

TUGAS AKHIR

**IMPLEMENTASI ALGORITMA GAUSSIAN NAIVE BAYES
UNTUK KLASIFIKASI SPAM *EMAIL***



YUGO FAHRENO

1910651022

PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS TEKNIK

UNIVERSITAS MUHAMMADIYAH JEMBER

2026

TUGAS AKHIR

IMPLEMENTASI ALGORITMA GAUSSIAN NAIVE BAYES UNTUK KLASIFIKASI SPAM *EMAIL*

Disusun Untuk Melengkapi Dan Memenuhi Syarat Kelulusan Guna Meraih Gelar

Sarjana Komputer

Teknik Informatika Universitas Muhammadiyah Jember



**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS MUHAMMADIYAH JEMBER**

2026

HALAMAN PERSETUJUAN SIDANG TUGAS AKHIR

**IMPLEMENTASI ALGORITMA GAUSSIAN NAIVE BAYES UNTUK
KLASIFIKASI SPAM *EMAIL***

Oleh:

Yugo Fahreno

1910651022

Telah disetujui bahwa Laporan Tugas Akhir ini untuk diajukan pada Sidang Tugas Akhir sebagai salah satu syarat kelulusan dan mendapatkan gelar Sarjana Komputer (S.Kom)

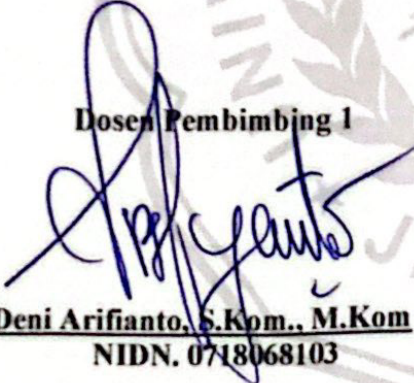
di

Universitas Muhammadiyah Jember

Disetujui Oleh:

Dosen Pembimbing 1

Dosen Pembimbing 2


Deni Arifianto, S.Kom., M.Kom
NIDN. 0718068103


Miftahur Rahman, S.Kom., M.Kom
NIDN. 0724039201

HALAMAN PENGESAHAN
IMPLEMENTASI ALGORITMA GAUSSIAN NAIVE BAYES UNTUK
KLASIFIKASI SPAM EMAIL

Oleh :
Yugo Fahreno
1910651022

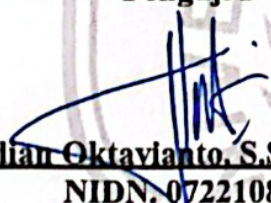
Telah mempertanggung jawabkan Laporan Tugas Akhir pada Sidang Tugas Akhir
tanggal 05 Bulan Juni Tahun 2026 sebagai salah satu syarat kelulusan dan
mendapatkan gelar Sarjana Komputer (S.Kom)

di
Universitas Muhammadiyah Jember


Disetujui Oleh,

Dosen Penguji :

Penguji I


Hardian Oktavianto, S.Si., M.Kom
NIDN. 0722108105

Penguji II


Daryanto, S.Kom., M.Kom
NIDN. 0730037703

Mengesahkan,


Dekan

Fakultas Teknik


Prof. Dr. Ir. Muhtar, S.T., M.T., IPM
* NIDN. 0010067301

Dosen Pembimbing

Pembimbing I


Deni Arifianto, S.Kom., M.Kom
NIDN. 0718068103


Pembimbing II


Miftahur Rahman, S.Kom., M.Kom
NIDN. 0724039201

Mengetahui,

Ketua Program Studi

Teknik Informatika


Rosita Yanuarti, S.Kom., M.Cs
NIDN. 0629018601

PERNYATAAN ORISINALITAS

Saya yang bertanda tangan di bawah ini:

Nama : Yugo Fahreno
NIM : 1910651022
Program Studi : Teknik Informatika
Judul Tugas Akhir : Implementasi Algoritma Gaussian Naive Bayes Untuk
Klasifikasi Spam Email

Dengan ini menyatakan bahwa karya ilmiah yang saya susun adalah murni hasil pemikiran, penelitian, dan penyusunan saya sendiri. Segala sumber informasi yang berasal atau dikutip dari karya orang lain, baik berupa publikasi, artikel, buku, maupun sumber lainnya, telah disebutkan dengan jelas di dalam daftar pustaka sesuai dengan kaidah penulisan ilmiah.

