BAB I
PENDAHULUAN

Bab pendahuluan ini membahas mengenai latar belakang masalah, rumusan masalah, tujuan dan manfaat, ruang lingkup, serta sistematika penulisan pada skripsi yang berjudul klasifikasi email phishing menggunakan metode *Support Vector Machine* (SVM) dan *Naive Bayes* dengan vektorisasi *TF-IDF*, *Count Vectorizer*, dan *Word2Vec*.

1.1 Latar Belakang

Email telah menjadi salah satu sarana komunikasi paling penting dan banyak digunakan dalam kehidupan sehari-hari. Baik untuk keperluan pribadi, pendidikan, maupun bisnis, email memungkinkan pengiriman pesan dan dokumen dengan cepat, efisien, dan lintas lokasi geografis. Hampir semua organisasi kini mengandalkan email sebagai media utama dalam pertukaran informasi dan koordinasi kerja (Alhuzali dkk., 2025). Namun, di balik kemudahan tersebut, email juga sering dimanfaatkan sebagai pintu masuk serangan siber, salah satunya melalui email phishing, yaitu bentuk penipuan digital dengan mengirim pesan palsu yang tampak sah untuk mencuri data pribadi seperti kredensial login, informasi keuangan, atau data sensitif lainnya (Li dkk., 2024).

Phishing merupakan salah satu bentuk rekayasa sosial di mana penyerang berpura-pura sebagai pihak resmi seperti bank, instansi pemerintah, atau perusahaan ternama agar korban percaya dan melakukan tindakan tertentu. Biasanya, pelaku mengirimkan tautan yang mengarah ke situs web palsu atau lampiran berbahaya, dengan tujuan mencuri informasi sensitif pengguna (Thakur dkk., 2023). Dalam banyak kasus, pengguna tertipu karena tampilan dan bahasa pesan yang dibuat sangat mirip dengan komunikasi asli. Hal ini membuat serangan phishing menjadi salah satu ancaman keamanan siber terbesar bagi individu maupun organisasi di seluruh dunia (Tausif dkk., 2024).

Kondisi keamanan email saat ini menunjukkan bahwa serangan phishing semakin nyata terlihat dalam kehidupan sehari-hari, terutama dengan meningkatnya penggunaan layanan digital seperti perbankan online, *e-commerce*, dan sistem kerja jarak jauh. Banyak pengguna menerima email yang mengatasnamakan bank, platform belanja, atau layanan

digital populer yang berisi peringatan palsu mengenai akun bermasalah, transaksi mencurigakan, atau permintaan verifikasi data. Pesan-pesan tersebut sering kali dibuat dengan bahasa yang mendesak dan tampilan profesional, sehingga mendorong korban untuk segera mengambil tindakan tanpa berpikir panjang (Somesha & Pais, 2022). Peningkatan volume email yang diterima membuat pemeriksaan email kurang teliti. Studi oleh Tausif dkk. (2024) menunjukkan bahwa faktor kelalaian pengguna dan kepercayaan terhadap tampilan visual email menjadi penyebab utama keberhasilan serangan phishing, bahkan pada pengguna yang telah memiliki pengetahuan dasar mengenai keamanan siber. Kondisi ini menegaskan bahwa email phishing bukan hanya persoalan teknis, tetapi juga berkaitan erat dengan perilaku pengguna, sehingga diperlukan sistem deteksi otomatis yang mampu bekerja secara konsisten dan akurat untuk membantu pengguna dalam mengenali ancaman tersebut.

