

**MENGOPTIMALKAN DETEKSI SPAM EMAIL MELALUI PENANGANAN
KETIDAKSEIMBANGAN KELAS DENGAN BOBOT KELAS DAN
HYPERPARAMETER MENGGUNAKAN GRIDSEARCHCV**

LAPORAN NON-REGULER SCIENTIST

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Teknik Komputer



disusun oleh

MUHAMMAD RIDHO NURSYAM

22.83.0788

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

**MENGOPTIMALKAN DETEKSI SPAM EMAIL MELALUI PENANGANAN
KETIDAKSEIMBANGAN KELAS DENGAN BOBOT KELAS DAN
HYPERPARAMETER MENGGUNAKAN GRIDSEARCHCV**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Teknik Komputer



disusun oleh

MUHAMMAD RIDHO NURSYAM

22.83.0788

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

HALAMAN PERSETUJUAN

SKRIPSI

MENGOPTIMALKAN DETEKSI SPAM EMAIL MELALUI PENANGANAN

KETIDAKSEIMBANGAN KELAS DENGAN BOBOT KELAS DAN

HYPERPARAMETER MENGGUNAKAN GRIDSEARCHCV

yang disusun dan diajukan oleh

Muhammad Ridho Nursyam

22.83.0788

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 24 Februari 2026

Dosen Pembimbing,



Muhammad Kopravi, S.Kom., M.Eng.
NIK. 190302454

HALAMAN PENGESAHAN
SKRIPSI
MENGOPTIMALKAN DETEKSI SPAM EMAIL MELALUI PENANGANAN
KETIDAKSEIMBANGAN KELAS DENGAN BOBOT KELAS DAN
HYPERPARAMETER MENGGUNAKAN GRIDSEARCHCV

yang disusun dan diajukan oleh

Muhammad Ridha Nursyam

22.83.0788

Telah dipertahankan di depan Dewan Penguji
pada tanggal 24 Februari 2026

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Melwin Syafrizal, S.Kom., M.Eng., Ph.D.
NIK. 190302105

Jeki Kuswanto, S.Kom., M.Kom.
NIK. 190302456

Muhammad Koprawi, S.Kom., M.Eng.
NIK. 190302454

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 24 Februari 2026

DEKAN FAKULTAS ILMU KOMPUTER



Prof. Dr. Kusriani, M.Kom.
NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : **Muhammad Ridho Nursyam**
NIM : **22.83.0788**

Menyatakan bahwa Skripsi dengan judul berikut:

**MENGOPTIMALKAN DETEKSI SPAM EMAIL MELALUI
PENANGANAN KETIDAKSEIMBANGAN KELAS DENGAN BOBOT
KELAS DAN HYPERPARAMETER MENGGUNAKAN GRIDSEARCHCV**

Dosen Pembimbing : **Muhammad Kopravi, S.Kom., M.Eng.**

1. Karya tulis ini adalah benar-benar **ASLI** dan **BELUM PERNAH** diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan **gagasan, rumusan dan penelitian SAYA** sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung-jawab **SAYA**, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini **SAYA** buat dengan sesungguhnya, apabila di kemudian hari terdapat **penyimpangan dan ketidakbenaran** dalam pernyataan ini, maka **SAYA** bersedia menerima **SANKSI AKADEMIK** dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 24 Februari 2026

Yang Menyatakan,



Muhammad Ridho Nursyam

HALAMAN PERSEMBAHAN

Dengan penuh rasa syukur ke hadirat Allah Subhanahu wa Ta'ala atas limpahan rahmat, karunia, serta kekuatan yang telah diberikan selama proses penulisan skripsi ini, izinkan penulis mempersembahkan karya sederhana ini kepada:

- Kedua orang tua tercinta, Ayah dan Ibu, yang selalu menjadi alasan penulis untuk terus bertahan dan berjuang. Terima kasih atas kasih sayang tanpa batas, doa yang tak pernah terputus, kesabaran yang luar biasa, serta dukungan moral maupun materiil yang mengiringi setiap langkah penulis. Kepercayaan dan keikhlasan kalian menjadi sumber kekuatan terbesar dalam menyelesaikan perjalanan akademik ini.
- Diri saya sendiri, yang telah berani memulai kembali setelah jeda panjang, memilih bangkit ketika banyak keraguan datang, dan tetap melangkah meskipun jalan terasa tidak mudah. Terima kasih telah bertahan, belajar dari pengalaman, dan tidak menyerah pada keadaan.
- Geng Pergerakan Madang Iqbal, Aji, King Ahsari, Bintang, dan Fayi yang telah menjadi teman seperjalanan, tempat berbagi cerita, tawa, keluh, serta semangat. Terima kasih atas kebersamaan, dukungan, dan energi positif yang kalian berikan selama masa perkuliahan hingga proses penyusunan skripsi ini.
- Seluruh pihak yang secara langsung maupun tidak langsung telah membantu penulis dalam menyelesaikan studi dan penulisan skripsi ini, baik melalui bimbingan, motivasi, maupun dukungan lainnya.