Saya menyadari bahwa dalam proses penyusunan Tugas Akhir ini, saya memanfaatkan bantuan teknologi kecerdasan buatan (AI) seperti *ChatGPT/Grammarly/Alat AI lain* hanya sebatas untuk:

1. Membantu dalam penyusunan tata bahasa, parafrasa, atau perbaikan redaksi.
2. Memberikan inspirasi ide awal atau kerangka berpikir, yang selanjutnya saya kembangkan secara mandiri.
3. Membantu dalam pengecekan konsistensi format, ejaan, dan tata tulis.

Saya menegaskan bahwa tidak ada bagian dari karya ilmiah ini yang seluruhnya dibuat oleh AI tanpa keterlibatan pemikiran kritis saya sendiri. Tanggung jawab penuh atas isi, keaslian, dan kebenaran karya ilmiah ini ada pada diri saya.

Apabila di kemudian hari terbukti bahwa pernyataan ini tidak benar, maka saya bersedia menerima segala konsekuensi sesuai dengan peraturan yang berlaku di Universitas Muhammadiyah Jember.

Jember, 2 Juli 2026
Yang membuat pernyataan,


Yugo Fahreno
1910651022

MOTTO

“Kesuksesan adalah sebagaimana kita memposisikan diri”

(Yugo Fahreno)



KATA PENGANTAR

Segala puji bagi Allah SWT Yang Maha Pengasih lagi Maha Penyayang. Atas limpahan rahmat, karunia, dan ridha-Nya, penulis dapat menyelesaikan skripsi yang berjudul “Implementasi Algoritma *Gaussian Naive Bayes* Untuk Klasifikasi *Spam* Email”. Penyusunan skripsi ini dilakukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer (S.Kom) pada Program Studi Teknik Informatika Universitas Muhammadiyah Jember.

Penyelesaian laporan Tugas Akhir ini tentunya tidak terlepas dari bantuan, dukungan, arahan, serta doa dari berbagai pihak. Oleh sebab itu, dengan segala kerendahan hati penulis ingin menyampaikan ucapan terima kasih kepada:

1. Allah SWT yang senantiasa melimpahkan rahmat, kesehatan, kemudahan, dan petunjuk-Nya sehingga penulis mampu menyelesaikan Tugas Akhir ini dengan baik hingga memperoleh gelar Sarjana Komputer.
2. Ibu Rosita Yanuarti, S.Kom., M.Cs. selaku Ketua Program Studi Teknik Informatika Universitas Muhammadiyah Jember.
3. Prof. Dr. Ir. Muhtar, S.T., M.T., IPM. selaku Dekan Fakultas Teknik Universitas Muhammadiyah Jember.
4. Bapak Deni Arifianto, S.Kom., M.Kom. selaku Dosen Pembimbing I serta Bapak Miftahur Rahman, S.Kom., M.Kom. selaku Dosen Pembimbing II yang telah memberikan bimbingan, arahan, masukan, serta meluangkan waktu dan tenaga selama proses penyusunan Tugas Akhir ini.
5. Bapak Hardian Oktavianto, S.Si., M.Kom. selaku Dosen Penguji I dan Bapak Daryanto, S.Kom., M.Kom. selaku Dosen Penguji II yang telah memberikan kritik, saran, dan evaluasi demi penyempurnaan Tugas Akhir ini.
6. Seluruh dosen Program Studi Teknik Informatika Universitas Muhammadiyah Jember yang telah memberikan ilmu pengetahuan, pengalaman, dan pembelajaran selama penulis menjalani masa studi.
7. Kedua orang tua, adik, dan seluruh keluarga yang selalu memberikan doa, dukungan, perhatian, motivasi, serta kesabaran kepada penulis selama menempuh pendidikan di Teknik Informatika Universitas Muhammadiyah Jember.

8. Ucapan terima kasih yang tulus penulis sampaikan kepada Anindha Sekar Nawang Wulan yang selalu menemani dan memberikan dukungan selama proses penyusunan Tugas Akhir ini. Terima kasih atas doa, perhatian, semangat, dan kesabaran yang terus diberikan kepada penulis dalam berbagai kondisi. Kehadirannya menjadi salah satu sumber kekuatan bagi penulis untuk tetap bertahan dan menyelesaikan setiap proses dengan penuh keyakinan.
9. Terima kasih kepada Muhammad Andi F., Iqbal Rafian Lubis, dan Loren Austin M. selaku teman seperjuangan sejak awal perkuliahan yang senantiasa memberikan motivasi dan semangat kepada penulis agar tidak menyerah dalam menyelesaikan Tugas Akhir ini.
10. Terima kasih kepada teman-teman UKM Musik Gudang Production atas kebersamaan, dukungan, serta pengalaman berharga yang diberikan kepada penulis selama masa perkuliahan hingga penyusunan Tugas Akhir ini selesai.
11. Semua pihak yang tidak dapat disebutkan satu per satu, penulis mengucapkan terima kasih atas segala doa, bantuan, dan dukungan yang telah diberikan selama proses penyelesaian Tugas Akhir ini.

