

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Diabetes adalah gangguan metabolisme yang disebabkan oleh masalah pada kemampuan tubuh untuk memproduksi insulin [1]. Pada kondisi ini, pankreas memproduksi insulin dalam jumlah yang tidak mencukupi, yang mengakibatkan ketidakseimbangan kadar gula darah dan meningkatkan konsentrasi gula darah [2]. Diabetes dapat disebabkan oleh berbagai faktor, termasuk faktor genetik seperti riwayat keluarga, serta faktor gaya hidup seperti merokok, pola makan yang tidak sehat, kurangnya aktivitas fisik, stres, dan kelebihan berat badan. [3].

Diabetes adalah masalah kesehatan yang signifikan di seluruh dunia. Federasi Diabetes Internasional (IDF) melaporkan pada tahun 2021 bahwa sekitar 537 juta orang di seluruh dunia mengidap diabetes [4]. Di Indonesia, jumlah penderita diabetes terus meningkat setiap tahunnya. Pada tahun 2021, jumlah kasus diabetes di kalangan orang dewasa akan mencapai 19 juta dari total populasi orang dewasa sebanyak 179 juta, dengan prevalensi 10,8% [5]. Penyakit ini dapat menjadi sangat berbahaya bagi penderitanya, dan jika tidak terdeteksi dengan cepat, akan menyebabkan komplikasi serius seperti neuropati (kerusakan saraf), nefropati (kerusakan ginjal), retinopati (kerusakan mata), serta komplikasi mikrovaskular dan makrovaskular. Dalam jangka waktu tertentu, komplikasi tersebut dapat mengganggu berbagai sistem organ tubuh manusia dan bahkan berpotensi menyebabkan kematian [6].

Dengan meningkatnya jumlah kasus diabetes, Upaya untuk mencegah dan menangani diabetes menjadi semakin penting seiring dengan meningkatnya jumlah kasus diabetes. Penanganan diabetes yang tepat sangat penting untuk mengurangi risiko komplikasi dan efek negatif lainnya. Dengan kemajuan teknologi saat ini, teknik machine learning telah banyak dimanfaatkan dalam bidang kesehatan, termasuk untuk deteksi penyakit dan prediksi penyakit diabetes. Salah satu algoritma machine learning yang populer dan efektif untuk masalah klasifikasi diabetes adalah Extreme Gradient Boosting (XGBoost).

Sebelumnya, telah dilakukan berbagai penelitian untuk mengidentifikasi diabetes. Pada tahun 2022, penelitian dilakukan menggunakan metode regresi logistik untuk mengklasifikasikan diabetes dan mencapai akurasi sebesar 77%. Namun, setelah menerapkan teknik SMOTE dan optimasi hyperparameter dengan GridSearchCV, kinerja model meningkat menjadi 82% [7].

Penelitian lain tahun 2021 yang dilakukan untuk mendeteksi diabetes menggunakan metode logika fuzzy untuk memprediksi diabetes pada tahun 2021 mencapai tingkat akurasi sebesar 96,47% [8]. Penelitian lainnya juga dilakukan tahun 2023 untuk mendeteksi penyakit diabetes menggunakan metode Support Vector Machine dan Random Forest dengan menerapkan teknik SMOTE (Synthetic Minority Oversampling Technique) untuk menyeimbangkan data diabetes. Hasil penelitian menunjukkan bahwa metode Random Forest dengan penerapan teknik SMOTE mencapai akurasi tertinggi, yaitu sebesar 95,8% [9].

