PERBANDINGAN KINERJA ALGORITMA SVM DAN C4. 5 DALAM KLASIFIKASI SPAM EMAIL

SKRIPSI NON-REGULER JURNAL

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana

Program Studi Informatika

disusun oleh

RAMADHANI KUSUMA HADI

19.11.3009

Kepada

FAKULTAS ILMU KOMPUTER

UNIVERSITAS AMIKOM YOGYAKARTA

YOGYAKARTA

2024

PERBANDINGAN KINERJA ALGORITMA SVM DAN C4. 5 DALAM KLASIFIKASI SPAM EMAIL

SKRIPSI NON-REGULER JURNAL

untuk memenuhi salah satu syarat mencapai derajat Sarjana

Program Studi Informatika

disusun oleh

RAMADHANI KUSUMA HADI

19.11.3009

Kepada

FAKULTAS ILMU KOMPUTER

UNIVERSITAS AMIKOM YOGYAKARTA

YOGYAKARTA

2024

HALAMAN PERSETUJUAN

SKRIPSI

PERBANDINGAN KINERJA ALGORITMA SVM DAN C4. 5 DALAM KLASIFIKASI SPAM EMAIL

yang disusun dan diajukan oleh

RAMADHANI KUSUMA HADI

19.11.3009

telah disetujui oleh Dosen Pembimbing Skripsi

pada tanggal 27 Juni 2024

Dosen Pembimbing,

Mulia Sulistiyono, M.kom

NIK. 190302248

HALAMAN PENGESAHAN

SKRIPSI

PERBANDINGAN KINERJA ALGORITMA SVM DAN C4. 5 DALAM KLASIFIKASI SPAM EMAIL

yang disusun dan diajukan oleh

RAMADHANI KUSUMA HADI

19.11.3009

Telah dipertahankan di depan Dewan Penguji
pada tanggal 27 Juni 2024

Susunan Dewan Penguji

Nama Penguji	Tanda Tangan
Yoga Prisyanto, S.Kom., M.Eng. **NIK. 190302412**	
Wiwi Widayani, M.Kom **NIK. 190302272**	
Mulia Sulistiyono, M.Kom **NIK. 190302248**	

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 27 Juni 2024

DEKAN FAKULTAS ILMU KOMPUTER

Hanif Al Fatta,S.Kom., M.Kom.
NIK. 190302096

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Ramadhani Kusuma Hadi
NIM : 19.11.3009

Menyatakan bahwa Skripsi dengan judul berikut:

PERBANDINGAN KINERJA ALGORITMA SVM DAN C4. 5 DALAM KLASIFIKASI SPAM EMAIL

Dosen Pembimbing : Mulia Sulistiyono, M.Kom

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 27 juni 2024

Yang Menyatakan,

Ramadhani Kusuma Hadi

HALAMAN PERSEMBAHAN

Alhamdulillah, Segala Puji Bagi ALLAH SWT. Kepada-Nya kita memeohon, memuji, dan meminta perlindungan, pengampunan serta petunjuk Kepada-Nya. Kita berlindung kepada ALLAH SWT dari kejahatan diri kita dan keburukan amal-amal kita. Dengan ini penulis mempersembahkan skripsi ini kepada semua pihak yang telah membantu da memberikan motivasi serta dukungan yang besar sehingga skripsi ini dapat diselesaikan dengan sebaik-baiknya.

Melalui skripsi ini saya persembahkan terimakasih serta syukur kepada:

1. Rasa syukur kepada ALLAH SWT yang telah memberikan rahmat berupa karunia untuk terus mengucap syukur dan sabar dalam menyelesaikan tugas akhir skripsi.
2. Kepada orang tua saya yang saya cintai, bapak Susilo Hariyanto dan Wiwit Sulastri semoga selalu diberikan kesehatan atas doa dan motivasi serta dukungannya dalam membimbing penulis.
3. Saudara penulis Yoga Nourma Hadi Pratama serta keluarga kecilnya dengan Khainunnifa dengan malaikat kecilnya kaivan yang selalu mendukung penulis
4. Bapak Mulia Sulistiyono, M.Kom yang telah menjadi pembimbing penulis dalam menyelesaikan skripsi ini.
5. Bapak/Ibu dosen yang telah memberikan ilmu dan pendidikan selama berkuliah di Amikom.
6. Teman-teman pengurus inti HMIF periode 2021/2022 yang Bersama-sama berjuang dalam mencapai kesuksesan yang penuh suka dan duka. Anggota pengurus dan Kader HMIF 2021/2022 yang menjadi pemberi dukungan serta semangat saat menjadi Pengurus Himpunan.
7. Anggota grub "SKRIPSHIT KELAR MAKSIMAL SMSTER 8" yang telah mensupport dan berkontribusi dalam pengerjaan skripsi ini.
8. 'dIa' sosok periang tapi ada kesedihan yang mendalam dibaliknya yang cukup kuat memberikan dorongan, motivasi, serta menyemangati sehingga dapat melewati semua rintangan yang ada, dan juga semangat terus kmu.

