

BAB I

PENDAHULUAN

1.1 LATAR BELAKANG

Penyakit ginjal kronis adalah proses *patologis* dengan berbagai *etiologi*, yang menyebabkan penurunan fungsi ginjal secara progresif dan tidak dapat dipulihkan [1]. Fungsi ginjal yaitu membuang sisa *metabolisme* dan mengatur keseimbangan cairan dalam tubuh manusia. Pasien diidentifikasi ginjal kronis ketika mereka memiliki laju *filtrasi glomerulus (GFR)* lebih rendah dari 60 ml/min/1,73 m², atau *GFR* lebih besar dari 60 ml/min/1,73 m², tetapi dengan bukti kerusakan struktur ginjal selama periode yang sama dengan atau lebih dari tiga bulan [2]. Gejala pada orang yang mungkin terkena penyakit ginjal kronis sulit untuk dikenali, kebanyakan pasien baru menyadari terkena penyakit ginjal kronis jika sudah dalam tahap stadium akhir. Penyakit ginjal kronis disebabkan oleh banyak faktor, di antaranya diabetes, hipertensi, *glomerulonefritis*, *sindrom nefrotik* dan kista ginjal [3].

Masalah lain yang dihadapi oleh penderita penyakit gagal ginjal kronis adalah mahalnya biaya perawatan dan belum adanya pengobatan yang dapat menyembuhkan penyakit gagal ginjal kronis. Menurut [4], Angka kejadian ginjal kronik di Indonesia pada tahun 2018 yaitu sebesar 0,38% dari 252.124.458 penduduk Indonesia dimana 713.783 orang menderita gagal ginjal kronik. Oleh karena itu, penting bagi setiap orang untuk lebih memperhatikan kesehatan ginjal. Sebagai langkah pencegahan yaitu dengan menjalani pola hidup sehat dan mengontrol penyakit yang dapat meningkatkan risiko terjadinya penyakit ginjal

kronis. Serta perlu dilakukannya pendeteksian penyakit ginjal kronis sejak dini agar bisa meminimalisir pasien rawat inap dan kematian akibat penyakit ginjal kronis.

Dalam beberapa tahun belakangan, telah dilakukan penelitian untuk menilai kinerja metode klasifikasi pada diagnosis penyakit ginjal kronis dengan menerapkan pendekatan *Machine Learning*. Hal ini dilakukan untuk mencari solusi yang efektif dalam mendeteksi dan mengatasi penyakit ginjal kronis. Sejumlah penelitian sebelumnya [5],[6],[7] telah mengadopsi pendekatan *Machine Learning* untuk mengevaluasi sejauh mana algoritma *Machine Learning* dapat mendeteksi penyakit ginjal kronis.

Beberapa penelitian sebelumnya yaitu penelitian [5], menggunakan metode algoritma *Decision Tree* diperoleh *accuracy* (98,48%), *Random forest* diperoleh *accuracy* (94,16%) dan *Logistic Regression* diperoleh *accuracy* (99,24%). Berdasarkan hasil pengujian didapatkan *accuracy* terbaik yaitu metode *Logistic Regression* dengan *accuracy* (99,24%). Selanjutnya penelitian [6], menggunakan metode algoritma *Ada boost*, *Naïve bayes*, *Gradien boosting*, *Linear Discriminant*, *ANN*, *K-nearest neighbors*, *Gaussian Naïve Bayes*, *random forest*, *support vector machine* dan *Decision Trees*. Di antara model pembelajaran mesin yang digunakan, metode *Gaussian Naïve Bayes* dan *Decision Trees* mencapai *accuracy* terbaik dengan (100%) keakuratan diagnosis.

Selanjutnya penelitian [7], menggunakan metode algoritma *Random Forest*, *Ada Boost*, *CatBoost*, *XGBoost*, *Gradient Boosting*, *MLP*, *Bagging*, *Extra Trees*, *K-Nearst Neighbors*, *LightGBM*, *SGB*, *Decision Tree*, *Support Vector Machine*

dan *Gaussian NB*. Berdasarkan hasil pengujian didapatkan *accuracy* terbaik yaitu metode *Extra Trees* dengan *accuracy* (98%).

Berdasarkan beberapa penelitian terdahulu yang memprediksi penyakit ginjal kronis, penelitian tersebut menggunakan *dataset* yang sama tetapi dengan metode yang berbeda-beda dan tingkat *accuracy* yang berbeda. Oleh karena itu penulis akan melakukan komparasi dengan menggunakan *dataset* yang sama untuk memperoleh tingkat *accuracy* berbeda dengan penggunaan metode *Random Forest*, *Support vector machine* dan *K-Nearest Neighbors*. Penggunaan metode *Random Forest*, *Support vector machine* dan *K-Nearest Neighbors* dapat menjadi tolak ukur yang tepat, terlebih lagi hingga saat ini belum ada kepastian mengenai metode yang paling akurat untuk memprediksi penyakit ginjal kronis. Metode *Random forest* memiliki kelebihan yakni metode dengan kinerja *generalisasi* yang baik sehingga sangat efisien mengklasifikasi suatu data terutama data yang terdapat *missing value*. Metode *Support vector machine* memiliki kelebihan yaitu efektif menangani *dataset* dengan data yang tidak seimbang untuk menghasilkan klasifikasi yang akurat, sedangkan metode *K-Nearest Neighbors* memiliki keunggulan dapat melakukan prediksi data sekitarnya tanpa adanya distribusi data.

