

BAB I PENDAHULUAN

1.1 Latar Belakang

Escherichia coli (*E. coli*) merupakan salah satu jenis bakteri yang umum ditemukan di lingkungan, khususnya pada saluran pencernaan manusia dan hewan. Bakteri ini termasuk dalam kelompok bakteri gram-negatif dan berbentuk batang. Keberadaannya di dalam usus manusia dan hewan berdarah panas berfungsi sebagai bagian dari flora normal yang membantu proses pencernaan dan pembentukan vitamin K. Namun, tidak semua strain *E. coli* bersifat menguntungkan. Meskipun sebagian besar strain *E. coli* bersifat non-patogen dan berperan penting dalam menjaga keseimbangan mikrobiota usus, beberapa strain lainnya bersifat patogen dan dapat menimbulkan berbagai penyakit serius seperti diare, infeksi saluran kemih, dan sindrom hemolitik uremik [1]. Dalam beberapa kasus, strain patogen *E. coli* dapat menghasilkan toksin berbahaya yang mampu merusak jaringan tubuh dan menyebabkan komplikasi kesehatan yang parah. Oleh karena itu, keberadaan strain-strain ini memerlukan perhatian khusus dari kalangan medis dan ilmuwan mikrobiologi.

Identifikasi serta klasifikasi strain *E. coli* menjadi aspek yang sangat penting dalam bidang mikrobiologi dan kesehatan, guna mendeteksi potensi bahaya dari strain patogen tersebut secara cepat dan akurat [2]. Pengklasifikasian yang tepat dapat membantu dalam penanganan penyakit infeksi yang ditimbulkan oleh *E. coli* dan juga mencegah penyebaran strain yang berpotensi menyebabkan wabah. Namun, metode identifikasi *E. coli* secara tradisional yang mengandalkan kultur bakteri dan teknik biokimia dinilai masih memiliki banyak keterbatasan, terutama dalam hal efisiensi waktu dan tingkat akurasi. Proses ini memerlukan tenaga ahli dan prosedur laboratorium yang cukup kompleks, serta rentan terhadap kesalahan manusia [3]. Selain itu, ketergantungan terhadap media pertumbuhan dan kondisi lingkungan dapat memengaruhi hasil akhir dari identifikasi tersebut.

Ketidaktepatan dalam proses klasifikasi dapat berdampak pada keterlambatan penanganan medis, sehingga diperlukan pendekatan yang lebih modern dan efisien.

Perkembangan teknologi dalam bidang kecerdasan buatan atau Artificial Intelligence dan pembelajaran mesin atau *Machine Learning* membuka peluang besar dalam pengolahan data biologis secara lebih cepat dan akurat. Machine learning memungkinkan komputer untuk belajar dari data historis tanpa perlu diprogram secara eksplisit, sehingga cocok digunakan dalam analisis data mikrobiologi yang kompleks. Salah satu pendekatan yang banyak digunakan dalam klasifikasi data biologis adalah metode *supervised learning*, seperti algoritma *Support Vector Machine* atau dapat disebut juga SVM dan Naïve Bayes [4]. SVM dikenal dengan kemampuannya menangani data non-linear menggunakan teknik kernel untuk pemetaan ke ruang dimensi lebih tinggi [5], sehingga menjadikannya pilihan yang andal dalam berbagai kasus klasifikasi yang kompleks. Di sisi lain, Naïve Bayes merupakan algoritma yang sangat sederhana namun efektif, karena memiliki efisiensi komputasi tinggi dalam data dengan atribut independen [6].

Dalam penelitian sebelumnya, algoritma Naïve Bayes yang dikombinasikan dengan teknik boosting seperti AdaBoost mampu meningkatkan akurasi klasifikasi bakteri *E. coli* dari 76% menjadi 94%, menunjukkan bahwa kolaborasi ini efektif dalam menangani dataset biologis kompleks [2]. Kombinasi metode ini memungkinkan peningkatan kinerja model klasifikasi secara signifikan tanpa harus meningkatkan kompleksitas struktur algoritma secara berlebihan. Sebaliknya, algoritma SVM juga menunjukkan performa luar biasa dalam klasifikasi non-linear dan bahkan mencapai akurasi hingga 100% pada studi lain di bidang klasifikasi berita palsu [8]. Hasil ini menunjukkan bahwa dengan parameter yang tepat, SVM mampu memberikan hasil klasifikasi yang sangat akurat dan konsisten. Oleh karena itu, membandingkan kedua algoritma ini menjadi relevan untuk dilakukan, mengingat masing-masing memiliki karakteristik yang berbeda. Naïve Bayes unggul dari segi kesederhanaan, kecepatan, dan efisiensi komputasi, sedangkan SVM lebih kuat dalam hal akurasi dan kestabilan prediksi, terutama pada data yang tidak terdistribusi secara linier. Dengan membandingkan performa keduanya dalam

kasus klasifikasi E. coli, diharapkan dapat diperoleh pemahaman yang lebih mendalam mengenai kelebihan dan keterbatasan masing-masing algoritma dalam konteks data biologis yang kompleks.