Extreme Gradient Boosting (XGBoost) adalah metode yang dikembangkan oleh Chen dan Guestrin (2016) yang menerapkan konsep Gradient Boosting (GB) yang efisien, cepat, dan terukur. XGBoost menggunakan pendekatan level-wise dalam pertumbuhannya, membentuk serangkaian pohon keputusan di mana setiap model bergantung pada model sebelumnya. Model awal dalam XGBoost biasanya memiliki kelemahan dalam inisialisasi prediksi, tetapi diperkuat melalui pembaharuan bobot pada setiap model yang dibuat [10]. Penelitian sebelumnya yang dilakukan pada tahun 2023 telah menunjukkan keunggulan XGBoost dalam memprediksi penyakit kardiovaskular (CVD) dibandingkan dengan beberapa algoritma lain seperti Decision Tree, K-Nearest Neighbors, Naïve Bayes, dan Random Forest. Dalam penelitian tersebut, XGBoost mencapai akurasi tertinggi sebesar 92,34% [11]. Penelitian lain pada tahun 2024 juga menggunakan metode XGBoost untuk mendeteksi dan menganalisis kanker payudara. Hasil akurasi yang diperoleh adalah 94,74% [12]. Pemilihan hyperparameter yang sesuai adalah kunci untuk meningkatkan kinerja dan akurasi model klasifikasi. Penyetelan hyperparameter memungkinkan peningkatan kinerja model menuju hasil optimal dengan mencoba berbagai kombinasi parameter dan nilai secara berulang [13].

Berdasarkan penelitian-penelitian yang telah dipaparkan sebelumnya, penelitian ini, akan menerapkan metode XGBoost untuk klasifikasi deteksi penyakit diabetes dengan melakukan hyperparameter tuning menggunakan GridSearchCV dan RandomSearchCV. Selain itu, penelitian ini juga akan mengevaluasi pengaruh teknik resampling menggunakan SMOTE (Synthetic Minority Over-sampling Technique) dalam mengatasi masalah ketidakseimbangan kelas pada data diabetes. Perbandingan kinerja algoritma XGBoost sebelum dan sesudah melakukan resampling dengan metode grid search dan random search akan dilakukan untuk mengetahui pendekatan terbaik dalam masalah klasifikasi deteksi penyakit diabetes.

Penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan model machine learning yang akurat dan efisien untuk deteksi dini penyakit diabetes, serta memberikan wawasan tentang pengaruh hyperparameter tuning dan teknik resampling terhadap kinerja model.

1.2 Rumusan Masalah

- a. Bagaimana cara mengoptimalkan kinerja algoritma XGBoost untuk klasifikasi penyakit diabetes menggunakan GridSearchCV?
- b. Bagaimana cara mengoptimalkan kinerja algoritma XGBoost untuk klasifikasi penyakit diabetes menggunakan RandomSearchCV?
- c. Bagaimana pengaruh penggunaan Teknik resampling SMOTE terhadap kinerja klasifikasi penyakit diabetes menggunakan metode XGBoost yang telah dioptimalkan dengan GridSearchCV dan RandomSearchCV?

1.3 Tujuan Penelitian

- a. Mengoptimalkan kinerja algoritma XGBoost untuk mendeteksi penyakit diabetes melalui hyperparameter tuning menggunakan metode GridSearchCV.
- b. Mengoptimalkan kinerja algoritma XGBoost untuk mendeteksi penyakit diabetes melalui hyperparameter tuning menggunakan metode RandomSearchCV.
- c. Mengevaluasi hasil pengaruh penggunaan Teknik resampling SMOTE terhadap kinerja klasifikasi penyakit diabetes menggunakan metode XGBoost yang telah dioptimalkan dengan GridSearchCV dan RandomSearchCV, serta membandingkan hasil sebelum dan sesudah dilakukan resampling.

1.4 Batasan Masalah

- a. Platform yang digunakan dalam penelitian ini adalah google colaboratory dengan Bahasa pemrograman Python untuk pengolahan data.
- b. Data yang digunakan berasal dari Kaggle dengan judul “Diabetes Dataset”.
- c. Penelitian ini akan menggunakan algoritma Extreme Gradient Boosting (XGBoost) untuk mendeteksi penyakit diabetes.

d. Penelitian ini hanya melakukan perbandingan optimasi hyperparameter Tuning menggunakan GridSearchCV dan RandomSearchCV serta pengaruh Teknik resampling terhadap kinerja model XGBoost dengan hyperparameter tuning.

