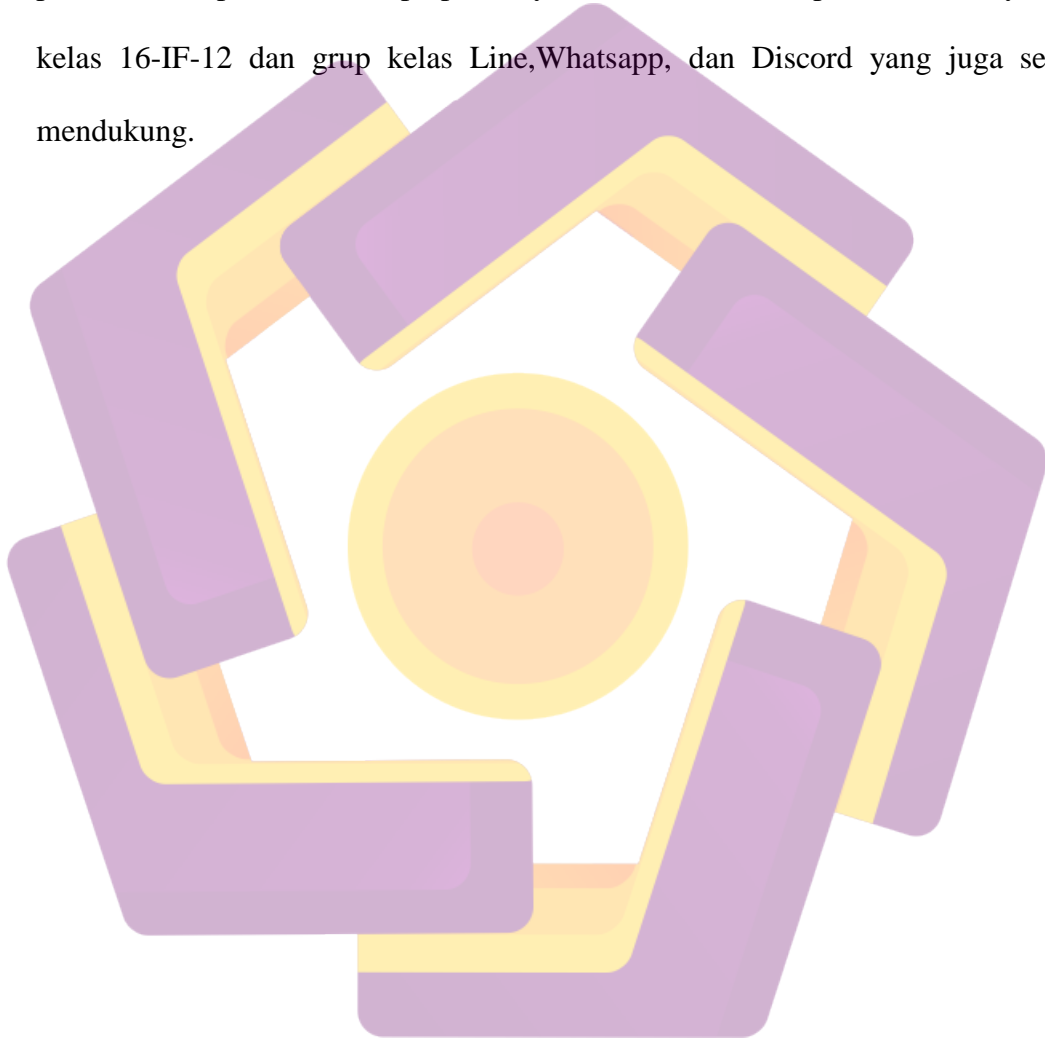
## MOTTO

”NEVER GIVE UP!” – A Japanese man from Shed Sheeran Stream



## PERSEMBAHAN

Skripsi ini penulis persembahkan kepada kedua orang tua dan adik yang selalu msensupport dan memotivasi penulis dalam kehidupan dan juga dalam penulisan skripsi ini. Tak lupa pula saya berterima kasih kepada kawan saya dari kelas 16-IF-12 dan grup kelas Line,Whatsapp, dan Discord yang juga selalu mendukung.



