

## **BAB I**

### **PENDAHULUAN**

#### **1.1 Latar Belakang**

Kulit merupakan organ tubuh pada manusia yang sangat penting karena terletak pada bagian luar tubuh yang berfungsi untuk menerima rangsangan seperti sentuhan, rasa sakit dan pengaruh lainnya (Putri *et al.*, 2019). Setiap individu memiliki kondisi kulit yang bervariasi, tergantung pada jenis kulit masing-masing. Beragam faktor dapat memengaruhi kondisi tersebut dan berpotensi menimbulkan penyakit, seperti kebersihan, pola makan, kondisi fisik, paparan bahan kimia, sistem imun, serta faktor lingkungan. (Wijaya *et al.*, 2023). Penyakit kulit merupakan salah satu jenis penyakit yang umum terjadi di negara-negara beriklim tropis seperti Indonesia. Berdasarkan Profil Kesehatan Indonesia Tahun 2023, penyakit kulit termasuk dalam kategori penyakit dengan angka kunjungan rawat jalan yang tinggi di berbagai rumah sakit di Indonesia (Iskandar & Salam, 2024).

Rumah Sakit Cut Meutia merupakan salah satu rumah sakit rujukan utama di wilayah Lhokseumawe dan Aceh Utara yang melayani ribuan pasien setiap tahunnya. Salah satu kelompok penyakit yang cukup sering ditangani di rumah sakit ini adalah penyakit kulit. Informasi pasien seperti diagnosis, gejala, hasil pemeriksaan, serta riwayat pengobatan terdokumentasi secara menyeluruh dalam jumlah besar di sistem informasi rumah sakit. Data tersebut memiliki potensi besar untuk dianalisis lebih lanjut guna mengungkap pola penyakit kulit yang sering terjadi, mengidentifikasi faktor-faktor pemicunya, serta menilai efektivitas intervensi medis yang telah dilakukan. Sayangnya, keterbatasan waktu dan sumber daya membuat analisis manual terhadap data ini menjadi tantangan tersendiri bagi tenaga medis.

Sejalan dengan kemajuan teknologi, khususnya di bidang kecerdasan buatan dan data mining, kini tersedia berbagai metode komputasional yang mampu membantu tenaga medis menganalisis data kesehatan secara lebih cepat dan akurat. Kemajuan tersebut dapat dimanfaatkan untuk analisis data kesehatan secara

mendalam, termasuk mengelompokkan jenis penyakit berdasarkan karakteristik tertentu melalui proses klasifikasi. Klasifikasi sendiri merupakan proses identifikasi dan pengelompokan suatu objek ke dalam kelompok atau kategori yang sama (Melfia *et al.*, 2025). Dalam konteks medis, penerapan teknik klasifikasi memungkinkan sistem untuk mengenali pola dari data gejala pasien dan mengelompokkannya ke dalam jenis penyakit tertentu. Hal ini menjadi dasar pengembangan sistem bantu diagnosis digital yang mampu meningkatkan efisiensi serta akurasi dalam proses penanganan penyakit.

Menurut penelitian yang dilakukan oleh Hikmayanti *et al.* (2023) dengan judul Implementasi Algoritma Logistic Regression untuk Klasifikasi Penyakit Stroke, hasil penelitian menunjukkan bahwa algoritma *Logistic Regression* memiliki performa yang sangat baik dalam mengklasifikasikan penyakit stroke, dengan akurasi mencapai 94%. Model ini juga menunjukkan nilai *precision* sebesar 94%, *recall* 100%, serta *F1-score* 97%, yang mengindikasikan keseimbangan antara ketepatan dan sensitivitas. Jika dibandingkan dengan algoritma lain seperti *Support Vector Machine* (SVM), yang hanya mencapai akurasi 76%, *Logistic Regression* terbukti lebih unggul dan efektif dalam mendeteksi kasus stroke berdasarkan dataset yang digunakan. Temuan ini mengindikasikan bahwa pendekatan berbasis *Logistic Regression* berpotensi besar untuk diterapkan dalam sistem deteksi dini guna mendukung penanganan penyakit stroke secara lebih cepat dan akurat.

Selain itu, penelitian oleh Apriliah *et al.* (2021) yang berjudul Prediksi Kemungkinan Diabetes pada Tahap Awal Menggunakan Algoritma Klasifikasi Random Forest, menunjukkan bahwa algoritma *Random Forest* merupakan metode yang paling efektif dalam memprediksi diabetes, dengan tingkat akurasi mencapai 97,88%. Selain akurasi yang tinggi, algoritma ini juga mencatat nilai *precision* sebesar 0,979, *recall* sebesar 0,979, *F-measure* sebesar 0,979, serta ROC sebesar 0,998, yang menunjukkan kemampuan klasifikasi yang sangat baik dan stabil. Temuan ini menegaskan bahwa algoritma *Random Forest* tidak hanya unggul dari sisi akurasi, tetapi juga konsisten dalam berbagai metrik evaluasi lainnya. Dengan performa tersebut, *Random Forest* berpotensi digunakan secara

efektif dalam diagnosis dini penyakit diabetes serta dapat diadaptasi untuk prediksi penyakit lainnya di masa depan.