Selain itu, studi lain yang membandingkan SVM dan Naïve Bayes dalam berbagai kasus menunjukkan hasil yang variatif. Misalnya, pada klasifikasi kesehatan mental mahasiswa, Naïve Bayes mampu mengungguli SVM dalam hal klasifikasi bias mental [9], sementara pada data teks berbahasa Arab, SVM terbukti lebih akurat dibanding Naïve Bayes [10]. Hal ini menunjukkan bahwa performa masing-masing algoritma sangat dipengaruhi oleh karakteristik dataset yang digunakan. Tidak ada satu algoritma yang secara mutlak lebih unggul dari yang lain dalam semua jenis data. Oleh karena itu, sangat penting untuk menguji kedua algoritma tersebut dalam konteks dan kondisi data yang spesifik.

Khusus untuk klasifikasi E. coli, algoritma Naïve Bayes telah digunakan untuk mengelompokkan strain bakteri gram-negatif berdasarkan tujuh fitur biologis seperti *mcg*, *gvh*, dan *alm1*, dan berhasil mencapai akurasi hingga 80.93% [11]. **Pada umumnya, proses klasifikasi ini bertujuan untuk memetakan setiap strain E. coli ke dalam kelas-kelas lokasi subselular tertentu, seperti cytoplasm (cp), inner membrane (im), periplasm (pp), outer membrane (om), serta kelas lain yang lebih spesifik seperti inner membrane lipoprotein (imL), inner membrane without signal (imS), outer membrane lipoprotein (omL), dan inner membrane with uncleaved signal (imU).** Setiap kelas merepresentasikan lokasi protein spesifik dari strain E. coli yang dapat memberikan informasi penting dalam analisis biologis dan medis. Penelitian-penelitian tersebut menunjukkan bahwa meskipun Naïve Bayes memiliki struktur yang sederhana, algoritma ini tetap mampu memberikan hasil yang kompetitif. Sementara itu, dalam pendekatan machine learning yang lebih kompleks, proses klasifikasi dataset E. coli menggunakan SVM ataupun Naïve Bayes sering melibatkan tahapan preprocessing, pengukuran probabilitas, dan evaluasi menggunakan confusion matrix untuk memastikan akurasi dan presisi model [12]. Tahapan-tahapan tersebut

sangat penting untuk mendapatkan hasil klasifikasi yang dapat diandalkan dan akurat.

Berdasarkan berbagai temuan dan perkembangan tersebut, penelitian ini bertujuan untuk **mengembangkan klasifikasi bakteri *E. coli*** dengan menerapkan dua algoritma machine learning, yaitu SVM dan Naïve Bayes. Klasifikasi yang dikembangkan mencakup rangkaian proses mulai dari preprocessing data, pelatihan model, evaluasi performa klasifikasi, hingga visualisasi hasil. Evaluasi dilakukan menggunakan metrik akurasi, presisi, recall, dan F1-score, serta menganalisis kelebihan dan kekurangan masing-masing algoritma. Dengan pendekatan yang sistematis, diharapkan hasil dari penelitian ini dapat memberikan rekomendasi praktis mengenai algoritma yang paling sesuai untuk klasifikasi bakteri *E. coli*. Selain itu, penelitian ini juga diharapkan dapat menjadi kontribusi nyata dalam upaya peningkatan kualitas deteksi mikrobiologi klinis dan pengambilan keputusan dalam dunia medis dan kesehatan masyarakat secara umum.

1.2 Rumusan Masalah

Untuk memastikan penelitian ini tetap terarah dan fokus pada tujuan yang ingin dicapai, diperlukan perumusan masalah yang jelas. Berikut adalah beberapa pertanyaan yang menjadi dasar penelitian ini:

1. Bagaimana klasifikasi bakteri *E. coli* dapat dikembangkan menggunakan algoritma SVM dan Naïve Bayes?
2. Bagaimana performa algoritma SVM dan Naïve Bayes dalam klasifikasi *E. coli* ditinjau dari akurasi, presisi, recall, dan f1-score?
3. Seberapa efektif klasifikasi berbasis machine learning yang dikembangkan dalam meningkatkan efisiensi dan akurasi identifikasi bakteri *E. coli* dibandingkan pendekatan konvensional?