KATA PENGANTAR

Puji syukur kehadirat ALLAH SWT karena berkat rahmat serta hidayah-Nya penulis diberikan kesehatan dan kekuatan serta iman yang kuat sehingga dapat menyelesaikan skripsi yang berjudul "**PERBANDINGAN KINERJA ALGORITMA SVM DAN C4. 5 DALAM KLASIFIKASI SPAM EMAIL**". Oleh karena itu penulis mengucapkan syukur karena dapat menyelesaikan skripsi ini. Skripsi ini diajukan kepada Program Studi S1 Informatika, Fakultas Ilmu Komputer, serta Universitas AMIKOM Yogyakarta

Penulis dalam menyelesaikan skripsi mendapatkan beberapa hambatan dalam berbagai hal, namun banyak pihak yang selalu membantu sehingga skripsi ini dapat diselesaikan dengan baik. Sengan begitu penulis hendak mengungkapkan terima kasih terhadap:

1. Bapak Prof. Dr. M Suyanto, M.M selaku Rektor Universitas AMIKOM Yogyakarta.
2. Bapak Hanif Al Fatta, S.Kom., M.Kom., Ph.D. selaku Dekan Fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta.
3. Ibu Windha Mega Pradya D, M.Kom., selaku Ketua Program Studi S1 Informatika Universitas AMIKOM Yogyakarta.
4. Bapak Mulia Sulistiyono, M.Kom selaku dosen pembimbing yang selalu memberikan saran dan arahan dalam proses penulisan skripsi ini
5. Bapak Susilo Hariyanto dan Ibu Wiwit Sulastri selaku orang tua serta Yoga Nourma Hadi Pratama beserta keluarga kecilnya dengan Khainunnisa Fasiha dengan malaikat kecilnya Kaivan dalam memberikan doa dan motivasi kepada penulis.

Yogyakarta, 27 Juni 2024

Penulis

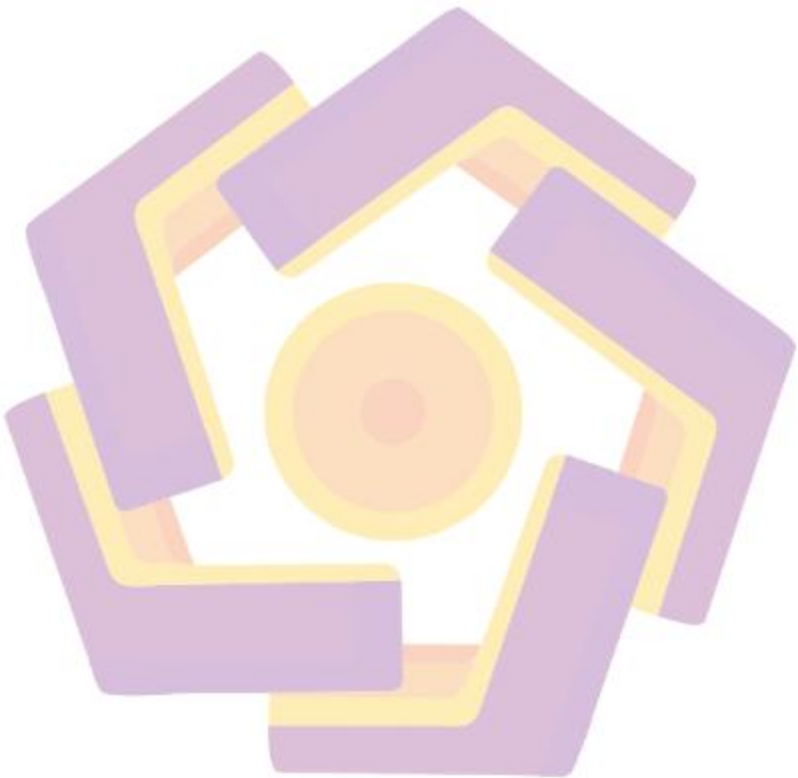
DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