Skripsi ini adalah wujud kecil dari perjuangan panjang yang penuh proses, doa, dan pengorbanan. Semoga karya ini dapat menjadi bentuk balas budi yang membawa manfaat dan menjadi pijakan awal untuk langkah selanjutnya.

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa, karena atas berkat rahmat dan karunia-Nya, penulis dapat menyelesaikan skripsi ini dengan judul “Mengoptimalkan Deteksi Spam Email melalui Penanganan Ketidakseimbangan Kelas dengan Bobot Kelas dan Hyperparameter Menggunakan GridSearchCV” sebagai salah satu syarat untuk memperoleh gelar Sarjana pada Program Studi Teknik Komputer di Universitas Amikom Yogyakarta.

Penulisan skripsi ini tidak lepas dari bimbingan, dukungan, dan bantuan berbagai pihak. Oleh karena itu, pada kesempatan ini penulis ingin mengucapkan terima kasih yang sebesar-besarnya kepada:

- Orang tua tercinta yang selalu memberikan doa, dukungan, dan kasih sayang yang tiada henti, tanpa mereka penulis tidak akan bisa menyelesaikan studi ini.
- Dosen pembimbing yang telah memberikan arahan, bimbingan, serta motivasi yang luar biasa selama proses penyusunan skripsi ini.
- Seluruh dosen dan staf akademik di lingkungan Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta, atas ilmu dan fasilitas yang telah diberikan.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna, baik dari segi isi maupun penyusunannya. Oleh karena itu, penulis mengharapkan kritik dan saran yang konstruktif demi perbaikan dan pengembangan pengetahuan di masa depan. Akhir kata, penulis berharap semoga skripsi ini dapat bermanfaat bagi para pembaca, terutama dalam pengembangan ilmu pengetahuan dan praktis di bidang cyber security.