Serangan phishing memiliki beberapa jenis yang berkembang semakin kompleks. Serangan *spear phishing* menargetkan individu tertentu dengan pesan yang dipersonalisasi menggunakan informasi pribadi korban, sementara *clone phishing* meniru email sah dengan mengganti tautan aslinya menjadi tautan berbahaya. Ada pula *Business Email Compromise* (BEC) yang meniru identitas pimpinan perusahaan untuk mengarahkan korban melakukan transfer dana ke rekening palsu(Li dkk., 2024). Jenis serangan ini sulit dideteksi karena tampilannya yang sangat mirip dengan komunikasi internal organisasi. Penelitian oleh Thakur dkk. (2023) juga menjelaskan bahwa peningkatan kemampuan pelaku phishing dalam meniru gaya bahasa manusia membuat metode deteksi tradisional seperti daftar hitam (*blacklist*) dan aturan berbasis kata kunci (*rule-based filtering*) semakin tidak baik.

Laporan Anti-Phishing Working Group (APWG) menyebutkan bahwa jumlah serangan phishing global meningkat lebih dari 40% dalam dua tahun terakhir, dengan mayoritas insiden terjadi melalui email (Tausif dkk., 2024). Kondisi ini menunjukkan bahwa ancaman phishing terus berkembang dan memerlukan pendekatan yang lebih adaptif. Metode deteksi konvensional hanya mampu mengenali pola lama, sementara pelaku kini menggunakan teknik yang lebih kompleks seperti personalisasi pesan dan pemanfaatan kecerdasan buatan untuk membuat email yang menyerupai komunikasi asli.

Salah satu pendekatan modern yang banyak digunakan untuk menghadapi tantangan ini adalah *machine learning*, karena mampu belajar dari data historis dan mengenali pola-pola baru yang belum pernah muncul sebelumnya. Dengan algoritma pembelajaran mesin, sistem dapat secara otomatis mengklasifikasikan email berdasarkan ciri-ciri tertentu dari teksnya (Al Tawil dkk., 2024). Pendekatan ini berkaitan dengan vektorisasi teks, yaitu proses mengubah teks menjadi representasi numerik agar dapat diproses oleh algoritma komputer. Beberapa metode umum yang digunakan antara lain *Term Frequency–Inverse Document Frequency* (TF-IDF), *Count Vectorizer*, dan *Word2Vec* dengan pendekatan *Continuous Bag of Words* (CBOW) (Mohammed & Ahmed, 2024). TF-IDF digunakan untuk menyoroti kata-kata penting dalam pesan phishing, sedangkan *Word2Vec* mampu memahami makna kontekstual antar kata sehingga sistem lebih cerdas dalam mengenali pesan berbahaya.

Selain teknik vektorisasi teks, penelitian ini juga menggunakan algoritma klasifikasi untuk menentukan kategori email. Algoritma klasifikasi seperti *Support Vector Machine* (SVM) dan *Naive Bayes* digunakan untuk menentukan apakah sebuah email termasuk phishing atau tidak. Menurut Al Tawil dkk. (2024), algoritma SVM bekerja dengan mencari *hyperplane* terbaik yang memisahkan dua kelas data dan terbukti efisien pada data teks berdimensi tinggi seperti hasil TF-IDF. Sementara itu, Naive Bayes menggunakan pendekatan probabilistik sederhana yang efisien dan cepat, serta memberikan performa tinggi pada dataset berukuran besar.

Penelitian yang dilakukan oleh Fares dkk. (2024) membandingkan performa beberapa algoritma *machine learning* dalam deteksi email phishing dengan menggunakan teknik TF-IDF sebagai representasi teks. Hasil eksperimen menunjukkan bahwa *Support Vector Machine* dengan TF-IDF mampu mencapai tingkat akurasi yang sangat tinggi, yaitu sekitar 98%, sementara Random Forest dengan TF-IDF memperoleh akurasi yang lebih rendah, berada pada kisaran 95%. Hasil ini mengindikasikan bahwa SVM memiliki kemampuan yang cukup baik dalam menangani data teks berdimensi tinggi yang dihasilkan oleh TF-IDF dibandingkan dengan Random Forest.