Sebagai penutup, penulis menyampaikan rasa terima kasih yang sebesar-besarnya kepada seluruh pihak yang telah membantu. Penulis menyadari bahwa karya ilmiah ini masih memiliki banyak kekurangan dan jauh dari kata sempurna. Oleh karena itu, kritik dan saran yang membangun sangat diharapkan demi perbaikan dan penyempurnaan penulisan di masa mendatang. Penulis berharap semoga karya sederhana ini dapat memberikan manfaat bagi para pembaca dan pihak-pihak yang membutuhkan.

Amin ya Rabbal 'alamin.

DAFTAR ISI

| | |
|---|-----------|
| HALAMAN SAMPUL..... | i |
| HALAMAN JUDUL | i |
| HALAMAN PERSETUJUAN SIDANG TUGAS AKHIR | ii |
| HALAMAN PENGESAHAN | iii |
| PERNYATAAN ORISINALITAS..... | iv |
| MOTTO..... | v |
| KATA PENGANTAR..... | vi |
| DAFTAR ISI..... | viii |
| DAFTAR GAMBAR..... | x |
| DAFTAR TABEL..... | xi |
| DAFTAR LAMPIRAN | xii |
| ABSTRAK | xiii |
| <i>ABSTRACT</i> | xiv |
| BAB I. PENDAHULUAN..... | 1 |
| 1.1 Latar Belakang..... | 1 |
| 1.2 Rumusan Masalah | 3 |
| 1.3 Tujuan Penelitian..... | 4 |
| 1.4 Manfaat Penelitian | 4 |
| 1.5 Batasan Masalah..... | 5 |
| BAB II. LANDASAN TEORI..... | 6 |
| 2.1 <i>Text Mining</i> | 6 |
| 2.2 <i>Random Over Sampling</i> | 6 |
| 2.3 <i>Preprocessing</i> | 7 |
| 2.4 TF-IDF | 9 |
| 2.5 Algoritma <i>Gaussian Naive Bayes</i> | 11 |
| 2.6 <i>K-Fold Cross-Validation</i> | 13 |
| 2.7 <i>Confusion Matrix</i> | 13 |
| 2.8 Google Colab..... | 15 |
| 2.9 Penelitian Terdahulu | 15 |
| BAB III. METODE PENELITIAN | 20 |
| 3.1 Tahapan Penelitian..... | 20 |

| | | |
|---|---|-----------|
| 3.2 | Pengumpulan Data | 21 |
| 3.3 | <i>Random Oversampling (Data Imbalance)</i> | 21 |
| 3.4 | Label Data..... | 21 |
| 3.5 | <i>Preprocessing Data</i> | 22 |
| 3.6 | TF-IDF | 27 |
| 3.7 | <i>K-Fold Cross Validation</i> | 31 |
| 3.8 | Implementasi <i>Algoritma Gaussian Naive Bayes</i> | 33 |
| 3.9 | <i>Confusion Matrix</i> | 36 |
| BAB IV HASIL DAN PEMBAHASAN..... | | 39 |
| 4.1 | Pengumpulan Data | 39 |
| 4.2 | <i>Random Over Sampling (Data Imbalance)</i> | 40 |
| 4.3 | <i>Preprocessing Data</i> | 42 |
| 4.4 | TF-IDF | 50 |
| 4.5 | Pembagian Data <i>K-Fold Cross-Validation</i> | 51 |
| 4.6 | Hasil Terbaik dari Seluruh Pengujian <i>K-Fold Cross Validation</i> | 59 |
| BAB V KESIMPULAN..... | | 61 |
| 5.1 | Kesimpulan..... | 61 |
| 5.2 | Saran..... | 62 |
| DAFTAR PUSTAKA | | 64 |
| LAMPIRAN..... | | 68 |

