

BAB I PENDAHULUAN

1.1 Latar Belakang

Escherichia coli (*E. coli*) merupakan bakteri gram negatif berbentuk batang yang secara alami menghuni saluran pencernaan manusia maupun hewan berdarah panas. Dalam keadaan normal, sebagian besar *strain E. coli* termasuk ke dalam flora usus yang tidak berbahaya dan justru memberikan kontribusi penting dalam proses pencernaan, seperti membantu pemecahan makanan dan sintesis vitamin. Namun demikian, terdapat beberapa *strain* patogenik yang memiliki kemampuan untuk menyebabkan berbagai gangguan kesehatan serius. Infeksi oleh *strain* patogen tersebut dapat menimbulkan penyakit mulai dari diare berat, infeksi saluran kemih (ISK), hingga komplikasi yang lebih parah seperti *sindrom uremik hemolitik* (HUS) yang berpotensi mengancam jiwa. Penularan *E. coli* ini umumnya terjadi melalui konsumsi makanan atau air yang terkontaminasi bakteri, serta melalui kontak langsung dengan hewan atau lingkungan yang tercemar oleh kotoran. Kondisi sanitasi yang buruk dan kurangnya pengelolaan lingkungan yang higienis memperbesar risiko penyebaran bakteri ini di masyarakat. Hal tersebut sejalan dengan hasil penelitian yang dilakukan di Rumah Sakit Umum Pusat Sanglah Bali, yang mengidentifikasi *E. coli* sebagai penyebab utama infeksi saluran kemih dengan prevalensi mencapai 39,76%[1]. Studi ini juga menemukan bahwa infeksi *E. coli* lebih sering terjadi pada wanita dan kelompok usia lanjut, menegaskan bahwa faktor biologis dan demografis turut berperan dalam kerentanan terhadap penyakit ini. Selain di fasilitas kesehatan, berbagai penelitian di Indonesia juga mengungkap tingginya tingkat kontaminasi *E. coli* pada makanan dan minuman yang beredar di masyarakat. Penelitian[2] di Kendari Beach menemukan bahwa 81,2% minuman tradisional terkontaminasi *E. coli*, sementara penelitian[3] di Semarang menunjukkan sepertiga jajanan sekolah dasar positif mengandung bakteri ini. Kasus serupa juga terjadi di penelitian[4] pada minuman kekinian di Mamuju, dengan tingkat kontaminasi mencapai lebih dari 90%, serta pada cinau hitam di pasar tradisional Semarang, di mana seluruh sampel yang diuji 100%

positif mengandung *E. coli*. Fakta-fakta ini memperlihatkan bahwa masalah sanitasi pangan masih menjadi tantangan besar di Indonesia, dan keberadaan *E. coli* patogen berpotensi menimbulkan ancaman serius bagi kesehatan masyarakat. Kondisi tersebut menegaskan pentingnya upaya pengendalian sanitasi dan peningkatan kesadaran akan kebersihan lingkungan sebagai langkah pencegahan penyebaran infeksi.

Identifikasi dan klasifikasi *E. coli* merupakan aspek krusial dalam dunia kesehatan karena berperan langsung dalam upaya pencegahan penyebaran infeksi serta pelaksanaan penanganan medis yang akurat dan tepat waktu. Metode identifikasi konvensional seperti kultur bakteri, uji biokimia, dan teknik molekuler seperti *polymerase chain reaction* (PCR) selama ini telah digunakan secara luas dan terbukti memberikan hasil yang akurat. Namun, prosedur-prosedur tersebut sering kali membutuhkan waktu yang cukup lama, biaya operasional yang tinggi, serta memerlukan keahlian khusus di laboratorium yang tidak selalu tersedia di semua fasilitas kesehatan, terutama di daerah dengan sumber daya terbatas. Oleh sebab itu, muncul kebutuhan akan metode alternatif yang dapat melakukan proses klasifikasi secara otomatis dengan kecepatan yang lebih tinggi dan efisiensi yang lebih baik, tanpa mengorbankan akurasi hasil diagnosis. Teknologi *machine learning* hadir sebagai solusi inovatif dalam memenuhi kebutuhan tersebut. Dengan kemampuan untuk memproses data dalam jumlah besar dan menemukan pola yang kompleks secara otomatis, metode ini berpotensi merevolusi cara identifikasi bakteri dilakukan. Penelitian [5] yang dilakukan memperkuat hal ini dengan membandingkan berbagai metode *ensemble* dalam model pohon keputusan untuk klasifikasi bakteri *E. coli*. Hasil penelitian sebelumnya menunjukkan bahwa penggunaan metode *ensemble* mampu meningkatkan akurasi klasifikasi secara signifikan dibandingkan dengan metode konvensional, sekaligus menawarkan efisiensi yang lebih tinggi dalam proses identifikasi.