Beberapa pendekatan dalam mengklasifikasikan email phishing telah dilakukan seperti dilakukan oleh Fahim dkk. (2024) yang memanfaatkan pendekatan *natural*

language processing dan *machine learning* untuk mendeteksi email phishing menggunakan Naive Bayes dengan dua teknik vektorisasi teks, yaitu *Count Vectorizer* dan TF-IDF. Berdasarkan hasil pengujian pada *Spambase dataset*, *Naive Bayes* dengan *Count Vectorizer* mampu mencapai akurasi sekitar 98%, sedangkan kombinasi *Naive Bayes* dengan TF-IDF menghasilkan akurasi yang sedikit lebih rendah, yaitu sekitar 97%. Temuan ini menunjukkan bahwa efektivitas teknik vektorisasi dapat berbeda tergantung pada karakteristik data dan distribusi fitur yang dihasilkan.

Penelitian lain yang dilakukan oleh Bountakas dkk. (2021) yang mengkaji performa beberapa algoritma *machine learning* dalam mendeteksi email phishing. Berdasarkan hasil eksperimen dan kajian yang dilakukan, kombinasi *Naive Bayes* dengan *Word2Vec* menunjukkan akurasi yang stabil dan tinggi, berada pada kisaran 96%. Sementara itu, *Naive Bayes* dengan TF-IDF juga memberikan performa yang kompetitif dengan tingkat akurasi antara 93%. Temuan ini menunjukkan bahwa perbedaan vektorisasi dapat mempengaruhi performa model dalam mengklasifikasikan email phishing.

Berdasarkan penelitian yang telah dilakukan sebelumnya, penelitian ini dilakukan untuk mengembangkan sistem klasifikasi email phishing dengan membandingkan tiga metode vektorisasi teks, yaitu TF-IDF, *Count Vectorizer*, dan *Word2Vec* (CBOW), menggunakan dua algoritma utama, SVM dan *Naive Bayes*. Penelitian ini bertujuan untuk menemukan kombinasi metode dan algoritma yang paling akurat, efisien, dan stabil dalam mendeteksi email phishing. Diharapkan hasil penelitian ini dapat membantu organisasi, perusahaan, dan pengguna individu dalam memperkuat sistem keamanan email serta meminimalkan risiko kebocoran data akibat serangan phishing.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan sebelumnya, rumusan masalah dalam penelitian ini adalah bagaimana pengaruh vektorisasi teks TF-IDF, *Count Vectorizer*, dan *Word2Vec* (CBOW) terhadap performa metode *Support Vector Machine* dan *Naive Bayes* dalam mengklasifikasikan email phishing.

1.3 Tujuan dan Manfaat

Tujuan dari penelitian ini adalah membandingkan performa metode *Support Vector Machine* dan *Naive Bayes* dengan tiga metode vektorisasi teks (TF-IDF, *Count Vectorizer*, dan Word2Vec) untuk menentukan kombinasi metode yang menghasilkan performa terbaik dalam klasifikasi email phishing.

Manfaat dari penelitian ini adalah tersedianya model klasifikasi yang mampu membedakan email phishing atau tidak menggunakan vektorisasi teks TF-IDF, *Count Vectorizer*, dan *Word2Vec* (CBOW) dan metode SVM serta Naive Bayes. Model tersebut diharapkan dapat digunakan untuk pengembangan sistem klasifikasi email phishing otomatis yang lebih akurat di masa mendatang, terutama dalam bidang seperti klasifikasi teks, pendidikan, dan pemrosesan bahasa alami.

