

**IMPLEMENTASI ALGORITMA SUPPORT VECTOR MACHINE (SVM)
KERNEL RADIAL BASIS FUNCTION (RBF) UNTUK KLASIFIKASI
KOMENTAR SPAM PADA INSTAGRAM**

SKRIPSI



disusun oleh

Nunik Sri Utami

15.11.8596

**PROGRAM SARJANA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2018**



**IMPLEMENTASI ALGORITMA SUPPORT VECTOR MACHINE (SVM)
KERNEL RADIAL BASIS FUNCTION (RBF) UNTUK KLASIFIKASI
KOMENTAR SPAM PADA INSTAGRAM**

SKRIPSI

untuk memenuhi sebagian persyaratan
mencapai gelar Sarjana
pada Program Studi Sistem Informasi



disusun oleh

Nunik Sri Utami

15.11.8596

**PROGRAM SARJANA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2018**

PERSETUJUAN

SKRIPSI

**IMPLEMENTASI ALGORITMA SUPPORT VECTOR MACHINE (SVM)
KERNEL RADIAL BASIS FUNCTION (RBF) UNTUK KLASIFIKASI
KOMENTAR SPAM PADA INSTAGRAM**

yang dipersiapkan dan disusun oleh

Nunik Sri Utami

15.11.8596

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 23 November 2018

Dosen Pembimbing,



Hartatik, S.T., M.Cs.

NIK. 190302232

PENGESAHAN

SKRIPSI

IMPLEMENTASI ALGORITMA SUPPORT VECTOR MACHINE (SVM) KERNEL RADIAL BASIS FUNCTION (RBF) UNTUK KLASIFIKASI KOMENTAR SPAM PADA INSTAGRAM

yang dipersiapkan dan disusun oleh

Nunik Sri Utami

15.11.8596

telah dipertahankan di depan Dewan Penguji

pada tanggal 17 November 2018

Susunan Dewan Penguji

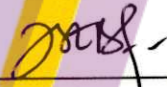
Nama Penguji

Tanda Tangan

Erni Seniwati, M.Cs
NIK. 190302231



Yuli Astuti, M.Kom
NIK. 190302146



Windha Mega PD, M.Kom
NIK. 190302185



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

Tanggal 28 November 2018

DEKAN FAKULTAS ILMU KOMPUTER



Krisnawan, S.Si, MT
NIK. 190302038

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.

