

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Ada banyak organ penting yang mendasari kehidupan manusia dan salah satunya adalah ginjal, karena ginjal merupakan organ penting untuk menjaga komposisi darah, mencegah penumpukan limbah dan mengatur keseimbangan cairan tubuh dan fungsi lainnya seperti menjaga kestabilan kadar elektrolit, memproduksi hormon dan enzim yang membantu tubuh mengatur tekanan darah, membentuk sel darah merah dan menjaga kesehatan tulang[1]. Namun, perubahan fungsi nefron yang melemah dapat mempengaruhi kondisi ginjal, karena fungsi ginjal secara keseluruhan didasarkan oleh fungsi nefron[2]. Nefron berfungsi dalam membersihkan darah dari zat yang tidak dikehendaki oleh tubuh, bagian dari metabolisme seperti kreatinin, ureum, dan asam urat[3]. Penurunan fungsi ginjal secara progresif menyebabkan penyakit ginjal kronis[4]. Sementara itu, menurut *World Health Organization* (WHO), penyakit ginjal kronis menyebabkan sekitar 850.000 kematian per tahun, hingga WHO memperhitungkan jumlahnya kurang lebih 36 juta orang di dunia pada tahun 2015, meninggal akibat penyakit ginjal kronis[5].

Penelitian [6] mengungkapkan bahwa tingginya angka kematian yang disebabkan oleh penyakit ginjal kronis, membutuhkan pencegahan secara dini, dengan mengetahui adanya potensi terserang penyakit ginjal kronis melalui teknik klasifikasi. Klasifikasi adalah bagian dari penambangan data atau sering disebut sebagai *data mining*, dimanfaatkan dalam menggambarkan pencarian informasi dari basis data[7]. Model klasifikasi dengan penerapan yang sederhana dan efektif yang dapat digunakan adalah *K-Nearest Neighbor* (K-NN)[8]. Namun, *K-Nearest Neighbor* menerapkan metode yang tidak melakukan proses *training*[9]. Metode *AdaBoost* merupakan metode yang menangani pembobotan data *training*[10]. Sehingga menjadi solusi dalam penanganan kekurangan yang dimiliki *K-Nearest*

Neighbor, dengan penelitian [11] membuktikan bahwa *AdaBoost* memperoleh nilai akurasi yang tinggi.

Cara kerja metode *AdaBoost* menangani kekurangan algoritma *K-Nearest Neighbor* yaitu dengan menyeimbangkan dan menggabungkan catatan kelas, memberikan faktor pembobotan berdasarkan kinerja hasil pelatihan sebelumnya, bermanfaat meningkatkan skor akurasi dengan meminimalkan kesalahan fungsi klasifikasi[12]. Penelitian ini membutuhkan alat untuk proses *data mining*, alat yang digunakan adalah *Rapidminer*, karena *Rapidminer* adalah perangkat lunak yang memproses data menggunakan prinsip algoritma *data mining*[13]. Pada tahap klasifikasi menggunakan *Rapidminer*, tahap *pre-processing* sebagai tahap penanganan *missing value* dari data yang diproses, karena adanya *missing value* pada data mempengaruhi hasil klasifikasi dan model klasifikasi tidak dapat berjalan berdasarkan proses *data mining* dengan benar[14].

Missing value akan menekan kinerja pemodelan dalam *data mining*, tiga persoalan dapat terjadi pada hasil model klasifikasi, yaitu pertama kerugian efisiensi, yang kedua kesulitan dalam menggunakan ekstraksi dan analisis data dan yang ketiga, masalah dapat muncul memutarbalikkan hasil modelnya dengan adanya perbedaan antara nilai yang hilang dan data lengkap[15]. Nilai yang hilang ditangani dengan menganalisis data yang ada dan mengubah data kategorikal menjadi nilai numerik menggunakan teknik *label encoding* dan *one-hot-encoding* yang akan memudahkan dalam perhitungan klasifikasi[16]. Pembagian data pelatihan dan pengujian menggunakan teknik *split data*[17]. Sementara itu, hasil dari algoritma adalah nilai dari *accuracy*, *recall* dan *precision* yang diperoleh dari hasil *confusion matrix*[18].