DAFTAR GAMBAR

| | |
|--|----|
| Gambar 3. 1 Tahapan Penelitian | 20 |
| Gambar 3. 2 Preprocessing Data | 22 |
| Gambar 4. 1 Kaggle Dataset | 39 |
| Gambar 4. 2 Data Excel | 40 |
| Gambar 4. 3 Grafik Jumlah Spam dan Non spam..... | 40 |
| Gambar 4. 4 Data Hasil Random Over Sampling..... | 41 |
| Gambar 4. 5 Grafik Setelah Proses Random Over Sampling | 41 |
| Gambar 4. 6 Data Uji 1 Pada Uji K-Fold 2..... | 51 |
| Gambar 4. 7 Data Uji 2 Pada Uji K-Fold 2..... | 52 |
| Gambar 4. 8 Data Uji 1 Pada Uji K-Fold 4..... | 53 |
| Gambar 4. 9 Data Uji 2 Pada Uji K-Fold 4..... | 53 |
| Gambar 4. 10 Data Uji 3 Pada Uji K-Fold 4..... | 53 |
| Gambar 4. 11 Data Uji 4 Pada Uji K-Fold 4..... | 54 |
| Gambar 4. 12 Data Uji 1 Pada Uji K-Fold 8..... | 55 |
| Gambar 4. 13 Data Uji 2 Pada Uji K-Fold 8..... | 55 |
| Gambar 4. 14 Data Uji 3 Pada Uji K-Fold 8..... | 56 |
| Gambar 4. 15 Data Uji 4 Pada Uji K-Fold 8..... | 56 |
| Gambar 4. 16 Data Uji 5 Pada Uji K-Fold 8..... | 56 |
| Gambar 4. 17 Data Uji 6 Pada Uji K-Fold 8..... | 57 |
| Gambar 4. 18 Data Uji 7 Pada Uji K-Fold 8..... | 57 |
| Gambar 4. 19 Data Uji 8 Pada Uji K-Fold 8..... | 57 |

DAFTAR TABEL

| | |
|---|----|
| Tabel 2. 1 <i>Confusion matrix</i> 2x2..... | 13 |
| Tabel 2. 2 Jadwal Penelitian..... | 15 |
| Tabel 3. 1 Data Email dan label | 22 |
| Tabel 3. 2 Hasil <i>Cleaning</i> | 23 |
| Tabel 3. 3 Hasil <i>Case Folding</i> | 24 |
| Tabel 3. 4 Hasil <i>Tokenizing</i> | 25 |
| Tabel 3. 5 Hasil <i>Stopword</i> | 26 |
| Tabel 3. 6 Hasil <i>Stemming</i> | 27 |
| Tabel 3. 7 TF (<i>Spam</i>) | 28 |
| Tabel 3. 8 TF (<i>Non Spam</i>)..... | 28 |
| Tabel 3. 9 Hasil IDF..... | 29 |
| Tabel 3. 10 Hasil Proses perhitungan TF-IDF | 30 |
| Tabel 3. 11 Tabel Representasi Data..... | 34 |
| Tabel 3. 12 Tabel Representasi Data..... | 36 |
| Tabel 3. 13 Tabel Representasi Data..... | 37 |
| Tabel 3. 14 Tabel Ringkasan..... | 38 |
| Tabel 4. 1 Hasil <i>Cleaning</i> | 42 |
| Tabel 4. 2 Hasil <i>Case Folding</i> | 43 |
| Tabel 4. 3 Hasil <i>Tokenizing</i> | 45 |
| Tabel 4. 4 Hasil <i>Stopword</i> | 47 |
| Tabel 4. 5 Hasil <i>Stemming</i> | 48 |
| Tabel 4. 6 Potongan Hasil Proses TF-IDF | 50 |
| Tabel 4. 7 <i>Confusion Matrix</i> 2-Fold Cross Validation | 52 |
| Tabel 4. 8 <i>Confusion Matrix</i> 4-Fold Cross Validation | 55 |
| Tabel 4. 9 <i>Confusion Matrix</i> 8-Fold Cross Validation | 59 |
| Tabel 4. 10 Hasil Terbaik dari Seluruh Pengujian K-Fold Cross Validation..... | 59 |

DAFTAR LAMPIRAN

| | |
|--|----|
| Lampiran 1 Skript Coding..... | 68 |
| Lampiran 2 Tabel Hasil <i>K-Fold</i> Cross Validation..... | 76 |
| Lampiran 3 QR Dataset..... | 81 |