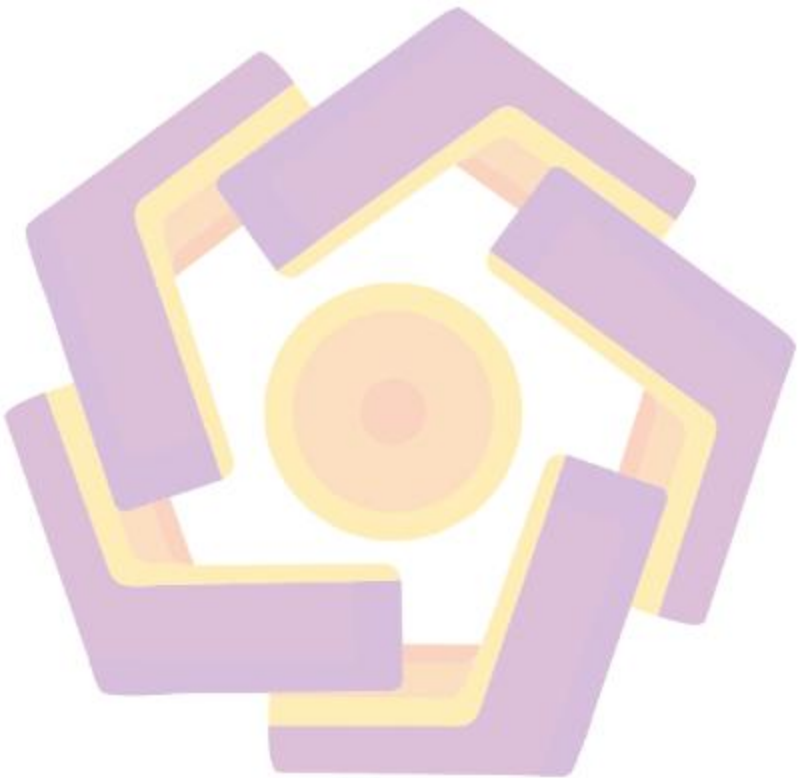
DAFTAR LAMBANG DAN SINGKATAN

SVM	Support Vector Machines
C4.5	Pohon Keputusan C4.5
TF-IDF	Term Frequency-Inverse Document Frequency



DAFTAR ISTILAH

Vektor	besaran yang mempunyai arah
Eigen Value	akar akar persamaan



INTISARI

Email merupakan surat elektronik yang memungkinkan seseorang untuk mengirimkan dan menerima pesan lebih dari satu orang. Email memiliki kelebihan diantara lain cepat, hemat dan dapat mengirimkan pesan dalam berbagai bentuk dokumen yang membuatnya populer di seluruh dunia. Banyaknya email yang dikirim setiap harinya sering disalahgunakan oleh beberapa pihak yang tidak bertanggungjawab untuk mengirim informasi iklan produk jasa dan berbagai informasi yang tidak berguna atau tidak diinginkan oleh pengguna, yang dimana itu disebut *SPAM*.

Untuk menyaring email *SPAM* digunakan banyak algoritma klasifikasi, algoritma tersebut diantaranya adalah *Support Vector Machine* (SVM) serta *Decission Tree* (C4. 5). Pada penelitian ini akan dilakukan perbandingan algoritma SVM dan C4. 5. Untuk melihat kinerja dari permodelan algoritma maka dilakukan pembagian uji *max_features* 1500 dan 3000 data serta menggunakan perbandingan rasio 90% : 10%, 80% : 20%,70% : 30% dan 60% : 40% yang bertujuan menilai data tersebut akan menghasilkan nilai akurasi tertinggi pada pengujian yang dilakukan.

Berdasarkan hasil pengujian yang menggunakan data sebanyak 10.000 baris data dan 2 kolom, menggunakan algoritma *Support Vector Machine* (SVM) serta *Decission Tree* (C4. 5) diperoleh hasil akurasi terbaik menggunakan *Max_features* 3000 dengan rasio 90% : 10% dimana SVM mendapatkan sebesar 98.20% dan C4. 5 sebesar 93.30%. Dari hasil uji coba tersebbut membuktikan bahwa algoritma *Support Vector Machine* (SVM) merupakan metode terbaik dalam melakukan proses klasifikasi *spam email*.

Kata kunci: Klasifikasi, Spam-email, SVM, C4.5.

ABSTRACT

Email is an electronic mail that allows one to send and receive messages from more than one person. Email has advantages including fast, economical and can send messages in various forms of documents which make it popular around the world. The number of emails sent every day is often misused by some irresponsible parties to send information on advertisements for service products and various information that is useless or unwanted by users, which is called SPAM.

To filter SPAM emails, many classification algorithms are used, including Support Vector Machine (SVM) and Decission Tree (C4. 5). In this study, a comparison of the SVM and C4. 5. To see the performance of the algorithm modeling, the max_features test division is 1500 and 3000 data and uses a ratio of 90% : 10%, 80% : 20%,70% : 30% and 60% : 40% which aims to assess the data that will produce the highest accuracy value in the test conducted.

Based on the test results using data as much as 10,000 rows of data and 2 columns, using the Support Vector Machine (SVM) and Decission Tree (C4. 5) algorithms, the best accuracy results were obtained using Max_features 3000 with a ratio of 80%: 20% where SVM gets 98.20% and C4. 5 amounted to 93.30%. From the test results, it proves that the Support Vector Machine (SVM) algorithm is the best method in performing the email spam classification process.

Keyword: *Classification, Spam-email, SVM, C4.5*