BAB I
PENDAHULUAN

1.1 Latar Belakang

Email telah menjadi alat komunikasi kunci dalam era digital saat ini, digunakan oleh individu dan organisasi untuk berbagai keperluan, namun meningkatnya penggunaan email juga membawa tantangan baru[1]. Salah satu masalah yang paling meresahkan adalah spam email, yang selain mengganggu penerima, juga memberikan beban pada infrastruktur penyedia layanan email. Ancaman terkait spam termasuk tautan berbahaya, serangan phishing, dan malware yang dapat merusak perangkat pengguna. Serangan phishing melalui email menjadi isu utama, dimana penjahat siber berupaya untuk mencuri informasi sensitif melalui email. Spam email juga mengurangi efisiensi komunikasi dan dapat mengganggu produktivitas, baik bagi individu maupun organisasi, sehingga biaya pengelolaan dan pencegahan spam bisa menjadi signifikan. Oleh karena itu, pengelolaan spam email menjadi perhatian penting dalam lingkungan digital yang semakin kompleks[2].

Dengan tantangan dan risiko yang semakin meningkat, kebutuhan akan solusi yang lebih cerdas dan efisien dalam mendeteksi dan memfilter spam email menjadi semakin mendesak. Di sinilah teknologi informasi, khususnya melalui pendekatan berbasis *machine learning*, dapat memainkan peran kunci dalam memberikan solusi untuk tantangan ini. Deteksi otomatis spam email, dengan akurasi yang tinggi dan waktu respons yang cepat, menjadi harapan bagi banyak pengguna dan penyedia layanan email untuk mengatasi masalah yang semakin kompleks ini[3].

Algoritma *machine learning* telah memberikan terobosan signifikan dalam berbagai aspek teknologi informasi, termasuk dalam bidang deteksi spam email. Dengan kemampuan untuk belajar dari data dan membuat prediksi atau klasifikasi berdasarkan data tersebut, *machine learning* menawarkan pendekatan yang lebih adaptif dan dinamis dibandingkan dengan metode tradisional. *Support Vector*

Machine (SVM) dan *Random Forest* adalah dua algoritma machine learning yang telah mendapatkan popularitas khusus dalam tugas klasifikasi teks, seperti deteksi spam email. Keduanya mewakili dua pendekatan berbeda dalam machine learning, tetapi keduanya telah menunjukkan kinerja yang mengesankan dalam berbagai aplikasi[4].

Support Vector Machine (SVM) adalah algoritma berbasis margin yang bertujuan untuk menemukan hyperplane terbaik yang memisahkan data antar kelas. Dalam konteks deteksi spam email, SVM dapat diandalkan untuk menemukan batasan yang memisahkan email spam dari email non-spam dengan cara yang optimal. Keunggulan SVM terletak pada kemampuannya untuk menangani data berdimensi tinggi dan efisiensinya dalam klasifikasi teks. Namun, SVM mungkin memerlukan waktu pelatihan yang lebih lama pada dataset yang sangat besar dan dapat membutuhkan penyetelan parameter yang cermat untuk optimal.

Random Forest Merupakan algoritma berbasis ensemble yang menggabungkan prediksi dari banyak pohon keputusan. Dengan melakukan ini, *Random Forest* cenderung memberikan kinerja yang lebih stabil dan mengurangi risiko overfitting, suatu fenomena di mana model terlalu spesifik terhadap data pelatihan dan kurang generalisasi dengan baik pada data baru. Kelebihan utama dari *Random Forest* adalah fleksibilitasnya dalam menangani fitur berdimensi tinggi dan kemampuannya untuk memberikan estimasi pentingnya fitur. Meskipun demikian, model ini mungkin lebih kompleks dan memerlukan lebih banyak sumber daya komputasi dibandingkan dengan model lainnya[5].

Berdasarkan latar belakang diatas, penelitian ini akan difokuskan pada optimasi algoritma *Support Vector Machine* (SVM) dan *Random Forest* untuk deteksi spam email. Langkah-langkah optimasi, termasuk penentuan parameter optimal untuk SVM seperti *kernel, C, dan gamma*, serta penyesuaian jumlah pohon dan fitur untuk Random Forest, akan menjadi fokus utama. Pemilihan fitur yang relevan juga akan diperhatikan menggunakan metode pemilihan fitur yang sesuai. Evaluasi kinerja kedua algoritma akan dilakukan dengan mempertimbangkan metrik-metrik seperti akurasi, *presisi, recall, dan F1-score*. Dengan pendekatan ini,

diharapkan penelitian ini dapat memberikan rekomendasi terkait algoritma yang paling efektif dalam konteks deteksi spam email.

