

# **TUGAS AKHIR**

**DETEKSI EMAIL SPAM MENGGUNAKAN  
ALGORITMA *NAIVE BAYES* BERDASARKAN  
POLA KATA DAN FRASA**



**Disusun Oleh :**

**HENRIAN REVINIELDY IGU**  
**1462100101**

**PROGRAM STUDI INFORMATIKA  
FAKULTAS TEKNIK  
UNIVERSITAS 17 AGUSTUS 1945 SURABAYA**

**2025**

# TUGAS AKHIR

**DETEKSI EMAIL SPAM MENGGUNAKAN  
ALGORITMA *NAIVE BAYES* BERDASARKAN  
POLA KATA DAN FRASA**



**Disusun Oleh :**

**HENDRIAN REVINIELDY IGU**  
**1462100101**

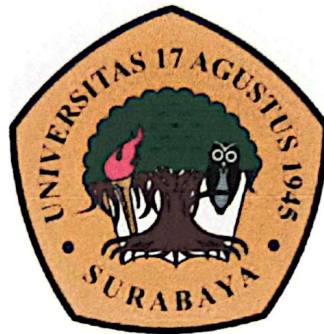
**PROGRAM STUDI INFORMATIKA  
FAKULTAS TEKNIK  
UNIVERSITAS 17 AGUSTUS 1945 SURABAYA**

**2025**

## TUGAS AKHIR

# DETEKSI EMAIL SPAM MENGGUNAKAN ALGORITMA *NAIVE BAYES* BERDASARKAN POLA KATA DAN FRASA

Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer  
di Program Studi Informatika



Oleh:

Hendrian Revinieldy Igu

1462100101

PROGRAM STUDI INFORMATIKA  
FAKULTAS TEKNIK  
UNIVERSITAS 17 AGUSTUS 1945 SURABAYA  
2025

**FINAL PROJECT**  
**SPAM EMAIL DETECTION USING NAIVE BAYES**  
**ALGORITHM BASED ON WORD AND PHRASE PATTERNS**

**Prepared as partial fulfilment of the requirement for the degree of Sarjana  
Komputer at Informatics Department**



**By:**

**Hendrian Revinieldy Igu**

**1462100101**

**INFORMATICS DEPARTMENT**  
**FACULTY OF ENGINEERING**  
**UNIVERSITAS 17 AGUSTUS 1945 SURABAYA**  
**2025**

**PROGRAM STUDI INFORMATIKA  
FAKULTAS TEKNIK  
UNIVERSITAS 17 AGUSTUS 1945 SURABAYA**

---

**LEMBAR PENGESAHAN TUGAS AKHIR**

**Nama** : Hendrian Revinieldy Igu  
**NBI** : 1462100101  
**Prodi** : S-1 Informatika  
**Fakultas** : Teknik  
**Judul** : DETEKSI EMAIL SPAM MENGGUNAKAN  
ALGORITMA *NAIVE BAYES* BERDASARKAN POLA  
KATA DAN FRASA


**Mengetahui/ Menyetujui,  
Dosen Pembimbing 1**




Supangat, M. Kom  
NPP. 20460.15.0675


**Mengetahui,**

**Dekan Fakultas Teknik Universitas 17  
Agustus 1945 Surabaya**



  
Dr. Irfan Ganiyo, M.Kes., IPU., ASEAN Eng.  
NPP. 20410.90.0197

**Ketua Program Studi Informatika  
University 17 Agustus 1945**



Aidil Primasetya Armin S.ST., M.T.  
NPP. 20460.16.0700

## PERNYATAAN KEASLIAN DAN PERSETUJUAN PUBLIKASI TUGAS AKHIR

Saya yang bertanda tangan di bawah ini

Nama : Hendrian Revinielly Igu  
NBI : 1462100101  
Fakultas / Program Studi : Teknik Informatika  
Judul Tugas Akhir : Deteksi Email Spam Menggunakan  
Algoritma *Naive Bayes* Berdasarkan Pola  
Kata dan Frasa