1.3 Batasan Masalah

Pada penelitian ini terdapat beberapa batasan yang diterapkan dalam ruang lingkup penelitian. Batasan ini bertujuan untuk memastikan bahwa proses pengembangan klasifikasi dilakukan secara terfokus dan efisien tanpa menyimpang dari tujuan utama penelitian. Berikut adalah batasan yang ditetapkan:

1. Penelitian ini hanya berfokus pada **pengembangan klasifikasi** untuk bakteri *E. coli* dan tidak mencakup analisis terhadap jenis bakteri lainnya.
2. Dataset yang digunakan berasal dari **sumber terbatas** dan tidak mencakup seluruh kemungkinan variasi strain *E. coli* yang ada.
3. Algoritma machine learning yang digunakan dalam klasifikasi ini terbatas pada **SVM dan Naïve Bayes**, tanpa mempertimbangkan algoritma lain.
4. Evaluasi performa klasifikasi dilakukan berdasarkan metrik akurasi, presisi, recall, dan F1-score, tanpa mempertimbangkan efisiensi komputasi atau waktu eksekusi model.
5. Klasifikasi dikembangkan dan diuji menggunakan bahasa pemrograman Python.
6. Penelitian ini tidak mencakup **implementasi klasifikasi** ke dalam bentuk perangkat lunak siap pakai atau aplikasi praktis dalam lingkungan laboratorium nyata.

1.4 Tujuan Penelitian

Penelitian ini bertujuan untuk menyelesaikan permasalahan yang telah dirumuskan serta memberikan kontribusi dalam pengembangan klasifikasi bakteri *E. coli*. Berikut adalah tujuan utama dari penelitian ini:

1. Mengembangkan klasifikasi bakteri *E. coli* menggunakan algoritma SVM dan Naïve Bayes.
2. Menganalisis performa masing-masing algoritma dalam klasifikasi ditinjau dari metrik akurasi, presisi, recall, dan F1-score.
3. Mengevaluasi efektivitas klasifikasi berbasis machine learning dalam meningkatkan efisiensi dan akurasi identifikasi bakteri *E. coli*.
4. Memberikan rekomendasi terhadap algoritma yang lebih sesuai untuk diterapkan dalam pengembangan klasifikasi bakteri *E. coli* berdasarkan hasil evaluasi.

1.5 Manfaat Penelitian

Penelitian ini diharapkan memberikan manfaat baik secara teoritis maupun praktis. Berikut adalah manfaat yang diharapkan dari penelitian ini:

1. Manfaat Teoritis

- Memberikan kontribusi terhadap pengembangan klasifikasi berbasis machine learning dalam bidang mikrobiologi, khususnya untuk identifikasi bakteri *E. coli*.
- Menambah wawasan mengenai performa algoritma SVM dan Naïve Bayes dalam membangun klasifikasi bakteri.
- Menjadi referensi bagi penelitian selanjutnya dalam pengembangan klasifikasi yang lebih akurat, efisien, dan dapat diimplementasikan secara praktis.

2. Manfaat Praktis

- Membantu praktisi mikrobiologi dan tenaga medis dalam mengidentifikasi strain bakteri *E. coli* melalui klasifikasi berbasis machine learning yang dikembangkan.
- Memberikan rekomendasi mengenai algoritma yang paling sesuai untuk diterapkan dalam pengembangan klasifikasi deteksi bakteri.
- Berpotensi untuk dijadikan dasar dalam pengembangan otomatis berbasis komputer yang dapat digunakan di laboratorium untuk mendukung diagnosa penyakit akibat *E. coli*.
- Memudahkan pengolahan dan analisis data mikrobiologi melalui klasifikasi yang terintegrasi dan efisien.

1.6 Sistematika Penulisan

Untuk memberikan gambaran mengenai isi skripsi, berikut adalah sistematika penulisan yang digunakan:

BAB I PENDAHULUAN

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

BAB II TINJAUAN PUSTAKA

Membahas dasar-dasar teori yang digunakan dalam penelitian, termasuk teori tentang *Escherichia coli*, konsep klasifikasi berbasis machine learning, serta penjelasan mengenai algoritma SVM dan Naïve Bayes.

BAB III METODE PENELITIAN

Menjelaskan metode penelitian yang digunakan, mulai dari pengumpulan data, tahapan pengembangan klasifikasi, preprocessing data, pelatihan model, evaluasi performa, hingga alat bantu yang digunakan.

BAB IV HASIL DAN PEMBAHASAN

Menyajikan hasil dari klasifikasi yang telah dikembangkan, analisis performa algoritma SVM dan Naïve Bayes, serta pembahasan mengenai kelebihan, kekurangan, dan efektivitas klasifikasi yang dihasilkan.

BAB V PENUTUP

Berisi kesimpulan dari hasil penelitian serta saran untuk pengembangan lebih lanjut, khususnya terkait peningkatan akurasi, efisiensi, dan implementasi klasifikasi dalam konteks mikrobio