Pada penelitian ini akan dilakukan penelitian mengenai “**Implementasi Algoritma K-Nearest Neighbor Berbasis AdaBoost Untuk Klasifikasi Penyakit Ginjal Kronis**”. Berdasarkan kelebihan dan kekurangan yang tercantum dalam algoritma *K-Nearest-Neighbor*, *AdaBoost* merupakan metode klasifikasi yang mengatasi kekurangan dari algoritma *K-Nearest-Neighbor*, dengan menggabungkan algoritma *K-Nearest Neighbor* dan *AdaBoost*, diharapkan hasil

accuracy, *recall* dan *precision* menggunakan algoritma *K-Nearest Neighbor* berbasis *AdaBoost* memperoleh nilai yang tinggi.

1.2 Perumusan Masalah

Berdasarkan latar belakang, perumusan masalah dalam penelitian ini adalah:

1. Pengaruh adanya penerapan *AdaBoost* pada algoritma *K-Nearest Neighbor* untuk klasifikasi penyakit ginjal kronis.
2. Hasil *accuracy*, *recall*, *precision* pada algoritma *K-Nearest Neighbor* berbasis *AdaBoost* untuk klasifikasi penyakit ginjal kronis.

1.3 Pertanyaan Penelitian

Pertanyaan yang diajukan dalam penelitian ini adalah:

1. Apa pengaruh penerapan *AdaBoost* pada algoritma *K-Nearest Neighbor* untuk klasifikasi penyakit ginjal kronis?
2. Berapa besar hasil *accuracy*, *recall*, *precision* pada algoritma *K-Nearest Neighbor* berbasis *AdaBoost* untuk klasifikasi penyakit ginjal kronis?

1.4 Tujuan Penelitian

Setelah menentukan rumusan masalah, adapun tujuan dari penelitian adalah:

1. Mengetahui pengaruh dari metode *AdaBoost* pada algoritma *K-Nearest Neighbor* untuk klasifikasi penyakit ginjal kronis.
2. Mengetahui hasil *accuracy*, *recall* dan *precision* pada algoritma *K-Nearest Neighbor* berbasis *AdaBoost* untuk klasifikasi penyakit ginjal kronis.

1.5 Batasan Masalah

Adapun batasan masalah pada penelitian ini adalah sebagai berikut:

1. Data diambil dari *UCI Machine Learning Repository*, pada bulan Juli tahun 2015 yaitu dataset *Chronic Kidney Disease*[19].
2. Data yang digunakan sebanyak 400 data dan dikelompokan berdasarkan kategori dengan 24 atribut dan memiliki satu *class* data yang diklasifikasi, ada

250 data yang termasuk ckd (*chronic kidney disease*) dan 150 data yang termasuk data notckd (*not chronic kidney disease*)[19].

3. Atribut terdiri dari 24 atribut (*age, blood pressure, specific gravity, albumin, sugar, red blood cells, pus cell, pus cell clumps, bacteria, blood glucose random, blood urea, serum creatinine, sodium, potassium, hemoglobin, packed cell volume, white blood cell count, red blood cell count, hypertension, diabetes mellitus, coronary artery disease, appetite, pedal edema, anemia*). Dan memiliki satu *class* (ckd/notckd)[19].
4. Algoritma yang digunakan untuk penelitian ini adalah algoritma *classifier* yaitu *K-Nearest Neighbor* dan *AdaBoost*.
5. *Split data* membagi data dengan 70% data latih dan 30% data uji.
6. Iterasi *AdaBoost* yang digunakan pada *Rapidminer* adalah 10.
7. Nilai *k* yang digunakan pada klasifikasi *K-Nearest Neighbor* adalah 3.

1.6 Manfaat Penelitian

Adapun manfaat dari penelitian ini adalah sebagai berikut:

1. Secara Akademis
 - a) Menambahkan pengetahuan terhadap pembaca mengenai algoritma *K-Nearest Neighbor* dan *AdaBoost*.
 - b) Membantu para peneliti sebagai bahan referensi.
2. Secara Praktis
 - a) Membantu ahli medis mengklasifikasi penyakit ginjal kronis untuk mencegah kasus penyakit ginjal kronis yang meningkat.