

BAB I PENDAHULUAN

1.1 Latar Belakang

Dalam era Komunikasi digital telah menjadi bagian integral dalam kehidupan sehari-hari, terutama melalui penggunaan email. Email menawarkan cara yang cepat dan efisien untuk bertukar informasi, baik untuk keperluan pribadi maupun bisnis. Dalam berbagai sektor seperti pendidikan, bisnis, dan pemerintahan, email digunakan sebagai alat utama untuk komunikasi. Kemudahan dan aksesibilitas yang ditawarkan oleh email membuatnya menjadi pilihan populer di berbagai belahan dunia.[1]

Namun, seiring dengan peningkatan penggunaan email, masalah spam mulai muncul. Spam email adalah pesan yang tidak diinginkan yang sering kali berisi konten iklan atau upaya penipuan. Spam tidak hanya mengganggu, tetapi juga dapat mengakibatkan risiko keamanan, seperti pencurian data atau penyebaran *malware*. Hal ini membuat deteksi spam menjadi sangat penting, khususnya untuk melindungi pengguna dari ancaman yang semakin canggih.[2]

Deteksi spam secara manual tidak mungkin dilakukan mengingat besarnya volume email yang dikirim setiap hari. Untuk mengatasi ini, banyak penelitian dilakukan dengan mengadopsi metode *machine learning*, seperti algoritma *Support Vector Machine* (SVM), untuk membantu dalam mengklasifikasikan email sebagai spam atau non-spam secara otomatis dan akurat.[3] Algoritma ini terbukti efektif dalam mengidentifikasi pola-pola yang dapat menunjukkan karakteristik spam email.

SVM merupakan salah satu algoritma yang memiliki kemampuan tinggi dalam klasifikasi data, terutama pada *dataset* yang kompleks. Namun, salah satu tantangan yang dihadapi adalah ketidakseimbangan kelas antara spam dan non-spam. Untuk itu, metode SMOTE digunakan untuk menyeimbangkan *dataset*

sehingga meningkatkan akurasi prediksi. Kombinasi kedua metode ini dapat memberikan solusi yang lebih andal dalam mendeteksi spam email.[4]

Penelitian oleh Sofyan, Rahaningsih, dan Dana (2024) [2] mengeksplorasi penerapan algoritma *Support Vector Machine* (SVM) untuk mendeteksi SMS spam berbahasa Indonesia. Hasilnya menunjukkan SVM efektif dalam mengklasifikasikan pesan spam dengan akurasi tinggi setelah proses *praproses* data. Namun, tantangan utama yang dihadapi adalah ketidakseimbangan kelas dalam *dataset*. Peneliti menyarankan penggunaan teknik SMOTE untuk meningkatkan distribusi kelas minoritas. Studi ini memberikan wawasan penting dalam pengembangan sistem deteksi spam di platform komunikasi digital. Hasilnya relevan untuk pengembangan lebih lanjut dalam deteksi spam berbasis bahasa Indonesia.

1.2 Rumusan Masalah

1. Bagaimana cara menangani ketidakseimbangan kelas pada *dataset* yang mengandung lebih banyak data non-spam daripada spam?
2. Bagaimana mengoptimalkan *hyperparameter* model SVM untuk meningkatkan akurasi prediksi?

1.3 Batasan Masalah

1. Penelitian ini menggunakan *dataset* teks email yang telah melalui *preprocessing* sederhana dan penanganan ketidakseimbangan data melalui teknik SMOTE untuk mengklasifikasi email spam menggunakan algoritma *Support Vector Machine*.
2. Model yang dihasilkan dievaluasi menggunakan metrik akurasi, *confusion matrix*, dan *classification report*, dengan batasan pada teknik *preprocessing*, penanganan ketidakseimbangan data, model klasifikasi, dan metrik evaluasi yang digunakan.
3. Penelitian ini dibatasi pada pengembangan sistem deteksi spam yang berfokus pada berbagai jenis email spam, seperti promosi dan iklan tidak diinginkan, penipuan (*phishing* dan *scam*).
4. *Dataset* yang di gunakan berupa data yang berbahasa Indonesia.

1.4 Tujuan Penelitian

1. Tujuan Penelitian menerapkan teknik SMOTE untuk mengatasi masalah ketidakseimbangan kelas pada *dataset* pesan email.
2. Penelitian ini juga bertujuan untuk membangun model klasifikasi berbasis SVM untuk mendeteksi email spam dengan akurasi tinggi.
3. Mengoptimalkan performa model SVM melalui *tuning hyperparameter* menggunakan *GridSearchCV*.

1.5 Manfaat Penelitian

1. Penelitian ini memberikan kontribusi dalam pengembangan teori tentang sistem otomatis untuk klasifikasi email spam yang lebih efektif, khususnya dengan penggunaan algoritma *Support Vector Machine* (SVM) dan teknik penanganan ketidakseimbangan data seperti SMOTE.
2. Penelitian ini menyediakan pendekatan untuk mengatasi ketidakseimbangan data dalam klasifikasi email yang dapat menjadi referensi bagi peneliti lain dalam penelitian klasifikasi pada domain berbeda, memperkaya pengetahuan tentang strategi penanganan data tidak seimbang.
3. Penelitian ini menekankan pentingnya proses *tuning hyperparameter* dalam meningkatkan performa model klasifikasi, sehingga dapat menjadi panduan bagi peneliti lain yang ingin mengoptimalkan algoritma SVM atau algoritma lainnya dalam tugas klasifikasi.

1.6 Sistematika Penulisan

BAB I PENDAHULUAN

Berisi latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan

BAB II TINJAUAN PUSTAKA

Bab ini menjelaskan teori-teori yang relevan dengan penelitian, termasuk algoritma *Support Vector Machine* (SVM), teknik *Synthetic Minority Over-sampling Technique* (SMOTE), serta kajian literatur terkait deteksi spam

email..

BAB III METODE PENELITIAN

Bab ini berisi menguraikan metode penelitian yang digunakan, meliputi, sumber data, proses pengumpulan data, teknik analisis data, desain penelitian, serta langkah-langkah penerapan SVM dan SMOTE untuk pengembangan sistem deteksi spam email.

BAB IV HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil pengolahan data dan implementasi algoritma, diikuti dengan analisis hasil berdasarkan akurasi, presisi, *recall*, dan F1-score, serta pembahasan terhadap kelebihan dan kekurangan metode yang digunakan.

BAB V PENUTUP

Bab ini berisi kesimpulan dari hasil penelitian yang menjawab rumusan masalah, serta menyampaikan saran untuk pengembangan penelitian lebih lanjut atau aplikasi praktis sistem deteksi spam email.