IMPLEMENTASI ALGORITMA *SUPPORT VECTOR MACHINE (SVM)* UNTUK KLASIFIKASI KONTEN BERITA BAHASA INDONESIA

SKRIPSI

disusun oleh

Zendi Apri Jeki

15.11.9300

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

2020

IMPLEMENTASI ALGORITMA *SUPPORT VECTORMACHINE (SVM)* UNTUK KLASIFIKASI KONTEN BERITA BAHASA INDONESIA

SKRIPSI

untuk memenuhi sebagian persyaratan
mencapai gelar Sarjana
pada Program Studi Informatika

disusun oleh

Zendi Apri Jeki

15.11.9300

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

PERSETUJUAN

SKRIPSI

IMPLEMENTASI ALGORITMA *SUPPORT VECTOR MACHINE (SVM)* UNTUK KLASIFIKASI KONTEN BERITA BAHASA INDONESIA

yang dipersiapkan dan disusun oleh

Zendi Apri Jeki

15.11.9300

telah disetujuai oleh Dosen Pembimbing Skripsi

pada tanggal 3 Mei 2019

Dosen Pembimbing,

Mardhiya Hayaty, S.T., M.Kom.

NIK. 190302108

PENGESAHAN

SKRIPSI

IMPLEMENTASI ALGORITMA *SUPPORT VECTOR MACHINE (SVM)* UNTUK KLASIFIKASI KONTEN BERITA BAHASA INDONESIA

yang dipersiapkan dan disusun oleh

Zendi Apri Jeki

15.11.9300

telah dipertahankan di depan Dewan Penguji
pada tanggal 19 Februari 2020

Susunan Dewan Penguji

Nama Penguji	Tanda Tangan
Sri Ngudi Wahyuni, S.T., M.Kom. **NIK. 190302060**	
Wiwi Widayani, M.Kom. **NIK. 190302272**	
Mardhiya Hayaty, S.T., M.Kom. **NIK. 190302108**	

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 19 Februari 2020

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 tanggungjawab saya pribadi.

Yogyakarta, 19 Februari 2020

Zendi Apri Jeki

NIM. 15.11.9300

MOTTO

Khoirunnas anfa'uhum linnas

“Sebaik-baik manusia adalah yang paling bermanfaat bagi manusia”.

(HR. Ahmad, ath-Thabrani, ad-Daruqutni)

Faidza farogh tafanshob

“Maka apabila kamu telah selesai (dari sesuatu urusan), kerjakanlah dengan sungguh-sungguh (urusan) yang lain”.

(Al-Insyirah Ayat 7)

Yarfa'illāhullażīna āmanụ mingkum wallażīna ụtul-'ilma darajāt

“Allah akan meninggikan orang-orang yang beriman di antaramu dan orang-orang yang diberi ilmu pengetahuan beberapa derajat”.

(Al-Mujadilah Ayat 11)

PERSEMBAHAN

Skripsi berjudul "Implementasi Algoritma *Support Vector Machine (Svm)* Untuk Klasifikasi Konten Berita Bahasa Indonesia" ini dipersembahkan kepada:

1. Allah SWT yang telah memberikan nikmat luar biasa kepada saya.
2. Kedua orang tua saya yang telah memberikan segala yang terbaik kepada anak-anaknya.
3. Universitas AMIKOM Yogyakarta sebagai perguruan tinggi tempat saya bernaung.
4. Ibu Mardhiya Hayaty, S.T., M.Kom. yang telah membimbing saya dalam penyusunan skripsi ini dengan penuh kesabaran.
5. Teman-teman IF-12 angkatan 2015 yang telah menemani saya selama perkuliahan.
6. Para sahabat saya yang tidak bisa saya sebutkan satu per satu.

KATA PENGANTAR

Segala puji bagi Allah SWT tuhan semesta alam yang telah memberikan kenikmatan luar biasa yang tak ada henti-hentinya sehingga skripsi ini dapat terselesaikan dengan sebaik-baiknya. Shalawat serta salam semoga selalu tercurah kepada nabi agung Nabi Muhammad SAW beserta keluarga, para sahabat dan umatnya hingga akhir zaman.

