PENERAPAN ALGORITMA C4.5 MENGGUNAKAN TF-IDF DAN N-GRAM UNTUK PREDIKSI *EMAIL* SPAM

SKRIPSI

Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana Komputer
Pada Program Studi Teknik Informatik

Oleh:

Bharaka Zulfa Maraghi

4611419062

PROGRAM STUDI SARJANA TEKNIK INFORMATIKA

FAKULKTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS NEGERI SEMARANG

2023

PERSETUJUAN PEMBIMBING

Skripsi berjudul "Penerapan Algoritma C4.5 Menggunakan TF-IDF dan N-Gram untuk Prediksi *Email* Spam" yang disusun oleh

Nama : Bharaka Zulfa Maraghi

NIM : 4611419062

Prodi : Teknik Informatika

Telah disetujui untuk diajukan ke sidang ujian skripsi.

Semarang, 14 Agustus 2023

Pembimbing,

Aji Purwinarko, S.Si., M.Cs.

NIP 198509102015041001

PENGESAHAN PENGUJI

Skripsi berjudul "Penerapan Algoritma C4.5 Menggunakan TF-IDF dan N-Gram untuk Prediksi *Email* Spam" yang disusun oleh

Nama : Bharaka Zulfa Maraghi

NIM : 4611419062

Prodi : Teknik Informatika

Telah dipertahankan dalam ujian sidang skripsi pada hari Kamis, 17 Agustus 2023.

Tim Penguji

Ketua Penguji Prof. Dr. Edy Cahyono, M.Si. NIP. 196412051990021001	
Sekretaris Dr. Alamsyah, S.Si., M.Kom. NIP. 197405172006041001	
Penguji 1 Riza Arifudin, S.Pd., M.Cs. NIP. 19800525005011001	
Penguji 2 Anggyi Trisnawan Putra, S.Si., M.Si. NIP. 198707062014041003	
Penguji 3/Pembimbing Aji Purwinarko, S.Si., M.Cs. NIP 198509102015041001	

PERNYATAAN

Skripsi yang ditulis berjudul "Penerapan Algoritma C4.5 Menggunakan TF-IDF dan N-Gram untuk Prediksi *Email* Spam" merupakan karya ilmiah asli dan bukan hasil plagiasi dari karya ilmiah orang lain. Pendapat atau temuan orang lain yang dikutip di dalam Skripsi ini telah ditulis berdasarkan kode etik ilmiah.

Semarang, 16 Agustus 2023

Yang menyatakan,

Bharaka Zulfa Maraghi

NIM. 4611419062

MOTTO DAN PERSEMBAHAN

MOTTO

- Hidup akan tragis jika tidak lucu. (Stephen Hawking)
- Kebahagiaan kita tergantung pada diri kita sendiri. (Aristoteles)

PERSEMBAHAN

Skripsi ini saya persembahkan kepada:

- Kedua orang tua saya, Bapak Joko Sri Mulyono dan Ibu Rini Murwanti yang selalu memberi dukungan, binaan, tempaan, kasih sayang, do'a, serta kesabaran dalam mendidik saya sedari kecil hingga saat ini.
- Untuk saudara saya, Ramadhani Azlam Satria Maraghi yang telah memberikan dukungan bagi penulis.
- Diri saya sendiri, yang sudah mau berusaha, belajar, dan berjuang sampai sekarang serta bertahan ditengah banyaknya terpaan dari segala sisi. Semoga dapat menjadikan pribadi yang lebih kuat untuk kehidupan kedepannya.
- Seluruh dosen Universitas Negeri Semarang yang senantiasa memberikan kepada saya bekal ilmu yang tak ternilai. Khususnya Bapak dan Ibu dosen jurusan Ilmu Komputer yang saya anggap seperti orang tua saya sendiri ketika berada di kampus.
- Seluruh keluarga yang tidak dapat disebutkan satu persatu yang telah memberikan doa dan dukungan moril selama proses penyusunan skripsi.
- Keluarga The kams, Amara, Fachrizal, Andhika, dan Ghamal.
- Sahabat-sahabat saya yang selalu menemani hari-hari saya dalam situasi dan kondisi apapun. Semoga tali persahabatan ini selalu terjaga.
- Orang spesial di hati saya yang sudah memberikan semangat dalam penulisan skripsi ini.
- Almamater Universitas Negeri Semarang.

ABSTRAK

Maraghi, Bharaka Z. 2023. Penerapan Algoritma C4.5 Menggunakan TF-IDF dan N-Gram untuk Prediksi *Email* Spam. Skripsi, Program Studi Teknik Informatika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Semarang. Pembimbing Aji Purwinarko, S.Si., M.Cs.

