PERBANDINGAN ALGORITMA PADA *MACHINE LEARNING* UNTUK KLASIFIKASI *EMAIL SPAM*

SKRIPSI

disusun oleh

Hery Iswanto

17.11.1713

PROGRAM SARJANA

PROGRAM STUDI INFORMATIKA

FAKULTAS ILMU KOMPUTER

UNIVERSITAS AMIKOM YOGYAKARTA

YOGYAKARTA

2021

PERBANDINGAN ALGORITMA PADA *MACHINE LEARNING* UNTUK KLASIFIKASI *EMAIL SPAM*

SKRIPSI

Untuk memenuhi sebagian persyaratan
mencapai gelar Sarjana
pada Program Studi Informatika

disusun oleh

Hery Iswanto

17.11.1713

PROGRAM SARJANA

PROGRAM STUDI INFORMATIKA

FAKULTAS ILMU KOMPUTER

UNIVERSITAS AMIKOM YOGYAKARTA

YOGYAKARTA

2021

PERSETUJUAN

SKRIPSI

PERBANDINGAN ALGORITMA PADA *MACHINE LEARNING* UNTUK KLASIFIKASI *EMAIL SPAM*

yang dipersiapkan dan disusun oleh

Hery Iswanto

17.11.1713

telah disetujui oleh Dosen Pembimbing Skripsi

pada tanggal 8 November 2021

Dosen Pembimbing

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

NIK. 190302231

PENGESAHAN

SKRIPSI

PERBANDINGAN ALGORITMA PADA *MACHINE LEARNING* UNTUK KLASIFIKASI *EMAIL SPAM*

yang dipersiapkan dan disusun oleh

Hery Iswanto

17.11.1713

telah dipertahankan di depan Dewan Penguji

pada tanggal 18 November 2021

Susunan Dewan Penguji

Nama Penguji	Tanda Tangan
Yuli Astuti, M.Kom **NIK. 190302146**	
Anna Baita, M.Kom **NIK. 190302290**	
Erni Seniwati, S.Kom, M.Cs **NIK. 190302231**	

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

Tanggal 18 November 2021

DEKAN FAKULTAS ILMU KOMPUTER

Hanif Al Fatta, M.Kom

NIK. 190302096

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, 18 November 2021

Hery Iswanto

NIM. 17.11.1713

MOTTO

"Didalam hidup ada saat untuk berhati-hati, atau berhenti berlari."

Taifun – Barasuara

PERSEMBAHAN

Puji syukur kepada Allah SWT yang telah memberikan saya sehat rohani maupun jasmani, nikmat yang tak terkira sehingga dapat menyelesaikan tugas akhir ini. Dengan ini saya persembahkan untuk :

1. Allah SWT yang telah memberikan kemudahan dan kelancaran dalam menyelesaikan skripsi ini.
2. Kedua orang tua yang selalu memberikan dukungan doa dan support kepada saya sehingga dapat berada dititik saat ini.
3. Untuk ibu dosen pembimbing saya Ibu Erni Seniwati, S.Kom, M.Cs terimakasih atas waktu dan ilmunya dalam membimbing saya menyelesaikan skripsi ini.
4. Untuk teman – teman kelas IF-12 yang sudah banyak membantu saat berkuliah di Universitas Amikom Yogyakarta.

Dan seluruh pihak yang tidak dapat saya sebutkan satu per satu, terimakasih atas segala bantuan dan doanya sehingga saya dapat menyelesaikan skripsi ini dengan baik.

KATA PENGANTAR

Assalamualaikum Warahmatullahi Wabarakatuh

Allhamdullilah Puji dan syukur senantiasa peneliti panjatkan kepada Allah SWT, karena berkat nikmat, rahmat, dan pertolongan-Nya peneliti dapat menyelesaikan laporan skripsi ini dengan baik. Laporan skripsi yang dibuat untuk memenuhi syarat memperoleh gelar kesarjanaan Strata-1 (S1) jurusan Informatika Universitas AMIKOM Yogyakarta diharapkan bisa menjadi salah satu referensi pembuatan skripsi di Universitas AMIKOM Yogyakarta serta dapat memberikan penambahan ide yang dapat dikembangkan dimasa depan.