Dalam bidang klasifikasi, algoritma *Decision Tree* banyak digunakan karena mampu menghasilkan model yang sederhana, mudah dipahami, dan dapat divisualisasikan dalam bentuk struktur hierarkis, sehingga alur pengambilan

keputusannya jelas. Namun, metode ini memiliki kelemahan berupa kecenderungan mengalami *overfitting* pada data latih yang dapat menurunkan akurasi saat diuji dengan data baru. Untuk mengatasi hal tersebut, dikembangkan algoritma *Random Forest* sebagai metode *ensemble* yang membangun banyak pohon keputusan sekaligus, lalu menggabungkan hasil prediksi setiap pohon agar diperoleh output yang lebih akurat, stabil, dan tahan terhadap *overfitting*. Kedua algoritma ini telah terbukti efektif dalam menyelesaikan permasalahan klasifikasi, termasuk dalam konteks data biologis seperti bakteri *E. coli* yang memiliki banyak atribut dan kompleksitas. Oleh karena itu, penelitian ini dilakukan untuk menerapkan sekaligus membandingkan kinerja algoritma *Decision Tree* dan *Random Forest* dalam klasifikasi bakteri *E. coli*, dengan evaluasi menggunakan metrik akurasi, *precision*, *recall*, dan *F1-score*. Hasil yang diperoleh diharapkan dapat memberikan kontribusi nyata dalam pengembangan sistem klasifikasi otomatis berbasis *machine learning*, yang berpotensi dimanfaatkan pada bidang biomedis, penelitian mikrobiologi, maupun sebagai pendukung proses diagnostik awal di laboratorium kesehatan.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan, maka rumusan masalah dalam penelitian ini adalah sebagai berikut:

Bagaimana menerapkan dan membandingkan algoritma *Decision Tree* dan *Random Forest* dalam mengklasifikasikan bakteri *E. coli*?

1.3 Batasan Masalah

Untuk memfokuskan ruang lingkup penelitian, maka batasan masalah dalam penelitian ini ditentukan sebagai berikut:

1. Penelitian hanya membahas klasifikasi data bakteri *Escherichia coli* menggunakan dua algoritma *machine learning*, yaitu *Decision Tree* dan *Random Forest*.
2. Dataset yang digunakan merupakan data numerik yang mewakili karakteristik biologis bakteri *E. coli*.

3. Proses evaluasi performa model menggunakan metrik akurasi, *precision*, *recall*, dan *F1-score*.
4. Penelitian ini tidak meliputi pengembangan atau penerapan sistem klasifikasi dalam bentuk perangkat lunak siap digunakan maupun aplikasi praktis yang dioperasikan langsung di lingkungan laboratorium sesungguhnya.

1.4 Tujuan Penelitian

Adapun tujuan penelitian ini meliputi:

1. Mengimplementasikan algoritma *Decision Tree* dan *Random Forest* untuk mengklasifikasikan data bakteri *Escherichia coli*.
2. Menganalisis performa antara algoritma *Decision Tree* dan *Random Forest* dengan menggunakan metrik evaluasi seperti akurasi, *precision*, *recall*, dan *F1-score*.
3. Memberikan rekomendasi terkait algoritma klasifikasi yang paling efektif dan akurat dalam mengolah data bakteri *E. coli* pada penerapan *machine learning*.

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan kontribusi yang berarti baik dari sisi teoritis maupun praktis. Adapun manfaat yang diharapkan dari penelitian ini adalah sebagai berikut:

1. Manfaat Teoritis

Penelitian ini diharapkan dapat memberikan ilmu pengetahuan, khususnya dalam bidang ilmu komputer dan bioteknologi, dengan menambah wawasan terkait penerapan algoritma *Decision Tree* dan *Random Forest* dalam klasifikasi data mikroorganisme, khususnya bakteri *E. coli*. Selain itu, hasil penelitian ini diharapkan menjadi acuan bagi pengembangan metode klasifikasi berbasis *machine learning* di bidang yang sama.

2. Manfaat Praktis

Secara praktis, penelitian ini menyediakan metode klasifikasi otomatis yang dapat membantu tenaga medis dan peneliti mikrobiologi dalam mempercepat proses identifikasi serta klasifikasi bakteri *E. coli*. Dengan demikian,

diharapkan penelitian ini dapat mendukung efisiensi dan akurasi pengambilan keputusan di lingkungan laboratorium maupun bidang kesehatan.

1.6 Sistematika Penulisan

Penulisan skripsi ini disusun dengan sistematika sebagai berikut:

Bab I PENDAHULUAN

Bab ini berisi latar belakang penelitian, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, serta sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Berisi kajian teori terkait bakteri *Escherichia coli*, konsep klasifikasi, penjelasan algoritma, serta penelitian terdahulu yang relevan.

BAB III METODE PENELITIAN

Menjelaskan metode penelitian yang digunakan, termasuk sumber data, proses pengolahan data, tahapan implementasi algoritma, serta teknik evaluasi model klasifikasi.

BAB IV HASIL DAN PEMBAHASAN

Menyajikan hasil pengujian dan analisis kinerja algoritma Decision Tree dan Random Forest.

BAB V KESIMPULAN DAN SARAN

Berisi kesimpulan serta saran yang dapat diberikan untuk pengembangan penelitian.