Dengan terselesaikannya skripsi ini yang berjudul "Implementasi Algoritma *Support Vector Machine (SVM)* Untuk Klasifikasi Konten Berita Bahasa Indonesia" penulis mengucapkan terima kasih yang sebanyak-banyaknya kepada:

1. Kedua orang tua penulis bernama Zainal dan Juliana, kakak penulis bernama Zofi Rizki, serta kedua adik penulis bernama Zaldi Yanto dan Ani Yunita yang telah mendukung penuh seluruh kegiatan penulis, sehingga mustahil penulis untuk menghitung kebaikannya.
2. Pasuk Nazirin, S.ST dukungan dari beliau tidak mampu penulis hitung.
3. Bapak Prof. Dr. M. Suyanto selaku Rektor Universitas AMIKOM Yogyakarta.
4. Bapak Sudarmawan, M.Kom. selaku Dekan Fakultas Sains dan Teknologi, dan Ketua Program Studi S1 Informatika.
5. Ibu Mardhiya Hayaty, S.T., M.Kom. selaku pembimbing skripsi ini yang selalu sabar dalam memberikan arahan-arahan serta kesempatan waktu bimbingan yang banyak.

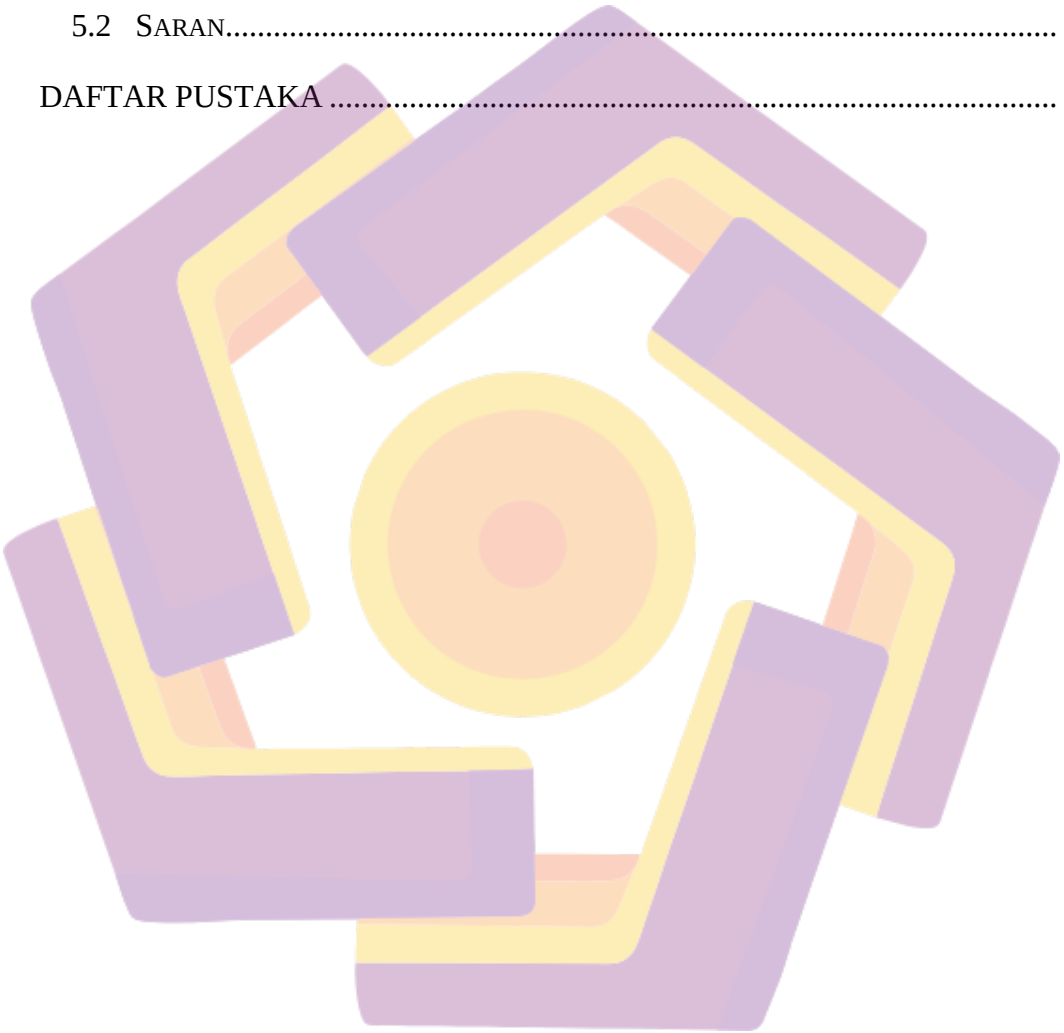
6. Seluruh dosen dan staff Universitas AMIKOM Yogyakarta yang pernah berkontribusi terhadap segala aktivitas penulis selama menjalani perkuliahan.
7. Teman-teman IF-12 angkatan 2015 yang telah menemani penulis selama perkuliahan.
8. Akhi Fail, Cakra Amiyantoro, Aditya Wiha, dan Subraga Islammada yang telah berkontribusi kepada penulis dalam menyelesaikan skripsi ini.
9. Mas Hanif dan Alfi Maghfiroh telah memberikan saran serta masukan.
10. Mas Arif, Ibed, dan Jihad teman seperjuangan mengerjakan skripsi.
11. Teman-teman Sholih di PPM Nur Baiturrahman dan KAMMI Amikom.

