ANALISIS KOMPARATIF ALGORITMA MACHINE LEARNING UNTUK DETEKSI EMAIL SPAM

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana

Program Studi Teknik Komputer

disusun oleh

HUSEIN ALI ASGHAR

18.83.0313

Kepada

FAKULTAS ILMU KOMPUTER

UNIVERSITAS AMIKOM YOGYAKARTA

YOGYAKARTA

2023

ANALISIS KOMPARATIF ALGORITMA MACHINE LEARNING UNTUK DETEKSI EMAIL SPAM

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana

Program Studi Teknik Komputer

disusun oleh

HUSEIN ALI ASGHAR

18.83.0313

Kepada

FAKULTAS ILMU KOMPUTER

UNIVERSITAS AMIKOM YOGYAKARTA

YOGYAKARTA

2023

HALAMAN PERSETUJUAN

SKRIPSI

ANALISIS KOMPARATIF ALGORITMA MACHINE LEARNING UNTUK DETEKSI EMAIL SPAM

yang disusun dan diajukan oleh

Husein Ali Asghar

18.83.0313

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 21 Agustus 2023

Dosen Pembimbing,

Banu Santoso, S.T., M.Eng.
NIK. 190302327

HALAMAN PENGESAHAN

SKRIPSI

ANALISIS KOMPARATIF ALGORITMA MACHINE LEARNING UNTUK DETEKSI EMAIL SPAM

yang disusun dan diajukan oleh

Husein Ali Asghar

18.83.0313

Telah dipertahankan di depan Dewan Penguji
pada tanggal 21 Agustus 2023

Susunan Dewan Penguji

Nama Penguji	Tanda Tangan
Arifiyanto Hadinegoro, S.Kom., M.T. **NIK. 190302289**	
Wahid Miftahul Ashari, S.Kom., M.T. **NIK. 190302452**	
Banu Santoso, S.T., M.Eng. **NIK. 190302327**	

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 21 Agustus 2023

DEKAN FAKULTAS ILMU KOMPUTER

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

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Husein Ali Asghar
NIM : 18.83.0313

Menyatakan bahwa Skripsi dengan judul berikut:

Analisis Komparatif Algoritma Machine Learning Untuk Deteksi Email Spam
Dosen Pembimbing : Banu Santoso, S.T., M.Eng.

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, 21 Agustus 2023

Yang Menyatakan,

Husein Ali Asghar

HALAMAN PERSEMBAHAN

Puji syukur kehadirat Allah SWT atas berkat dan rahmat yang telah diberikan kepada saya sehingga dapat menyelesaikan skripsi ini. Oleh karena itu dengan rasa syukur dan bangga saya ucapkan terimakasih kepada:

1. Allah SWT, berkat rahmat kasih sayang dan pertolongan-Nya saya bisa sampai sejauh ini.
2. Kedua orang tua saya, Bapak dan Ibu, yang selalu mendukung dan mendidik saya sejak kecil hingga saat ini. Mereka selalu memberikan dukungan moril dan doa yang tulus sehingga saya dapat menyelesaikan masa studi S1 dengan baik. Tanpa dukungan dan doa mereka, saya tidak akan dapat mencapai titik ini.
3. Dosen pembimbing saya, Bapak Banu Santoso, S.T., M.Eng. yang selalu memacu dan membimbing saya dalam proses menyelesaikan skripsi.
4. Bapak/Ibu dosen pembimbing, pengajar, serta penguji di Universitas AMIKOM Yogyakarta yang selalu memberikan pengalaman ilmu dengan tulus dan mengarahkan saya sehingga saya dapat menyelesaikan masa studi S1dengan baik. Saya berharap segala ilmu yang telah diberikan oleh Bapak/Ibu akan menjadi ladang amal yang memberikan pahala jariyah dan selalu diberikan keberkahan ilmu dan rezeki oleh Tuhan yang Maha Esa.
5. Bapak Ali Reza Mulachela, Bapak A.M Safwan, dan Bapak Alm. Didiek Sri Wiyono yang telah mendukung saya baik secara materiil maupun moril selama menempuh pendidikan S1.
6. Serta semua pihak yang tidak bisa penulis sebutkan satu persatu.

KATA PENGANTAR

Dengan rasa syukur yang mendalam kepada Tuhan yang Maha Esa karena dengan rahmat-Nya saya dapat menyelesaikan penyusunan skripsi ini. Judul skripsi yang diajukan adalah "Analisis Komparatif Algoritma Machine Learning Untuk Deteksi Email Spam ". Skripsi ini diajukan untuk memenuhi persyaratan kelulusan mata kuliah skripsi di Fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta.