## KATA PENGANTAR

Puji dan syukur penulis panjatkan kepada Allah SWT karena ridho dan hidayahnya skripsi yang sudah disusun dapat terselesaikan sebagai syarat memperoleh gelar Sarjana dari program studi Informatika. Selain itu, Penulis tidak lupa untuk menyadari bahwa skripsi yang telah tersusun merupakan hasil dari bimbingan, motivasi dari diri sendiri dan juga orang lain, serta bantuan dari seluruh pihak yang berkaitan dengan penulis sendiri. Pada Kesempatan ini, penulis menyampaikan rasa terima kasih kepada

1. Kedua orang tua penulis, Mansur Makassar dan Rocsmawati Achmad karena telah memotivasi, memberikan support dan membantu dari segala segi dalam penulisan skripsi ini.
2. Bapak Prof. Dr. M. Suyanto, M.M, selaku rektor UNIVERSITAS AMIKOM YOGYAKARTA
3. Ibu Eny Nurnilawati, S.E, M.M selaku Wakil Rektor II Bidang Perencanaan, Administrasi Umum dan Keuangan
4. Bapak Achmad Fauzi, S.E, M.M selaku Wakil Rektor III Bidang Kemahasiswaan dan Alumni
5. Bapak Arief Setyanto, S.Si, MT, Ph.D selaku Wakil Rektor IV Bidang Kerjasama dan Pengembangan
6. Bapak Sudarmawan, MT selaku Dekan Fakultas Sains dan Teknologi, dan Ketua Program Studi S1 Informatika
7. Ibu Dina Maulina, M.Kom sebagai Dosen Pembimbing yang telah meluangkan waktu memberikan petunjuk dan koreksi dalam penyusunan skripsi ini hingga skripsi ini selesai.



## DAFTAR ISI

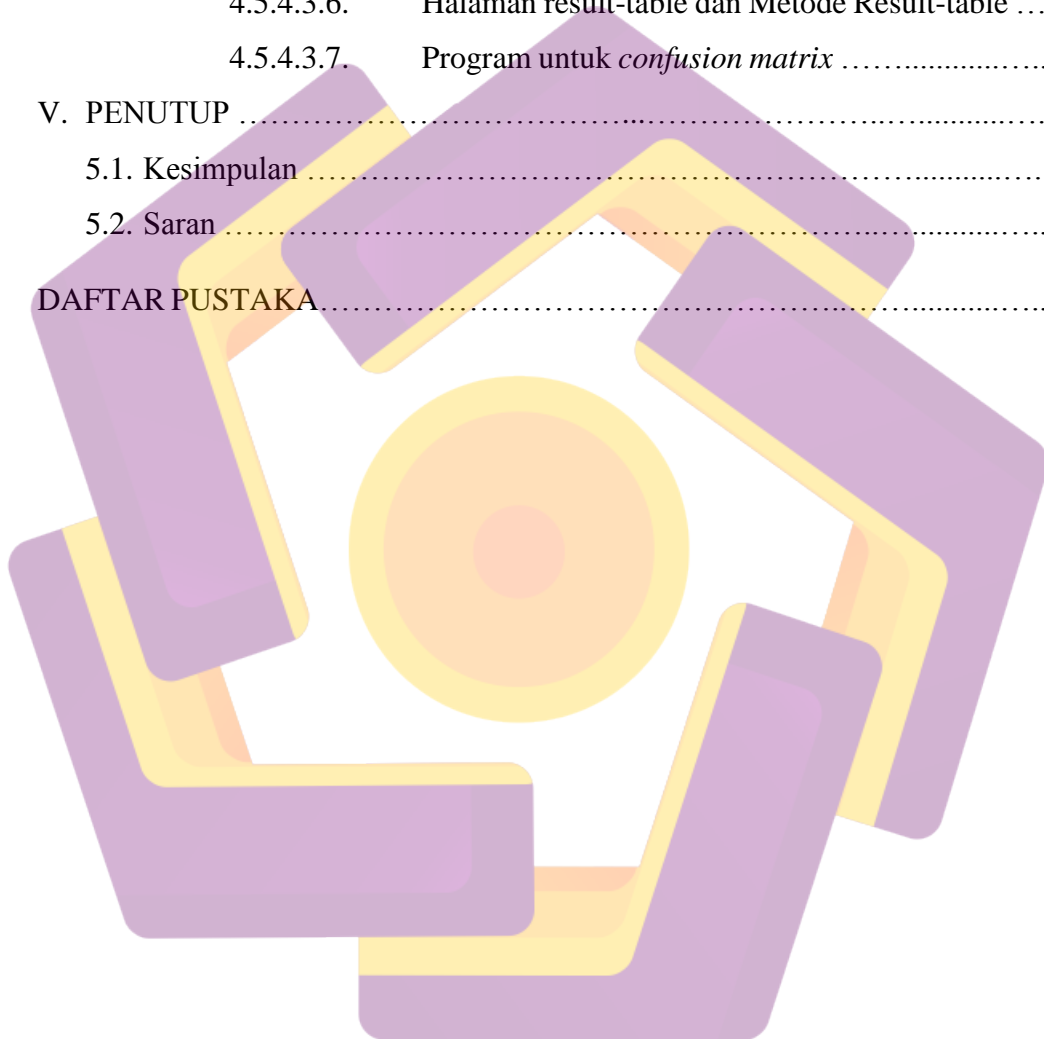
JUDUL .....	i
PERSETUJUAN .....	ii
PENGESAHAN .....	iii
PERNYATAAN .....	iv
MOTTO .....	v
PERSEMBAHAN .....	vi
KATA PENGANTAR .....	vii
DAFTAR ISI .....	viii
DAFTAR TABEL .....	xiii
DAFTAR GAMBAR .....	xv
DAFTAR LAMPIRAN .....	xvii
INTISARI .....	xix
ABSTRACT .....	xx
<b>I. PENDAHULUAN</b> .....	<b>1</b>
I.1. Latar Belakang .....	1
I.2. Rumusan Masalah .....	3
I.3. Batasan Masalah .....	3
I.4. Maksud dan Tujuan Penelitian .....	4
I.5. Manfaat Penelitian .....	4
I.6. Metodologi Penelitian .....	5
I.6.1. Sumber Data .....	5
I.6.2. Tahapan Penelitian .....	6
I.7. Sistematika Penulisan .....	6
<b>II. LANDASAN TEORI</b> .....	<b>9</b>
2.1. Tinjauan Pustaka .....	9
2.2. Cyberbullying .....	14
2.3. Youtube Data API .....	15
2.4. Klasifikasi .....	16
2.5. Support Vector Machine .....	17
2.5.1. SVM Non-Linear/Kernel Trick .....	23