Menyatakan dengan sesungguhnya bahwa:

1. Tugas Akhir dengan judul diatas bukan merupakan tiruan atau duplikasi dari Tugas Akhir yang sudah dipublikasikan dan atau pernah dipakai untuk mendapatkan gelar Sarjana Teknik di lingkungan Universitas 17 Agustus 1945 Surabaya maupun di Perguruan Tinggi atau Instansi manapun, kecuali bagian yang sumber informasinya dicantumkan sebagaimana mestinya.
2. Tugas Akhir dengan judul diatas bukan merupakan plagiarisme, pencurian hasil karya milik orang lain, hasil kerja orang lain untuk kepentingan saya karena hubungan material maupun non – material, ataupun segala kemungkinan lain yang pada hakekatnya bukan merupakan karya tulis tugas akhir saya secara orisinal dan otentik.
3. Demi pengembangan ilmu pengetahuan, saya memberikan hak atas Tugas Akhir ini kepada Universitas 17 Agustus 1945 Surabaya, untuk menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.
4. Pernyataan ini saya buat dengan kesadaran sendiri dan tidak atas tekanan ataupun paksaan dari pihak maupun demi menegakan integritas akademik di institusi ini dan bila kemudian hari diduga kuat ada ketidaksesuaian antara fakta dengan kenyataan ini, saya bersedia diproses oleh tim Fakultas yang dibentuk untuk melakukan verifikasi, dengan sanksi terberat berupa pembatalan kelulusan/kesarjanaan.



Hendrian Revinielly Igu  
1462100101



UNIVERSITAS  
17 AGUSTUS 1945  
SURABAYA

BADAN PERPUSTAKAAN  
JL. SEMOLOWARU 45 SURABAYA  
TELP. 031 593 1800 (Ext. 311)  
e-mail : [perpus@untag-sby.ac.id](mailto:perpus@untag-sby.ac.id)

## LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Sebagai Civitas Akademik Universitas 17 Agustus 1945 Surabaya, saya yang bertanda tangan di bawah ini:

Nama : Hendrian Revinielly Igu  
NBI/ NPM : 1462100101  
Fakultas : Teknik  
Program Studi : Informatika  
Jenis Karya : Skripsi/ Tesis/ ~~Disertasi/ Laporan Penelitian/ Praktek\*~~

Demi perkembangan ilmu pengetahuan, saya menyetujui untuk memberikan kepada Badan Perpustakaan Universitas 17 Agustus 1945 Surabaya **Hak Bebas Royalti Noneksklusif (Nonexclusive Royalty-Free Right)**, atas karya saya yang berjudul:

**Deteksi Email Spam Menggunakan Algoritma *Naive Bayes* Berdasarkan Pola Kata dan Frasa**

Dengan **Hak Bebas Royalti Noneksklusif (Nonexclusive Royalty - Free Right)**, Badan Perpustakaan Universitas 17 Agustus 1945 Surabaya berhak menyimpan, mengalihkan media atau memformatkan, mengolah dalam bentuk pangkalan data (database), merawat, mempublikasikan karya ilmiah saya selama tetap tercantum

Dibuat di : Universitas 17 Agustus 1945 Surabaya  
Pada tanggal : 30 Juli 2025

Yang Menyatakan,

( Hendrian Revinielly Igu )

## KATA PENGANTAR

Puji dan syukur ke hadirat Tuhan Yang Maha Esa, atas rahmat, kekuatan, dan terang Roh Kudus yang diberikan selama proses penyusunan Tugas Akhir ini hingga dapat diselesaikan dengan baik. Tugas akhir yang berjudul “Deteksi Email Spam Menggunakan Algoritma Naïve Bayes Berdasarkan Pola Kata dan Frasa”. Tugas akhir ini diajukan untuk memenuhi salah satu syarat dalam memperoleh gelar Sarjana Komputer pada Program Studi Informatika, Fakultas Teknik, Universitas 17 Agustus 1945 Surabaya.

