

BAB I

PENDAHULUAN

Upaya untuk mengurangi jumlah penderita penyakit ginjal jinak bertujuan untuk memberikan pengobatan dini yang cepat dan tepat guna mencegah perkembangan penyakit ginjal menjadi lebih serius. Secara umum, banyak orang tidak mengetahui bahwa ginjal memainkan peran besar dalam mengatur tubuh kita[1]. Ginjal berfungsi untuk memproduksi urin dan menjaga keseimbangan cairan; misalnya, dalam suhu dingin, tubuh cenderung lebih sering buang air kecil, sedangkan dalam kondisi panas, tubuh mengalami kekurangan cairan yang lebih besar[2]. Dr. Zulkhair Ali, Ketua Pernefri (Perhimpunan Nefrologi Indonesia), menekankan bahwa disfungsi ginjal dapat menyebabkan gagal ginjal[3]. Kurangnya informasi kepada masyarakat mengenai gejala penyakit ginjal turut memperburuk masalah ini. Oleh karena itu, sangat penting untuk menerapkan langkah- langkah pencegahan dengan mengidentifikasi individu yang berisiko mengalami penyakit ginjal[3]– [5]. Pencegahan idealnya dimulai sejak tahap awal, di mana pasien menjalani skrining untuk mengetahui ada atau tidaknya faktor risiko penyakit ginjal. Dengan kemajuan teknologi, inovasi telah merambah ke berbagai sektor, termasuk industri kesehatan. Dalam institusi kesehatan, ketepatan analisis dan efisiensi waktu sangat penting untuk pengambilan keputusan yang tepat. Dalam dunia medis, klasifikasi penyakit berbasis kecerdasan buatan (AI) telah berkembang pesat untuk meningkatkan akurasi diagnosis dan membantu tenaga medis dalam pengambilan keputusan[2]–[9]. Namun, salah satu tantangan utama dalam penerapan model klasifikasi adalah ketidakseimbangan data dalam dataset medis[2], [10]–[15]. Data pasien yang positif mengidap penyakit ginjal kronis (PGK) sering kali jauh lebih sedikit dibandingkan dengan data pasien sehat, sehingga model machine learning cenderung lebih bias terhadap kelas mayoritas. Ketidakseimbangan ini dapat mengurangi kemampuan model dalam mengidentifikasi pasien berisiko tinggi secara akurat[8], [16], [17].

Berbagai metode telah dikembangkan untuk menangani masalah ketidakseimbangan data ini. Salah satu metode yang paling banyak digunakan adalah Synthetic Minority Over-sampling Technique (SMOTE)[18]–[20]. Teknik ini bekerja dengan menghasilkan sampel sintesis berdasarkan kedekatan data minoritas untuk meningkatkan representasi kelas tersebut dalam dataset. Dengan menggunakan SMOTE, diharapkan kinerja model klasifikasi dapat meningkat dalam mendeteksi pasien yang sebenarnya menderita Penyakit Ginjal Kronis (PGK). Beberapa penelitian sebelumnya telah membuktikan efektivitas SMOTE dalam menangani dataset yang tidak seimbang. Sebagai contoh, studi oleh Garcia et al. menunjukkan bahwa penerapan SMOTE pada dataset medis dapat meningkatkan nilai recall tanpa mengorbankan akurasi secara signifikan[21]. Studi lain menemukan bahwa kombinasi SMOTE dengan algoritma ensemble learning, seperti Random Forest dan Gradient Boosting, dapat memberikan hasil yang lebih akurat dibandingkan dengan metode oversampling lainnya.

Selain itu, penelitian menunjukkan bahwa meskipun SMOTE efektif dalam meningkatkan kinerja klasifikasi, pemilihan parameter yang tepat dalam teknik ini sangat memengaruhi hasil akhir[11], [20]. Jika tidak disesuaikan dengan baik, SMOTE dapat menyebabkan overfitting pada model, sehingga mengurangi kemampuan generalisasi terhadap data baru. Oleh karena itu, diperlukan penelitian lebih lanjut untuk mengeksplorasi implementasi SMOTE yang optimal dalam kasus CKD, guna menghasilkan model klasifikasi yang lebih andal dan akurat. Mengingat urgensi deteksi dini CKD dan tantangan dalam klasifikasi data medis yang tidak seimbang, penelitian ini berfokus pada penerapan SMOTE dalam optimalisasi model klasifikasi PGK.

Melalui penelitian ini, diharapkan dapat diperoleh pemahaman yang lebih baik tentang dampak ketidakseimbangan data terhadap model klasifikasi, serta efektivitas SMOTE dalam meningkatkan akurasi diagnosis CKD secara otomatis.