2.6. Preprocessing .....	24
2.6.1. Data Cleaning .....	24
2.6.2. Normalization/Normalisasi .....	24
2.6.3. Case Folding .....	25
2.6.4. Tokenization .....	25
2.6.5. Stemming .....	26
2.7. Term Frequency – Inverse Document Frequency (TF-IDF) .....	26
2.8. Evaluasi .....	27
2.9. Validasi .....	29
2.10. Bahasa Pemrograman .....	29
2.10.1. Python .....	29
2.10.2. Scikit-Learn .....	30
2.11. Metode Waterfall .....	30
2.12. Analisis Kebutuhan Sistem .....	32
2.13. Analisis Kelayakan Sistem .....	33
2.13.1. Kelayakan Teknis .....	33
2.13.2. Kelayakan Operasional .....	33
2.13.3. Kelayakan Ekonomi .....	34
2.14. Flowchart .....	34
2.15. Unified Model Language (UML) .....	35
2.15.1. Use Case Diagram .....	36
2.15.2. Activity Diagram .....	37
2.15.3. Sequence Diagram .....	39
2.15.4. Class Diagram .....	41
2.16. Testing .....	41
2.16.1. Blackbox Testing .....	41
2.16.2. Whitebox Testing .....	42
III. ANALISIS DAN PERANCANGAN .....	43
3.1. Analisis Identifikasi Masalah .....	43
3.2. Analisis Kebutuhan .....	43
3.2.1. Kebutuhan Non-Fungsional .....	43