Yogyakarta, 15 Desember 2025

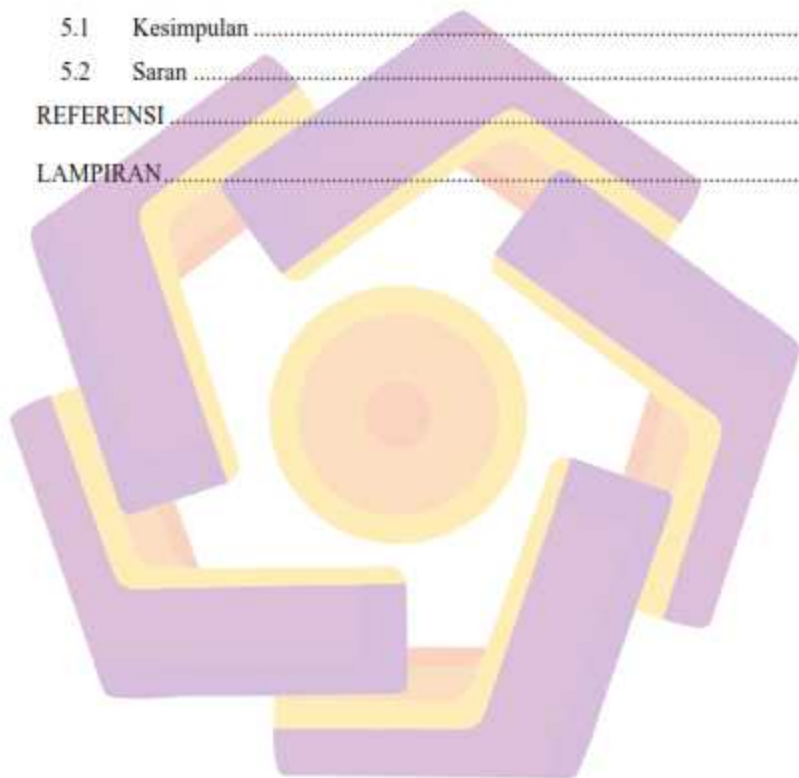
Penulis,

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI.....	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR.....	vi
DAFTAR ISI.....	vii
DAFTAR TABEL.....	x
DAFTAR GAMBAR.....	xi
DAFTAR LAMPIRAN.....	xii
INTISARI.....	xiii
<i>ABSTRACT</i>	xiv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	2
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian.....	3
1.5 Manfaat Penelitian.....	4
1.6 Sistematika Penulisan.....	5
BAB II TINJAUAN PUSTAKA.....	7
2.1 Studi Literatur.....	7
2.2 Dasar Teori.....	12
2.2.1 Supervised Learning.....	12

2.2.2	Machine Learning	12
2.2.3	Email	13
2.2.4	Support Vector Machine	14
2.2.5	XGBoost	15
2.2.6	Multinomial Naïve Bayes	16
2.2.7	Random Forest	16
2.2.8	Handling Class Imbalance with Class Weights	17
2.2.9	Evaluation Matrix	18
BAB III METODE PENELITIAN		21
3.1	Objek Penelitian	21
3.2	Alur Penelitian	21
3.2.1	Dataset	21
3.2.2	Data Pre-Processing	22
3.2.3	Data Split	25
3.2.4	Class Imbalance With Class Weight	26
3.2.5	Application Of Algorithm	26
3.2.6	GridSearchCV	28
3.2.7	Evaluasi Model	28
3.3	Alat dan Bahan	29
BAB IV HASIL DAN PEMBAHASAN		31
4.1	Distribusi Dataset	31
4.2	Praprocessing	32
4.3	Clean Dataset	33
4.4	Performance Sebelum Tuning of Machine Learning Algorithm	34

4.5	Performance Setelah Tuning of Machine Learning Algorithm	39
4.6	Perbandingan Akurasi Sebelum dan Setelah Penyetelan pada Model ...	43
4.7	Perbandingan Metode dan Hasil Terbaik Studi Ini dengan Penelitian Lain	46
BAB V PENUTUP		48
5.1	Kesimpulan	48
5.2	Saran	48
REFERENSI		50
LAMPIRAN		53



DAFTAR TABEL

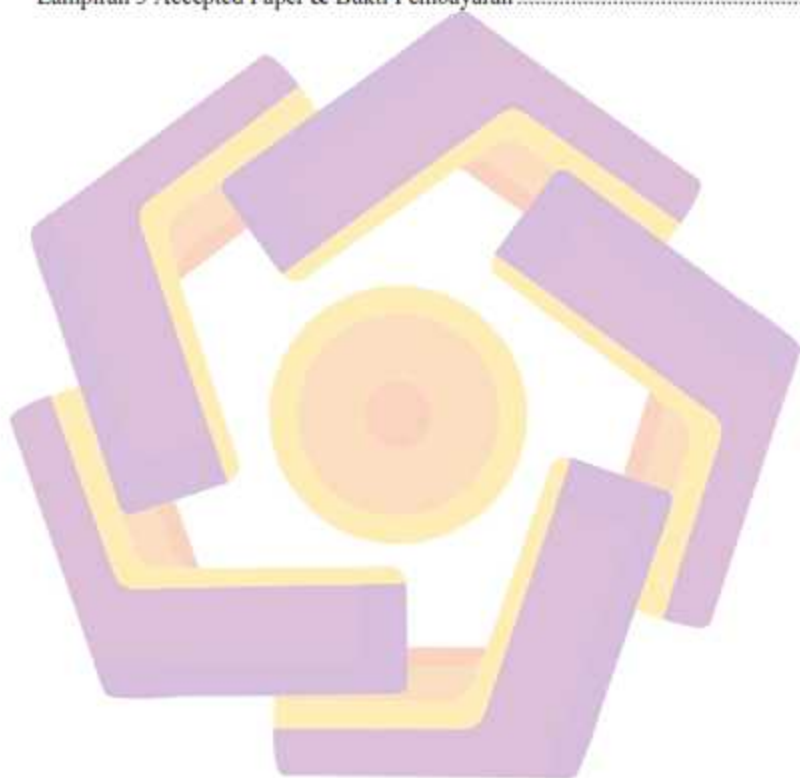
Table 2.1 Keaslian Penelitian	10
Table 3.1 Distribusi Kumpulan Data	23
Table 4.1 Dataset Sebelum Praprocessing	34
Table 4.2 Dataset Setelah Praprocessing	35
Table 4.3 Perbandingan Kinerja Model Klasifikasi Sebelum Hyperparameter GridSearchCV	45
Table 4.4 Perbandingan Kinerja Model Klasifikasi Setelah Hyperparameter GridSearchCV	46
Table 4.5 Perbandingan Akurasi Model Deteksi Spam dengan Metode dan Algoritma yang Berbeda	48

