ANALISIS KLASIFIKASI EMAIL SPAM MENGGUNAKAN ALGORITMA NAÏVE BAYES

SKRIPSI

Oleh
Azan Rahman
210210102

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK DAN KOMPUTER
UNIVERSITAS PUTERA BATAM
TAHUN 2025

ANALISIS KLASIFIKASI EMAIL SPAM MENGGUNAKAN ALGORITMA NAÏVE BAYES

SKRIPSI

Untuk memenuhi salah satu syarat
Memperoleh gelar sarjana

Oleh
Azan Rahman
210210102

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK DAN KOMPUTER
UNIVERSITAS PUTERA BATAM
TAHUN 2025

SURAT PERNYATAAN ORISINALITAS

Yang bertanda tangan di bawah ini saya:

Nama : Azan Rahman

NPM : 210210102

Fakultas : Teknik dan Komputer

Program studi : Teknik Informatika

Menyatakan bahwa **"SKRIPSI"** yang saya buat dengan judul:

ANALISIS KLASIFIKASI SPAM EMAIL MENGGUNAKAN ALGORITMA NAÏVE BAYES

Adalah hasil karya sendiri dan bukan "duplikasi" dari karya orang lain. Sepengetahuan saya, di dalam naskah Skripsi ini tidak terdapat karya ilmiah atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis dikutip didalam naskah ini dan disebutkan dalam sumber kutipan dan daftar pustaka.

Apabila ternyata di dalam naskah Skripsi ini dapat dibuktikan terdapat unsur-unsur Plagiasi, saya bersedia naskah Skripsi ini digugurkan dan gelar akademik yang saya peroleh dibatalkan, serta diproses sesuai dengan peraturan perundang-undangan yang berlaku. Demikian pernyataan ini saya buat dengan sebenarnya tanpa ada paksaan dari siapapun

Batam, 7 Februari 2025

Azan Rahman

ANALISIS KLASIFIKASI SPAM EMAIL MENGGUNAKAN ALGORITMA NAÏVE BAYES

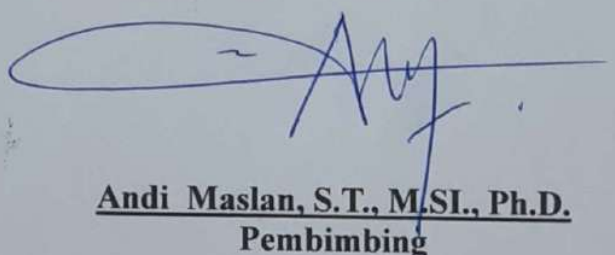
SKRIPSI

Untuk memenuhi salah satu syarat
Memperoleh gelar Sarjana

Oleh
Azan Rahman
210210102

Telah disetujui oleh Pembimbing pada tanggal
seperti tertera dibawah ini

Batam, 7 Februari 2025



Andi Maslan, S.T., M.SI., Ph.D.
Pembimbing

ABSTRAK

Perkembangan teknologi informasi dan komunikasi telah membawa perubahan signifikan dalam cara manusia berkomunikasi, termasuk dalam penggunaan surat elektronik (email). Namun, meningkatnya penggunaan email juga diikuti oleh maraknya penyebaran email spam, yang dapat mengganggu pengguna dan menimbulkan risiko keamanan. Penelitian ini bertujuan untuk menganalisis klasifikasi email spam dengan menerapkan algoritma Naïve Bayes. Proses penelitian diawali dengan tahap *preprocessing* data guna meningkatkan kualitas teks sebelum dilakukan klasifikasi dimulai dari tokenisasi, lemmatisasi, penghapusan kata tidak penting (stopword), serta konversi teks menggunakan Term Frequency-Inverse Document Frequency (TF-IDF) agar dapat direpresentasikan dalam bentuk numerik. Selanjutnya, algoritma Naïve Bayes digunakan untuk mengklasifikasikan email ke dalam kategori spam atau non-spam. Kinerja model dianalisis menggunakan confusion matrix serta sejumlah metrik evaluasi, termasuk akurasi, presisi, recall, dan F1-score. Hasil eksperimen menunjukkan bahwa model awal memperoleh akurasi sebesar 88%, namun memiliki nilai *False Negative* yang cukup tinggi akibat ketidakseimbangan kelas pada dataset, di mana jumlah email non-spam lebih banyak dibandingkan email spam. Untuk mengatasi permasalahan ini, diterapkan teknik penyeimbangan kelas menggunakan Synthetic Minority Over-sampling Technique (SMOTE). Setelah penerapan teknik ini, model mengalami peningkatan performa dengan akurasi mencapai 98% serta penurunan signifikan pada nilai *False Negative*. Hasil penelitian ini menunjukkan bahwa algoritma Naïve Bayes memiliki kinerja yang baik dalam mengklasifikasikan email spam setelah dilakukan penyesuaian terhadap ketidakseimbangan data.

Kata Kunci: *Naïve Bayes, klasifikasi spam, email, preprocessing, evaluasi model*

ABSTRACT

The advancement of information and communication technology has significantly transformed the way people communicate, including the use of electronic mail (email). However, the increasing use of email has also led to the widespread distribution of spam emails, which can disrupt users and pose security risks. This study aims to analyze spam email classification by implementing the Naïve Bayes algorithm. The research process begins with data preprocessing to enhance text quality before classification. This includes tokenization, lemmatization, stopword removal, and text conversion using Term Frequency-Inverse Document Frequency (TF-IDF) to represent text in numerical form. Subsequently, the Naïve Bayes algorithm is utilized to classify emails into spam or non-spam categories. The model's performance is evaluated using a confusion matrix and several evaluation metrics, including accuracy, precision, recall, and F1-score. Experimental results indicate that the initial model achieved an accuracy of 88%, but exhibited a high False Negative rate due to class imbalance in the dataset, where the number of non-spam emails exceeded spam emails. To address this issue, a class balancing technique using the Synthetic Minority Over-sampling Technique (SMOTE) was applied. After implementing this technique, the model demonstrated improved performance, achieving an accuracy of 98% and a significant reduction in False Negative values. The findings of this study indicate that the Naïve Bayes algorithm performs well in classifying spam emails after addressing data imbalance issues.

Keywords: *Naïve Bayes, spam classification, email, preprocessing, model evaluation*

KATA PENGANTAR

Segala puji bagi Allah yang telah melimpahkan segala rahmat dan karuniaNya, sehingga penulis dapat menyelesaikan laporan tugas akhir yang merupakan salah satu persyaratan untuk menyelesaikan program studi strata satu (S1) pada Program Studi Teknik Informatika Universitas Putera Batam.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna. Karena itu, kritik dan saran akan senantiasa penulis terima dengan senang hati. Dengan segala keterbatasan, penulis menyadari pula bahwa skripsi ini takkan terwujud tanpa bantuan, bimbingan, dan dorongan dari berbagai pihak. Untuk itu, dengan segala kerendahan hati, penulis menyampaikan ucapan terima kasih kepada:

1. Rektor Universitas Putera Batam;
2. Dekan Fakultas Teknik dan Komputer;
3. Ketua Program Studi Teknik Informatika, Andi Maslan S.T., M.SI., Ph.D.;
4. Bapak Andi Maslan. S.T., M.SI.,Ph.D. selaku pembimbing skripsi pada Program Studi Teknik Informatika Universitas Putera Batam;
5. Dosen dan Staff Universitas Putera Batam;

Semoga Allah membalas kebaikan dan selalu mencurahkan hidayah serta taufik-Nya, Amin.

Batam, 7 Februari 2025

Penulis (Azan rahman)

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

DAFTAR RUMUS