3.2.2. Kebutuhan Fungsional .....	44
3.3. Analisis Kelayakan Sistem .....	45
3.3.1. Kelayakan Teknik .....	45
3.3.2. Kelayakan Hukum .....	45
3.3.3. Kelayakan Ekonomi .....	45
3.4. Gambaran Umum Proses .....	45
3.5. Analisis Algoritma .....	46
3.5.1. Pengambilan Data .....	46
3.5.1.1. Pengambilan Data Training .....	46
3.5.1.2. Pengambilan Data Testing .....	47
3.5.2. Preprocessing .....	48
3.5.2.1. Data Cleaning .....	48
3.5.2.2. Case Folding .....	49
3.5.2.3. Stopword Removal .....	50
3.5.2.4. Stemming .....	51
3.5.2.5. Normalization .....	52
3.5.2.6. Tokenization .....	54
3.5.3. TF-IDF .....	55
3.5.4. Klasifikasi Support Vector Machine .....	58
3.5.5. Evaluasi .....	71
3.6. Analisis Perancangan Sistem .....	72
3.6.1. Use Case .....	72
3.6.2. Activity Diagram .....	73
3.6.3. Sequence Diagram .....	75
3.6.4. Class Diagram .....	76
3.6.5. User Interface .....	77
IV. IMPLEMENTASI DAN PEMBAHASAN .....	81
4.1. Implementasi Sistem .....	81
4.2. Persiapan Data .....	81
4.2.1. Pengambilan Data .....	81
4.2.1.1. Pengambilan Data Training .....	81

4.2.1.2.	Pengambilan Data Testing .....	83
4.2.2.	Preprocessing Data Training .....	84
4.3.	Implementasi Program .....	85
4.3.1.	Tampilan Program .....	85
4.4.	Uji Coba Program .....	87
4.4.1.	White-box Testing .....	87
4.4.1.1.	Input VideoID .....	88
4.4.1.2.	Tombol Lihat Video .....	89
4.4.1.3.	Tombol Lihat Seluruh Komentar .....	90
4.4.1.4.	Tombol Hasil pada Navbar .....	91
4.4.1.5.	Tombol Tentang pada Navbar .....	92
4.4.2.	Black-box Testing .....	93
4.4.3.	Evaluasi Akurasi Klasifikasi .....	94
4.4.3.1.	Uji Coba dengan SVM Linear .....	95
4.4.3.2.	Uji Coba dengan SVM Polynomial .....	97
4.4.3.3.	Uji coba dengan SVM RBF .....	101
4.4.3.4.	Uji coba dengan SVM Sigmoid .....	103
4.5.	Pembahasan Program .....	105
4.5.1.	Koneksi dengan Youtube Data API .....	105
4.5.2.	Instalasi Program .....	111
4.5.3.	Menjalankan Program .....	112
4.5.4.	Kode Program .....	112
4.5.4.1.	Kelas Youtube_Mining .....	112
4.5.4.2.	Preprocessing .....	114
4.5.4.2.1.	Metode remove_punctuation .....	114
4.5.4.2.2.	Metode stopwords_removal .....	114
4.5.4.2.3.	Metode Stemming_Create .....	114
4.5.4.2.4.	Metode normalization dan normalization_per_word .....	115
4.5.4.2.5.	Kelas classification dan metode svm .....	116
4.5.4.3.	Koding untuk halaman dan Flask .....	117

4.5.4.3.1.	Metode mynavbar .....	117
4.5.4.3.2.	Halaman Home dan Metode Index .....	117
4.5.4.3.3.	Metode formtest .....	118
4.5.4.3.4.	Halaman About dan Metode about .....	118
4.5.4.3.5.	Halaman Result dan Metode Result .....	118
4.5.4.3.6.	Halaman result-table dan Metode Result-table ...	120
4.5.4.3.7.	Program untuk <i>confusion matrix</i> .....	120
V. PENUTUP .....		122
5.1. Kesimpulan .....		122
5.2. Saran .....		123
DAFTAR PUSTAKA .....		124



## DAFTAR TABEL

Tabel 2.1 Penelitian lain serta hasil penelitian .....	10
Tabel 2.2 Jenis-Jenis cyberbullying .....	15
Tabel 2.3 Simbol Flowchart .....	34
Tabel 2.4 Simbol Use Case .....	36
Tabel 2.5 Simbol <i>Activity Diagram</i> .....	37
Tabel 2.6 Simbol <i>Sequence Diagram</i> .....	39
Tabel 3.1 Tabel Contoh komentar yang telah diambil .....	47
Tabel 3.2 Tabel komentar setelah data cleaning .....	49
Tabel 3.3 Tabel Case Folding Komentar .....	49
Tabel 3.4 Stopword Removal Pada Komentar .....	51
Tabel 3.5 Hasil Stemming Komentar .....	51
Tabel 3.6 Normalisasi Komentar Pertama .....	53
Tabel 3.7 Hasil Normalisasi .....	54
Tabel 3.8 Tokenisasi Komentar .....	54
Tabel 3.9 Perhitungan Df Seluruh Kata .....	56
Tabel 3.10 Perhitungan Idft Seluruh Kata .....	57
Tabel 3.11 Perhitungan TF-IDF .....	58
Tabel 3.12 Pelabelan Dokumen/Komentar .....	59
Tabel 3.13 Perhitungan Kernel Linear .....	61
Tabel 3.14 Perhitungan Kernel Polynomial .....	62
Tabel 3.15 Perhitungan Kernel Sigmoid .....	63