Oleh karena itu, saya ingin mengucapkan terima kasih yang sebesar besarnya kepada :

1. Allah SWT, berkat rahmat kasih sayang dan pertolongan-Nya saya bisa sampai sejauh ini.
2. Bapak Prof. Dr. M. Suyanto, MM. selaku Rektor Universitas AMIKOM Yogyakarta.
3. Bapak Hanif Al Fatta, S.Kom., M.Kom. selaku Dekan Fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta.
4. Bapak Dony Ariyus, M.Kom. Selaku Kaprodi Teknik Komputer Universitas AMIKOM Yogyakarta.
5. Bapak Banu Santoso, S.T., M.Eng. selaku dosen pembimbing saya.
6. Dosen penguji dan segenap Dosen serta Karyawan Universitas AMIKOM Yogyakarta yang telah berbagi ilmu dan pengalamannya.
7. Kedua orang tua yang telah mendoakan, mendukung dan memberikan semangat kepada saya.
8. Semua pihak yang telah membantu dan tidak dapat disebutkan satu persatu.

Pada akhir kata pengantar ini, penulis menyadari bahwa tidak ada yang sempurna dan pasti terdapat kesalahan dalam penyusunan skripsi. Oleh karena itu, penulis meminta maaf atas segala kesalahan yang mungkin terjadi. Penulis berharap agar skripsi ini dapat bermanfaat bagi pembaca dan dapat dijadikan sebagai referensi untuk pengembangan ilmu yang lebih baik.

Yogyakarta, 21 Januari 2023

Penulis

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR LAMBANG DAN SINGKATAN

FPR	False Positive Rate
MCC	Matthew Correlation Coefficient
MNB	Multinomial Naïve Bayes
OCR	Optical Character Recognition
RF	Random Forest
SVM	Support Vector Machine
TNR	True Negative Rate
TPR	True Positive Rate

INTISARI

Penelitian ini bertujuan untuk melakukan perbandingan terhadap beberapa metode klasifikasi untuk klasifikasi surel spam dan gambar spam berdasarkan ekstraksi fitur N-gram. Penelitian ini menggunakan algoritma klasifikasi machine learning yaitu Multinomial Naïve Bayes, Random Forest dan Support Vector Machine. Dalam melakukan klasifikasi gambar spam, peneliti menggunakan metode Optical Character Recognition dengan menggunakan library Python yaitu pyteserract. Penelitian ini menggunakan open-source dataset yang berasal dari Kaggle untuk deteksi surel spam dan Spam Hunter untuk deteksi gambar spam.

Hasil dari penelitian ini menunjukkan bahwa algoritma Random Forest merupakan pemodelan machine learning terbaik dalam mendeteksi surel spam dengan tingkat akurasi 0,99, dan nilai Matthew Correlation Coefficient 0,999. Algoritma pemodelan machine learning dengan performa terbaik untuk mendeteksi gambar spam adalah algoritma Support Vector Machine dan algoritma Multinomial Naïve Bayes dengan tingkat akurasi yang sama yaitu 0,98 dan nilai Matthew Correlation Coefficient yang sama yaitu 0,976. Metode ektraksi fitur N-gram terbaik pada penelitian ini adalah 1-gram.

Dari hasil penelitian ini, dapat disimpulkan bahwa algoritma Random Forest dengan ektraksi fitur 1-gram memiliki performa terbaik dalam mendeteksi surel spam dan algoritma Multinomial Naïve Bayes serta Support Vector Machine dengan ekstraksi fitur 1-gram memiliki performa terbaik dalam mendeteksi gambar spam.

Kata kunci: *N-gram, Random Forest, Support Vector Machine, Multinomial Naïve Bayes,* Deteksi *Spam*

ABSTRACT

This study aims to compare several classification methods for the classification of spam emails and spam images based on N-gram feature extraction. This study uses a machine learning classification algorithm, namely Multinomial Naïve Bayes, Random Forest and Support Vector Machine. In classifying spam images, researchers use the Optical Character Recognition method using the Python library, namely pyteserract. This study uses open-source datasets from Kaggle for spam email detection and Spam Hunter for image spam detection.

The results of this study indicate that the Random Forest algorithm is the best machine learning model for detecting spam emails with an accuracy level of 0.99, and a Matthew Correlation Coefficient value of 0.999. The machine learning modeling algorithms with the best performance for detecting spam images are the Support Vector Machine algorithm and the Multinomial Naïve Bayes algorithm with the same accuracy level of 0.98 and the same Matthew Correlation Coefficient value of 0.976. The best N-gram feature extraction method in this study is 1-gram.

From the results of this study, it can be concluded that the Random Forest algorithm with 1-gram feature extraction has the best performance in detecting spam emails and the Multinomial Naïve Bayes algorithm and the Support Vector Machine with 1-gram feature extraction has the best performance in detecting spam images.

***Keyword**: N-gram, Random Forest, Support Vector Machine, Multinomial Naïve Bayes, Spam Detection*