

**ANALISIS KARAKTER ORANG DENGAN ALGORITMA SVM PADA LINI
MASA AKUN TWITTERNYA**

SKRIPSI



disusun oleh

Aji Kamaludin

16.11.0051

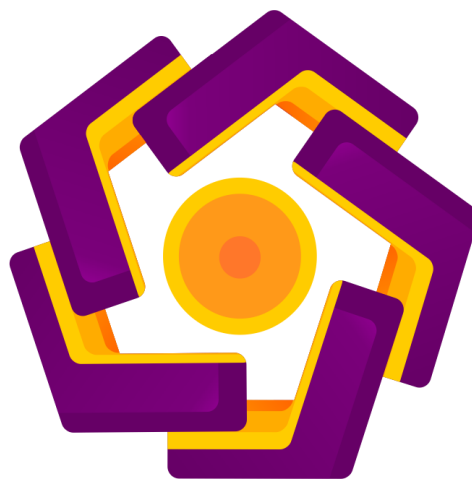
**PROGRAM SARJANA
PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2019

**ANALISIS KARAKTER ORANG DENGAN ALGORITMA SVM PADA LINI
MASA AKUN TWITTERNYA**

SKRIPSI

untuk memenuhi sebagian persyaratan
mencapai gelar Sarjana
pada Program Studi Teknik Informatika



disusun oleh

Aji Kamaludin

16.11.0051

**PROGRAM SARJANA
PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2019**

PERSETUJUAN

SKRIPSI

ANALISIS KARAKTER ORANG DENGAN ALGORITMA SVM PADA LINIMASA AKUN TWITTERNYA

yang dipersiapkan dan disusun oleh

Aji Kamaludin

16.11.0051

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 22 Oktober 2019

Dosen Pembimbing,



Anggit Dwi Hartanto, M.Kom.

NIK. 190302163

PENGESAHAN

SKRIPSI

ANALISIS KARAKTER ORANG DENGAN ALGORITMA SVM PADA LINIMASA AKUN TWITTERNYA

yang dipersiapkan dan disusun oleh

Aji Kamaludin

16.11.0051

telah dipertahankan di depan Dewan Penguji
pada tanggal 18 Oktober 2019

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Erni Seniwati, S.Kom, M.Cs.
NIK. 190302231



Anggit Dwi Hartanto, M.Kom
NIK. 190302163



Rizqi Sukma Kharisma, M.Kom
NIK. 190302215



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 25 Oktober 2019



DEKAN FAKULTAS ILMU KOMPUTER

Krisnawati, S.Si, M.T.
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 tanggung jawab saya pribadi.