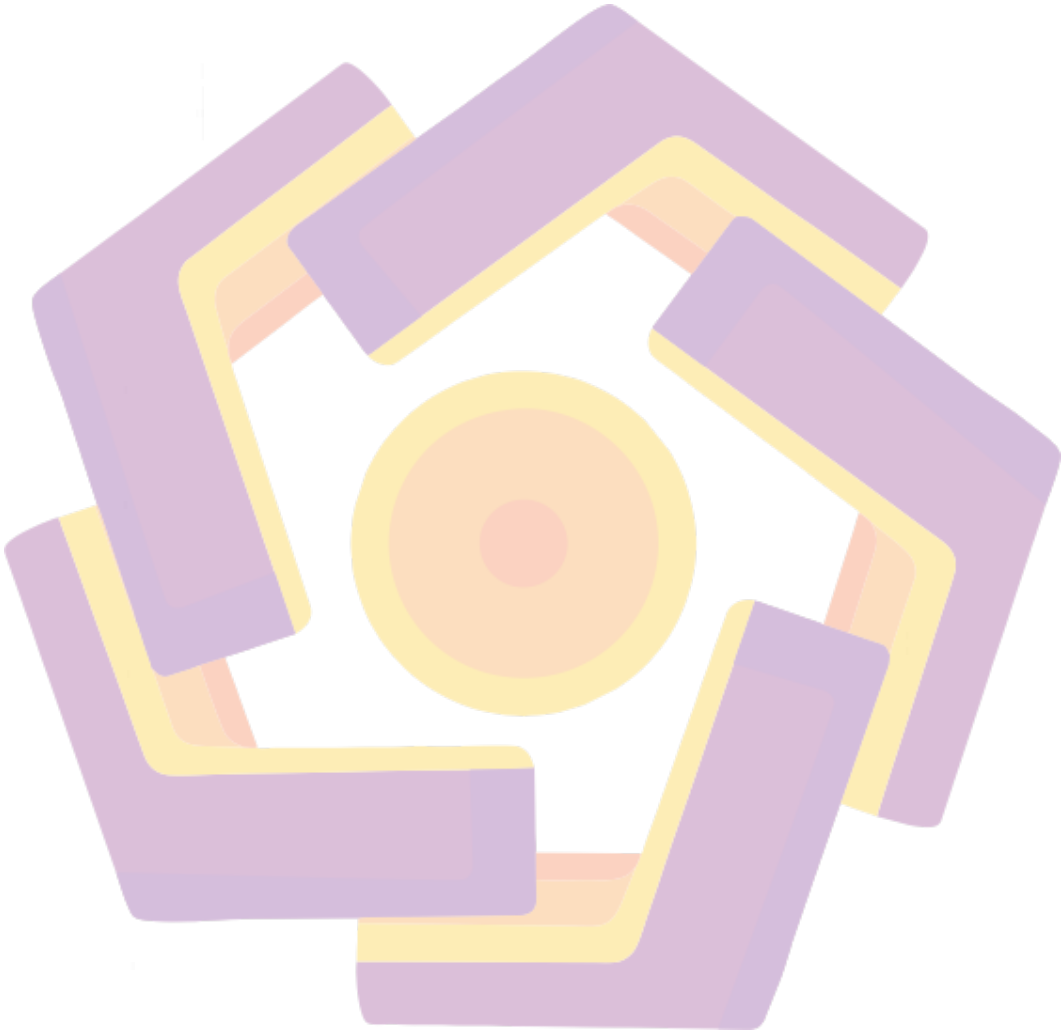
Yogyakarta, 23 November 2018



NUNIK SRI UTAMI
NIM. 15.11.8596

MOTTO

Punggung pisaupun bila diasah akan menjadi tajam



PERSEMBAHAN

Sujud syukur ku persembahkan pada ALLAH yang maha kuasa, berkat dan rahamat detak jantung, denyut nadi, nafas dan putaran roda kehidupan yang diberikan-Nya hingga saat ini saya dapat mempersembahkan skripsi ku pada orang-orang tersayang:

1. Kepada kedua orang tua yang sangat aku sayangi.
2. Kepada Mas Nunung Asrori, Mbak Nuri Wulandari, Mbak Nungki Widiyastuti yang aku sayangi.
3. Kepada Bu Hartatik selaku pembimbing yang sudah memberika Motivasi dan bimbingan selama saya melakukan penelitian.
4. Kepada Teman-teman yang selalu mensupport dan memberikan dukungan moril.
5. Teruntuk *kamu* yang selalu ada untuk ku. Lafyuu.
6. Buat Grup Gombel yang selalu bingung berdiskusi.
7. Terimakasih untuk diriku sendiri yang selalu bisa melawan rasa malas.



KATA PENGANTAR

السَّلَامُ عَلَيْكُمْ وَرَحْمَةُ اللَّهِ وَبَرَكَاتُهُ

Segala puji dan syukur kami ucapkan kehadiran Allah SWT, Tuhan Yang Maha Esa, Maha Pengasih lagi Maha Penyayang. Sholawat serta salam tidak lupa kami ucapkan kepada Nabi Besar, Muhammad SAW. Alhamdulillah rabbil`alamin, peneliti dapat menyelesaikan skripsi ini sebagai syarat untuk menamatkan gelar Sarjana (S-1) di Jurusan Teknik Informatika, Universitas AMIKOM Yogyakarta.

Proses penelitian dan penulisan skripsi ini tentu tidak terlepas dari banyaknya pihak yang memberikan bantuan sehingga skripsi ini dapat terselesaikan. Tidak lupa, peneliti juga mengucapkan banyak terima kasih untuk semua pihak yang terlibat secara langsung maupun tidak dalam proses pengerjaan skripsi ini. Semoga skripsi ini dapat melengkapi studi kajian Ilmu Komputer selanjutnya, khususnya dalam bidang *Artificial Intelligence*. Terakhir, semoga skripsi ini bermanfaat bagi peneliti maupun orang lain di masa depan.

وَسَّلَامٌ عَلَيْكُمْ وَرَحْمَةُ اللَّهِ وَبَرَكَاتُهُ

DAFTAR ISI

HALAMAN JUDUL.....	i
PERSETUJUAN.....	iii
PENGESAHAN.....	iv
PERNYATAAN.....	v
MOTTO.....	vi
PERSEMBAHAN.....	vii
KATA PENGANTAR.....	viii
DAFTAR TABEL.....	xii
DAFTAR GAMBAR.....	xiii
INTISARI.....	xiv
ABSTRACT.....	xv
BAB I Pendahuluan.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	4
1.3 Batasan Masalah.....	4
1.4 Maksud dan Tujuan Penelitian.....	5
1.5 Manfaat Penelitian.....	5
1.6 Metodologi Penelitian.....	5
1.6.1 Metode Pengumpulan Data.....	5
1.6.2 Metode Analisis.....	6
1.6.3 Metode Perancangan.....	6
1.6.4 Implementasi.....	6
1.6.5 Pengujian.....	6
1.7 Sistematika Penulisan.....	7
BAB II LANDASAN TEORI.....	9
2.1 Tinjauan Pustaka.....	9
2.2 Dasar Teori.....	11
2.2.1 Spam.....	11
2.2.2 Komentar Spam.....	12
2.2.3 Instagram dan Instagram API.....	13
2.2.4 Text Mining.....	14

