

BAB I

PENDAHULUAN

1.1. Latar belakang

Penyakit ginjal kronik (PGK) adalah kerusakan ginjal baik struktur atau fungsinya yang berlangsung selama 3 bulan atau lebih [1]. PGK kini menjadi salah satu penyebab utama kematian di abad ke-21. Peningkatan faktor risiko seperti obesitas dan diabetes melitus turut berkontribusi terhadap bertambahnya jumlah pasien PGK. Pada tahun 2017, diperkirakan sebanyak 843,6 juta orang di seluruh dunia terdampak oleh penyakit ini [2]. Jumlah kasus PGK terus meningkat seiring dengan bertambahnya prevelensi kasus hipertensi, obesitas, dan diabetes melitus yang merupakan faktor-faktor risiko PGK [2]. Hal ini menunjukkan bahwa PGK menjadi masalah kesehatan global yang mendesak untuk ditangani secara serius.

Sebagian besar peningkatan beban kasus PGK diperkirakan terjadi di Asia. Berdasarkan data Jaminan Kesehatan Nasional (JKN) mulai dari tahun 2017 hingga tahun 2020 menunjukkan peningkatan tiga kali lipat data jumlah pasien baru dan aktif penyakit ginjal tahap akhir (PGTA) serta sedang menjalani dialisis [3]. Kemudian di Asia Selatan, prevalensi diabetes diperkirakan meningkat lebih dari 150% antara tahun 2000 dan 2035. Sementara Cina dan India, diproyeksikan tetap menjadi negara dengan jumlah penderita diabetes tertinggi pada tahun 2035 (total 251,7 juta orang). Selain itu, peningkatan kelompok PGK dengan etiologi yang tidak diketahui telah dilaporkan di beberapa bagian Asia Selatan termasuk Sri Lanka dan India [4].

Salah satu upaya untuk menangani permasalahan diagnosa PGK adalah dengan menerapkan teknologi kecerdasan buatan atau *artificial intelligence* (AI). AI merupakan salah satu ilmu komputer yang dapat membuat suatu komputer mampu melakukan tugas dengan cara yang mirip bahkan sebaik manusia. Salah satu bentuk AI yang cocok digunakan untuk menangani permasalahan diagnosa PGK adalah sistem pakar. Sistem pakar merupakan suatu sistem komputer yang mampu melakukan penalaran dan kemampuan berpikir seperti seorang ahli dalam menyelesaikan suatu permasalahan tertentu.

Teknologi kecerdasan buatan mampu melakukan pendeteksian penyakit secara cepat dan akurat. Beberapa penelitian terdahulu mengenai klasifikasi PGK dan non-PGK menggunakan data *Chronic Kidney Disease* (CKD) Rumah Sakit Karaikudi, Tamilnadu India yang diperoleh dari UC Irvine Machine Learning Repository [5] pernah dilakukan pada penelitian [6] dengan menggunakan metode *Naïve Bayes Classifier* (NBC) dan diperoleh akurasi 96.43%. Kemudian, pada penelitian [7] dengan menggunakan dataset yang sama melakukan perbandingan metode pembelajaran *Support Vector Machine* (SVM) dan *Decision Tree* (DT), hasil penelitian tersebut menunjukkan bahwa akurasi klasifikasi dengan metode SVM lebih tinggi yakni 97,5% dibandingkan hasil klasifikasi menggunakan metode DT yakni 92,5%. Pada penelitian [8] menerapkan metode *K-Nearest Neighbor* (K-NN) untuk mengklasifikasikan data PGK menjadi dua kelas yakni PGK atau non-PGK dengan hasil akurasi akhir adalah 85,83%.

