

BAB I

PENDAHULUAN

1.1 Latar Belakang

Gagal ginjal kronis (GGK) adalah kondisi medis di mana ginjal secara bertahap kehilangan kemampuannya untuk menyaring limbah dan cairan dari darah, yang biasanya berlangsung selama beberapa bulan hingga bertahun-tahun. Penyebab utama GGK meliputi diabetes, hipertensi, dan penyakit ginjal lainnya. Pada tahap awal, GGK sering tidak menunjukkan gejala yang jelas, sehingga banyak penderita tidak menyadari kondisi mereka sampai ginjal mengalami kerusakan signifikan. Gejala yang sering muncul meliputi kelelahan, pembengkakan kaki, kesulitan tidur, dan penurunan nafsu makan. Pengobatan GGK bertujuan untuk mengelola gejala dan memperlambat perkembangan penyakit, sedangkan pada kasus yang lebih parah penderita mungkin memerlukan dialisis atau transplantasi ginjal [1].

Salah satu metode terapi untuk pasien GGK stadium akhir adalah hemodialisis, atau yang sering dikenal sebagai cuci darah. Hemodialisis bertujuan menggantikan sebagian fungsi ginjal dalam membuang limbah metabolismik dan kelebihan cairan dari tubuh. Proses ini dilakukan menggunakan mesin dialisis yang menyaring darah, mengeluarkan zat-zat yang tidak diperlukan, serta menjaga keseimbangan cairan dan elektrolit. Hemodialisis biasanya dilakukan beberapa kali dalam seminggu tergantung tingkat keparahan kondisi pasien. Meskipun efektif sebagai solusi sementara, hemodialisis tidak dapat menyembuhkan GGK dan hanya membantu mempertahankan fungsi tubuh. Pasien yang menjalani terapi ini kerap menghadapi tantangan tambahan seperti kelelahan, risiko infeksi, dan berbagai komplikasi lainnya. Oleh karena itu, pemantauan rutin terhadap tingkat keparahan pasien hemodialisis menjadi hal yang sangat penting untuk memastikan perawatan yang tepat dan meningkatkan kualitas hidup pasien [2].

Perkembangan teknologi informasi dan komputasi telah mendorong pemanfaatan metode berbasis kecerdasan buatan untuk membantu prediksi dan pemantauan kondisi medis. Model pembelajaran mesin dapat memanfaatkan data klinis pasien untuk memprediksi tingkat keparahan secara akurat, sehingga tenaga medis dapat melakukan intervensi dini dan personalisasi perawatan. Dua algoritma

yang dapat digunakan dalam penelitian ini adalah Extreme Gradient Boosting (XGBoost) dan K-Nearest Neighbors Enhanced (K-NN Enhanced). XGBoost merupakan algoritma *ensemble boosting* yang menggabungkan sejumlah pohon keputusan untuk menghasilkan prediksi yang akurat dan stabil. Algoritma ini dikenal efektif dalam menangani data dengan jumlah fitur yang besar, memiliki mekanisme *regularization* untuk mengurangi risiko *overfitting*, serta mampu melakukan seleksi fitur secara otomatis [3].

K-NN Enhanced merupakan pengembangan dari algoritma K-Nearest Neighbors (K-NN) standar dengan menambahkan mekanisme *distance-weighted voting*, di mana tetangga yang lebih dekat memiliki bobot kontribusi lebih besar dalam menentukan kelas prediksi dibandingkan tetangga yang lebih jauh. Selain itu, parameter jumlah tetangga (k) dan metrik jarak pada K-NN Enhanced dioptimalkan melalui proses pencarian parameter, sehingga performa prediksi dapat lebih maksimal dibandingkan K-NN konvensional [4].

