

Optimalisasi Model Klasifikasi Diabetes Menggunakan Ensemble Learning Adaboost, Gradient Boosting, dan XGBoost

Setyawan Wibisono*¹, Wiwien Hadikurniawati², Heribertus Yulianton³,
Endang Lestariningsih⁴, Taufiq Dwi Cahyono⁵

^{1,2,3,4}Fakultas Teknologi Informasi dan Industri, Teknik Informatika, Universitas Stikubank,
Semarang, Indonesia

⁵Fakultas Teknik, Universitas Semarang, Jl. Soekarno-Hatta, Semarang, Indonesia

*¹setyawan@edu.unisbank.ac.id, ²wiwien@edu.unisbank.ac.id, ³heri@edu.unisbank.ac.id,

⁴endang_lestariningsih @edu.unisbank.ac.id, ⁵taufiq_dc@usm.ac.id

Abstrak

Diabetes mellitus adalah penyakit kronis yang memengaruhi jutaan orang secara global dan membutuhkan metode diagnosis dini untuk mencegah komplikasi. Penelitian ini bertujuan untuk mengoptimalkan prediksi diabetes dengan membandingkan tiga metode ensemble learning: AdaBoost, Gradient Boosting, dan XGBoost. Dataset yang digunakan adalah Diabetes Health Indicators, yang menggabungkan indikator kesehatan seperti tekanan darah, kolesterol, dan kebiasaan gaya hidup. Tahapan penelitian meliputi pemrosesan data, pengembangan model, serta evaluasi performa menggunakan metrik akurasi, presisi, recall, F1-score, dan AUC (Area Under the Curve). Hasil menunjukkan bahwa Gradient Boosting unggul dalam akurasi dan AUC, menandakan kemampuan yang lebih baik dalam mendeteksi diabetes secara konsisten dibandingkan dengan dua metode lainnya. AdaBoost memperlihatkan keseimbangan yang baik antara presisi dan recall, menjadikannya cocok untuk skenario yang memerlukan pengendalian kesalahan positif dan negatif secara proporsional. Sementara itu, XGBoost menawarkan efisiensi pemrosesan yang optimal dengan performa yang kompetitif. Gradient Boosting direkomendasikan untuk aplikasi klinis yang membutuhkan akurasi tinggi, sedangkan AdaBoost dapat menjadi alternatif ketika keseimbangan prediksi menjadi prioritas. Penelitian ini berkontribusi dalam pengembangan alat prediksi diabetes yang lebih akurat, efektif, dan dapat diterapkan di sektor kesehatan untuk mendukung upaya deteksi dini.

Kata Kunci - ensemble learning, AdaBoost, Gradient Boosting, XGBoost

Abstract

Diabetes mellitus is a chronic disease that affects millions of people globally and requires early diagnostic methods to prevent complications. This study aims to optimize diabetes prediction by comparing three ensemble learning methods: AdaBoost, Gradient Boosting, and XGBoost. The dataset used is the Diabetes Health Indicators, which includes health indicators such as blood pressure, cholesterol levels, and lifestyle habits. The research process involves data preprocessing, model development, and performance evaluation using metrics like accuracy, precision, recall, F1-score, and AUC (Area Under the Curve). The results show that Gradient Boosting excels in accuracy and AUC, indicating a better ability to consistently detect diabetes compared to the other two methods. AdaBoost demonstrates a good balance between precision and recall, making it suitable for scenarios where controlling both positive and negative errors is crucial. Meanwhile, XGBoost offers optimal processing efficiency with competitive performance. Gradient Boosting is recommended for clinical applications requiring high accuracy, while AdaBoost can be an alternative when balanced predictions are prioritized. This study contributes to the development of more accurate and effective diabetes prediction tools, which can be applied in the healthcare sector to support early detection efforts.

Keywords— ensemble learning, AdaBoost, Gradient Boosting, XGBoost

1. PENDAHULUAN

Diabetes mellitus telah menjadi salah satu masalah kesehatan global yang semakin mendesak. Berdasarkan laporan International Diabetes Federation (IDF) pada tahun 2021, terdapat lebih dari 536,6 juta orang di seluruh dunia yang menderita diabetes, dan angka ini diproyeksikan akan meningkat

menjadi 783,2 juta pada tahun 2045. Kondisi ini semakin mengkhawatirkan karena hampir setengah dari populasi dewasa yang mengidap diabetes tidak menyadari kondisi kesehatan mereka, yang menyebabkan keterlambatan diagnosis dan pengobatan, sehingga mengarah pada komplikasi serius di kemudian hari [1].