Tabel 3.16 Perhitungan Kernel RBF .....	63
Tabel 3.17 Perhitungan Polynomial X8 .....	69
Tabel 3.18 Perhitungan Sigmoid X8 .....	69
Tabel 3.19 Perhitungan RBF X8 .....	70
Tabel 3.20 Confusion Matrix .....	72
Tabel 4.1 Komentar Positif dan Cyberbully .....	83
Tabel 4.2 Komentar setelah Preprocessing .....	84
Tabel 4.3 Uji Coba Menu Home .....	93
Tabel 4.4 Uji Coba Menu Hasil .....	93
Tabel 4.5 Hasil Klasifikasi dengan SVM Linear .....	96
Tabel 4.6 Confusion Matrix SVM Linear .....	97
Tabel 4.7 Hasil Klasifikasi dengan SVM Polynomial .....	98
Tabel 4.8 Confusion Matrix SVM Polynomial .....	101
Tabel 4.9 Akurasi Klasifikasi dengan SVM RBF .....	101
Tabel 4.10 Confusion Matrix SVM RBF .....	103
Tabel 4.11 Akurasi Klasifikasi dengan SVM Sigmoid .....	103
Tabel 4.12 Confusion Matrix SVM Sigmoid .....	105

## DAFTAR GAMBAR

Gambar 2.1 <i>Flowchart</i> Klasifikasi .....	17
Gambar 2.2 Konsep pencarian hyperplane .....	18
Gambar 2.3 Margin dari Hyperplane .....	19
Gambar 2.4 Confusion Matrix .....	28
Gambar 2.5 Alur Metode Waterfall .....	31
Gambar 3.1 <i>Flowchart</i> proses sistem .....	46
Gambar 3.2 <i>Flowchart</i> Pengambilan <i>Data Testing</i> .....	48
Gambar 3.3 <i>Flowchart Preprocessing</i> .....	48
Gambar 3.4 Proses normalisasi kata .....	53
Gambar 3.5 Proses TF-IDF .....	55
Gambar 3.6 Use Case Diagram .....	73
Gambar 3.7 Activity Diagram .....	74
Gambar 3.8 Sequence Diagram .....	75
Gambar 3.9 Class Diagram .....	76
Gambar 3.10 <i>Wireframe</i> Halaman Home .....	77
Gambar 3.11 <i>Wireframe</i> Halaman Result .....	78
Gambar 3.12 <i>Wireframe</i> Halaman Seluruh Komentar .....	79
Gambar 3.13 <i>Wireframe</i> Halaman About .....	80
Gambar 4.1 Tampilan Halaman Home .....	85
Gambar 4.2 Tampilan Halaman Hasil .....	86
Gambar 4.3 Tampilan Halaman Lihat Seluruh Komentar .....	86