2.2.5	Preprocessing	13
2.2.6	Text Transformation dan TF-IDF Weighting	14
2.2.7	Data Mining	15
2.2.8	Klasifikasi	15
2.2.9	Support Vector Machine (SVM).....	16
2.2.10	<i>Validasi dan Evaluasi</i>	24
2.2.11	Optimisasi Hyperplane.....	26
BAB III ANALISIS DAN PERANCANGAN		27
3.1	Analisis Sistem	27
3.1.1	Analisis Kebutuhan Fungsional	27
3.1.2	Analisis Kebutuhan Non-Fungsional	27
3.2	Gambaran Umum Sistem	28
3.3	Analisis Algoritma	28
3.3.1	Pengumpulan Data	28
3.3.2	Pelabelan Komentar	30
3.3.3	Pre-processing	32
3.3.4	Pembobotan TF-IDF	38
3.3.5	Support Vector Machine (SVM).....	45
3.4	Analisis Kebutuhan Data.....	50
3.4.1	Perancangan Dataset	50
3.4.2	Perancangan File Pengumpulan Komentar	51
3.5	Perancangan Antarmuka.....	52
3.5.1	Perancangan Antarmuka Halaman Stemming	52
3.5.2	Perancangan Antarmuka Halaman Tokenizing.....	53
3.5.3	Perancangan Antarmuka Halaman Text Mining.....	53
3.5.4	Perancangan Antarmuka Halaman My Result	54
BAB IV IMPLEMENTASI DAN PEMBAHASAN		55
4.1	Deskripsi Implementasi	55
4.2	Implementasi Pengumpulan Data Komentar.....	55
4.3	Implementasi Pelabelan Data	57
4.4	Implementasi Preprocessing.....	58
4.4.1	Komentar Cleansing.....	58

4.4.2	Case Folding	59
4.4.3	Tokenizing.....	60
4.5	Vektor Komentar	61
4.5.1	Implementasi TF-IDF.....	61
4.5.2	Implementasi Pelatihan Support Vector Machine.....	63
4.6	Evaluasi dan Validasi	65
4.6.1	Evaluasi Data Testing Terhadap Data Training.....	65
4.6.2	Pengujian Skenario.....	68
4.7	Implementasi Antarmuka	70
4.7.1	Implementasi Antarmuka Halaman Stemming.....	70
4.7.2	Implementasi Antarmuka Halaman Tokenizing	71
4.7.3	Implementasi Antarmuka Halaman Text Mining	72
4.7.4	Implementasi Antarmuka Halaman TF-IDF	73
4.7.5	Implementasi Antarmuka Halaman My Result.....	74
BAB V KESIMPULAN.....		75
5.1	Kesimpulan.....	75
5.2	Saran	76
DAFTAR PUSTAKA		77

DAFTAR TABEL

Tabel 3. 1 Komentar yang telah diambil.....	29
Tabel 3. 2 Komentar Berlabel.....	30
Tabel 3. 3 Profil Hasil Pelabelan Instagram 10 Artis Indonesia.....	31
Tabel 3. 4 Komentar Cleansing.....	33
Tabel 3. 5 Komentar Case Folding.....	35
Tabel 3. 6 Komentar Tokenizing.....	36
Tabel 3. 7 Dokumen Komentar.....	39
Tabel 3. 8 TF Dokumen Pertama.....	40
Tabel 3. 9 TF Dokumen Kedua.....	40
Tabel 3. 10 DF Dokumen Pertama.....	42
Tabel 3. 11 DF Dokumen Kedua.....	42
Tabel 3. 12 TF-IDF Dokumen Pertama.....	43
Tabel 3. 13 TF-IDF Dokumen Kedua.....	44
Tabel 3. 14 Sampel Data Training.....	46
Tabel 3. 15 Perhitungan Kernel RBF.....	47
Tabel 3. 16 Hasil Confusion Matrix Data Komentar Artis 1.....	48
Tabel 3. 17 Hasil Accuracy, Recall, Precision, F-Measure.....	49
Tabel 4. 1 Hasil Komentar Cleaning.....	59
Tabel 4. 2 Hasil Case Folding.....	59
Tabel 4. 3 Hasil Script Tokenizing.....	60
Tabel 4. 4 Hasil k-Fold Cross Validation.....	65