Yogyakarta, 30 September 2019



AJI KAMALUDIN

NIM 16.11.0051

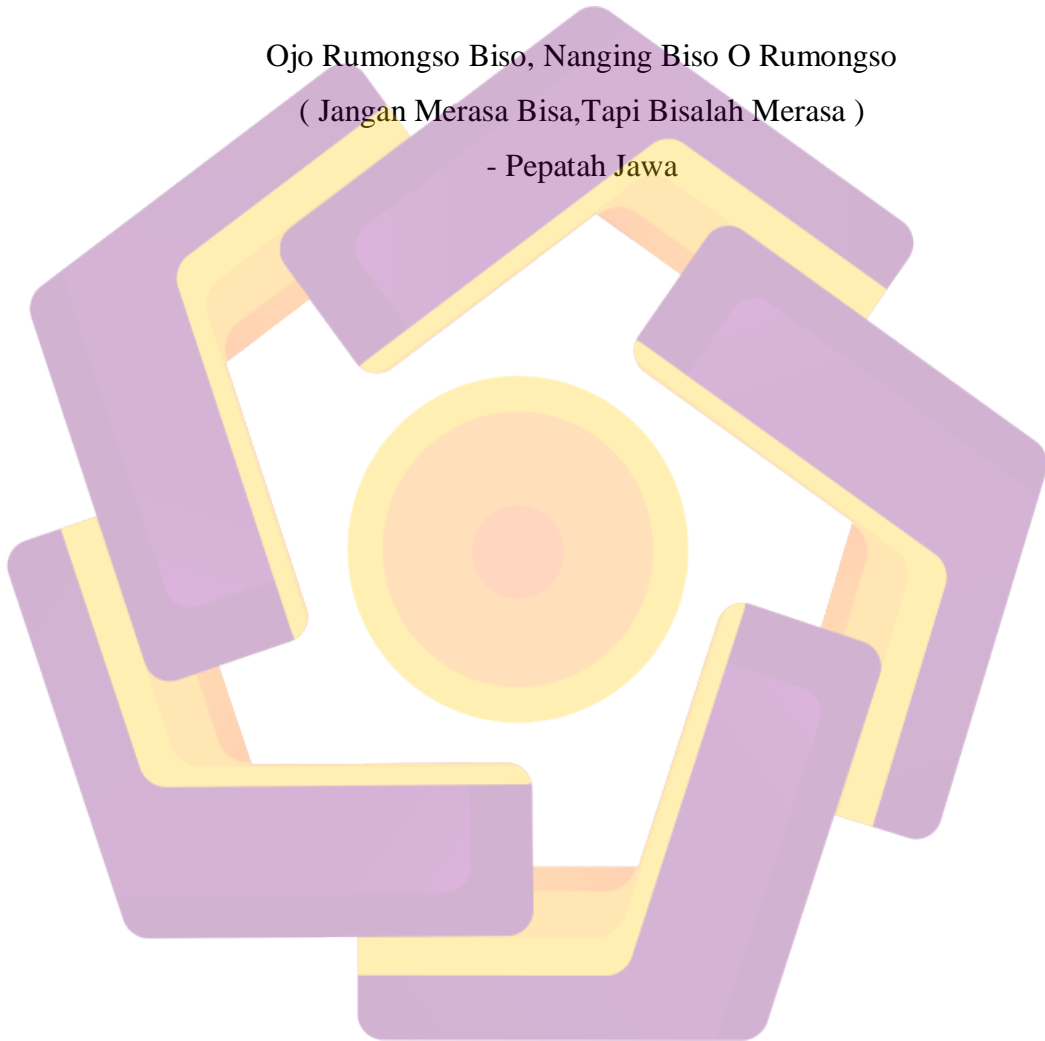
MOTTO

Jangan Paksa-kan Kehendak Pada Orang Lain Berikan Pilihan Dan Peluang Pada Mereka

- Suro Dhemit

Ojo Rumongso Biso, Nanging Biso O Rumongso
(Jangan Merasa Bisa, Tapi Bisalah Merasa)

- Pepatah Jawa



KATA PENGANTAR

Puji syukur kepada ALLAH tuhan Yang Maha Esa, penulis ucapkan karena penelitian dengan judul “ Analisis Karakter Orang dengan Algoritma SVM pada Lini Masa Akun Twitternya ” telah dapat penulis selesaikan. Penelitian ini disusun untuk memenuhi syarat guna memperoleh gelar Sarjana Komputer pada Fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta. Topik ini diambil karena belum pernah ada sebelumnya dan merupakan hal yang menarik bagi peneliti karena banyak sekali orang melakukan cuitan di twitter setiap harinya merupakan hasil dari buah pikiran dan watak yang ada dalam diri mereka.

Selesainya skripsi ini penulis akui tidak terlepas dari bantuan beberapa pihak. Untuk itu penulis mengucapkan terima kasih Kepada :

1. Kedua Orang Tua Penulis.
2. Kepada Bapak dan Ibu Pembimbing, Bapak Anggit Dwi Hartanto, M.Kom, Ibu Sumarni Adi, S.Kom, M.Cs, Prof. Dr. Ema Utami, S.Si., M.Kom.
3. Kepada Sahabat dan teman-teman yang selalu membantu dan memberikan dukungan untuk menyelesaikan penelitian ini yang tidak dapat penulis tuliskan disini karena terlalu banyak.
4. Untuk teman yang hebat dan pernah menyesal untuk mengecewakanya, Vreseliana Ayuningtyas.