1.4 Ruang Lingkup

Ruang lingkup dalam penelitian ini adalah sebagai berikut:

1. Pengumpulan data diperoleh di kaggle.com dengan tema email phishing.
2. Data yang digunakan adalah dataset CEAS_08 email yang dikumpulkan dalam rentang waktu 2007-2008, kemudian dataset SpamAssasin email yang dikumpulkan dalam rentang waktu 2002-2003, lalu dataset Enron email yang dikumpulkan pada tahun2005, dan yang terakhir dataset Phishing_Email email yang dikumpulkan pada tahun 2023.
3. Dataset CEAS_08, SpamAssasin, Enron fitur yang digunakan hanya *subject*, *body*, dan label, untuk dataset Phishing_Email fitur Email_Text dan Email_Type.
4. Metode yang digunakan dalam penelitian ini adalah *Support Vector Machine* dan *Naive Bayes* dengan vektorisasi teks TF-IDF*, Count Vectorizer,* dan *Word2Vec* dengan pendekatan *Continuous Bag of Words* (CBOW).
5. Evaluasi model klasifikasi dilakukan menggunakan metrik evaluasi seperti *accuracy, precision, recall,* dan *F1-score* untuk menilai kinerja klasifikasi.

1.5 Sistematika Penulisan

Sistematika penulisan dibuat untuk memberikan gambaran umum dan jelas mengenai penulisan serta pembahasan pada penelitian klasifikasi email phishing menggunakan metode *Support Vector Machine* dan *Naive Bayes* dengan vektorisasi TF-IDF, *Count Vectorizer*, dan *Word2Vec*. Berikut adalah sistematika penulisan yang digunakan.

BAB I PENDAHULUAN

Pendahuluan menyajikan tentang latar belakang, rumusan masalah, tujuan dan manfaat, ruang lingkup, dan sistematika penulisan pada penelitian klasifikasi email phishing menggunakan metode *Support Vector Machine* dan *Naive Bayes* dengan vektorisasi TF-IDF, *Count Vectorizer*, dan *Word2Vec*.

BAB II TINJAUAN PUSTAKA

Bab ini menyajikan teori-teori yang mencakup *state-of-the-art*, definisi email phishing, klasifikasi email phishing, pra-pemrosesan data, *machine learning*, *K-fold Cross Validation*, metode *Support Vector Machine*, metode *Naive Bayes,* vektorisasi teks, serta metrik evaluasi model. Teori-teori tersebut disusun untuk mendukung pelaksanaan dan penyusunan penelitian klasifikasi email phishing menggunakan model *Support Vector Machine* dan *Naive Bayes* dengan vektorisasi TF-IDF, *Count Vectorizer*, dan *Word2Vec*.

BAB III METODOLOGI PENELITIAN

Bab ini menyajikan tentang tahapan penelitian, pengumpulan data, pra-pemrosesan data, pembagian data latih, dan data uji, pembangunan dan model klasifikasi email phishing dan skenario pelatihan model, serta skenario pengujian model dan evaluasi yang merupakan metode dan proses yang terjadi pada penelitian klasifikasi email phishing menggunakan metode *Support Vector Machine* dan *Naive Bayes* dengan vektorisasi TF-IDF, *Count Vectorizer*, dan *Word2Vec*.

BAB IV HASIL DAN PEMBAHASAN

Bab ini membahas tentang pembahasan yang meliputi lingkungan dan perangkat yang digunakan untuk penelitian, data penelitian hasil pra-pemrosesan data, menampilkan hasil skenario yang mana model dilatih

dengan *K-Fold Cross Validation* untuk membandingkan kombinasi vektorisasi dengan model mana yang terbaik. Kemudian menampilkan hasil skenario pengujian akhir yang menggunakan data uji, analisis hasil pelatihan dan pengujian, dan interpretasi hasil pengujian pada penelitian klasifikasi email phishing menggunakan metode *Support Vector Machine* dan *Naive Bayes* dengan vektorisasi TF-IDF, *Count Vectorizer*, dan *Word2Vec*.

BAB V PENUTUP

Bab ini berisi kesimpulan dari klasifikasi email phishing menggunakan metode *Support Vector Machine* dan *Naive Bayes* dengan vektorisasi TF-IDF, *Count Vectorizer*, dan *Word2Vec* yang telah dilakukan beserta saran penulis.