Salah satu tantangan dalam pengolahan data medis adalah tingginya dimensi fitur (*high dimensionality*), yang dapat menurunkan kinerja model, khususnya pada algoritma berbasis jarak seperti K-NN Enhanced. Fenomena ini dikenal dengan istilah *curse of dimensionality*, di mana semakin banyak fitur yang digunakan, jarak antar data menjadi semakin sulit dibedakan. Untuk mengatasi permasalahan ini, digunakan teknik reduksi dimensi seperti PCA. PCA mentransformasikan data berdimensi tinggi menjadi sejumlah komponen utama yang mempertahankan sebagian besar variasi informasi dari data asli. Dengan mengurangi dimensi data, PCA dapat mengurangi *noise*, menghilangkan redundansi fitur, dan mempercepat proses pelatihan model [5].

Dalam penelitian ini, PCA diterapkan pada kedua algoritma, baik XGBoost maupun K-NN Enhanced, untuk mengevaluasi pengaruhnya terhadap performa prediksi. Beberapa penelitian terdahulu menunjukkan bahwa PCA sering memberikan peningkatan performa yang signifikan pada K-NN standar, sementara pada XGBoost efeknya relatif kecil karena algoritma ini telah memiliki mekanisme seleksi fitur internal. Namun, belum banyak studi yang membandingkan performa XGBoost dan K-NN Enhanced dalam prediksi tingkat keparahan pasien hemodialisis, khususnya pada skenario dengan dan tanpa PCA [6].

Menurut Vista dan Sabilla [7], perbandingan tiga metode reduksi fitur yang meliputi PCA, Linear Discriminant Analysis, serta gabungan PCA dan Linear Discriminant Analysis menghasilkan peningkatan rata-rata akurasi sebesar 1,9 persen, recall sebesar 2,2 persen, dan precision sebesar 0,93 persen dibandingkan model baseline. Metode reduksi fitur terbaik diperoleh dari kombinasi PCA dan LDA dengan akurasi tertinggi sebesar 93,5 persen, recall sebesar 91,1 persen, dan precision sebesar 97,7 persen. Penelitian lain yang dilakukan oleh Sari Junita menunjukkan bahwa integrasi PCA ke dalam XGBoost dapat meningkatkan akurasi dan efisiensi secara signifikan dibandingkan XGBoost tanpa PCA, terutama dalam prediksi penyakit diabetes. PCA membantu mengurangi *noise* serta menghilangkan fitur yang kurang relevan, sehingga meningkatkan *precision* dan akurasi meskipun pada beberapa kasus nilai recall sedikit menurun.

Data yang digunakan dalam penelitian ini diperoleh dari Rumah Sakit Islam (RSI) Jemursari Surabaya, yang memiliki unit hemodialisis dengan fasilitas modern, tenaga medis berpengalaman, dan sistem pencatatan medis yang baik. Parameter klinis yang digunakan meliputi kadar ureum, kreatinin, tekanan darah, hemoglobin, dan indikator lainnya yang relevan. Penelitian ini dirancang untuk menguji empat skenario utama, yaitu XGBoost tanpa PCA, XGBoost dengan PCA, K-NN Enhanced tanpa PCA, dan K-NN Enhanced dengan PCA. Masing-masing skenario akan diuji dengan variasi parameter tertentu untuk mendapatkan konfigurasi terbaik [8].

Berdasarkan uraian tersebut, penelitian ini difokuskan pada perbandingan kinerja dua algoritma prediksi, yaitu XGBoost dan K-NN Enhanced, baik dengan maupun tanpa PCA, dalam memprediksi tingkat keparahan pasien hemodialisis. Diharapkan hasil penelitian ini dapat memberikan kontribusi bagi pengembangan sistem pendukung keputusan di bidang medis, sehingga proses pemantauan dan perencanaan perawatan pasien dapat dilakukan secara lebih akurat, cepat, dan efisien.

Selain mengevaluasi model klasifikasi, penelitian ini juga menyertakan prototipe antarmuka pengguna grafis untuk mendemonstrasikan proses prediksi pada data baru. Antarmuka tersebut berfungsi pada mode inferensi murni, sehingga pengguna non-teknis dapat mengunggah data dan memperoleh label prediksi tanpa

melalui proses pelatihan ulang. Penambahan ini dimaksudkan sebagai jembatan translasi hasil penelitian ke penggunaan praktis.