Penyusunan tugas akhir ini tentunya tidak terlepas dari bantuan berbagai pihak yang telah memberikan dukungan baik secara moril maupun materil. Oleh karena itu, dengan segala kerendahan hati, penulis ingin mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Bapak, Mama, dan keluarga tercinta, yang selalu memberikan doa, dukungan, dan kasih sayang yang tak ternilai serta menjadi sumber kekuatan dan semangat dalam menyelesaikan Tugas Akhir ini.
2. Bapak Supangat, M.Kom selaku Dosen Pembimbing, atas bimbingan, arahan, dan motivasi yang diberikan selama proses penyusunan tugas akhir ini.
3. Bapak Aidil Primasetya Armin S.T., MT, sebagai Kepala Program Studi Teknik Informatika Universitas 17 Agustus 1945 Surabaya..
4. Bapak Dr. Ir. Sajiyo, M.Kes., IPM selaku Dekan Fakultas Teknik Universitas 17 Agustus 1945 Surabaya.
5. Seluruh dosen dan staf pengajar Program Studi Informatika yang telah memberikan ilmu dan pengalaman selama masa perkuliahan.
6. Teman-teman seperjuangan Informatika angkatan 2021, yang telah menjadi bagian dari perjalanan ini, terima kasih atas kebersamaan, tawa, dan semangat yang saling menguatkan.
7. Gracell Yunikrista Hamin, terima kasih atas kasih yang tak terukur, ketulusan yang tak pernah surut, dan kesabaran yang mendampingi dalam tiap langkah perjalanan ini. Dalam lelah dan ragu, kehadiranmu adalah pelita yang menenangkan, tempat berpulang dari segala tekanan, serta pengingat bahwa setiap perjuangan tak harus ditempuh seorang diri. Terima kasih telah menjadi bahu untuk bersandar, telinga untuk berbagi, dan hati yang setia menemani hingga titik akhir pencapaian ini.
8. Kevin, Def, Caron, dan Krisen, terima kasih atas tawa yang tak pernah habis, obrolan panjang yang jadi pelipur lelah, dan semangat yang tak pernah gagal menguatkan. Kalian adalah bagian dari perjalanan ini yang tidak tergantikan tempat berbagi cerita, saling mengingatkan untuk terus maju, dan menjadi rumah di tengah segala tekanan. Terima kasih telah hadir bukan hanya sebagai teman, tetapi sebagai keluarga yang selalu ada dalam setiap langkah perjuangan ini.

9. Untuk keluarga kecil “Loyo Tapi Ayo”, terima kasih sudah jadi tempat tertawa, mengeluh, bercanda, bahkan saling menyemangati di tengah rasa malas dan capek yang kadang tak tertahankan. Meski sering loyo, tapi semangat kalian untuk terus ayo melanjutkan langkah meski terseok telah menjadi energi yang tak ternilai dalam perjalanan ini. Bersama kalian, setiap beban terasa lebih ringan, dan setiap hari terasa lebih berarti.
10. Untuk kelompok “kerja praktek sapi” yang luar biasa, yang telah menemani sejak semester pertama hingga hari ini, terima kasih atas kebersamaan yang konsisten, tawa di tengah tugas yang menumpuk, dan semangat saling dorong saat motivasi nyaris habis. Kalian bukan hanya rekan kerja, tapi sudah jadi bagian penting dari perjalanan panjang ini tempat saling belajar, bertumbuh, dan menua bersama di bawah tekanan deadline. Perjalanan ini tidak akan sama tanpamu, dan aku bersyukur pernah berjalan sejauh ini bersama kalian.
11. Sahabat-sahabat terbaik, yang tidak pernah lelah memberikan semangat, membantu dalam kesulitan, dan menjadi tempat berbagi di saat senang maupun sulit.
12. Diri sendiri, terima kasih atas keteguhan, ketabahan, dan keyakinan untuk terus melangkah meskipun dalam keterbatasan. Terima kasih telah percaya bahwa perjalanan ini akan sampai di akhir dengan hasil yang layak dibanggakan.
13. Untuk band kesayangan saya, yang lagunya selalu jadi pelarian saat dunia terasa terlalu berat, terima kasih telah menciptakan karya yang bukan hanya enak didengar, tapi juga menyembuhkan. Dalam setiap lirik dan nada, aku menemukan ketenangan, semangat, dan pengingat bahwa aku tidak sendiri. Musik kalian telah menemani malam-malam penuh overthinking, pagi-pagi penuh harapan, dan momen-momen penting dalam hidupku. Terima kasih telah menjadi suara dari perasaan yang tak selalu bisa kuungkapkan sendiri.
14. Dan seluruh pihak yang tidak dapat disebutkan satu per satu, yang telah membantu dan memberikan dukungan, baik secara langsung maupun tidak langsung. Segala bantuan, doa, dan perhatian yang diberikan sangat berarti bagi penulis.

