



**PENERAPAN ALGORITMA NAIVE BAYES UNTUK DETEKSI SPAM
PADA EMAIL**

LAPORAN TUGAS AKHIR

MUHAMMAD REIZA

41520010059

**UNIVERSITAS
MERCU BUANA**

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA
2025**



**PENERAPAN ALGORITMA NAIVE BAYES UNTUK DETEKSI SPAM
PADA EMAIL**

LAPORAN TUGAS AKHIR

MUHAMMAD REIZA

41520010059

Diajukan sebagai salah satu syarat untuk memperoleh gelar sarjana

**UNIVERSITAS
PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA
2025**

HALAMAN PERNYATAAN KARYA SENDIRI

Saya yang bertanda tangan di bawah ini:

Nama : Muhammad Reiza
NIM : 41520010059
Program Studi : Teknik Informatika
Judul Laporan Skripsi : Penerapan Algoritma Naive Bayes Untuk Deteksi Spam Pada Email

Menyatakan bahwa Laporan Skripsi ini adalah hasil karya saya sendiri dan bukan plagiat, serta semua sumber baik yang dikutip maupun dirujuk telah saya nyatakan dengan benar. Apabila ternyata ditemukan di dalam Laporan Skripsi saya terdapat unsur plagiat, maka saya siap mendapatkan sanksi akademis yang berlaku di Universitas Mercu Buana.



Jakarta, 7 Juli 2025



Muhammad Reiza

UNIVERSITAS
MERCU BUANA

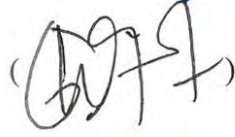
HALAMAN PENGESAHAN

Laporan Skripsi ini diajukan oleh:

Nama : Muhammad Reiza
NIM : 41520010059
Program Studi : Teknik Informatika
Judul Laporan Skripsi : Penerapan Algoritma Naïve Bayes Untuk Deteksi Spam Pada Email

Telah berhasil dipertahankan pada sidang di hadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Strata 1 pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer Universitas Mercu Buana.

Disahkan oleh:

Pembimbing : Sabar Rudiarto, S.Kom., M.Kom. ()
NIDN : 0309036902
Ketua Penguji : Wawan Gunawan, S.Kom., M.T., M.Kom. ()
NIDN : 0424108104
Penguji 1 : Roy Mubarak, S.T., M.Kom ()
NIDN : 0310027402
Penguji 2 : Siti Maesaroh, S.Kom., M.T.I ()
NIDN : 0413059003

UNIVERSITAS
MERCU BUANA

Jakarta, 21 Juli 2025

Mengetahui,

Dekan



Dr. Bambang Jokonowo, S.Si., MTI
NIDN : 0320037002

Ketua Program Studi



Dr. Hadi Santoso, S.Kom., M.Kom
NIDN : 0225067701

KATA PENGANTAR

Puji syukur kehadiran Tuhan yang Maha Esa, atas segala rahmat dan ridha-Nya sehingga penulis dapat menyelesaikan tugas akhir yang merupakan salah satu persyaratan kelulusan Program Studi Strata Satu (S1) pada jurusan Teknik Informatika, Universitas Mercu Buana.

Penulis menyadari bahwa tugas akhir ini masih jauh dari sempurna, karena kesempurnaan sejatinya hanya milik Tuhan yang Maha Esa. Oleh karena itu, saran dan masukan yang membangun senantiasa penulis terima dengan senang hati. Serta berkat dukungan, motivasi, bantuan, bimbingan, dan doa dari banyak pihak, penulis mengucapkan terima kasih kepada:

1. Prof. Dr. Andi Adriansyah, M.Eng. selaku Rektor Universitas Mercu Buana.
2. Dr. Bambang Jokonowo, S.Si., MTI selaku Dekan Fakultas Ilmu Komputer.
3. Dr. Hadi Santoso, S.Kom., M.Kom. selaku Ketua Program Studi Teknik Informatika Universitas Mercubuana.
4. Sabar Rudiarto, M.Kom. selaku dosen pembimbing tugas akhir yang telah memberikan pengarahan, motivasi, menyediakan waktu, tenaga, dan pikiran sehingga selama pembuatan tugas akhir ini terjadwal dengan baik.
5. Kedua Orang Tua saya yang selalu mensuport dan mendukung saya selama menjalani masa studi sebagai mahasiswa Universitas Mercubuana..
6. Pasangan saya yang selalu berbagi memberikan dukungan.

Akhir kata, penulis berharap semoga Tuhan yang Maha Esa membalas kebaikan dan selalu mencurahkan rahmat, hidayah, serta panjang umur kepada kita semua, aamiin. Terima Kasih.

Jakarta, 21 Juli 2025

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademik Universitas Mercu Buana, saya yang bertanda tangan di bawah ini:

Nama : Muhammad Reiza
NIM : 41520010059
Program Studi : Teknik Informatika
Judul Laporan Skripsi : Penerapan Algoritma Naive Bayes Untuk Deteksi Spam Pada Email

Demi pengembangan ilmu pengetahuan, dengan ini memberikan izin dan menyetujui untuk memberikan kepada Universitas Mercu Buana **Hak Bebas Royalti Non-Eksklusif (*Non-exclusive Royalty-Free Right*)** atas karya ilmiah saya yang berjudul di atas beserta perangkat yang ada (jika diperlukan).

Dengan Hak Bebas Royalti Non-Eksklusif ini Universitas Mercu Buana berhak menyimpan, mengalihmedia/format-kan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan Laporan Magang/Skripsi/Tesis/Disertasi saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 7 Juli 2025

Yang menyatakan,

UNIVERSITAS
MERCU BUANA



Muhammad Reiza

ABSTRAK

Nama : Muhammad Reiza
NIM : 41520010059
Program Studi : Teknik Informatika
Judul Laporan Skripsi : Penerapan Algoritma Naive Bayes Untuk Deteksi Spam Pada Email
Dosen Pembimbing : Sabar Rudiarto, S.Kom., M.Kom.

