

BAB I

PENDAHULUAN

1.1 Latar Belakang

Penyakit gagal ginjal kronis merupakan masalah kesehatan global yang serius dengan tingkat prevalensi yang terus meningkat. Pada tahun 2017 terdapat 697,5 juta kasus penyakit gagal ginjal kronis di seluruh dunia, dengan peningkatan prevalensi sebesar 29,3% sejak tahun 1990 [1]. Di Indonesia, berdasarkan Riset Kesehatan Dasar 2018 [2], prevalensi penyakit gagal ginjal kronis mencapai 3,8 per 1000 penduduk, dengan estimasi peningkatan menjadi 4,17% pada tahun 2021 [3].

Penyakit gagal ginjal kronis tidak hanya mempengaruhi kualitas hidup penderitanya, tetapi juga membebani sistem kesehatan secara signifikan. Biaya perawatan penyakit gagal ginjal kronis, terutama untuk pasien yang memerlukan dialisis atau transplantasi ginjal, sangat tinggi dan terus meningkat setiap tahun [4]. Selain itu, penyakit gagal ginjal kronis juga meningkatkan risiko penyakit kardiovaskular dan kematian dini, dengan estimasi menyebabkan 1,2 juta kematian secara langsung pada tahun 2017 [1].

Keterlambatan diagnosis menjadi salah satu faktor utama yang memperburuk prognosis penyakit gagal ginjal kronis. Hal ini disebabkan oleh beberapa faktor seperti kompleksitas gejala dimana gejala awal penyakit gagal ginjal kronis seringkali tidak spesifik dan dapat diabaikan oleh pasien [5], keterbatasan akses kesehatan terutama di negara berkembang, akses terhadap fasilitas kesehatan yang memadai masih menjadi tantangan [6], serta kurangnya kesadaran baik di kalangan masyarakat maupun tenaga kesehatan akan pentingnya deteksi dini penyakit gagal ginjal kronis masih perlu ditingkatkan [7].

Data Mining adalah proses analisis data yang bertujuan untuk menemukan pola, tren, dan informasi berharga dari kumpulan data yang besar dan kompleks. Proses ini melibatkan beberapa langkah penting, termasuk pengumpulan, pembersihan, dan transformasi data, serta penerapan teknik analisis seperti

klasifikasi, clustering, dan regresi untuk mengekstrak informasi yang berguna. Dalam konteks penyakit gagal ginjal kronis, teknik *Data Mining* dapat digunakan untuk menganalisis data pasien, termasuk riwayat medis, hasil laboratorium, dan faktor risiko lainnya [8]. Dengan menerapkan Algoritma *Machine Learning*, *Data Mining* dapat membantu dalam mengidentifikasi faktor-faktor yang berkontribusi terhadap perkembangan penyakit serta memprediksi kemungkinan progresivitasnya [9].

Klasifikasi adalah salah satu teknik dalam *Data Mining* yang digunakan untuk mengelompokkan data ke dalam kategori atau kelas yang telah ditentukan sebelumnya. Proses klasifikasi bertujuan untuk memprediksi kelas atau label dari data baru berdasarkan pola yang ditemukan dalam data pelatihan yang sudah ada. Teknik klasifikasi dapat membantu dalam menentukan status kesehatan pasien berdasarkan data yang tersedia [10]. Dengan memanfaatkan *Data Mining*, beberapa Algoritma ML, seperti *Random Forest* dan *Gradient Boosting*, telah menunjukkan performa yang menjanjikan dalam berbagai kasus prediksi medis [11].

Salah satu Algoritma klasifikasi yang paling banyak digunakan saat ini adalah *Random Forest*, yang dikenal efektif dalam mengatasi masalah *overfitting*. Dengan memanfaatkan pendekatan *ensemble* yang menggabungkan prediksi dari berbagai pohon keputusan, *Random Forest* mampu memberikan hasil prediktif yang akurat dan stabil, terutama pada dataset dengan banyak fitur dan data yang hilang. Studi oleh [12] menunjukkan bahwa, meskipun *XGBoost* unggul dalam beberapa aspek, *Random Forest* tetap kompetitif dalam klasifikasi permukaan perkotaan yang kompleks, menunjukkan akurasi yang tinggi pada dataset multi-sensor yang memerlukan pemrosesan fitur yang beragam.

Sementara itu, *Gradient Boosting* juga dikenal unggul dalam klasifikasi data yang kompleks, terutama ketika digunakan pada dataset yang kecil dan presisi tinggi. Algoritma ini mampu menghasilkan akurasi yang lebih baik pada data dengan karakteristik khusus, meskipun lebih rentan terhadap *overfitting* jika pengaturannya kurang optimal [13]. Penelitian oleh [14] menekankan bahwa *Gradient Boosting* mampu memberikan performa tinggi, terutama dalam konteks

klasifikasi dengan variasi fitur yang besar dan kebutuhan akan interpretasi yang akurat. Oleh karena itu, baik Algoritma *Random Forest* maupun *Gradient Boosting* diimplementasikan pada penelitian ini.