Penulis menyadari bahwa tugas akhir ini masih memiliki kekurangan dan keterbatasan. Oleh karena itu, kritik dan saran yang membangun sangat diharapkan untuk penyempurnaan di masa mendatang. Akhir kata, semoga tugas akhir ini dapat memberikan manfaat bagi pembaca serta menjadi kontribusi nyata dalam pengembangan ilmu pengetahuan di bidang teknologi informasi, khususnya dalam keamanan email dan klasifikasi berbasis machine learning.

Surabaya, 25 Juni 2025

Penulis

## ABSTRAK

Nama : Hendrian Revinieldy Igu  
Program Studi : Teknik Informatika  
Judul : Deteksi Email Spam Menggunakan Algoritma Naive Bayes Berdasarkan Pola Kata dan Frasa

Perkembangan teknologi digital yang pesat telah meningkatkan penggunaan email sebagai media komunikasi utama, baik untuk kepentingan pribadi, bisnis, maupun institusi. Namun, seiring meningkatnya penggunaan email, jumlah email spam juga terus bertambah. Spam email sering kali berisi iklan tidak relevan, penipuan, hingga ancaman keamanan seperti phishing dan malware, yang dapat merugikan pengguna secara ekonomi maupun privasi. Oleh karena itu, diperlukan sebuah sistem yang mampu mendeteksi dan memfilter email spam secara efektif.

Penelitian ini bertujuan untuk mengembangkan sistem deteksi email spam berbasis web menggunakan algoritma Naïve Bayes yang dikenal efisien dan cepat dalam klasifikasi teks. Fokus utama dari penelitian ini adalah pada analisis pola kata dan frasa dalam email spam untuk meningkatkan akurasi klasifikasi. Dataset yang digunakan diambil dari sumber publik (Kaggle) dan telah melalui tahapan preprocessing seperti case folding, tokenization, stopword removal, dan stemming agar data siap digunakan dalam proses pembelajaran mesin. Proses ekstraksi fitur menggunakan metode TF-IDF untuk merepresentasikan bobot kata dalam dokumen. Sistem ini diuji dengan membandingkan hasil klasifikasi pada data uji dan menunjukkan bahwa model mampu mengidentifikasi email spam dengan akurasi lebih dari 90%. Hal ini membuktikan bahwa algoritma Naïve Bayes cukup andal dalam mengklasifikasikan email berdasarkan fitur tekstual, terutama dengan dukungan preprocessing yang tepat dan pemilihan fitur yang relevan. Implementasi sistem dalam bentuk aplikasi berbasis web juga memberikan kemudahan dalam pengujian dan interaksi pengguna, baik untuk klasifikasi otomatis maupun input manual.

Penelitian ini diharapkan dapat memberikan kontribusi nyata dalam pengembangan sistem keamanan informasi, khususnya dalam deteksi dini terhadap ancaman email spam di lingkungan digital Indonesia.