Berdasarkan beberapa hasil penelitian terdahulu terkait pengklasifikasian PGK atau non-PGK menggunakan dataset [5] dengan penerapan metode pembelajaran yang beragam dapat diketahui bahwa hasil klasifikasi terbaik ditunjukkan pada penelitian yang menggunakan metode SVM dengan nilai akurasi 97,5%. Keunggulan metode SVM dalam menyelesaikan kasus klasifikasi juga ditunjukkan dalam penelitian lain [9] mengenai prediksi pasien hipertensi menggunakan metode SVM, NB, *artificial neural network* (ANN), dan *instance-based k-nearest neighbors* (IBK). Pada penelitian ini menunjukkan performa metode SVM lebih unggul dibandingkan metode lain yang diujikan dalam penelitian tersebut dengan hasil akurasi tertinggi adalah 83,41%.

Pada penelitian [6], [7], [8] menggunakan data penelitian yang sama yakni data [5], dimana dalam data penelitian tersebut memuat nilai data yang hilang atau biasa disebut dengan *missing value*. Dalam penelitian [6], [7] untuk mengatasi *missing value* pada data penelitian tersebut dilakukan dengan cara menghapus baris data yang memuat *missing value* atau yang biasa disebut dengan *Listwise deletion*. Namun, metode penanganan *missing value* tersebut kurang efektif sebab jumlah data sampel berkurang dan dapat mempengaruhi proses pelatihan model. Metode yang dinilai lebih baik digunakan untuk mengatasi masalah *missing value* adalah metode imputasi [10]. Pada penelitian [8] yang melakukan klasifikasi data dengan

metode K-NN dan menerapkan metode imputasi *Mean*, *Median*, serta *Modus* menunjukkan hasil akurasi 83,41%, dimana nilai akurasi tersebut lebih kecil dibandingkan nilai akurasi pada penerapan metode lain di penelitian lainnya.

Selain metode imputasi *Mean*, *Median*, dan *Modus* terdapat metode imputasi lain yakni Imputasi K-NN. Metode Imputasi K-NN merupakan salah satu metode yang paling populer untuk menyelesaikan permasalahan *missing value* [11]. Dalam penelitian [11], melakukan perbandingan metode Imputasi K-NN dengan metode imputasi statistik sederhana lainnya seperti *Mean*, *Median*, serta *Modus*. Dari penelitian tersebut diperoleh hasil akurasi RMSE atau MAPE pada hasil Imputasi K-NN menunjukkan hasil yang jauh lebih baik daripada semua metode imputasi lain yang diujikan. Metode imputasi K-NN juga terbukti efektif dalam meningkatkan performa model klasifikasi, menurut penelitian [12] imputasi K-NN mampu mengatasi data hilang dalam jumlah besar yakni 96% *missing value* dari keseluruhan data penelitian.

Berdasarkan permasalahan mengenai PGK dan pemaparan penelitian terdahulu yang telah dijelaskan sebelumnya, maka pada penelitian ini melakukan klasifikasi faktor-faktor PGK dan non-PGK menggunakan data *Chronic Kidney Disease* yang diperoleh dari UC Irvine Machine Learning Repository [5] dengan menerapkan metode klasifikasi SVM serta penanganan *missing value* menggunakan Imputasi K-NN. Penelitian ini dilakukan dengan tujuan untuk membantu tenaga medis dalam mendiagnosa PGK dan non-PGK secara tepat akurat berdasarkan faktor-faktor yang dialami terduga pasien PGK.

1.2. Rumusan masalah

Berdasarkan pemaparan latar belakang di atas maka, dapat dirumuskan rumusan masalah sebagai berikut:

1.2.1. Permasalahan

Kasus kejadian PGK terus mengalami peningkatan, dimana hal tersebut memerlukan tindakan pengelolaan untuk mencegah dan menurunkan kasus kejadian PGK. Akan tetapi, dalam pengelolaan PGK terjadi kendala-kendala semacam kurangnya fasilitas pelayanan kesehatan yang mendukung

pemeriksaan serta ketidakmampuan tenaga kesehatan dalam mengidentifikasi PGK secara dini. Guna mengatasi permasalahan tersebut dibutuhkan sebuah sistem untuk membantu pengklasifikasian PGK secara cepat dan akurat. Dalam data penelitian yang digunakan pada penelitian ini juga memuat *missing value* yang perlu dilakukan penanganan sehingga diperoleh model klasifikasi yang handal.

