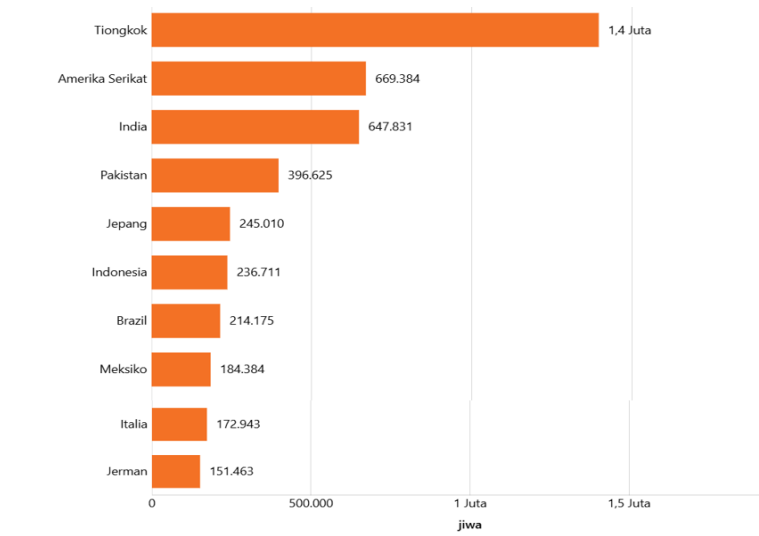


BAB I PENDAHULUAN

1.1 Latar Belakang

Diabetes melitus adalah satu penyakit tidak menular yang menjadi ancaman Kesehatan global. Menurut data World Health Organization (WHO), jumlah penderita diabetes terus meningkat setiap tahunnya diperkirakan 346 juta pertahun 2022 dan International Diabetes Federation (IDF) memperkirakan terdapat 589 juta orang diseluruh dunia mengidap diabetes, di Indonesia sendiri terdapat 20 juta lebih kasus diabetes pada orang dewasa [1][2] . Penyakit ini diperkirakan akan menjadi penyebab utama kematian di banyak negara berkembang, dilansir dari IDF pada tahun 2021 terdapat 6,7 juta kasus kematian diseluruh dunia akibat diabetes. Dapat dilihat pada Gambar 1. 1, Tiongkok menjadi negara dengan jumlah angka kematian tertinggi mencapai 1,4 juta orang dan Indonesia berada di peringkat ke enam dengan jumlah angka kematian mencapai 236 ribu jiwa.



Gambar 1. 1 Diagram Diabetes [2]

Seiring dengan peningkatan jumlah penderita tersebut, kebutuhan akan sebuah potensi untuk pendeteksian dini semakin mendesak. Dalam konteks inilah pendekatan berbasis machine learning menjadi solusi yang banyak dikembangkan dalam era digital saat ini, terlebih lagi pendekatan berbasis data telah digunakan secara luas untuk membantu prediksi dan klasifikasi penyakit, termasuk diabetes

[3]. Namun, Klasifikasi pasien diabetes menjadi tantangan tersendiri, terutama dalam membedakan antara tipe diabetes yang berbeda. Metode tradisional diagnosis diabetes yang didasarkan pada parameter klinis dasar sering kali tidak cukup sensitif dalam membedakan berbagai jenis diabetes, sehingga dapat mengakibatkan misdiagnosis dan kesalahan dalam perencanaan terapi. Sebagai respons terhadap keterbatasan tersebut, penggunaan algoritma machine learning modern seperti XGBoostA semakin dipertimbangkan.

XGBoost menawarkan keunggulan struktur ensemble boosting teratur yang mampu menangani dataset tabular kompleks serta teknik imbalance handling seperti SMOTE untuk meningkatkan performa klasifikasi. XGBoost (Extreme Gradient Boosting) yang telah terbukti unggul dalam berbagai kompetisi data science, ini mampu menangani data yang tidak seimbang dan memberikan hasil yang lebih baik dibandingkan dengan metode lain seperti Random Forest dan Naive Bayes. Penelitian menunjukkan bahwa XGBoost dapat mencapai akurasi hingga 98% dalam klasifikasi diabetes, dengan precision dan recall yang sangat tinggi[3]. Dengan meningkatnya jumlah pasien diabetes, penggunaan teknologi dan algoritma canggih seperti XGBoost menjadi sangat relevan. Penelitian ini bertujuan untuk mengeksplorasi lebih lanjut penerapan XGBoost dalam klasifikasi pasien diabetes, serta dampaknya terhadap pengambilan keputusan klinis dan manajemen kesehatan.

Beberapa metode klasifikasi telah diterapkan sebelumnya dalam upaya klasifikasi pasien diabetes, seperti Naive Bayes, Random Forest, dan Logistic Regression. Namun, masing-masing metode memiliki keterbatasan. Naive Bayes, misalnya, sangat bergantung pada asumsi independensi antar fitur, yang dalam data medis sering tidak terpenuhi [4]. Random Forest meskipun kuat, sering kali menghasilkan model yang sulit diinterpretasikan dan bisa memakan waktu lebih lama dalam pelatihan ketika jumlah pohon sangat besar [5]. Logistic Regression bersifat linier dan memiliki keterbatasan dalam memodelkan relasi non-linier kompleks antar fitur, yang umum dalam data kesehatan [6]. Oleh karena itu, dibutuhkan pendekatan lain yang mampu mengatasi kompleksitas dan karakteristik dataset dengan lebih baik.

Penelitian ini diharapkan mampu memberikan peningkatan kinerja klasifikasi pasien diabetes menggunakan algoritma XGBoost dengan target akurasi minimal 90%, recall di atas 85% untuk kasus positif (pasien diabetes), serta waktu komputasi yang lebih efisien dibandingkan metode baseline lainnya. Hasil ini akan mendukung pengembangan sistem diagnosis otomatis yang lebih presisi dan dapat diintegrasikan ke dalam sistem pendukung keputusan klinis [7].