Didasarkan atas permasalahan dan pembahasan yang telah dipaparkan, masih terdapat kesenjangan dalam penelitian yang membandingkan efektivitas berbagai Algoritma *Machine Learning*, khususnya *Random Forest* dan *Gradient Boosting* dalam konteks prediksi penyakit gagal ginjal kronis. Penelitian ini bertujuan untuk mengisi kesenjangan tersebut dengan melakukan analisis komparatif yang komprehensif. Melalui penelitian ini, perbandingan Algoritma *Random Forest* dan *Gradient Boosting* digunakan untuk memprediksi penyakit gagal ginjal kronis sehingga, hasil penelitian ini nantinya dapat memberikan kontribusi signifikan dalam pengembangan sistem pendukung keputusan klinis yang dapat membantu tenaga medis dalam mendiagnosis penyakit gagal ginjal kronis secara lebih dini dan akurat, sehingga dapat meningkatkan kualitas perawatan pasien dan mengurangi beban ekonomi akibat penyakit gagal ginjal kronis.

1.2 Rumusan Masalah

Berdasarkan latar belakang penelitian ini maka permasalahan dapat dirumuskan adalah:

1. Bagaimana perbandingan kinerja antara Algoritma *Random Forest* dan *Gradient Boosting* dalam memprediksi diagnosis penyakit gagal ginjal kronis?
2. Apa saja faktor-faktor yang berkontribusi terhadap perkembangan penyakit gagal ginjal kronis yang dapat diidentifikasi melalui teknik *Data Mining*?

1.3 Batasan Masalah

Berdasarkan latar belakang dan rumusan masalah yang sudah dipaparkan sebelumnya, ada batasan yang telah ditentukan untuk mempermudah dalam proses pembuatan penelitian ini, yaitu:

1. Dataset yang digunakan dalam penelitian ini terbatas pada data medis yang

bersumber dari Kaggle.com.

2. Algoritma yang dibandingkan hanya terbatas pada *Random Forest* dan *Gradient Boosting*.
3. Variabel target dalam penelitian ini adalah status penyakit ginjal kronis (Positif atau Negatif).
4. Evaluasi performa Algoritma akan menggunakan metrik akurasi, presisi, recall, dan F1-score.

1.4 Tujuan Penelitian

Adapun tujuan dari penyusunan penelitian ini yang akan dicapai adalah:

1. Menghasilkan model prediksi penyakit gagal ginjal kronis menggunakan Algoritma *Random Forest* dan *Gradient Boosting* berdasarkan dataset *Chronic Kidney Disease data* dari Kaggle.
2. Menganalisis performa kedua Algoritma (*Random Forest* dan *Gradient Boosting*) dalam memprediksi penyakit gagal ginjal kronis dengan menggunakan metrik evaluasi seperti accuracy, precision, recall, dan F1-score.
3. Menentukan Algoritma terbaik antara *Random Forest* dan *Gradient Boosting* untuk prediksi penyakit gagal ginjal kronis.

1.5 Manfaat Penelitian

Adapun manfaat dalam penyusunan penelitian ini adalah:

1. Memberikan pemahaman yang lebih mendalam tentang perbandingan Algoritma *Random Forest* dan *Gradient Boosting* dalam memprediksi penyakit gagal ginjal kronis untuk meningkatkan efektivitas dan efisiensi diagnosis.
2. Penelitian ini memberikan panduan bagi peneliti dalam menggunakan metode *Machine Learning* untuk memprediksi penyakit gagal ginjal kronis secara lebih efektif.
3. Hasil penelitian ini akan memberikan panduan praktis pada pihak terkait, seperti rumah sakit, dokter, dan tenaga medis untuk mengidentifikasi serta

menangani kasus penyakit gagal ginjal kronis pada tahap awal sebelum mencapai kondisi yang lebih serius.

1.6 Sistematika Penulisan

BAB I PENDAHULUAN

Bab I pendahuluan ini memberikan arahan tentang masalah yang akan dibahas mengenai latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Bab II tinjauan pustaka tentang pembahasan penelitian sebelumnya yang sudah dilakukan peneliti lain sebagai referensi penelitian ini dijalankan dan berisikan dasar teori-teori yang akan digunakan dalam penelitian.

BAB III METODE PENELITIAN

Bab III menguraikan tentang alur penelitian dalam mendapatkan hasil dan juga pengujian yang dilakukan dengan menerapkan metode *Random Forest* dan *Gradient Boosting*.

BAB IV HASIL DAN PEMBAHASAN

Bab IV tentang tahapan yang dilakukan dalam penerapan alur yang telah dirancang meliputi rancangan implementasi di dalam kode untuk menghasilkan klasifikasi.

BAB V PENUTUP

Bab V berisi kesimpulan yang diperoleh dari perumusan masalah dan saran yang diharapkan dari penelitian metode *Random Forest* dan *Gradient Boosting*. Bab ini juga membahas bagaimana hasil penelitian ini bisa membantu dalam prediksi penyakit gagal ginjal kronis di masa yang akan datang.