

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Kesehatan merupakan hal paling utama dalam hidup, sehingga manusia melakukan banyak cara untuk memperoleh kondisi badan yang sehat disetiap harinya. Dengan mendapatkan badan yang sehat manusia dapat melakukan aktifitas sehari-hari lebih menyenangkan. Tetapi kenyatannya banyak manusia ketika sudah menginjak dewasa mulai terserang penyakit seperti Penyakit Ginjal Kronis, dimana diantaranya disebabkan oleh pola hidup tidak sehat. Ginjal adalah salah satu organ yang sangat penting yang berfungsi untuk menjaga komposisi darah dengan mencegah menumpuknya limbah atau kotoran dan mengatur keseimbangan cairan di dalam tubuh, menjaga tingkatan elektrolit seperti sodium, potasium dan fosfat tetap stabil, serta memproduksi hormon dan enzim yang membantu tubuh dalam mengendalikan tekanan darah, membuat sel darah merah dan menjaga tulang tetap kuat dan sehat (Kemenkes, 2017)[1].

Penyakit Ginjal Kronis (*Chronic Kidney Disease*) merupakan masalah kesehatan yang banyak diderita oleh masyarakat di seluruh dunia, khususnya pada negara-negara yang memiliki tingkat penghasilan rendah ataupun menengah. Penyakit ginjal kronis adalah keadaan terjadinya penurunan fungsi ginjal yang menyebabkan ginjal tidak dapat membuang racun dan produk sisa dari darah, yang ditandai adanya protein dalam urin serta penurunan laju filtrasi glomerulus. Penyakit ini bersifat progresif dan umumnya tidak dapat pulih kembali (*irreversible*). Gejala penyakit ini umumnya adalah tidak ada nafsu makan, mual,

muntah, pusing, sesak nafas, rasa lelah, edema pada kaki dan tangan, serta uremia (Almatsier, 2006). [2]

Penyakit ginjal kronis dilihat dari kaca mata global mengalami peningkatan jumlah prevalensi sebesar 87% dari tahun 1990 sampai 2016 [3]. Menurut *World Health Organization* (WHO), penyakit gagal ginjal kronis berkontribusi pada beban penyakit dunia dengan angka kematian sebesar 850.000 jiwa per tahun (Pongsibidang, 2016). Menurut studi *Global Burden Disease 2010* yang dilakukan oleh *International Society of Nephrology*, penyakit ginjal kronis telah diangkat sebagai penyebab penting kematian di seluruh dunia dengan jumlah kematian meningkat sebesar 82,3% dalam dua dekade terakhir (J. Adhkrishnan et al, 2014; R. Lozano et al, 2010). Di Indonesia sendiri peningkatan penderita penyakit ini mencapai angka 20%. Pusat data dan informasi Perhimpunan Rumah Sakit Seluruh Indonesia (PDPERSI) yang menyatakan jumlah penderita gagal ginjal kronik diperkirakan sekitar 50 orang per satu juta penduduk. Berdasarkan data dari Indonesia Renal Registry, suatu kegiatan registrasi dari perhimpunan nefrologi Indonesia, pada tahun 2008 jumlah pasien hemodialisa (cuci darah) mencapai 2260 orang dari 2146 orang pada tahun 2007[4]. Penyakit ginjal kronik juga dapat menyebabkan penyakit mematikan lainnya seperti diabetes, tekanan darah tinggi, sakit jantung dan lupus[5]. Hal-hal tersebut menunjukkan bahwa penyakit ginjal kronis memerlukan perhatian lebih, salah satunya dengan penanganan cepat melalui sistem prediksi yang akurat.[2]

Beberapa penelitian telah dilakukan untuk memprediksi Penyakit Ginjal Kronis (PGK) atau Chronic Kidney Diseases (CKD), diantaranya dilakukan oleh Hilda Amalia (2018) tentang Perbandingan Metode Data Mining SVM dan *Neural*

