

BAB I

PENDAHULUAN

1.1 Latar Belakang

Penyakit Ginjal Kronis (Chronic Kidney Disease/CKD) merupakan gangguan fungsi ginjal yang bersifat progresif dan *irreversibel*, ditandai dengan penurunan laju filtrasi glomerulus (LFG) dan/atau adanya kerusakan ginjal yang berlangsung lebih dari tiga bulan. Penurunan fungsi ginjal ini terjadi secara bertahap akibat proses patologi kronis yang menyebabkan kehilangan nefron secara permanen sehingga berkontribusi terhadap peningkatan morbiditas dan mortalitas global [1], [2]. CKD umumnya berkembang secara perlahan dan sering kali tidak menimbulkan gejala klinis pada tahap awal karena mekanisme kompensasi ginjal masih mampu mempertahankan fungsi filtrasi yang adekuat [2]. Akibatnya, sebagian besar kasus baru terdiagnosis pada stadium lanjut ketika fungsi ginjal telah menurun secara signifikan [1]. Pada tahap lanjut, CKD meningkatkan risiko terjadinya komplikasi serius seperti penyakit kardiovaskular, gangguan metabolik, serta progresi menuju gagal ginjal terminal yang memerlukan terapi pengganti ginjal berupa dialisis atau transplantasi [1], [2]. Di Indonesia, CKD juga menimbulkan beban yang signifikan terhadap sistem pelayanan kesehatan. Berdasarkan Profil Kesehatan Indonesia Tahun 2024, gagal ginjal termasuk dalam kelompok penyakit *katastrofik* dengan jumlah kasus layanan gagal ginjal yang ditangani dan dibiayai melalui Program Jaminan Kesehatan Nasional (JKN) mencapai lebih dari 1,4 juta kasus pada tahun 2024, dengan total pembiayaan sekitar Rp2,7 triliun [3]. Tingginya beban layanan ini menunjukkan bahwa CKD tidak hanya berdampak pada kualitas hidup pasien, tetapi juga memberikan tekanan besar terhadap sistem pembiayaan dan pelayanan kesehatan nasional [3].

Permasalahan utama dalam penanganan Penyakit Ginjal Kronis (CKD) terletak pada keterlambatan dalam proses identifikasi dan deteksi dini penyakit. CKD sering kali tidak menunjukkan gejala klinis yang spesifik pada tahap awal sehingga diagnosis umumnya ditegakkan ketika telah terjadi penurunan fungsi

ginjal yang signifikan [1], [2]. Metode konvensional yang digunakan dalam praktik klinis, seperti estimasi laju filtrasi glomerulus (*estimated glomerular filtration rate/eGFR*), kadar kreatinin serum, dan *blood urea nitrogen* (BUN), memiliki keterbatasan karena parameter tersebut baru mengalami perubahan bermakna setelah terjadi kehilangan fungsi ginjal yang cukup besar [1]. Selain itu, kadar kreatinin serum dapat dipengaruhi oleh faktor usia, jenis kelamin, massa otot, dan status nutrisi sehingga sensitivitasnya dalam mendeteksi gangguan ginjal tahap awal menjadi terbatas [2], [4]. Keterbatasan ini menyebabkan peluang intervensi dini sering terlewatkan, padahal deteksi dan penatalaksanaan CKD pada tahap awal terbukti dapat memperlambat progresivitas penyakit, menurunkan risiko komplikasi kardiovaskular, serta mengurangi angka mortalitas [1].

Perkembangan *machine learning* dalam bidang kesehatan menghadirkan pendekatan yang lebih adaptif untuk memprediksi CKD berdasarkan data klinis pasien [5]. Algoritma *machine learning* mampu mengolah data klinis dan laboratorium dalam jumlah besar untuk mengidentifikasi pola kompleks dan hubungan nonlinier yang sulit ditangkap oleh metode statistik tradisional [6]. Pendekatan ini dilaporkan mampu meningkatkan akurasi prediksi CKD dibandingkan metode konvensional [7], [8]. Di antara berbagai algoritma yang digunakan, *K-Nearest Neighbors* (KNN) dan *Naive Bayes* merupakan metode yang paling sering diterapkan dalam penelitian CKD karena memiliki struktur yang relatif sederhana, efisiensi komputasi yang baik, serta kemampuan klasifikasi yang memadai pada data medis [7].

Meskipun KNN dan *Naive Bayes* telah banyak digunakan dalam prediksi CKD, berbagai penelitian menunjukkan bahwa kinerja kedua algoritma tersebut dapat berbeda secara signifikan tergantung pada karakteristik data dan asumsi model yang digunakan [8]. Kinerja KNN sangat dipengaruhi oleh pemilihan jumlah tetangga (k) dan distribusi data [9]. Sementara itu, *Naive Bayes* bergantung pada asumsi independensi antar fitur yang tidak selalu terpenuhi pada data klinis [6]. Namun demikian, kajian yang secara khusus membandingkan kinerja algoritma KNN dan *Naive Bayes* dalam memprediksi Penyakit Ginjal Kronis

menggunakan dataset terkini masih relatif terbatas. Oleh karena itu, penelitian ini dilakukan untuk menganalisis dan membandingkan kinerja algoritma KNN dan *Naive Bayes* dalam memprediksi CKD guna memperoleh algoritma dengan performa terbaik sebagai dasar pengembangan sistem pendukung keputusan di bidang kesehatan.

