

BAB II

TINJAUAN PUSTAKA

2.1 Penelitian Terdahulu

Penelitian-penelitian terdahulu yang memiliki relevansi literatur dengan berkaitan penelitian judul “Klasifikasi diagnosa penyakit anemia berdasarkan data hematologi menggunakan XGBoost” dapat dilihat pada tabel 2.1 .

Tabel 2. 1 Penelitian Terdahulu

No	Penulis	Judul	Metode	Hasil
1	Ningrum, R.A., Priyanto, A., & Athiyah, U. (2021)	Implementasi Algoritma Fuzzy Tsukamoto Untuk Diagnosis Penyakit Anemia (Studi Data: Rekam Medis Pasien Ibu RSIA Bunda Arif Purwokerto).	Fuzzy Tsukamoto	Berdasarkan 40 data rekam medis yang di ujikan 34 data sesuai dan 6 yang tidak sesuai sehingga di peroleh presentase keberhasilan 85%
2	Naftali, S., & Arita, W. (2020)	Sistem Pakar Untuk Diagnosis Penyakit Anemia Menggunakan Teorema Bayes.	Teorema Bayes	Menggunakan 50 data yang telah diujikan 45 pasien yang sesuai dan yang tidak sesuai 5 pasien sehingga diperoleh presentase keberhasilan 90%
3	Sirait, M.T.T.B., Fathonah, N.S., &	PEMANFAATAN ALGORITMA ADASYN DAN SUPPORT VECTOR	Support Vector Machine	Hasil pengujian menunjukkan bahwa akurasi prediksi model sebelum menggunakan

	Fauzan, M.N. (2024)	MACHINE DALAM MENINGKATKAN AKURASI PREDIKSI KANKER PARU- PARU		ADASYN hanya mencapai 59% karena ketidakseimbangan kelas yang signifikan. Setelah menerapkan ADASYN, akurasi meningkat drastis menjadi 99%.
4	Yutri Amelia. (2023)	Perbandingan Metode Machine Learning Untuk Mendeteksi Penyakit Jantung,	Random Forest, Support Vector Machine (SVM), XGBoost, K-Nearest Neighbor (KNN), Decision Tree, Logistic Regression, Multi-Layer Perceptron Classifier (MLP).	dataset dari UCI Machine Learning Repository dengan 303 data. Penggunaan algoritma XGBoost, Random Forest, SVM, KNN, MLP, Logistic Regression, dan Gradient Boosting, Evaluasi menggunakan Confusion Matrix untuk menghitung akurasi, presisi, recall, dan F1-Score. Akurasi tertinggi diperoleh dari XGBoost (95,08%).
5	Danang, W., & Bambang, P.H.	Analisis Perbandingan Performa Algoritma	XGBoost & LightGBM	Dataset yang memiliki total 569 data tentang kanker

		XGBoost dan LightGBM pada Klasifikasi Kanker Payudara	payudara, dibagi menjadi 70:30 data train dan data test. Model XGBoost mengungguli dengan hasil akurasi sebesar 97,03%, dan recall 97,40%. Sedangkan LightGBM dengan akurasi
--	--	---	--

2.2 Anemia

Anemia adalah kondisi di kadar hemoglobin (Hb) atau sel darah merah (eritrosit) dan PCV atau hematokrit dalam darah berada di bawah nilai normal, mengakibatkan gangguan pengiriman oksigen ke jaringan tubuh[15]. eritrosit berfungsi sebagai untuk mengetahui jenis anemia melalui morfologi. indeks eritrosit terdiri dari *Mean Corpuscular Hemoglobin Concentration* (MCHC) *Mean Corpuscular Hemoglobin* (MCH), dan *Mean Corpuscular Value* (MCV). Untuk mengetahui Indeks Eritrosit dilakukan kalkulasi kadar hematokrit jumlah eritrosit dan hemoglobin.

Gejala anemia tergantung pada keperluan oksigen seseorang. Gejala lebih tidak terasa apabila anemia terjadi secara perlahan. Macam-macam gejala anemia yang timbul berupa sesak napas, hiperdinamik, berkunang-kunang lemas dan lesu. Untuk anemia yang lebih berat biasanya timbul kebingungan, alergi, gagal ginjal, gagal jantung serta komplikasi lainnya[16].

2.3 ADASYN (Adaptive Synthetic Sampling)

Adaptive Synthetic Sampling (ADASYN) adalah salah satu teknik oversampling yang dirancang untuk mengatasi masalah ketidakseimbangan data. Kondisi ini dapat menyebabkan model cenderung bias terhadap kelas mayoritas dan mengabaikan kelas minoritas, yang berakibat pada menurunnya akurasi prediksi

untuk kelas minoritas [17]. ADASYN secara fungsi berfokus pada sampel minoritas yang sulit diklasifikasikan. Dengan itu, ADASYN menghasilkan lebih banyak sampel sintetis di sekitar data minoritas yang berada di dekat batas keputusan (decision boundary), sehingga membantu model untuk lebih memahami dan mempelajari karakteristik data minoritas yang kompleks [18].

2.4 Hyperparameter Tuning

Hyperparameter tuning adalah proses penting dalam pelatihan model machine learning yang bertujuan untuk mengoptimalkan kinerja algoritma. Dengan melakukan hyperparameter tuning dapat mencegah terjadinya overfitting, mempercepat waktu pelatihan, dan meningkatkan hasil evaluasi model [19]. Salah satu metode yang sering digunakan adalah GridSearch Cross-Validation (GridSearch CV), yaitu teknik pencarian kombinasi hyperparameter terbaik melalui evaluasi model secara sistematis.

2.4.1 Gridsearch CV (Cross Validation)

Gridsearch CV adalah teknik optimasi hyperparameter dalam machine learning yang digunakan untuk menemukan kombinasi hyperparameter terbaik untuk suatu model [20]. Prosesnya melibatkan hyperparameter parameter yang ditentukan sebelum proses pelatihan model dan juga melibatkan cross validation untuk memprediksi model supaya optimal pada data yang terbatas[21]. Metode akan mencoba semua hyperparameter yang telah ditentukan oleh pengguna, setelah itu memilih parameter yang memiliki hasil nilai eror terkecil[22].

2.5 eXtreme Gradient Boosting (XGBoost)

XGBoost adalah salah satu metode *machine learning* yang dikembangkan oleh Chen dan Guestrin pada 2016. Algoritma ini didasarkan pada konsep dasar Gradient Boosting, yaitu teknik penggabungan banyak model sederhana untuk membentuk satu model prediksi yang kuat [23]. XGBoost dirancang dengan keunggulan utama dalam mengurangi kompleksitas guna menghindari overfitting dan memberikan kinerja yang lebih baik. Maka karena itu, dapat memberikan solusi optimal untuk masalah klasifikasi maupun regresi [24].

Konsep dari algoritma XGBoost adalah menyesuaikan keputusan dibangun dengan model yang bergantung pada model sebelumnya. Pendekatan ini menggunakan pertumbuhan berbasis level (level-wise growth), yang berarti pengembangan struktur pohon dilakukan secara merata pada setiap tingkatannya untuk menjaga keseimbangan antara kompleksitas model dan akurasi prediksi. Pada tahap awal, model pertama yang dihasilkan mungkin masih memiliki banyak kesalahan atau prediksi yang lemah [25]. Namun, seiring bertambahnya iterasi, bobot kesalahan dari model sebelumnya akan diperhitungkan dan diperbaiki oleh model berikutnya, sehingga keseluruhan model menjadi semakin baik dan akurat dalam memprediksi data.