Berdasarkan hasil temuan tersebut, dapat disimpulkan bahwa algoritma *Logistic Regression* dan *Random Forest* memiliki potensi yang signifikan dalam mendukung proses klasifikasi penyakit, khususnya dalam bidang medis yang memerlukan akurasi tinggi dan pengambilan keputusan yang cepat. Dalam konteks penyakit kulit yang memiliki beragam gejala dan faktor penyebab, penerapan algoritma klasifikasi menjadi semakin penting karena dapat membantu mengidentifikasi jenis penyakit secara lebih sistematis, akurat, dan berbasis pada data historis. Pemanfaatan algoritma klasifikasi ini juga sangat relevan dalam mendukung sistem bantu diagnosis medis di fasilitas layanan kesehatan rumah sakit. Sistem tersebut memungkinkan tenaga medis untuk memperoleh rekomendasi diagnosis secara cepat dan berbasis data ketika menghadapi pasien dengan gejala tertentu. Dengan demikian, proses diagnosis tidak hanya menjadi lebih efisien, tetapi juga lebih tepat sasaran dalam menentukan tindakan medis yang sesuai.

Melihat potensi yang ditawarkan oleh kedua algoritma tersebut serta pentingnya deteksi dini penyakit kulit di fasilitas kesehatan, maka penelitian ini mengangkat judul “*Perbandingan Metode Logistic Regression dan Random Forest dalam Klasifikasi Penyakit Kulit Multikelas*”. Judul ini dipilih untuk mencerminkan fokus utama penelitian, yakni mengevaluasi dan membandingkan performa kedua algoritma dalam klasifikasi penyakit kulit berdasarkan data gejala yang bersumber dari data rekam medis pasien di Rumah Sakit Umum Cut Meutia. Melalui proses pelatihan dan pengujian model, Tujuan dari penelitian ini adalah untuk mengetahui algoritma mana yang menunjukkan hasil terbaik dari segi akurasi, precision, recall, dan F1-score. Temuan dari penelitian ini diharapkan dapat memberikan kontribusi tidak hanya bagi pengembangan keilmuan di bidang pengolahan data kesehatan, tetapi juga dapat diimplementasikan secara nyata untuk mendukung sistem pelayanan medis, khususnya dalam diagnosis dan penanganan penyakit kulit secara lebih optimal dan efisien.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, maka rumusan masalah yang akan dibahas dalam penelitian ini adalah:

1. Bagaimana tingkat akurasi yang dihasilkan oleh metode *Logistic Regression* dan *Random Forest* dalam mengklasifikasikan jenis penyakit kulit?
2. Metode mana yang menunjukkan tingkat efektivitas dan efisiensi antara *Logistic Regression* dan *Random Forest* dalam penerapan klasifikasi penyakit kulit berdasarkan dataset yang digunakan?

## 1.3 Tujuan Penelitian

Tujuan utama dari penelitian ini adalah

1. Menganalisis performa metode *Logistic Regression* dan *Random Forest* dalam mengklasifikasikan jenis penyakit kulit berdasarkan dataset yang digunakan.
2. Membandingkan tingkat akurasi, presisi, recall, dan F1-score dari kedua metode untuk menentukan metode yang lebih efektif dan efisien dalam klasifikasi penyakit kulit.
3. Memberikan rekomendasi metode terbaik untuk digunakan dalam pengembangan sistem klasifikasi penyakit kulit berbasis kecerdasan buatan.

## 1.4 Manfaat Penelitian

Manfaat dari penelitian ini adalah:

1. Mendukung perkembangan pengetahuan di bidang kecerdasan buatan dan data mining, terutama dalam cara menganalisis dan menerapkan metode klasifikasi pada data kesehatan.
2. Menyediakan informasi bagi praktisi kesehatan atau pengembang aplikasi medis tentang metode yang lebih efektif dalam mengklasifikasikan penyakit kulit, sehingga dapat membantu dalam pembuatan sistem diagnosis berbasis teknologi.
3. Memberikan dukungan terhadap pengembangan sistem berbasis kecerdasan buatan dalam rangka meningkatkan kecepatan, akurasi, dan efisiensi diagnosis

penyakit kulit.

### **1.5 Ruang Lingkup dan Batasan Masalah**

Merujuk pada rumusan masalah yang telah diuraikan sebelumnya, ruang lingkup pembahasan dalam tugas akhir ini dibatasi pada hal-hal berikut:

1. Penelitian ini hanya membandingkan dua metode, yaitu *Logistic Regression* dan *Random Forest*, tanpa melibatkan metode lain.
2. Penelitian tidak membahas secara rinci penyebab penyakit kulit atau langkah-langkah pengobatannya, tetapi berfokus pada aspek teknis klasifikasi data penyakit kulit.
3. Data yang digunakan dalam penelitian ini berasal dari Rumah Sakit Cut Meutia dan diambil berdasarkan informasi yang tercantum dalam rekam medis pasien.
4. Proses klasifikasi dalam penelitian ini dibatasi hanya pada 8 jenis penyakit kulit yaitu abses, dermatitis, pemvigus vulgaris, selulitis, impetigo, ukus dekubitus, urtikaria dan psoriasis.
5. Implementasi dilakukan menggunakan perangkat lunak atau *library* yang mendukung metode *Logistic Regression* dan *Random Forest*, seperti Python (*Scikit-learn*) atau *tools* serupa.