1.2 Perumusan Masalah

1.2.1 Permasalahan

Diabetes merupakan penyakit kronis yang memerlukan diagnosis dini untuk mencegah komplikasi serius. Saat ini, diagnosis diabetes masih banyak mengandalkan pemeriksaan laboratorium secara manual, yang memakan waktu dan biaya. Selain itu, metode konvensional memiliki tingkat akurasi yang bervariasi dan seringkali kurang efisien dalam menangani data medis yang kompleks.

Secara ilmiah dari sudut pandang Informatika, permasalahan ini berkaitan dengan bagaimana memanfaatkan algoritma machine learning, khususnya XGBoost, untuk meningkatkan akurasi klasifikasi pasien diabetes, serta mengevaluasi pengaruh teknik normalisasi data terhadap performa model klasifikasi

1.2.2 Metode Usulan

Untuk menjawab permasalahan yang telah diidentifikasi, penelitian ini mengusulkan pendekatan sebagai berikut: menggunakan algoritma XGBoost untuk membangun model klasifikasi dan melakukan tuning hyperparameter untuk mengoptimalkan kinerja model, serta menggunakan metrik evaluasi untuk kinerja masing-masing model dan melakukan analisis perbandingan antara model yang menggunakan normalisasi dan tidak untuk menentukan dampak normalisasi terhadap kinerja model.

1.2.3 Pertanyaan Penelitian

Berdasarkan permasalahan dan metode yang diusulkan, pertanyaan penelitian yang diajukan adalah:

1. Bagaimana performa XGBoost dalam mengklasifikasi pasien diabetes berdasarkan dataset yang tersedia?
2. Bagaimana perbandingan metrik evaluasi (akurasi, presisi, recall, F1-score) dan ROC-AUC antara model yang menggunakan normalisasi dan yang tidak?

1.3 Tujuan dan Manfaat

1.3.1 Tujuan Penelitian

1. Penelitian ini bertujuan untuk mengevaluasi kinerja model XGBoost dalam klasifikasi pasien diabetes
2. Untuk membandingkan metrik evaluasi model dan ROC antara model yang memakai normalisasi dan yang tidak

1.3.2 Manfaat Penelitian

Penelitian ini menunjukkan bahwa algoritma XGBoost mampu memberikan akurasi yang lebih tinggi dibandingkan metode lain. Dengan kemampuan untuk mengklasifikasi pasien diabetes secara akurat penelitian ini berkontribusi pada deteksi dini penyakit. Penelitian ini juga bisa menjadi referensi bagi peneliti selanjutnya untuk memperluas pengetahuan di bidang pembelajaran mesin dan aplikasi dalam kesehatan serta mendorong penelitian lebih lanjut dalam klasifikasi penyakit lainnya.

1.4 Batasan Masalah

Agar pembahasan penelitian ini tidak menyimpang dari tujuan, maka diperlukan batasan dalam penelitian ini adalah sebagai berikut:

1. Agar pembahasan penelitian ini tidak menyimpang dari tujuan, maka diperlukan batasan dalam penelitian ini adalah sebagai berikut:
2. Penelitian ini hanya menggunakan algoritma XGBoost untuk klasifikasi diabetes
3. Penelitian ini akan fokus pada klasifikasi biner, yaitu mendeteksi apakah seseorang menderita diabetes atau tidak.
4. Penelitian ini hanya akan membandingkan dua teknik normalisasi, yaitu minmax dan robustscaler.
5. Penelitian ini akan fokus pada metrik evaluasi tertentu, yaitu akurasi, presisi, recall, f1-score dan ROC-AUC.

1.5 Sistematika Proposal

BAB I Pendahuluan :

Bab ini berisi latar belakang penelitian mengenai tingginya penyakit diabetes dan pentingnya deteksi dini menggunakan pendekatan berbasis data. Diuraikan pula permasalahan yang dimiliki data dalam klasifikasi medis serta kebutuhan akan algoritma yang tepat. Bab ini mencakup rumusan masalah, metode usulan, pertanyaan penelitian, tujuan dan manfaat, serta batasan masalah.

BAB II Tinjauan Pustaka :

Bab ini menyajikan kajian literatur dan teori-teori yang mendasari penelitian, mencakup penjelasan mengenai algoritma *XGBoost*, teknik normalisasi, serta evaluasi model menggunakan metrik *Accuracy*, *F1-score*, *Recall*, *Precision* dan *AUC*. Juga disajikan penelitian terdahulu yang relevan sebagai acuan yang melahirkan ide untuk penelitian ini..

BAB III Metodologi Penelitian :

Bab ini menjelaskan metodologi yang digunakan dalam penelitian, meliputi tahapan-tahapan mulai dari pemilihan dataset, *preprocessing*, pelatihan model menggunakan *XGBoost* hingga evaluasi kinerja model menggunakan metrik klasifikasi.

BAB IV Hasil dan Pembahasan :

Bab ini menyajikan hasil uji dari algoritma dan teknik yang digunakan, serta pembahasan terhadap hasil evaluasi pada model berdasarkan nilai *akurasi*, *F1-score*, *precision*, *recall* dan *ROC-AUC*. Disajikan pula perbandingan hasil antara penggunaan normalisasi dan tidak

BAB V Penutup

Bab ini memuat kesimpulan dari hasil penelitian yang menjawab pertanyaan penelitian, serta saran untuk pengembangan penelitian selanjutnya, seperti penggunaan dataset yang lebih besar atau algoritma yang lebih canggih, terutama yang berbasis deep learning.