Diabetes, terutama pada tahap awal, sering kali sulit terdeteksi karena gejala yang tidak jelas atau tampak samar. Akibatnya, banyak individu yang berisiko tinggi tidak mendapatkan diagnosis atau perawatan yang tepat waktu. Oleh karena itu, pengembangan metode prediktif yang lebih akurat untuk mengidentifikasi individu dengan risiko tinggi terkena diabetes menjadi sangat penting. Teknologi *machine learning* menawarkan potensi besar dalam mendeteksi pola-pola risiko diabetes secara lebih efisien, sehingga memungkinkan diagnosis dini yang lebih baik dan intervensi yang lebih tepat.

Namun, penerapan algoritma klasifikasi konvensional sering kali menemui tantangan, terutama dalam menghadapi kompleksitas data medis yang sangat bervariasi. Data kesehatan memiliki karakteristik unik yang menuntut metode pemodelan yang lebih canggih untuk menangani berbagai fitur yang kompleks. Untuk mengatasi masalah ini, *teknik ensemble learning* muncul sebagai solusi yang lebih kuat. Teknik ini menggabungkan prediksi dari beberapa model *machine learning* untuk meningkatkan akurasi dan stabilitas prediksi.

Dalam penelitian ini, akan diterapkan teknik ensemble learning, seperti AdaBoost, Gradient Boosting, dan XGBoost, guna meningkatkan kinerja model prediksi risiko diabetes. Penggunaan teknik ini tidak hanya memperbaiki akurasi, tetapi juga menghasilkan model yang lebih andal dan konsisten. Dengan integrasi ini, penelitian ini bertujuan untuk mengembangkan solusi prediktif yang dapat membantu profesional kesehatan dan individu dalam memahami risiko diabetes secara lebih dini, sehingga memungkinkan intervensi yang lebih cepat dan efektif.

Penelitian ini memiliki beberapa tujuan. Pertama, untuk mengembangkan model prediktif berbasis *ensemble learning* yang dapat mengidentifikasi individu dengan risiko tinggi terkena diabetes. Kedua, untuk mengevaluasi performa model yang dikembangkan dengan membandingkan akurasi, stabilitas, dan konsistensi antara berbagai teknik *ensemble learning* dengan algoritma klasifikasi konvensional. Ketiga, untuk memberikan kontribusi praktis dalam upaya meningkatkan kemampuan diagnosa dini diabetes, sehingga dapat mendukung upaya global dalam mengurangi dampak jangka panjang penyakit ini.

Penelitian ini menggunakan *Diabetes Health Indicators Dataset* yang tersedia secara publik melalui platform *UCI Machine Learning Repository*. Data ini mencakup berbagai indikator kesehatan yang relevan dengan risiko diabetes, seperti tekanan darah, kadar kolesterol, dan aktivitas fisik. Proses penelitian mencakup beberapa tahapan utama, termasuk pembersihan data, pengembangan model prediktif, evaluasi performa, serta validasi hasil. Diharapkan model yang dikembangkan dapat memberikan prediksi yang lebih akurat dan dapat diterapkan dalam situasi praktis.

Dengan pendekatan ini, penelitian diharapkan dapat memberikan kontribusi yang signifikan dalam peningkatan deteksi dini diabetes, serta memberikan alat yang lebih efektif bagi tenaga medis dan individu untuk memahami risiko kesehatan mereka. Hasil penelitian ini juga akan divalidasi secara menyeluruh untuk memastikan bahwa model prediktif yang dikembangkan dapat diterapkan dalam skenario nyata dan memberikan manfaat dalam manajemen serta pencegahan diabetes.