Metode *Random forest* adalah algoritma *ensemble learning* yang menggabungkan beberapa klasifikasi lemah, hasil akhirnya dipilih berdasarkan *voting* atau diambil rata-ratanya [8]. Metode *Support vector machine* adalah metode untuk memisahkan (mendeskriminasi) kumpulan data kedalam beberapa

kelompok [9]. Metode *K-Nearest Neighbors* merupakan teknik klasifikasi yang melakukan prediksi berdasarkan pada sejumlah data terdekat.

Berdasarkan penelitian [5],[6],[7] penulis akan melakukan komparasi metode *Random forest*, *Support Vector Machine* dan *K-Nearest Neighbors* untuk menemukan metode mana yang lebih optimal dalam memprediksi penyakit ginjal kronis. Maka pada penelitian ini mengusulkan “**KOMPARASI METODE *RANDOM FOREST*, *SUPPORT VECTOR MACHINE* DAN *K-NEARS NEIGHBORS* UNTUK KLASIFIKASI PENYAKIT GINJAL KRONIS**”.

1.2 RUMUSAN MASALAH

Berikut merupakan perumusan masalah pada latar belakang yang telah dijabarkan sebelumnya :

1. Bagaimana mengimplementasikan metode *Random Forest*, *Support Vector Machine*, dan *K-Nearest Neighbors* untuk klasifikasi penyakit ginjal kronis?
2. Bagaimana perbandingan kinerja terhadap metode *Random Forest*, *Support Vector Machine*, dan *K-Nearest Neighbors*?

1.3 BATASAN MASALAH

Untuk memastikan bahwa penelitian ini berjalan lancar dan mencapai tujuannya, penulis menetapkan batasan masalah penelitian meliputi :

1. Penelitian ini menggunakan data sekunder dari *UCI Machine Learning Repository*.
2. Menggunakan 25 atribut yaitu, *Age*, *Blood pressure*, *Blood glucose random*, *Blood urea*, *Serum creatinin*, *Potassium*, *Red blood cell count*,

Packed cell volume, White blood cell count, Haemoglobin, sodium, Sugar, Hypertension, Anemia, Pedal edema, Appetite, Specific Gravity, Bacteria, Coronary artery disease, Albumin, Diabetes mellitus, Red blood cells, Pus cell clumps Pus cell dan Class.

3. Data yang diambil berjumlah 400 data dan 25 atribut, terdapat dua kelas yang digunakan yaitu, kelas ginjal kronis (*ckd*) dan bukan ginjal kronis (*notckd*).
4. Pengukuran pada metode berdasarkan nilai *accuracy, precision, recall*, dan *F1-Score*.
5. Bahasa pemrograman yang digunakan yaitu *python*.

1.3 TUJUAN PENELITIAN

Berlandaskan pada permasalahan penelitian, maka penelitian ini menetapkan beberapa tujuan yaitu :

1. Mengetahui tingkat *accuracy* penyakit ginjal kronis menggunakan metode *Random Forest, Support Vector Machine* dan *K-Nearest Neighbors*.
2. Membandingkan kinerja metode *Random Forest, Support Vector Machine* dan *K-Nearest Neighbors* untuk klasifikasi penyakit gagal ginjal kronis.

1.4 MANFAAT PENELITIAN

Beberapa manfaat dari penelitian ini yaitu :

1. Membantu pembaca mengetahui faktor-faktor penyebab penyakit ginjal kronis.
2. Memperluas pengetahuan peneliti tentang bagaimana kinerja metode *Random Forests*, *Support Vector Machines*, dan *K-Nearest Neighbors* untuk menemukan informasi penting yang tersembunyi dalam data.
3. Dapat mengetahui metode yang paling optimal untuk memprediksi penyakit ginjal kronis diantara ketiga metode yaitu *Random Forest*, *Support Vector Machine* dan *K-Nearest Neighbors*.

1.5 SISTEMATIKA PENULISAN

Adapun struktur penulisan tugas akhir ini dijabarkan oleh penulis dalam beberapa bab yaitu :

BAB I : PENDAHULUAN

Pada bab ini diberikan penjelasan secara menyeluruh terkait latar belakang masalah, perumusan masalah, keterbatasan, tujuan penelitian, serta manfaat dilakukannya penelitian ini dan sistematika penulisannya.

BAB II : LANDASAN TEORI

Dalam bab ini memaparkan landasan teoritis, berbagai teori yang

berkaitan dengan masalah yang diteliti. Teori-teori ini, yang bersumber dari artikel ilmiah, buku, dan publikasi relevan lainnya, berfungsi untuk mengatasi masalah pada penelitian.

BAB III : METODOLOGI PENELITIAN

Bab ini memaparkan metode pengumpulan data, langkah-langkah penelitian dan metode analisis yang melibatkan pendekatan serta penyelesaian masalah yang dilakukan untuk mendukung penelitian.

BAB IV : ANALISIS DAN HASIL

Pada bab ini memaparkan gambaran analisis dan hasil penelitian, yaitu menganalisis *dataset* yang diperoleh dari *UCI Machine Learning Repository* dengan melakukan pengambilan data, kemudian melakukan *Preprocessing*, pembagian data, dan pemodelan. Setelah itu, mengaplikasikan perbandingan algoritma *Random Forest*, *Support Vector Machine* dan *K-Nearest Neighbors* berdasarkan tingkat *accuracy*, *precision*, *recall*, dan *F1-Score*. Hasil yang telah dilakukan diinterpretasikan menjadi informasi mengenai perbandingan hasil *accuracy* dari masing-masing algoritma untuk klasifikasi penyakit ginjal kronis.

BAB V : KESIMPULAN

Bab ini mencakup rangkuman kesimpulan dan saran dari keseluruhan hasil penelitian.