Penulis menyadari skripsi ini masih banyak kekurangan, namun penulis berharap skripsi ini dapat memberikan manfaat bagi para pembacanya. Dan semoga Allah SWT membalas kebaikan semua orang yang telah memberikan dukungan dalam bentuk apapun kepada penulis.

Yogyakarta, 19 Februari 2020

Zendi Apri Jeki

DAFTAR ISI



DAFTAR TABEL

DAFTAR GAMBAR

INTISARI

Kebutuhan analisis *text mining* sangat diperlukan dalam menangani teks berjumlah besar dan tidak terstruktur (Big data). Salah satu kegiatan penting dalam *text mining* adalah klasifikasi atau kategorisasi teks. Analisis *text mining* dilakukan agar mempermudah kita dalam mengambil informasi atau mengolah informasi yang begitu banyak dari dunia internet atau digital, salah satunya dengan melakukan klasifikasi. Kategorisasi teks memiliki berbagai pendekatan antara lain pendekatan *probabilistic, support vector machine, artificial neural network*, atau *decision tree classification*. Dalam pembelajaran statistik. *Support Vector Machine* dipilih karena metode ini memiliki kelebihan dalam bidang klasifikasi dengan bantuan fungsi kernel.

Pada penelitian ini *support vector machine* akan mengelompokkan berita berdasarkan kategori menjadi 3 bagian atau *class* yaitu : politik, kuliner dan sepakbola. Kernel pada *Support Vector Mechine* akan di kombinasikan dengan pembobotan *tf-idf* dan *preprocessing* data.

Dengan trik kernel, metode pembobotan, dan *preprocessing* data diharapkan dapat membantu klasifikasi teks dengan baik serta mampu meningkatkan akurasi. Pengujian dilakukan menggunkan *confusion matrix* klasifikasi dengan metode *support vektor machine* dengan kernel *linear* mendapat rata-rata akurasi 92,7%, kemudian dengan kernal *rbf* mendapat rata-rata akurasi sebesar 92,3%.

Kata Kunci: *tf-idf, text mining, svm, preprocessing*, klasifikasi, konten berita

ABSTRACT

The need for text mining analysis is very much needed in handling large and unstructured text (Big data). One of the important activities in text mining is the classification or categorization of texts. Text mining analysis is carried out to make it easier for us to retrieve information or manage so much information from the internet or digital world, one of them is by doing classification. Text categorization has various approaches including probabilistic approaches, support vector machines, artificial neural networks, or decision tree classification. In learning statistics. Support Vector Machine was chosen because this method has advantages in the field of classification with the help of kernel functions.

In this study support vector machine will group news by category into 3 parts or classes, namely: politics, culinary and football. The kernel in Support Vector Mechine will be combined with tf-idf weighting and data preprocessing.

With kernel tricks, weighting methods, and data preprocessing is expected to help text classification well and be able to improve accuracy. Tests carried out using confusion matrix classification with the support vector machine method with linear kernels get an average accuracy of 92.7%, then with kernal rbf get an average accuracy of 92.3%.

Keyword: *tf-idf, text mining, svm, preprocessing, classification, news content*