2. METODE PENELITIAN

2.1 Obyek Penelitian

Penelitian ini menggunakan data dari *Diabetes Health Indicators Dataset* untuk mengidentifikasi individu berisiko diabetes. Dataset ini, yang tersedia di *UCI Machine Learning Repository*, mencakup 253.680 data dengan 22 fitur kesehatan, demografi, dan gaya hidup. Awalnya, kolom *Diabetes_012* memiliki tiga kelas (0, 1, dan 2) yang kemudian diubah menjadi dua kelas: kelas 0 untuk individu tanpa diabetes (termasuk pre-diabetes) dan kelas 1 untuk individu dengan diabetes penuh. Transformasi ini memungkinkan analisis yang lebih sederhana dengan dua kategori utama: tanpa diabetes (termasuk pre-diabetes) dan dengan diabetes penuh. Setelah transformasi, data dibersihkan untuk menghapus entri yang tidak lengkap atau duplikat, memastikan keandalan model prediktif yang akan dibangun. Proses pembersihan ini penting guna mengurangi potensi kesalahan dan meningkatkan akurasi prediksi diabetes dalam penelitian

2.2. Ensemble Learning

Teknik *ensemble learning* dalam *machine learning* adalah penggabungan hasil dari beberapa model *machine learning* yang berbeda untuk meningkatkan kinerja keseluruhan prediksi [2]. *Ensemble*

learning sering digunakan karena dapat meningkatkan kinerja model klasifikasi secara signifikan, dengan mengurangi *overfitting*, dan membuat prediksi yang lebih stabil [3]. Pendekatan *ensemble learning* yang direncanakan dalam penelitian ini adalah Adaptive Boosting (AdaBoost), Gradient Boosting, dan XGBoost.

Algoritma AdaBoost adalah metode pembelajaran mesin yang efektif untuk tugas klasifikasi, bekerja dengan cara menggabungkan beberapa pembelajar lemah untuk membentuk model klasifikasi yang kuat dan andal [4]. Dalam penelitian terkait diabetes, AdaBoost telah diterapkan untuk mengembangkan model klasifikasi yang mampu membedakan antara individu yang menderita diabetes dan yang tidak [5]. Selain itu, AdaBoost digunakan dalam merancang model prediksi yang lebih akurat pada sistem monitoring glukosa darah, di mana peningkatan akurasi sangat penting untuk deteksi dini dan penanganan diabetes [6].

Penelitian menunjukkan bahwa selain AdaBoost, algoritma Gradient Boosting juga banyak digunakan dalam membangun model prediksi diabetes mellitus. Dengan menggabungkan beberapa "*weak learner*" menjadi "*strong learner*", Gradient Boosting meningkatkan akurasi prediksi secara signifikan. Dalam studi perbandingan, Gradient Boosting terbukti memberikan akurasi lebih tinggi dibandingkan Decision Tree dan Random Forest. [7]. Penelitian menunjukkan bahwa pendekatan Hybrid Gradient Boosting, yang dikombinasikan dengan model lain seperti K-Nearest Neighbors, Gaussian Naive Bayes, dan Logistic Regression menggunakan *ensemble*, menghasilkan akurasi dan performa yang unggul dibandingkan algoritma lain [8]. Selain itu, penelitian yang membandingkan Gradient Boosting Classifier dengan Random Forest untuk klasifikasi diabetes menunjukkan bahwa Gradient Boosting Classifier tidak hanya memiliki akurasi yang lebih tinggi tetapi juga nilai AUC yang lebih baik dalam mendeteksi diabetes [9].

Metode *ensemble boosting* lain yang dapat digunakan dalam klasifikasi adalah XGBoost. Telah terbukti efektif dalam membangun model prediksi risiko diabetes, dengan akurasi yang lebih tinggi melalui pemilihan fitur yang optimal pada dataset "*Early Stage Diabetes Risk Prediction*" [10]. Dalam studi prediksi diabetes tahap awal berdasarkan data kuesioner dari *Sylhet Diabetes Hospital* di Bangladesh, XGBoost menunjukkan kinerja yang unggul dengan waktu komputasi lebih cepat dibandingkan Random Forest, menjadikannya pilihan efisien untuk pemodelan yang membutuhkan respons cepat [11]. XGBoost digunakan untuk mengembangkan sistem prediksi diabetes tahap awal dengan dataset *Pima Indians*, dan terbukti mencapai akurasi tertinggi dibandingkan dengan Logistic Regression, SVM, Random Forest, KNN, dan Decision Trees [12]. Dalam penelitian lain, *fine-tuning* XGBoost pada dataset Pima Indians Diabetes kembali membuktikan keunggulannya dengan mencapai akurasi tertinggi untuk deteksi dini diabetes, menjadikannya alat prediksi yang andal dalam skenario klinis [13].