Atas segala bantuan serta amal baik semua pihak diatas, semoga mendapat ridlo Allah SWT. Penulis sangat menyadari bahwa penulisan tugas akhir ini masih kurang sempurna mengingat kurangnya kemampuan dan pengetahuan penulis. Oleh karena itu, saran dan ktirik yang membangun dari pembaca sangat penulis harapkan demi kesempurnaan dan kebaikan tugas akhir ini. Penulis berharap bahwa penulisan tugas akhir ini dapat bermanfaat bagi penulis, pembaca, maupun penelitian di masa depan.

Wasalamualaikum Warahmatullahi Wabarakatuh

18 November 2021

Hery Iswanto

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR PERSAMAAN

INTISARI

Email merupakan bukti perkembangan teknologi di dunia dalam menggantikan peran surat konvensional sebagai sarana komunikasi. Fitur dan fasilitas yang diberikan oleh *email* memberikan kenyamanan dan kemudahan untuk pengguna mengirimkan berbagai informasi dalam waktu singkat meskipun terpaut jarak yang jauh. Kemampuan *email* yang dapat mengirim berbagai informasi ke penerima dengan mudah dan tanpa biaya ini sering disalahgunakan oleh berbagai pihak untuk mengirim informasi berisikan iklan produk atau jasa, dan berbagai informasi lainnya yang tidak diinginkan oleh pengguna *email*. Informasi inilah yang sering disebut dengan *email spam*.

Untuk mencegah *email spam*, terdapat banyak metode dalam melakukan penyaringan *email* untuk mengklasifikasikan jenis *email* kedalam *spam* atau *ham*. Pada penelitian ini akan melakukan perbandingan metode *Naïve Bayes, Random Forest,* dan *K-Nearest Neighbor* yang bertujuan untuk mencari metode klasifikasi yang baik untuk klasifikasi *email*. Sebelum melakukan proses klasifikasi, penelitian ini menambahkan proses *pre-processing* dan pembobotan dengan *TF-IDF*.

Berdasarkan hasil pengujian yang menggunakan 1998 sampel data, nilai akurasi tertinggi didapatkan pada metode *Naïve Bayes* dan *Random Forest* yaitu sebesar 83,5% diikuti dengan metode *K-Nearest Neighbor* sebesar 82,75%. Dengan penambahan uji coba menggunakan grafik *AUROC*, metode yang memilki nilai tertinggi didapatkan pada metode *Naïve Bayes* yaitu sebesar 0,859, metode *K-Nearest Neighbor* sebesar 0,845 dan metode *Random Forest* sebesar 0,840. Dari uji coba tersebut membuktikan bahwa metode *Naïve Bayes* merupakan metode terbaik dalam melakukan proses klasifikasi *email spam*.

Kata Kunci : *Email spam*, Klasifikasi, *Naïve Bayes, Random Forest, K-Nearest Neighbor*

ABSTRACT

Email is an evidence to technological developments in the world in replace the role of conventional letters as a form of communication. The features and facilities provided by e-mail provide comfort and convenience for users to send various information in a short time even though they are far apart. The ability of email to convey numerous types of information to recipients easily and cheaply is frequently abused by various parties to send information such as product or service advertisements, and as well as other information that email users do not want. This information is often referred to as spam email.

To prevent spam emails,there are many methods in filtering emails to classify email types into spam or ham. The study will compare the Methods of Naïve Bayes, Random Forest, and K-Nearest Neighbor which aims to find a good classification method for emailclassification. Before conducting the classification process, the study added a pre-processing and weighting process with TF-IDF.

Based on the results of tests using 1998 sample data, the highest accuracy value was obtained in the Naïve Bayes and Random Forest methods of 83.5% followed by the K-Nearest Neighbor method of 82.75%. With the addition of trials using AUROCcharts, the method that has the highest value is obtained in the Naïve Bayes method of 0.859, the K-Nearest Neighbor method of 0.845 and the Random Forest method of 0.840. From the trial proved that the Naïve Bayes method is the best method in the process of classifying spam emails

Keyword : *Spam Email,* Classification, *Naïve Bayes, Random Forest, K-Nearest Neighbor*