1.2.2. Metode usulan

Sebagai solusi permasalahan yang telah dijelaskan pada sub bab permasalahan maka sebagai solusi, dalam penelitian ini menerapkan metode klasifikasi SVM sebagai salah satu metode pembelajaran mesin yang handal dalam menangani kasus klasifikasi. Kemudian untuk menangani permasalahan *missing value* pada data penelitian ini menerapkan metode Imputasi K-NN.

1.2.3. Pertanyaan penelitian

Berdasarkan permasalahan dan solusi yang telah diuraikan, maka diperoleh pertanyaan penelitian ini sebagai berikut:

Seberapa akurat penerapan metode SVM dengan imputasi K-NN dalam mengklasifikasikan PGK dan non-PGK?

1.3. Tujuan dan manfaat

Dari rumusan masalah yang telah dituliskan sebelumnya, terdapat tujuan dan manfaat dari penelitian yang dilakukan saat ini. Berikut ini pemaparan terkait tujuan dan manfaat dari penelitian ini:

1.3.1. Tujuan

Adapun tujuan yang ingin dicapai dalam penelitian ini adalah untuk mengetahui seberapa akurat penerapan metode SVM dengan Imputasi K-NN dalam mengklasifikasikan PGK dan non-PGK.

1.3.2. Manfaat

1. Bagi bidang medis, hasil penelitian ini dapat membantu dalam melakukan klasifikasi PGK sehingga proses diagnosa menjadi lebih cepat dan akurat dalam memberikan pelayanan dan pengobatan terbaik

pada pasien PGK.

2. Bagi bidang pendidikan, penelitian ini dapat menjadi acuan pembelejaran dan pengembangan penelitian berikutnya dalam mengatasi permasalahan serupa di masa depan.

1.4. Batasan masalah

Agar pembahasan penelitian berfokus sesuai tujuan penelitian maka, peneliti merumuskan batasan-batasan masalah sebagai berikut:

1. Penelitian ini berfokus pada klasifikasi PGK dan non-PGK.
2. Data penelitian diperoleh dari Rumah Sakit Karaikudi, Tamilnadu India pada tahun 2015 melalui UC Irvine Machine Learning Repository dengan jumlah data sebanyak 400 *record* data dan memiliki 22 fitur. Data penelitian dapat diakses pada link <https://archive.ics.uci.edu/dataset/336/chronic+kidney+disease>.
3. Menggunakan metode SVM sebagai metode klasifikasi data PGK.
4. Menerapkan metode imputasi K-NN untuk penanganan *missing value* pada data PGK.

1.5. Sistematika penulisan

Adapun sistematika penulisan proposal penelitian ini terdiri dari 3 bab, berikut penjelasan terkait sistematika penulisan proposal penelitian ini:

1.5.1. BAB I Pendahuluan

Bab ini memuat latar belakang yang membahas tentang PGK, hasil literasi penelitian terdahulu mengenai keunggulan dan kekurangan metode dari penelitian terdahulu. Pada sub bab berikutnya terdapat rumusan masalah yang mencakup permasalahan, metode usulan, dan pertanyaan penelitian. Kemudian, terdapat sub bab tujuan dan manfaat penelitian. Sub bab berikutnya adalah batasan masalah dan terakhir ditutup dengan sub bab sistematika penulisan.

1.5.2. BAB II Kajian Pustaka

Pada bab ini akan menjelaskan teori penunjang terkait topik, tahapan-tahapan dan metode yang digunakan dalam penelitian ini. Teori penunjang yang dimaksud

adalah teori tentang definisi data mining, klasifikasi, PGK, SVM, dan Imputasi K-NN yang diperoleh dari proses literasi.

1.5.3. BAB III Metode Usulan

Pada bab ini menjelaskan secara rinci tentang alur yang terdapat dalam penelitian yang meliputi dataset yang akan digunakan, tahapan-tahapan dalam klasifikasi menggunakan algoritma SVM dengan penanganan *missing value* menggunakan metode Imputasi K-NN.