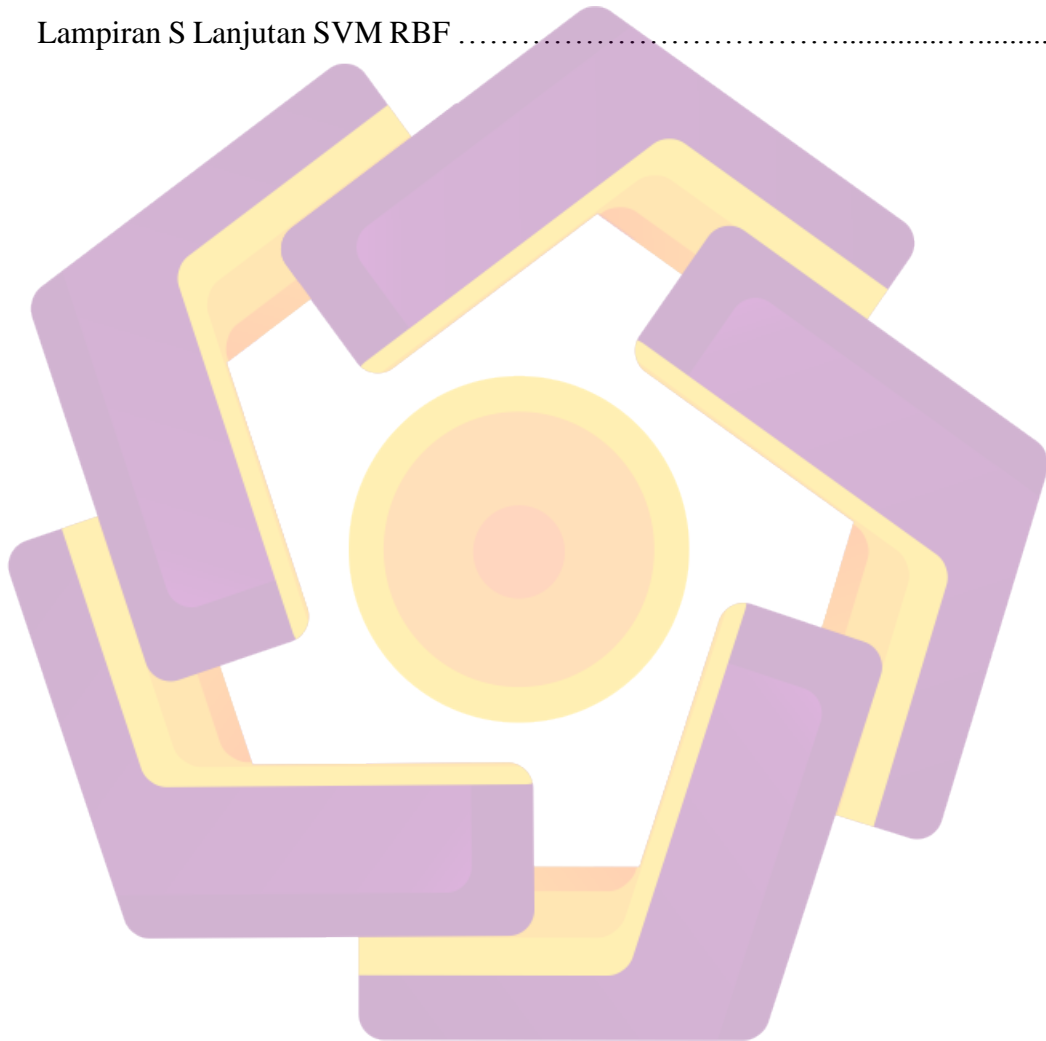


Gambar 4.4 Tampilan Halaman Tentang .....	87
Gambar 4.5 Ilustrasi Input VideoID .....	88
Gambar 4.6 Ilustrasi Tombol Lihat Video .....	89
Gambar 4.7 Ilustrasi Tombol Lihat Seluruh Komentar .....	90
Gambar 4.8 Ilustrasi Tombol Hasil pada Navbar .....	91
Gambar 4.9 Ilustrasi Tombol Tentang pada Navbar .....	92
Gambar 4.10 Hasil Prediksi Kelas .....	95
Gambar 4.11 Hasil Evaluasi SVM Linear .....	97
Gambar 4.12 Hasil Evaluasi SVM Polynomial .....	100
Gambar 4.13 Hasil Evaluasi SVM RBF $C=1$ dan $\gamma=1$ .....	102
Gambar 4.14 Hasil Evaluasi Sigmoid .....	104
Gambar 4.15 Tampilan Developer Console .....	106
Gambar 4.16 Tampilan Select a Project .....	106
Gambar 4.17 Tampilan Form New Project .....	107
Gambar 4.18 Tampilan Dashboard Youtube Data API .....	108
Gambar 4.19 Pencarian Youtube Data API .....	108
Gambar 4.20 Halaman Deskripsi API .....	109
Gambar 4.21 Halaman Dashboard Youtube Data API .....	109
Gambar 4.22 Halaman Add Credentials .....	110
Gambar 4.23 Mengisi Form Credentials .....	110
Gambar 4.24 Developer_Key .....	111
Gambar 4.25 Eksekusi Program .....	112