Deteksi spam pada email telah menjadi isu krusial di era digital. Dengan tingginya volume email yang saya terima setiap hari, saya sering berhadapan dengan pesan tidak diinginkan atau berpotensi berbahaya, seperti upaya *phishing* atau penyebaran *malware*. Kondisi ini menggarisbawahi urgensi sistem deteksi spam yang efektif dan andal. Oleh karena itu, penelitian ini saya tujukan untuk menerapkan Algoritma Naive Bayes sebagai metode klasifikasi guna otomatis mendeteksi spam dalam email. Dalam metodologi penelitian ini, saya mulai dengan mengumpulkan dataset email terlabel (spam/non-spam). Data kemudian melalui serangkaian proses pra-pemrosesan dan ekstraksi fitur penting. Saya mengaplikasikan CountVectorizer untuk mengubah teks email menjadi representasi numerik berupa matriks frekuensi kata, yang menjadi input bagi Algoritma Naive Bayes. Model ini saya latih untuk menghitung probabilitas kata dan menentukan kategori email. Proses pelatihan dan pengujian saya lakukan dengan membagi dataset secara proporsional. Hasil evaluasi menunjukkan bahwa Algoritma Naive Bayes mencapai kinerja sangat baik, dibuktikan dengan akurasi **0.9856 (98.56%)**, presisi **0.9567 (95.67%)**, *recall* **0.9412 (94.12%)**, dan F1-Score **0.9489 (94.89%)**. Meskipun saya menemukan 5 *false positives* (email sah yang salah klasifikasi) dan 10 *false negatives* (spam yang lolos) dalam *confusion matrix*, angka ini menunjukkan efektivitas tinggi model dalam mengidentifikasi email spam. Temuan ini memberikan wawasan berharga bagi pengembang sistem filter spam dan saya harapkan menjadi referensi studi lanjutan guna meningkatkan efektivitas deteksi spam.

Kata kunci: Deteksi Spam, Email, Algoritma Naive Bayes, Klasifikasi Teks, Pembelajaran Mesin.

ABSTRACT

Name : Muhammad Reiza
NIM : 41520010059
Study program : Teknik Informatika
Thesis Report Title : Application of Naive Bayes Algorithm for Spam
Detection in Email
Supervisor : Sabar Rudiarto, S.Kom., M.Kom.

Spam email detection has become a crucial issue in the digital age. With the high volume of emails I receive daily, I often encounter unwanted or potentially dangerous messages, such as phishing attempts or malware dissemination. This situation underscores the urgency of effective and reliable spam detection systems. Therefore, I directed this research towards implementing the Naive Bayes Algorithm as a classification method to automatically detect spam in emails. In this research methodology, I began by collecting a labeled email dataset (spam/non-spam). The data then underwent a series of preprocessing and essential feature extraction steps. I applied CountVectorizer to transform email text into a numerical representation in the form of a word frequency matrix, which served as input for the Naive Bayes Algorithm. I trained this model to calculate word probabilities and determine email categories. I performed the training and testing processes by proportionally splitting the dataset. Evaluation results show that the Naive Bayes Algorithm achieved very good performance, evidenced by an accuracy of **0.9856 (98.56%)**, precision of **0.9567 (95.67%)**, recall of **0.9412 (94.12%)**, and an F1-Score of **0.9489 (94.89%)**. Although I found 5 false positives (legitimate emails misclassified) and 10 false negatives (spam leaked) in the confusion matrix, these figures indicate the model's high effectiveness in identifying spam emails. These findings provide valuable insights for spam filter developers, and I hope they will serve as a reference for further studies to enhance spam detection effectiveness.

Keywords: Spam Detection, Email, Naive Bayes Algorithm, Text Classification, Machine Learning.

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERNYATAAN.....	ii
HALAMAN PENGESAHAN.....	iii
KATA PENGANTAR.....	iv
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS.....	v
ABSTRAK	vi
ABSTRACT	vii
DAFTAR ISI.....	viii
DAFTAR TABEL	x
DAFTAR GAMBAR.....	xi
DAFTAR LAMPIRAN	xii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Perumusan Masalah	2
1.3 Tujuan Penelitian.....	3
1.4 Manfaat Penelitian.....	3
1.5 Batasan Masalah	4
BAB II TINJAUAN PUSTAKA.....	6
2.1 Penelitian Terdahulu	6
2.2 Teori Pendukung.....	18
BAB III METODE PENELITIAN	22
3.1 Jenis Penelitian.....	22
3.2 Tahapan Penelitian	23
BAB IV PEMBAHASAN	30
4.1 Gambaran Umum Sistem Deteksi Spam	30
4.2 Hasil Pengujian Model	30
4.3 Analisis Confusion Matrix	31
4.4 Interpretasi Hasil	33
BAB V KESIMPULAN DAN SARAN	34
5.1 Kesimpulan.....	34

5.2 Saran	35
DAFTAR PUSTAKA.....	37
LAMPIRAN.....	40



DAFTAR TABEL

Tabel 2.1 Penelitian Terkait

6



DAFTAR GAMBAR

Gambar 3.1 Tahapan Penelitian

12



DAFTAR LAMPIRAN

Lampiran 1 Kartu Asistensi	18
Lampiran 2 Curriculum Vitae	21
Lampiran 3 Surat Pernyataan HAKI	22
Lampiran 4 Sertifikat BNSP	24
Lampiran 5 Hasil Turnitin	25