2.3. Metode Pengembangan

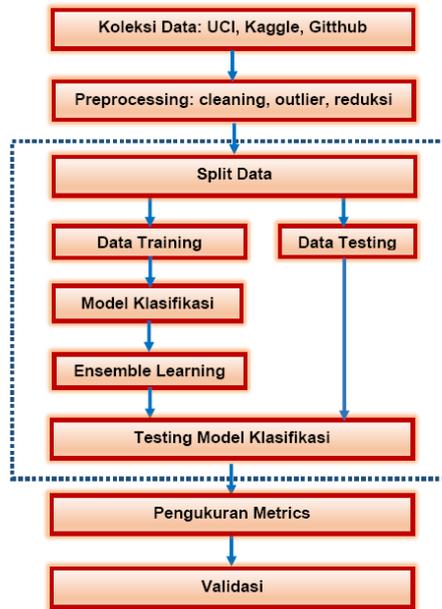
Dalam rangkaian proses penelitian, akan eksperimen komputasi untuk menguji hipotesis, mengidentifikasi faktor risiko, dan mengoptimalkan kinerja model. Selain itu, akan dilakukan dokumentasi semua langkah yang diambil dan hasil. Pada gambar 1 diperlihatkan tahapan penelitian yang dilakukan.

Penelitian ini dimulai dengan langkah pendahuluan, yang mencakup studi literatur mendalam tentang *diabetes mellitus*, teknik klasifikasi, dan metode *ensemble learning*. Tahap ini juga melibatkan pemahaman yang komprehensif tentang peningkatan jumlah penderita diabetes dan urgensi deteksi dini untuk mencegah komplikasi penyakit. Pada tahap pengumpulan data, penelitian ini mengidentifikasi sumber data yang relevan, termasuk data klinis dan demografis, serta mengumpulkan dataset publik yang berfokus pada klasifikasi *diabetes mellitus* dari platform seperti *UCI Machine Learning Repository*. Eksperimen dirancang untuk beberapa dataset publik guna memastikan generalisasi hasil.

Selanjutnya, tahap *preprocessing* data dilakukan untuk membersihkan dataset dari nilai-nilai yang hilang, serta melakukan transformasi data seperti normalisasi fitur numerik, *encoding* fitur kategorikal, dan penggabungan fitur sesuai kebutuhan. Setelah data siap, model klasifikasi dibangun menggunakan beberapa teknik *ensemble learning*, yaitu AdaBoost (Adaptive Boosting), Gradient Boosting, dan XGBoost (Extreme Gradient Boosting). Untuk mengukur performa model, evaluasi dilakukan menggunakan metrik akurasi, presisi, *recall*, *F1-score*, dan AUC.

Analisis dan interpretasi hasil mencakup evaluasi prediksi risiko diabetes untuk mengidentifikasi kekuatan dan kelemahan model yang dikembangkan, serta interpretasi faktor-faktor utama yang memengaruhi prediksi. Langkah terakhir meliputi kesimpulan yang merangkum temuan utama dan memberikan rekomendasi untuk pengembangan pendekatan pencegahan yang lebih efektif

dalam upaya penanggulangan *diabetes mellitus*.



Gambar 1. Tahapan Penelitian

3. HASIL DAN PEMBAHASAN

3.1. Evaluasi Model Klasifikasi

Hasil eksperimen yang diberikan menunjukkan evaluasi dari tiga model klasifikasi, yaitu AdaBoost, Gradient Boosting, dan XGBoost, berdasarkan beberapa metrik kinerja yang umum digunakan untuk menilai performa model, yaitu akurasi, presisi, *recall*, F1 score, dan AUC (*Area Under the Curve*). Tabel 1 menampilkan hasil evaluasi dari setiap algoritma.

Tabel 1. Hasil Eksperimen

Model	Akurasi	Presisi	<i>Recall</i>	F1 Score	AUC
AdaBoost	0.866426	0.836700	0.866426	0.838777	0.831143
Gradient Boosting	0.867510	0.837262	0.867510	0.836020	0.834408
XGBoost	0.866761	0.835799	0.866761	0.835363	0.832147

Model AdaBoost menunjukkan kinerja yang kuat dalam tugas klasifikasi, dengan akurasi 0,866426 atau sekitar 86,64%, menunjukkan prediksi yang benar untuk mayoritas sampel. Presisi model mencapai 0,836700, berarti 83,67% dari semua prediksi positif benar-benar merupakan sampel positif, menandakan kemampuan yang baik dalam meminimalkan kesalahan positif. *Recall* model juga sama tinggi, yaitu 0,866426, yang berarti model berhasil mendeteksi sekitar 86,64% dari total data positif, menunjukkan efektivitas dalam mengidentifikasi contoh positif dalam data.