Network untuk Klasifikasi Penyakit Ginjal Kronis dimana hasil yang lebih akurat untuk memprediksi penyakit ginjal kronis (PGK) yaitu SVM, penelitian selanjutnya dilakukan oleh Warid Yunus (2018) tentang Algoritma *K-Nearest Neighbor* Berbasis *Particle Swarm Optimization* Untuk Prediksi Penyakit Ginjal Kronik dimana dengan optimasi menggunakan PSO yang diterapkan pada metode k-NN menghasilkan peningkatan yang signifikan pada akurasi, selanjutnya penelitian dilakukan oleh Toni Arifin dan Daniel Ariesta (2019) tentang Prediksi Penyakit Ginjal Kronis Menggunakan Algoritma *Naïve Bayes Classifier* Berbasis *Particle Swarm Optimization* dimana adanya peningkatan akurasi ketika algoritma *Naïve Bayes* berbasis *Particle Swarm Optimization*, selanjutnya penelitian dilakukan oleh I Gede Aditya Mahardika Pratama (2022) tentang Diagnosis Penyakit Ginjal Kronis dengan Algoritma C4.5, K-Means dan BPSO dimana penerapan seleksi fitur dengan BPSO mampu meningkat kinerja klasifikasi C4.5 dengan nilai accuracy, precision, recall dan f-measure berturut-turut. Selanjutnya penelitian yang dilakukan oleh Sheva Naufal Rifqi tentang Optimasi *Random Forest* Untuk Diagnosis Penyakit Ginjal Kronik Dengan Menggunakan *Particle Swarm Optimization* dimana dengan *Particle Swarn Optimization* (PSO) berhasil mengoptimasi akurasi pemodelan yang dilakukan dengan metode *Random Forest* sebesar 0.867%.

Teknik klasifikasi merupakan salah satu teknik yang dapat digunakan dalam mendiagnosis penyakit ginjal kronis. Dalam penelitian ini menggunakan Algoritma C4.5 dan Algoritma Random Forest. Algoritma C4.5 merupakan salah satu algoritma klasifikasi yang dapat digunakan untuk membentuk pohon keputusan. Pohon keputusan terbentuk atas akar, cabang dan simpul daun [8]. Algoritma C4.5

merupakan pengembangan dari ID3, pada algoritma C4.5 menggunakan gain ratio sebagai kriteria untuk melakukan pemisahan[6]. [6]. *Random forest* (RF) adalah suatu algoritma yang digunakan pada klasifikasi data dalam jumlah yang besar. *Random Forest* (RF) merupakan salah satu algoritma *Machine Learning* yang bekerja dengan mengombinasikan sejumlah algoritma *Decision Tree* dalam pengambilan keputusannya. RF merupakan jenis *Machine Learning* ansambel yang dapat digunakan untuk masalah klasifikasi maupun regresi. Selain kuat terhadap *overfitting*, RF juga memiliki kinerja yang lebih baik jika dibandingkan dengan beberapa *Machine Learning* lain seperti *Support Vector Machine* (SVM) dan *Discriminant Analysis*. [7]

Pendekatan metode penelitian yang dilakukan adalah dengan menggunakan model data mining *decision tree* C4.5 untuk melakukan klasifikasi dan mendapatkan hasil akurasi terbaik dengan pendekatan pembagian data latih dan data uji dengan tiga kategori pendekatan. Hasil yang diharapkan dalam penelitian ini adalah berupa informasi pembentukan pohon keputusan dan hasil uji akurasi terbaik menggunakan data latih sebanyak 70% menghasilkan tingkat akurasi sebesar 95,71%. [8]

Azhima dkk (2022) dalam penelitiannya tentang *hybrid machine learning* untuk memprediksi penyakit jantung, disimpulkan bahwa *hybrid model* dapat meningkatkan akurasi. Klasifikasi dengan *random forest* memberikan nilai akurasi sebesar 83.16% dan klasifikasi dengan *logistic regression* memberikan skor akurasi 77.88%. Dengan *hybrid model*, skor akurasi klasifikasi sebesar 84.48% yang berarti skor akurasi *random forest* meningkat sebesar 1.32% [5].

Penelitian mengenai Penyakit Ginjal Kronis memang sudah pernah dilakukan dan cukup banyak, namun menggunakan algoritme C4.5 dan algoritme *Random Forest* masih sedikit, faktanya Algoritme C4.5 dapat menghasilkan pohon keputusan yang mudah diinterpretasikan, memiliki tingkat akurasi yang dapat diterima, efisien dalam menangani atribut bertipe diskret dan dapat menangani atribut bertipe diskret dan numerik [10]. Sedangkan Algoritme *Random Forest* dapat mengatasi *noise* dan *missing value* serta dapat mengatasi data dalam jumlah yang besar.

Dari uraian diatas penulis mencoba untuk meneliti Penyakit Ginjal Kronis atau Chronic Kidney Diseases (CKD). menggunakan Algoritme *C4.5* dan Algoritme *Random Forest* dan melakukan komparasi Penyakit Ginjal Kronis berdasarkan dataset yang di peroleh dari internet (<https://www.kaggle.com/datasets/mansoordaku/ckdisease>).