DAFTAR GAMBAR

Gambar 3.1 Alur Penelitian	10
Gambar 4.1 Perbandingan Distribusi Kelas	24
Gambar 4.2 Model Matrix Confusion Sebelum Penyetelan GridSearchCV	36
Gambar 4.3 Model ROC Sebelum Penyetelan GridSearchCV	37
Gambar 4.4 Model K-Fold Sebelum Penyetelan GridSearchCV	38
Gambar 4.5 Model Kinerja Sebelum Penyetelan GridSearchCV	39
Gambar 4.6 Model Matrix Confusion Setelah Penyetelan GridSearchCV	40
Gambar 4.7 Model ROC Setelah Penyetelan GridSearchCV	42
Gambar 4.8 Model K-Fold Setelah Penyetelan GridSearchCV	43
Gambar 4.9 Model Kinerja Setelah Penyetelan GridSearchCV	44

DAFTAR LAMPIRAN

Lampiran 1 Letter Of Acceptance (LOA).....	54
Lampiran 2 Lembar Review	55
Lampiran 3 Accepted Paper & Bukti Pembayaran	55



INTISARI

Spam email merupakan masalah besar dalam komunikasi digital yang dapat mengganggu produktivitas, membebani sumber daya jaringan, dan menimbulkan ancaman keamanan. Penelitian ini berfokus pada pengoptimalan deteksi email spam menggunakan pendekatan pembelajaran mesin dengan mengatasi ketidakseimbangan kelas melalui pembobotan kelas dan penyetelan hyperparameter menggunakan GridSearchCV. Untuk meningkatkan akurasi dan sensitivitas model, kombinasi dataset yang beragam diterapkan untuk menyediakan cakupan data pelatihan yang lebih luas. Model yang digunakan dalam penelitian ini meliputi Support Vector Machine (SVM), Random Forest, Multinomial Naive Bayes (MNB), dan XGBoost. Evaluasi dilakukan berdasarkan metrik seperti akurasi, presisi, recall, dan F1-score, sebelum dan sesudah penyetelan hyperparameter. Hasil eksperimen menunjukkan bahwa SVM menghasilkan akurasi tertinggi setelah penyetelan, mencapai 97,10%, dibandingkan dengan 96,73% sebelum penyetelan hyperparameter. Selain itu, Random Forest, MNB, dan XGBoost juga menunjukkan peningkatan yang signifikan, dengan masing-masing model mencapai kinerja yang lebih baik setelah penyetelan. Secara keseluruhan, penelitian ini menunjukkan bahwa penggabungan dataset dan penyesuaian bobot kelas dapat secara signifikan meningkatkan kemampuan model dalam mendeteksi spam, serta memberikan dasar untuk mengimplementasikan model tersebut dalam sistem deteksi spam email yang lebih efektif.

Kata kunci: Deteksi Spam, Pembelajaran Mesin, Ketidakseimbangan Kelas, GridSearchCV, Spam Email.

ABSTRACT

Email spam is a major problem in digital communication that can disrupt productivity, burden network resources, and pose a security threat. This research focuses on optimizing spam email detection using a machine learning approach by addressing class imbalance through class weighting and hyperparameter tuning using GridSearchCV. To improve model accuracy and sensitivity, a combination of diverse datasets is applied to provide a wider scope of training data. The models used in this study include Support Vector Machine (SVM), Random Forest, Multinomial Naive Bayes (MNB), and XGBoost. Evaluation is carried out based on metrics such as accuracy, precision, recall, and F1-score, before and after hyperparameter tuning. The experimental results show that SVM produces the highest accuracy after tuning, reaching 97.10%, compared to 96.73% before hyperparameter tuning. In addition, Random Forest, MNB, and XGBoost also show significant improvements, with each model achieving better performance after tuning. Overall, this study shows that dataset merging and class weight adjustment can significantly improve the model's ability to detect spam, as well as provide a basis for implementing the model in a more effective email spam detection system.

Keyword: *Spam Detection, Machine Learning, Class Imbalance, GridSearchCV, Email Spam.*