1.2 Tujuan Penelitian

Sebagai tindak lanjut dari rumusan masalah, penelitian ini menetapkan tujuan operasional yang akan dicapai melalui perancangan eksperimen dan evaluasi model. Tujuan tersebut meliputi:

1. Membandingkan kinerja XGBoost dan K-NN Enhanced dengan PCA dalam memprediksi tingkat keparahan pasien hemodialisis
2. Menganalisis pengaruh PCA terhadap performa kedua algoritma pada data berdimensi tinggi
3. Menentukan model berkinerja terbaik berdasarkan akurasi, precision, recall, dan F1-score
4. Membangun prototipe antarmuka pengguna berbasis Python untuk menampilkan hasil prediksi menggunakan model terbaik.

1.3 Manfaat Penelitian

Penelitian ini diharapkan memberi kontribusi teoretis sekaligus manfaat praktis bagi pengembangan sistem pendukung keputusan klinis berbasis data. Adapun manfaat yang ditargetkan adalah:

1. Memberikan kontribusi ilmiah dalam pengembangan model pembelajaran mesin untuk prediksi tingkat keparahan pasien hemodialisis.
2. Menyediakan pendekatan efisien dalam menangani data medis berdimensi tinggi melalui penerapan Principal Component Analysis (PCA).
3. Menyajikan analisis komparatif kinerja antara model XGBoost dan K-NN Enhanced dengan penerapan PCA sebagai dasar pengambilan keputusan medis berbasis data.
4. Menyediakan prototipe antarmuka yang memudahkan demonstrasi dan validasi hasil penelitian oleh pengguna non-teknis

1.4 Rumusan Masalah

Untuk menajamkan fokus kajian dan memastikan keterukuran hasil, penelitian ini merumuskan pokok-pokok permasalahan yang hendak dijawab. Secara khusus, pertanyaan penelitian adalah sebagai berikut:

1. Bagaimana penerapan metode XGBoost dan K-NN Enhanced dengan PCA dalam memprediksi tingkat keparahan pasien hemodialisis
2. Bagaimana perbandingan kinerja antara XGBoost dan K-NN Enhanced berdasarkan metrik akurasi, precision, recall, dan F1-score
3. Bagaimana pengaruh penerapan PCA terhadap performa kedua algoritma pada data medis berdimensi tinggi
4. Bagaimana merancang antarmuka pengguna untuk mendemonstrasikan proses prediksi menggunakan model terbaik hasil penelitian

1.5 Batasan Masalah

Agar penelitian tetap fokus, terukur, dan sesuai sumber daya yang tersedia, ruang lingkup dibatasi pada beberapa aspek berikut. Batasan penelitian ini mencakup:

1. Sumber Data. Data berasal dari Rumah Sakit Islam (RSI) Jemursari Surabaya dan satu dataset pendukung dari Kaggle.
2. Parameter Klinis. Analisis dibatasi pada parameter yang tersedia (misalnya ureum, kreatinin, tekanan darah, hemoglobin, dan fitur terkait hemodialisis).
3. Metode & Skenario. Hanya menggunakan dua algoritma, yaitu XGBoost dan K-Nearest Neighbors Enhanced (K-NN Enhanced), masing-masing diuji dengan dan tanpa PCA (PCA 10 komponen) menggunakan GridSearchCV 10-fold; algoritma lain tidak dibahas.
4. Keterbatasan Waktu & Data. Waktu dan sumber daya terbatas, sehingga ukuran data dan variasi pengujian dibatasi; tidak menggunakan data real-time/longitudinal.
5. Antarmuka Pengguna (GUI). GUI disediakan untuk inferensi saja, tidak melakukan pelatihan ulang atau tuning, tidak menampilkan metrik evaluasi, dan tidak terintegrasi dengan sistem klinis. Evaluasi kinerja model sepenuhnya disajikan pada bab hasil eksperimen.

Halaman ini sengaja dikosongkan