Yogyakarta, 30 September 2019

Aji Kamaludin

DAFTAR ISI

HALAMAN JUDUL	i
PERSETUJUAN	ii
PENGESAHAN	iii
PERNYATAAN	iv
MOTTO	v
KATA PENGANTAR	vi
DAFTAR ISI	vii
DAFTAR TABEL	ix
DAFTAR GAMBAR	x
DAFTAR LAMPIRAN	xii
INTISARI	xiii
ABSTRACT	xiv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah	3
1.4 Maksud dan Tujuan Penelitian	4
1.5 Manfaat Penelitian	4
1.6 Metodologi Penelitian	5
1.7 Sistematika Penulisan	6
BAB II LANDASAN TEORI	8
2.1 Tinjauan Pustaka	8

2.2 Dasar Teori	11
BAB III ANALISIS DAN PERANCANGAN	24
3.1 Analisis Sistem	24
3.2 Gambaran Umum	25
3.3 Analisis Algoritma	26
3.4 Analisis Kebutuhan Data	43
3.5 Perancangan Antarmuka	44
BAB IV IMPLEMENTASI DAN PEMBAHASAN	53
4.1 Deskripsi Implementasi	53
4.2 Implementasi Pengumpulan Data Cuitan Akun	53
4.3 Implementasi Pelabelan Data	56
4.4 Implementasi Preprocessing	57
4.5 Implementasi Pembobotan TF-IDF	58
4.6 Implementasi Support Vector Machine	59
4.7 Evaluasi dan Validasi	60
4.8 Implementasi Antarmuka	62
4.9 Pengujian Sistem	67
BAB V PENUTUP	69
5.1 Kesimpulan	69
5.2 Saran	70
DAFTAR PUSTAKA	72
LAMPIRAN	74

DAFTAR TABEL

Tabel 2.1 Perbandingan Pustaka	9
Tabel 2.2 Macam fungsi kernel linear	19
Tabel 2.3 Macam fungsi kernel non-linear	21
Tabel 2.4 <i>Confusion Matrix</i>	22
Tabel 3.1 Cuitan HvenSunshine14	26
Tabel 3.2 Dataset Kata	27
Tabel 3.3 Dataset Akun	28
Table 3.4 Contoh <i>Case Folding</i>	29
Table 3.5 Contoh Tweet <i>Cleansing</i>	31
Table 3.6 Contoh <i>Tokenizing</i>	33
Tabel 3.7 Term Kalimat pada dokumen pertama	35
Tabel 3.8 Hasil Pembobotan TF-IDF	35
Tabel 3.9 Contoh Data	37
Tabel 3.10 Hasil <i>Confusion Matrix</i> Seluruh Data	41
Tabel 3.11 Hasil <i>Accuracy, Recall, Recision</i>	42
Tabel 4.1 Hasil Pengujian	67

DAFTAR GAMBAR

Gambar 2.1 Persamaan TF-IDF	14
Gambar 2.2 Data non linear	20
Gambar 3.1 Alur proses sistem	26
Gambar 3.2 Flowchart <i>Preprocessing</i>	28
Gambar 3.3 Flowchart <i>Preprocessing Case Folding</i>	29
Gambar 3.4 Flowchart <i>Preprocessing Tweet Cleansing</i>	31
Gambar 3.5 Flowchart <i>Preprocessing Tokenizing</i>	33
Gambar 3.6 Flowchart <i>Preprocessing Pembobotan TF-IDF</i>	34
Gambar 3.7 Plot Contoh Data	38
Gambar 3.8 <i>Hyperplane</i>	40
Gambar 3.9 Dataset Akun, Cuitan dan Label	43
Gambar 3.10 Data Cuitan dalam satu Akun	44
Gambar 3.11 Flowchart Pengambilan Cuitan	45
Gambar 3.12 Perancangan menu sistem	46
Gambar 3.13 Rancangan Halaman <i>Scraping</i>	47
Gambar 3.14 Rancangan Halaman detail pengguna	48
Gambar 3.15 Rancangan Halaman <i>Preprocessing</i>	49
Gambar 3.16 Rancangan Halaman detail pengguna	50
Gambar 3.17 Rancangan halaman klasifikasi	51
Gambar 3.18 Rancangan halaman hasil klasifikasi	52
Gambar 4.1 kode program scrap tweet 1	54