1.2. Identifikasi Masalah

Berdasarkan latar belakang yang telah diuraikan, peneliti menemukan pemikiran yang terjadi pada penelitian sebelumnya yaitu penelitian dengan menggunakan berbagai metode klasifikasi meliputi C4.5, *K-Nearest Neighbor*, *K-Means*, *Random Forest*, *Artificial Neural Network* masih dapat dilakukan peningkatan dalam hal akurasi yang dihasilkan. Hal tersebut dapat dilihat karena terjadinya beberapa masalah seperti ketidakseimbangan data akibat terlalu banyaknya atribut dalam sebuah dataset. Oleh sebab itu, untuk mengatasi permasalahan tersebut, penelitian ini dimaksudkan untuk meningkatkan akurasi dengan menerapkan metode klasifikasi dalam objek penelitian ini yaitu menggunakan algoritma C4.5 dan *Random Forest* terhadap klasifikasi penyakit ginjal kronis (PGK) atau Chronic

Kidney Diseases (CKD) dengan Seleksi Fitur untuk mengoptimalkan klasifikasi pada metode algoritma C4.5 dan *Random Forest*.

1.3. Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, penulis menyusun rumusan masalah dalam penelitian ini adalah bagaimana cara menghasilkan model algoritma yang memiliki akurasi yang sangat tinggi dengan melakukan seleksi fitur pada algoritma Decision Tree C4.5 dan algoritma *Random Forest*

1.4. Batasan Masalah

Penelitian ini memiliki beberapa batasan masalah sebagai berikut.

- a. Menggunakan data public Penyakit Ginjal Kronis yang diperoleh dari situs kaggle <https://www.kaggle.com/datasets/mansoordaku/ckdisease>.
- b. Menggunakan algoritme C4.5 dan algoritme *Random Forest* dalam mendiagnosis Penyakit Ginjal Kronis
- c. Menggunakan tool google colaboratory, dimana memungkinkan siapa saja menulis dan mengeksekusi kode python arbitrer melalui browser, dan sangat cocok untuk machine learning, analisis data, serta pendidikan.[11]

1.5. Tujuan Penelitian

Penelitian ini memiliki beberapa tujuan penelitian sebagai berikut:

- a. Untuk mengukur tingkat akurasi dari penggunaan algoritme C4.5 dan algoritme *Random Forest* dalam mendiagnosis Penyakit Ginjal Kronis.
- b. Melakukan seleksi fitur dan dilakukan pengujian pada algoritme C4.5 dan algoritme *Random Forest* dalam mendiagnosis Penyakit Ginjal Kronis.

- c. Membandingkan nilai hasil pengujian pada algoritme C4.5 dan algoritme *Random Forest* sebelum dan sesudah seleksi fitur.

1.6. Manfaat Penelitian

Penelitian ini memiliki beberapa manfaat penelitian sebagai berikut.

- a. Tersedianya Prototipe tool yang dapat digunakan untuk mendeteksi penyakit ginjal kronis.
- b. Hasil klasifikasi pada algoritme C4.5 dan algoritme *Random Forest* menggunakan seleksi fitur dapat menjadi acuan pada penelitian berikutnya.

1.7. Sistematika Penulisan Laporan

Penyajian laporan tesis ini menggunakan sistematika penulisan sebagai berikut.

- a. Bab 1 Pendahuluan, pada bab ini berisi Latar Belakang Masalah, Identifikasi Masalah, Perumusan Masalah, Batasan Masalah, Tujuan Penelitian, Manfaat Penelitian, Sistematika Penulisan Laporan.
- b. Bab 2 Landasan Teori, pada bab ini berisi Penelitian Terkait, Data Mining, Klasifikasi, Algoritme C4.5, Algoritme *Random Forest*, Seleksi Fitur dan Metode Evaluasi.
- c. Bab 3 Metode Penelitian, pada bab ini berisi Metode Penelitian, Waktu Penelitian, Kebutuhan/Aplikasi yang digunakan.
- d. Bab 4 Hasil Penelitian dan Pembahasan, pada bab ini berisi Hasil Penelitian Penyakit Ginjal Kronis menggunakan Algoritme C4.5 dan Algoritme *Random Forest* serta adanya seleksi fitur dalam mendiagnosis Penyakit Ginjal Kronis.
- e. Bab 5 Kesimpulan dan Saran, pada bab ini berisi Kesimpulan dari penelitian yang telah dilakukan dan saran-saran untuk pengembangan penelitian kedepan.