1.2 Rumusan Masalah

Berdasar latar belakang diatas maka penulis ingin menyelesaikan masalah yaitu "Bagaimana Optimasi Algoritma *Support Vector Machine* Dan *Random Forest* Terhadap Deteksi Spam Email?"

1.3 Batasan Masalah

Dalam penelitian ini terdapat beberapa batasan masalah yaitu:

1. Klasifikasi menggunakan *Support Vector Machine* Dan *Random Forest*.
2. Dataset yang digunakan dari Kaggle dengan url: https://www.kaggle.com/datasets/ashfakyeafi/spam-email-classification/data
3. Bahasa pemograman yang digunakan adalah python.
4. Evaluasi kinerja dengan *cofussion matrix*.
5. Optimasi menggunakan *RandomSearchCV*

1.4 Tujuan Penelitian

Dengan adanya berbagai algoritma machine learning yang telah dikembangkan, pertanyaan muncul tentang efektivitas dan efisiensi masing-masing dalam mendeteksi email spam. Penelitian ini bertujuan untuk mengetahui pengaruh optimasi algoritma *Support Vector Machine* dan *Random Forest* dalam mendeteksi spam email.

1.5 Manfaat Penelitian

1. Manfaat Teoritis

a. Pemahaman Mendalam: Penelitian ini akan memberikan pemahaman yang lebih mendalam tentang bagaimana algoritma *Support Vector Machine* dan *Random Forest* bekerja dalam konteks deteksi spam email, serta kelebihan dan keterbatasan masing-masing algoritma.

b. Kontribusi Literatur: Hasil dari penelitian ini akan memperkaya literatur ilmiah dalam bidang *machine learning*, khususnya dalam aplikasinya untuk deteksi spam email. Ini akan menjadi referensi bagi peneliti lain yang tertarik pada topik serupa.

c. Pendekatan Metodologis: Penelitian ini dapat memberikan panduan metodologis untuk mendesain eksperimen dan evaluasi kinerja algoritma *machine learning* dalam konteks aplikasi lainnya.

2. Manfaat Praktis

a. Optimalisasi Deteksi Spam: Dengan mengetahui algoritma mana yang memiliki kinerja terbaik dalam kondisi atau dataset tertentu, penyedia layanan email dan perusahaan keamanan siber dapat mengoptimalkan sistem deteksi spam mereka untuk meningkatkan efektivitas dan efisiensi.

b. Pengembangan Alat: Hasil dari penelitian ini dapat digunakan sebagai dasar untuk mengembangkan alat atau perangkat lunak yang dapat membantu dalam deteksi otomatis spam email dengan akurasi yang lebih tinggi.

c. Pendidikan dan Pelatihan: Pengetahuan dari penelitian ini dapat diintegrasikan ke dalam modul pelatihan atau kurikulum pendidikan untuk meningkatkan keterampilan profesional dalam bidang keamanan siber dan pengolahan teks.

1.6 Sistematika Penulisan

Sistematika penulisan merupakan sebuah panduan umum dalam menyusun sebuah penelitian dalam bentuk karya tulis ilmiah. Dalam hal ini sistematika penulisannya antara lain sebagai berikut:

BAB I PENDAHULUAN, berisi Latar belakang masalah, rumusan masalah, Batasan masalah, tujuan penelitian, manfaat penelitian dan sistematika penulisan

BAB II TINJAUAN PUSTAKA, berisi tinjauan pustaka, dasar-dasar teori yang digunakan, dan penelitian terdahulu

BAB III METODE PENELITIAN, didalamnya terdapat tinjauan umum tentang

objek penelitian, alur penelitian, alat dan bahan penelitian

BAB IV HASIL DAN PEMBAHASAN, bab ini tentang hasil serta pembahasan penelitian

BAB V PENUTUP, berisi kesimpulan dan saran yang dapat peneliti rangkum selama proses penelitian