**Kata Kunci:** Email Spam, Naïve Bayes, Deteksi Spam, Klasifikasi Teks, Pola Kata, TF-IDF, Preprocessing

## ABSTRACT

Nama : Hendrian Revinielly Igu  
Major : Informatics Engineering  
Undergraduate : Spam Email Detection Using Naive Bayes Algorithm  
Thesis Title : Based on Word and Phrase Patterns

The rapid advancement of digital technology has significantly increased the use of email as a primary communication tool across personal, business, and institutional domains. However, this rise in email usage has also been accompanied by a growing volume of spam emails. These unsolicited messages often contain irrelevant advertisements, fraud attempts, and even dangerous content such as phishing links and malware attachments, posing risks to user security and privacy. Therefore, an efficient and reliable spam detection system is essential to mitigate these threats.

This research aims to develop a web-based spam email detection system using the Naïve Bayes algorithm, which is known for its simplicity, speed, and effectiveness in text classification tasks. The focus of this study lies in identifying and analyzing word and phrase patterns commonly found in spam emails to enhance classification accuracy. The dataset used is obtained from public sources (Kaggle) and has undergone several preprocessing stages including case folding, tokenization, stopword removal, and stemming to ensure data consistency and model readiness. TF-IDF is used for feature extraction to weigh the importance of terms within the email corpus.

Evaluation results show that the developed model achieved an accuracy of over 90% in detecting spam emails. This demonstrates that the Naïve Bayes algorithm is a reliable choice for text-based email classification, especially when supported by effective preprocessing and feature selection techniques. The implementation of the system as a web application provides users with a practical interface for both automated and manual spam classification.

This study is expected to contribute to the field of information security by offering an accessible and accurate spam detection solution, particularly relevant for addressing the increasing email threats in Indonesia's digital landscape.

**Keywords:** Spam Email, Naïve Bayes, Spam Detection, Text Classification, Word Patterns, TF-IDF, Preprocessing

## DAFTAR ISI

LEMBAR PENGESAHAN TUGAS AKHIR .....	i
PERNYATAAN KEASLIAN DAN PERSETUJUAN PUBLIKASI TUGAS AKHIR .....	iii
LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS.....	v
KATA PENGANTAR.....	vii
ABSTRAK .....	ix
ABSTRACT .....	xi
DAFTAR ISI.....	xiii
DAFTAR GAMBAR.....	xv
DAFTAR TABEL.....	xvii
<b>BAB 1 PENDAHULUAN.....</b>	<b>1</b>
1.1 Latar Belakang.....	1
1.2 Perumusan Masalah .....	2
1.3 Batasan Masalah .....	2
1.4 Tujuan Penelitian.....	3
1.5 Manfaat Penelitian.....	3
<b>BAB 2 KAJIAN PUSTAKA DAN DASAR TEORI .....</b>	<b>5</b>
2.1 Data Mining.....	5
2.2 Spam email .....	6
2.3 Klasifikasi .....	6
2.4 Preprocessing .....	7
2.5 Feature extraction.....	7
2.6 Klasifikasi Naïve Bayes dalam spam email .....	8
2.7 Naïve Bayes.....	8
2.8 Pola Kata dan Frasa dalam Klasifikasi Email .....	10
2.9 Confusion Matrix.....	11
2.10 Penelitian Sebelumnya.....	11
<b>BAB 3 METODE PENELITIAN .....</b>	<b>27</b>
3.1 Bahan dan Perangkat Penelitian.....	27
3.2 Obyek Penelitian .....	27
3.3 Tahapan Penelitian .....	27
3.4 Tabel Penjelasan Tahapan Penelitian .....	30
3.5 Skenario Pengujian.....	32
<b>BAB 4 HASIL DAN PEMBAHASAN .....</b>	<b>35</b>
4.1 Implementasi Sistem .....	35
4.1.1 Halaman Login.....	35
4.1.2 Halaman Dashboard.....	36
4.1.3 Halaman Dataset.....	36

