BAB I
PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi informasi dan komunikasi membawa perubahan besar dalam cara manusia berinteraksi dan bertukar informasi. Email menjadi salah satu media komunikasi yang paling banyak digunakan, baik dalam bisnis, pendidikan, maupun kebutuhan personal [1]. Seiring perkembangan, email menjadi alat komunikasi internet yang mudah dan cepat [2]. Namun, intensitas penggunaan email dapat menimbulkan dampak negatif berupa meningkatnya spam yang berisiko menyebabkan penipuan, *phishing*, dan penyebaran *malware* yang merugikan penerima maupun institusi [2], [3].

Berbagai pendekatan *machine learning* telah digunakan untuk mengklasifikasikan email sebagai spam atau *ham* [2], [3], [4]. *Naive Bayes (NB)* dan *Support Vector Machine (SVM)* dikenal efektif dalam klasifikasi teks, namun memiliki keterbatasan [5]. *NB* lemah karena asumsi independensi fitur sering tidak sesuai dengan data nyata [2], sedangkan *SVM* sensitif terhadap parameter dan membutuhkan komputasi tinggi [6]. Beberapa penelitian mencoba mengkombinasikan berbagai model untuk mengatasi keterbatasan masing-masing algoritma dan mengeksplorasi strategi kombinasi melalui metode *ensemble learning* [7], [8], [9].

Penelitian sebelumnya mencoba mengombinasikan *NB* dan *SVM* menggunakan *ensemble* terbukti lebih baik dibanding algoritma tunggal [10]. Metode *ensemble learning* yang umum diterapkan meliputi *Bagging, Boosting, Stacking,* dan *Voting* [11]. *Stacking, bagging, dan boosting* dapat meningkatkan performa, meski memiliki kelemahan masing-masing seperti rawan *overfitting*, sensitif terhadap *noise*, atau memerlukan komputasi tinggi [12].

Melihat permasalahan tersebut, diperlukan sebuah pendekatan yang mampu menggabungkan kelebihan masing-masing algoritma sekaligus menutupi kelemahannya. Metode *ensemble voting* menjadi salah satu solusi potensial karena dapat meningkatkan akurasi dan stabilitas hasil klasifikasi dengan mengkombinasikan prediksi dari beberapa model dasar [11]. Penerapan *ensemble voting* pada kombinasi *Naive Bayes* dan *SVM* diharapkan mampu menghasilkan model deteksi spam email yang lebih efektif, akurat, dan andal dalam mendukung sistem keamanan informasi.

1.2 Rumusan Masalah

1. Apakah metode *ensemble voting* mampu meningkatkan *precision* dan stabilitasi deteksi email spam dibandingkan algoritma tunggal?
2. Bagaimana *ensemble voting* dapat mengatasi kelemahan *Naive Bayes* dan *SVM* dalam klasifikasi email spam?

1.3 Batasan Masalah

Penelitian ini dibatasi oleh beberapa hal berikut :

1. Penelitian hanya difokuskan pada deteksi email spam berbasis teks, tanpa mempertimbangkan lampiran atau gambar di email.
2. Data yang digunakan berupa dataset email spam publik dari *Kaggle*, sehingga tidak mencakup data pribadi maupun data *real-time* dari server email.
3. Algoritma *machine learning* yang digunakan dibatasi pada *Naive Bayes* dan *Support Vector Machine (SVM)*, yang kemudian digabungkan menggunakan metode *ensemble voting*.

1.4 Tujuan Penelitian

Adapun yang menjadi tujuan dari penelitian ini antara lain :

1. Mengetahui apakah metode *ensemble voting* mampu meningkatkan *precision* dan stabilitas klasifikasi email spam dibandingkan algoritma tunggal.
2. Menganalisis bagaimana metode *ensemble voting* dapat mengatasi

kelemahan *Naive Bayes (NB)* dan *Support Vector Machine (SVM)* dalam klasifikasi email spam.

3. Mengimplementasikan metode *ensemble voting (hard voting* dan *soft voting)* dengan mengombinasikan *Naive Bayes* dan *SVM* untuk deteksi email spam.
4. Mengevaluasi performa metode *ensemble voting* menggunakan metrik evaluasi seperti *accuracy, precision, recall, dan F1-Score*, serta membandingkannya dengan model tunggal.
5. Menghasilkan sistem deteksi email spam berbasis *machine learning* yang dapat dimanfaatkan sebagai dasar dalam pengembangan sistem keamanan informasi.

1.5 Manfaat Penelitian

Hasil penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

a. Manfaat Teoritis

1. Menambah kajian ilmiah dalam bidang *machine learning* khususnya pada penerapan metode *ensemble voting* dalam klasifikasi teks.
2. Memberikan referensi bagi penelitian selanjutnya terkait pengembangan metode deteksi spam email berbasis kombinasi algoritma.
3. Menjadi dasar untuk perbandingan performa antara metode tunggal dan metode *ensemble*, sehingga memperkaya literatur akademis di bidang klasifikasi teks dan keamanan informasi.

b. Manfaat Praktis

1. Memberikan kontribusi dalam pengembangan sistem deteksi email spam yang lebih akurat dan stabil, sehingga dapat mendukung peningkatan keamanan informasi pada penggunaan email.
2. Dapat dimanfaat oleh akademisi, peneliti, dan praktisi IT sebagai

referensi dalam merancang sistem keamanan berbasis *machine learning*.

3. Menjadi pijakan bagi organisasi atau pengembang aplikasi keamanan informasi untuk mengintegrasikan metode *ensemble voting* dalam sistem deteksi spam.

1.6 Sistematika Penulisan

Sistematika penulisan skripsi ini terdiri dari lima bab utama yang disusun secara runtut dan terstruktur sebagai berikut :

BAB I PENDAHULUAN

Bab ini berisi latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, serta sistematika penulisan. Bagian ini menjelaskan alasan dilakukannya penelitian, ruang lingkup penelitian, serta arah dan kontribusi penelitian.

BAB II TINJAUAN PUSTAKA

Berisi tinjauan pustaka, dasar-dasar teori yang digunakan dalam penelitian, seperti konsep email dan spam, algoritma *machine learning*, *Naive Bayes*, *Support Vector Machine*, *ensemble learning*, serta metode *voting*. Selain itu, juga disajikan tinjauan penelitian terdahulu yang relevan untuk memperkuat landasan penelitian.

BAB III METODE PENELITIAN

Bab ini menjelaskan metode yang digunakan dalam penelitian, meliputi sumber dan jenis data, tahapan *preprocessing data*, metode *feature extraction*, pembangunan model *Naive Bayes* dan *SVM*, penerapan *ensamble voting (hard voting* dan *soft voting)*, serta metode evaluasi dengan metrik *accuracy, precision, recall, dan F1-Score*.

BAB IV HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil implementasi model deteksi email spam menggunakan *Naive Bayes, SVM* dan *ensemble voting*. Hasil pengujian

kemudian dianalisis dan membandingkan dengan performa model performa model tunggal dan *ensemble voting* untuk mengetahui tingkat *precision*, konsistensi, dan reliabilitas model.

BAB V PENUTUP

Berisi kesimpulan dan saran yang dapat peneliti rangkum selama proses penelitian berlangsung. Kesimpulan dibuat berdasarkan tujuan penelitian dan hasil yang dicapai, sedangkan saran diberikan untuk pengembangan penelitian di masa yang akan datang.