F1 Score dari AdaBoost adalah 0,838777, menandakan keseimbangan baik antara presisi dan *recall*, serta menunjukkan bahwa model ini memiliki performa konsisten dalam akurasi prediksi positif dan deteksi keseluruhan data positif. AUC (*Area Under the Curve*) sebesar 0,831143 menunjukkan kemampuan yang solid dalam membedakan antara kelas positif dan negatif, dengan nilai AUC di atas 0,8 yang mendekati 1, menandakan kemampuan diskriminatif yang baik.

Metrik-metrik ini menunjukkan bahwa model AdaBoost dapat diandalkan dalam klasifikasi, dengan potensi perbaikan khususnya dalam hal presisi. Hasil ini menjadikan AdaBoost pilihan yang kompeten untuk aplikasi yang memerlukan keseimbangan akurasi dan kemampuan deteksi kelas positif yang tinggi.

Model Gradient Boosting menunjukkan performa yang sedikit lebih unggul dibandingkan AdaBoost dalam berbagai metrik evaluasi. Dengan akurasi sebesar 0,867510 (86,75%), Gradient Boosting lebih sering memberikan prediksi yang benar dibandingkan AdaBoost. Presisi Gradient Boosting mencapai 0,837262, menunjukkan bahwa sekitar 83,73% dari semua prediksi positif yang

dibuat benar-benar merupakan sampel positif, sehingga model ini lebih efektif dalam mengurangi kesalahan tipe I (*false positive*).

Recall model ini sebesar 0,867510, menunjukkan bahwa Gradient Boosting mampu mengidentifikasi sekitar 86,75% dari data positif sebenarnya, sedikit lebih baik dari AdaBoost, yang menandakan sensitivitas lebih tinggi terhadap kelas positif. *F1 Score* Gradient Boosting berada pada 0,836020, sedikit lebih tinggi dari AdaBoost, menunjukkan keseimbangan presisi dan *recall* yang lebih baik, sehingga menghasilkan prediksi yang lebih andal.

AUC Gradient Boosting sebesar 0,834408, sedikit di atas AUC AdaBoost, menunjukkan kemampuan lebih baik dalam membedakan kelas positif dan negatif. Perbedaan kecil dalam AUC ini mengindikasikan bahwa Gradient Boosting sedikit lebih efektif dalam mengidentifikasi sampel positif dan negatif secara akurat. Secara keseluruhan, Gradient Boosting unggul dalam akurasi, presisi, *recall*, *F1 Score*, dan AUC, menjadikannya pilihan yang lebih kuat untuk tugas klasifikasi ini meskipun kedua model tetap sangat kompeten.

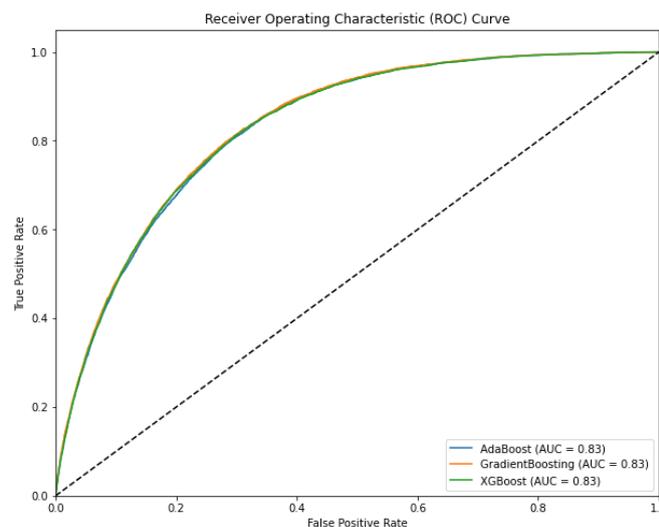
Model XGBoost menunjukkan performa yang hampir sama dengan AdaBoost dan Gradient Boosting dalam akurasi, dengan nilai 0.866761 atau 86.68%. Ini menunjukkan bahwa XGBoost dapat diandalkan dalam memberikan prediksi yang tepat secara keseluruhan, sebanding dengan Gradient Boosting dan sedikit lebih rendah dari AdaBoost.