4.1.4	Halaman Text Mining Naive Bayes .....	37
4.1.5	Halaman Inisialisasi Proses .....	37
4.1.6	Halaman Performa.....	38
4.1.7	Halaman Prediksi.....	40
4.1.8	Halaman TF-IDF.....	41
4.2	Implementasi Algoritma Naive Bayes .....	42
4.2.1	Case Folding .....	42
4.2.2	Tokenization .....	42
4.2.3	Stopword .....	43
4.2.4	Stemming .....	44
4.2.5	Pembobotan Kata.....	44
4.2.6	Klasifikasi Naive Bayes.....	46
4.3	Pengujian Sistem .....	49
4.3.1	Pengujian dan Hasil Klasifikasi .....	49
4.3.2	Pengujian Black Box Testing .....	50
<b>BAB 5 PENUTUP .....</b>		<b>58</b>
5.1	<b>KESIMPULAN.....</b>	<b>59</b>
5.2	<b>SARAN.....</b>	<b>59</b>
<b>DAFTAR PUSTAKA.....</b>		<b>61</b>

## DAFTAR GAMBAR

Gambar 3. 1 Tahapan Penelitian .....	28
Gambar 3. 2 Tahap Preprocessing Data .....	29
Gambar 4. 1 Halaman Login .....	35
Gambar 4. 2 Halaman Dashboard .....	36
Gambar 4. 3 Halaman Dataset.....	36
Gambar 4. 4 Halaman Add.....	37
Gambar 4. 5 Halaman Text Mining Naive Bayes .....	37
Gambar 4. 6 Halaman Inisialisasi Proses .....	38
Gambar 4. 7 Halman Performa .....	38
Gambar 4. 8 Data Training.....	39
Gambar 4. 9 Data Testing .....	39
Gambar 4. 10 Proses Testing.....	40
Gambar 4. 11 Hasil Akurasi .....	40
Gambar 4. 12 Halaman Prediksi .....	41
Gambar 4. 13 Halaman TF-IDF .....	41
Gambar 4. 14 Pengujian Verifikasi Username dan Password I.....	52
Gambar 4. 15 Pengujian Verifikasi Username dan Password II .....	53
Gambar 4. 16 Pengujian Verifikasi Username dan Password III.....	53
Gambar 4. 17 Pengujian Verifikasi Username dan Password IV.....	54
Gambar 4. 18 Pengujian Prediksi Naive Bayes I .....	54
Gambar 4. 19 Pengujian Prediksi Naive Bayes II.....	55
Gambar 4. 20 Pengujian Prediksi Naive Bayes III.....	55
Gambar 4. 21 Pengujian Dataset I.....	56
Gambar 4. 22 Pengujian Dataset I.....	56
Gambar 4. 23 Pengujian Non-Functional.....	57
Gambar 4. 24 Pengujian Non-Functional.....	57

## DAFTAR TABEL

Tabel 3. 1 Tabel Penjelasan Tahapan Penelitian .....	30
Tabel 3. 2 Skenario Pengujian.....	32
Tabel 3. 3 Skenario Pengujian Black Box Testing .....	33
Tabel 4. 1 Case Coding .....	42
Tabel 4. 2 Tokenization.....	42
Tabel 4. 3 Stopword .....	43
Tabel 4. 4 Stemming .....	44
Tabel 4. 5 Term Frequency (TF).....	45
Tabel 4. 6 Inverse Document Frequery (IDF).....	45
Tabel 4. 7 TF-IDF .....	45
Tabel 4. 8 Perhitungan Manual .....	46
Tabel 4. 9 TF-IDF Tiap Dokumen .....	47
Tabel 4. 10 Format Akhir Tabel.....	48
Tabel 4. 11 Pembagian Data .....	49
Tabel 4. 12 Data Traning dan Testing.....	49
Tabel 4. 13 Confusion Matrix Pada Data Testing .....	49
Tabel 4. 14 Perhitungan Evaluasi Kinerja (SPAM vs HAM) .....	50
Tabel 4. 15 Skenario Uji Coba Black Box Testing.....	50