## DAFTAR LAMPIRAN

Lampiran A Tabel Jaro Winkler Komentar Kedua .....	1
Lampiran B Tabel Jaro Winkler Komentar Ketiga .....	2
Lampiran C Tabel Jaro Winkler Komentar Keempat .....	3
Lampiran D Tabel Jaro Winkler Komentar Kelima .....	4
Lampiran E Tabel Jaro Winkler Komentar Keenam .....	6
Lampiran F Tabel Jaro Winkler Komentar Ketujuh .....	7
Lampiran G Tabel Jaro Winkler Komentar Kedelapan .....	8
Lampiran H <i>Listing</i> Metode Comment_Mining .....	10
Lampiran I <i>Listing</i> Metode Video_Description .....	11
Lampiran J <i>Listing</i> Metode remove_punctuation .....	11
Lampiran K <i>Listing</i> Metode stopword_removal .....	12
Lampiran L <i>Listing</i> Metode Stemming_Create .....	12
Lampiran M <i>Listing</i> Metode Stemming_Create .....	12
Lampiran N <i>Listing</i> Metode Normalization_per_Word .....	13
Lampiran H <i>Listing</i> Kelas Classification .....	15
Lampiran I <i>Listing</i> Navigation/Navbar .....	16
Lampiran J <i>Listing</i> Index .....	16
Lampiran K <i>Listing</i> Formtest .....	16
Lampiran L <i>Listing</i> About .....	17
Lampiran M <i>Listing</i> Result .....	17
Lampiran N <i>Listing</i> Result-Table .....	19

Lampiran O <i>Listing Confusion Matrix</i> .....	20
Lampiran P Lanjutan SVM <i>Linear</i> .....	21
Lampiran Q Lanjutan SVM <i>Polynomial</i> .....	27
Lampiran R Lanjutan SVM <i>Sigmoid</i> .....	29
Lampiran S Lanjutan SVM RBF .....	31



## INTISARI

Youtube merupakan media sosial dengan pengguna terbanyak di Indonesia. Perkembangan media sosial memiliki efek buruk seperti *cyberbullying*. Pada penelitian ini, penulis mengimplementasikan *Support Vector Machine* agar dapat digunakan untuk filtrasi dan memperlihatkan hasil filtrasi serta mengukur kinerja untuk mengetahui fungsi kernel yang terbaik. Fungsi kernel yang digunakan meliputi fungsi *linear*, *polynomial*, *sigmoid*, dan RBF. Selain itu, SVM diimplementasikan kedalam bentuk aplikasi website.

Uji coba dilakukan dengan menggunakan *whitebox testing* dan *blackbox testing* serta memprediksi komentar youtube sesuai dengan kelas (*cyberbullying* atau *non-cyberbullying*) menggunakan empat fungsi kernel. Hasil uji coba yang menunjukkan bahwa skenario SVM kernel RBF  $\gamma = 1$  dan  $C = 1$  menghasilkan akurasi yang terbaik dengan 83,33% tingkat Akurasi yang merupakan angka yang cukup baik. Adapun recall bernilai 80.0%, precision bernilai 85.7%, dan f-measure bernilai 82,7%. Confusion Matrix dari skenario itu menunjukkan bahwa dari 15 komentar *non-cyberbullying*, dua diantaranya diklasifikasi sebagai *cyberbullying*. Sementara itu, dari 15 komentar *cyberbullying*, 3 dianggap sebagai komentar *non-cyberbullying*.

**Kata Kunci** – Klasifikasi, *Cyberbullying*, Youtube, SVM, *Kernel*



## ABSTRACT

Youtube is a social media with the most users in Indonesia. The development of social media has bad effects such as cyberbullying. In this study, the authors implemented the Support Vector Machine so that it can be used for filtration and show the results of filtration and measure performance to find out the best kernel function. The kernel functions used include linear, polynomial, sigmoid, and RBF functions. In addition, SVM is implemented in the form of a website application.

The trial was conducted using whitebox testing and blackbox testing and predicting youtube comments according to the class (cyberbullying or non-cyberbullying) using four kernel functions. The results of trials that show that the RBF kernel SVM scenario  $\gamma = 1$  and  $C = 1$  produce the best accuracy with 83.33% accuracy rate which is a pretty good number. The recall is worth 80.0%, the precision is worth 85.7%, and the f-measure is 82.7%. The Confusion Matrix of the scenario shows that of the 15 non-cyberbullying comments, two of them are classified as cyberbullying. Meanwhile, out of 15 cyberbullying comments, 3 are considered non-cyberbullying comments.

**Keywords** – Classification, Cyberbullying, Youtube, SVM, Kernel