Kata kunci: klasifikasi teks, *email* spam, *machine learning,* C4.5, TF-IDF, N-gram

Email adalah sarana komunikasi populer dalam jaringan internet karena kemudahan dan kecepatan penggunaannya. Namun, penyalahgunaan *email* seperti email spam dapat merugikan orang lain dan mengganggu penggunaan *email* secara efisien. *Email* spam berisi konten iklan, penipuan, dan virus yang dikirim oleh pengguna yang tidak diminta, yang dapat menyebabkan pemborosan waktu, penyimpanan server email, dan *bandwidth*. Dalam penanganan masalah *email* spam, teknologi text mining dapat digunakan untuk mengenali *email* spam dengan lebih efisien dan akurat. Penelitian ini bertujuan untuk mengimplementasikan algoritma C4.5 dengan menerapkan metode TF-IDF dan N-gram dalam melakukan prediksi *email* spam. Penggunaan TF-IDF digunakan untuk merepresentasikan teks *email* menjadi representasi numerik, sementara analisis N-gram digunakan untuk memahami pola dan hubungan antar kata-kata dalam teks. Penelitian dilakukan dengan menggunakan *dataset Ling-Spam Dataset*. Hasil penelitian menunjukkan bahwa penerapan algoritma C4.5 dengan TF-IDF dan N-gram berhasil mencapai 97,92% untuk akurasi, 100% untuk presisi, dan 95,83% untuk *recall* dalam prediksi *email* spam. Penggabungan metode TF-IDF dan N-gram memberikan kemampuan lebih baik dalam mengenali pola khusus dari *email* spam, sehingga meningkatkan kualitas prediksi. Hasil dari penelitian ini menunjukkan potensi penggunaan TF-IDF dan N-gram dalam meningkatkan kualitas prediksi email spam. Penelitian ini memberikan kontribusi penting dalam pengembangan sistem prediksi *email* spam yang lebih handal dan efisien, dengan tujuan melindungi pengguna dari ancaman *email* spam yang berbahaya.

PRAKATA

Puji syukur saya panjatkan atas kehadirat Allah Subhanahu wa ta'ala yang telah memberikan rahmat, taufik, dan hidayah-Nya sehingga penulis dapat melakukan penelitian dan menyelesaikan skripsi yang berjudul "Penerapan Algoritma C4.5 Menggunakan TF-IDF dan N-Gram untuk Prediksi *Email* Spam" dengan lancar. Penyusunan skripsi ini digunakan untuk memenuhi salah satu persyaratan akademik dalam menyelesaikan Program Sarjana (S1) pada Program Studi Teknik Informatika, Program Studi Teknik Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Semarang.

Penulis menyadari bahwa tanpa bimbingan, motivasi, bantuan, dan dukungan dari berbagai pihak maka skripsi ini tidak mungkin dapat terselesaikan seperti pada saat ini. Oleh karena itu, penulis ingin menyampaikan rasa terima kasih kepada seluruh pihak yang berperan dalam proses penyusunan skripsi ini, diantaranya:

1. Bapak Prof. Dr. S Martono, M.Si., Rektor Universitas Negeri Semarang.
2. Bapak Prof. Dr. Edy Cahyono, M.Si., Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Semarang.
3. Bapak Dr. Alamsyah, S.Si., M.Kom., Koordinator Program Studi Teknik Informatika Universitas Negeri Semarang.
4. Bapak Aji Purwinarko, S.Si., M.Cs. selaku Dosen Wali sekaligus Dosen Pembimbing yang telah meluangkan waktu untuk mendampingi dan memberikan arahan penulis dalam proses penyusunan skripsi.
5. Bapak Riza Arifudin, S.Pd., M.Cs., selaku Dosen Penguji 1 dan Bapak Anggyi Trisnawan Putra, S.Si., M.Si., selaku Dosen Penguji 2 yang telah memberikan masukan berupa kritik dan sarannya dalam memperbaiki skripsi ini.
6. Seluruh dosen dan jajaran staf akademik di lingkungan Universitas Negeri Semarang yang telah membantu penulis dalam mengurus dokumen administrasi.

7. Kedua orang tua saya, Bapak Joko Sri Mulyono dan Ibu Rini Murwanti yang telah memberikan dukungan doa, moril, dan materil yang tiada henti-hentinya.
8. Saudara saya yang telah memberikan semangat dan motivasi bagi penulis dalam menyusun skripsi ini.
9. Sahabat dan kerabat terdekat yang turut memberikan semangat dan dukungan selama proses penyusunan skripsi ini.
10. Teman-teman di Program Studi Teknik Informatika angkatan 2019 Universitas Negeri Semarang yang turut membantu, memberikan dukungan, dan motivasi bagi penulis hingga terselesaikannya skripsi ini.
11. Seluruh pihak yang berperan dalam membantu proses dari awal penyusunan skripsi hingga skripsi ini dapat terselesaikan yang tidak dapat disebutkan satu per satu.

Demikian pengantar yang dapat penulis sampaikan. Segala kelebihan berasal dari Allah Swt. dan banyak kekurangan dari penulis. Oleh karena itu penulis memohon maaf apabila terdapat kekurangan dalam penulisan skripsi ini dan penulis berharap saran serta kritik yang membangun dari pembaca untuk menjadikan skripsi ini lebih baik lagi. Semoga skripsi ini dapat bermanfaat bagi para pembaca. Akhir kata, penulis menyampaikan terima kasih.

Semarang, 16 Agustus 2023

Yang menyatakan,

Bharaka Zulfa Maraghi

NIM. 4611419062

DAFTAR ISI