DAFTAR GAMBAR

Gambar 2. 1 Data non linear	22
Gambar 2. 2 Ilustrasi K-fold cross validation	24
Gambar 3. 1 Proses Preprocessing	32
Gambar 3. 2 Proses TF-IDF	38
Gambar 3. 3 Perancangan File Training	50
Gambar 3. 4 Proses Scrapping Data Instagram	51
Gambar 3. 5 Perancangan File Pengumpulan Komentar	52
Gambar 3. 6 Perancangan Antarmuka Halaman Stemming	52
Gambar 3. 7 Perancangan Antarmuka Halaman Tokenizing	53
Gambar 3. 8 Perancangan Antarmuka Halaman Text Mining	54
Gambar 3. 9 Perancangan Antarmuka Halaman My Result	54
Gambar 4. 1 Script Fungsi Mengambil Foto	56
Gambar 4. 2 Script Mengambil Komentar	57
Gambar 4. 3 Script Komentar Cleansing	58
Gambar 4. 4 Script Case Folding	59
Gambar 4. 5 Script Tokenizing	60
Gambar 4. 6 Script Menghitung TF	62
Gambar 4. 7 Script Menghitung IDF	62
Gambar 4. 8 Script Menghitung TF-IDF	63
Gambar 4. 9 Script Konfigurasi Model	64
Gambar 4. 10 Script Proses Training Data	64
Gambar 4. 11 Script Prediksi Data Testing	64
Gambar 4. 12 Script Confusion Matrix	65
Gambar 4. 13 Halaman Stemming	71
Gambar 4. 14 Halaman Tokenizing	72
Gambar 4. 15 Halaman Text Mining	73
Gambar 4. 16 Tampilan Dialogbox	73
Gambar 4. 17 Halaman TF-IDF	74
Gambar 4. 18 Halaman My Result	74

INTISARI

Instagram paling terkenal gambar dan video media berbagi didasarkan pada *web* & aplikasi *mobile*. Instagram pengguna dapat memiliki posting gambar yang dapat berkomentar oleh pengikut mereka. Tokoh masyarakat Indonesia seperti aktor, aktris, musisi menggunakan Instagram untuk mempromosikan kegiatan mereka kepada pengikut mereka. Sayangnya, ada banyak komentar spam di Instagram yang memerlukan perhatian khusus dan harus dihapus.

Penelitian ini memperoleh data komentar untuk dijadikan dataset yaitu dari tokoh-tokoh artis/aktor Indonesia yang memiliki lebih dari satu juta pengikut. Dengan menggunakan *preprocessing* (*cleansing data*, *tokenization*, *casefolding*, dan *stopword removal*), dan *TF-IDF* untuk pembobotan.

Penelitian ini akan membuat suatu *system* dalam mengklasifikasikan Spam dengan menggunakan *text mining* menggunakan metode *Support Vector Machine* (SVM), yang nantinya diharapkan dapat memilah komentar spam pada Instagram agar informasi yang diperoleh lebih akurat dan terpercaya. Dengan menggunakan algoritma *Support Vector Machine* (SVM) akan menghasilkan sebuah akurasi yang optimal dalam pengklasifikasian komentar spam pada Instagram.

Kata-kunci : *Support Vector Machine*, *Rbf*, *Classification*, dan *Text mining*.

ABSTRACT

Instagram most famous image and video media sharing based on web & mobile applications. Instagram users can have post a picture that can be commented by their followers. Indonesia public figures like actors, actresses, musicians using Instagram to promote their activities to their followers. Unfortunately, there are a lot of spam comments on Instagram which require special attention and should be removed.

This study made the comments to get data dataset i.e. the figures of the artist/actor Indonesia which has over one million followers. Using preprocessing (data cleansing, tokenization, casefolding, and stopword removal), and TF-IDF for weighting.

This research will build a system in classifying Spam using text mining method using Support Vector Machine (SVM), which was later expected to sort out the spam comments on Instagram so that the information obtained was more accurate and reliable. By using the algorithm of Support Vector Machine (SVM) will produce an optimal classification accuracy of spam comments on Instagram.

Keyword : *Support Vector Machine, Rbf, Classification, and Text mining.*