1.2 Rumusan Masalah

Rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana kinerja algoritma *K-Nearest Neighbors* (KNN) dan *Naive Bayes* dalam memprediksi Penyakit Ginjal Kronis (CKD) ?
2. Bagaimana pengaruh penerapan *hyperparameter optimization* terhadap peningkatan performa model *K-Nearest Neighbors*, dan *Naive Bayes* dalam memprediksi Penyakit Ginjal Kronis (CKD)?
3. Model manakah yang memiliki hasil prediksi Penyakit Ginjal Kronis (CKD) paling optimal berdasarkan metrik evaluasi klasifikasi, yaitu *accuracy*, *precision*, *recall*, dan *F1-score*, dengan fokus utama pada *recall*?

1.3 Batasan Masalah

Batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Penelitian ini dibatasi pada dataset CKD dari *UCI Machine Learning Repository* yang dikumpulkan dalam periode sekitar dua bulan.
2. Penelitian ini hanya menguji dua algoritma *machine learning*, yaitu *K-Nearest Neighbors* (KNN) dan *Naive Bayes*, dengan penerapan *hyperparameter optimization* menggunakan teknik *Grid Search* untuk meningkatkan kinerja model dalam memprediksi Penyakit Ginjal Kronis (CKD).
3. Evaluasi kinerja model dibatasi pada metrik *accuracy*, *precision*, *recall*, dan *F1-score* dengan fokus utama *matrix recall*, serta implementasi model ke dalam sistem berbasis web sebagai simulasi

prediksi secara komputasional.

1.4 Tujuan Penelitian

Tujuan penelitian ini adalah sebagai berikut:

1. Mengimplementasikan dan membandingkan algoritma *K-Nearest Neighbors* (KNN) dan *Naive Bayes* untuk menentukan algoritma yang paling akurat dan efisien diantara kedua algoritma tersebut dalam memprediksi Penyakit Ginjal Kronis (CKD) berdasarkan matrix evaluasi, dengan fokus utama pada *recall*
2. Menganalisis pengaruh penerapan *hyperparameter optimization* menggunakan teknik *Grid Search* terhadap peningkatan performa masing-masing model *machine learning* dalam memprediksi Penyakit Ginjal Kronis (CKD).

1.5 Manfaat Penelitian

Manfaat yang diharapkan dari penelitian ini :

1. Untuk mendukung proses prediksi Penyakit Ginjal Kronis (CKD) secara lebih akurat, sehingga dapat membantu upaya deteksi dini dan pengambilan keputusan awal dalam penanganan pasien, serta memungkinkan pemanfaatan sumber daya kesehatan secara lebih tepat dan efisien.
2. Untuk menghasilkan model prediksi berbasis algoritma *machine learning* yang andal dan efisien, sehingga dapat mendukung upaya deteksi dini dan pengambilan keputusan awal secara komputasional berdasarkan data klinis pasien.
3. Untuk memberikan dukungan dalam proses pengambilan keputusan, khususnya sebagai bahan pertimbangan dalam perencanaan dan penentuan prioritas penanganan Penyakit Ginjal Kronis (CKD) melalui pemanfaatan hasil prediksi yang dihasilkan oleh model.

1.6 Sistematika Penulisan

Sistematika penulisan skripsi ini disusun untuk memberikan gambaran mengenai struktur dan alur pembahasan yang disajikan. Skripsi ini terdiri atas lima bab, yang masing-masing bab diuraikan sebagai berikut:

BAB I PENDAHULUAN

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

BAB II TINJAUAN PUSTAKA

Bab ini memuat kajian pustaka dan landasan teori yang mendukung penelitian. Pembahasan meliputi konsep Penyakit Ginjal Kronis (CKD), *machine learning*, *data preprocessing*, *hyperparameter optimization*, , serta teori klasifikasi. Selain itu, bab ini juga menguraikan algoritma *K-Nearest Neighbors* (KNN) dan *Naive Bayes*, serta metrik evaluasi kinerja model yang digunakan dalam penelitian.

BAB III METODE PENELITIAN

Bab ini menjelaskan metode penelitian yang digunakan, meliputi alur penelitian, sumber dan karakteristik dataset, tahapan *preprocessing data*, pembagian data latih dan data uji, proses pemodelan menggunakan algoritma *K-Nearest Neighbors* (KNN) dan *Naive Bayes*, penerapan *hyperparameter optimization* dengan teknik *Grid Search*, serta metode evaluasi kinerja model.

BAB IV HASIL DAN PEMBAHASAN,

Bab ini menyajikan hasil penelitian yang diperoleh dari proses pengujian dan evaluasi model. Pembahasan mencakup hasil *preprocessing data*, kinerja masing-masing algoritma sebelum dan sesudah penerapan *Grid Search*, perbandingan performa KNN dan *Naive Bayes* berdasarkan metrik *accuracy*, *precision*, *recall*, dan *F1-score*, serta analisis terhadap hasil yang diperoleh.

BAB V PENUTUP

Bab ini berisi kesimpulan yang merangkum hasil penelitian sesuai dengan tujuan yang telah ditetapkan, serta saran yang dapat dijadikan sebagai bahan

pertimbangan untuk penelitian selanjutnya, khususnya terkait pengembangan metode, algoritma, maupun penggunaan dataset yang lebih luas.