Gambar 4.2 kode program scrap tweet 2	54
Gambar 4.3 kode program scrap tweet 3	55
Gambar 4.4 Proses scrap cuitan linimasa akun	56
Gambar 4.5 Pemilihan Akun pada <i>preprocessing</i>	57
Gambar 4.6 Output <i>Preprocessing</i>	58
Gambar 4.7 Implementasi TF-IDF	58
Gambar 4.8 Pemilihan Akun pada klasifikasi SVM	59
Gambar 4.9 Hasil Klasifikasi SVM	60
Gambar 4.10 Hasil Evaluasi Algoritma dengan <i>Confusion Matrix</i>	61
Gambar 4.11 Halaman <i>Scraping</i>	62
Gambar 4.12 Halaman detail pengguna	63
Gambar 4.13 Halaman <i>preprocessing</i>	64
Gambar 4.14 Halaman detail pengguna	64
Gambar 4.15 Halaman klasifikasi	65
Gambar 4.16 Halaman Hasil Klasifikasi	66

DAFTAR LAMPIRAN

Lampiran 1 Dataset Latih	76
Lampiran 2 Dataset Uji	79



INTISARI

Twitter sebagai salah satu media sosial terbesar dengan jumlah pengguna pada 326 juta pengguna aktif pada Januari 2019 khususnya di Indonesia [1], setiap hari lebih dari jutaan cuitan dipublikasi oleh penggunanya, dimana mereka mencurahkan segala unek - unek dengan menggunakan bahasa sehari - hari.

Maka dari itu penelitian ini mencoba menganalisis Tweet untuk mendapatkan karakter dari tiap-tiap akun tersebut dengan pendekatan karakter DISC. Algoritma klasifikasi yang akan digunakan dalam penelitian ini adalah Support Vector Machine dengan pembobotan TF-IDF pada data yang dimana didalamnya terdapat tahap Preprocessing seperti Cleansing, dan Case Folding, selain itu aplikasi ini juga dikembangkan menggunakan PHP dan Python.

Sedangkan hasil dari aplikasi ini sendiri menggunakan 109 akun sebagai dataset yang telah divalidasi oleh psikolog dan didapatkan tingkat akurasi sekitar 36,3636%. Dengan demikian kita dapat mengetahui karakter dari tiap-tiap akun yang diklasifikasikan.

Kata Kunci : Analisis Karakter, SVM - Support Vector Machine, Twitter, TF-IDF, DISC

ABSTRACT

Twitter as one of the largest social media with 326 million active users in January 2019 especially in Indonesia [1], every day more than millions of tweets are published by users, where they pour out all their gossip using everyday language.

Therefore this study tries to analyze Tweets to get the characters from each of these accounts with the DISC character approach. The classification algorithm that will be used in this study is Support Vector Machine with TF-IDF weighting on the data where there are Preprocessing stages such as Cleansing, and Case Folding, besides this application is also developed using PHP and Python.

While the results of this application alone use 109 accounts as a dataset that has been validated by psychologists and obtained an accuracy rate of around 36.3636%. Thus we can find out the character of each classified account.

Keywords: *Character Analysis, SVM - Support Vector Machine, Twitter, TF-IDF, DISC*