Presisi XGBoost sebesar 0.835799, sedikit lebih rendah dibandingkan Gradient Boosting, menunjukkan bahwa sekitar 83.58% dari prediksi positifnya benar-benar positif. Meski demikian, presisi XGBoost tetap lebih tinggi dari AdaBoost, menandakan kecenderungan yang baik dalam menghindari kesalahan tipe I (*false positive*). *Recall* XGBoost adalah 0.866761, setara dengan akurasinya, menunjukkan kemampuannya dalam mendeteksi sekitar 86.68% dari data positif, lebih baik daripada AdaBoost dan menunjukkan efisiensi yang kuat dalam mendeteksi kelas positif tanpa kehilangan banyak data.

F1 Score XGBoost berada di 0.835363, sedikit lebih tinggi dari AdaBoost, yang mencerminkan keseimbangan antara presisi dan *recall*. AUC untuk XGBoost adalah 0.832147, sedikit lebih tinggi daripada AdaBoost namun sedikit lebih rendah dari Gradient Boosting. AUC ini menunjukkan kemampuan XGBoost dalam membedakan kelas positif dan negatif, hampir setara dengan Gradient Boosting. XGBoost memberikan hasil yang sangat baik dan kompetitif dalam tugas klasifikasi, dengan keseimbangan yang solid antara presisi, *recall*, dan kemampuan diskriminasi kelas.

3.2. ROC (Receiver Operating Characteristic)

Gambar 2 menunjukkan kurva ROC (*Receiver Operating Characteristic*) yang menggambarkan performa tiga algoritma klasifikasi yang berbeda: AdaBoost, Gradient Boosting, dan XGBoost. Kurva ROC adalah grafik yang memplot hubungan antara *True Positive Rate* (TPR) dan *False Positive Rate* (FPR) dari suatu model klasifikasi pada berbagai *threshold*. *True Positive Rate*, yang juga dikenal sebagai sensitivitas atau *recall*, menunjukkan proporsi sampel positif yang benar-benar terklasifikasi sebagai positif. Sementara itu, *False Positive Rate* menunjukkan proporsi sampel negatif yang salah terklasifikasi sebagai positif.



Gambar 2. ROC

Pada sumbu x (horizontal) terdapat *False Positive Rate*, yang berkisar dari 0 hingga 1, dan pada sumbu y (vertikal) terdapat *True Positive Rate*, juga berkisar dari 0 hingga 1. Garis diagonal putus-putus dari kiri bawah ke kanan atas adalah garis acuan yang menunjukkan model acak atau tidak berdaya, berarti bahwa model yang mengikuti garis ini tidak lebih baik dari tebakan acak. Semakin jauh kurva ROC dari garis acuan ini dan semakin dekat ke sudut kiri atas, semakin baik performa model tersebut.

Dalam grafik ini, ketiga algoritma, yaitu AdaBoost, Gradient Boosting, dan XGBoost, memiliki AUC (*Area Under Curve*) pada kisaran yang hampir sama, yaitu sekitar 0.83. AUC adalah metrik yang sering digunakan untuk mengukur seberapa baik suatu model mampu membedakan antara kelas positif dan negatif. Nilai AUC berkisar antara 0 hingga 1, dengan 0.5 menunjukkan performa model yang setara dengan tebakan acak. Nilai AUC sebesar 0.83, seperti yang terlihat pada ketiga model ini, menunjukkan bahwa ketiga model memiliki performa yang baik dalam mengklasifikasikan data dan mampu membedakan antara kelas positif dan negatif dengan cukup baik.

Meskipun ketiga model memiliki nilai AUC yang sama, bentuk kurva ROC masing-masing model sedikit berbeda, yang mungkin menunjukkan perbedaan kecil dalam kinerja pada berbagai tingkat *threshold*. Namun, karena perbedaan ini sangat kecil dan ketiga kurva hampir tumpang tindih, dapat disimpulkan bahwa performa ketiga model sangat mirip dalam hal diskriminasi antara kelas positif dan negatif.

Secara keseluruhan, grafik ini menunjukkan bahwa baik AdaBoost, Gradient Boosting, maupun XGBoost adalah model yang cukup andal untuk tugas klasifikasi pada dataset ini, dengan performa yang setara satu sama lain dalam hal nilai AUC dan kemampuan untuk memisahkan kelas positif dari kelas negatif.

3.3. Cross-Validation

Cross-Validation adalah teknik yang digunakan untuk menguji model *machine learning* dengan membagi dataset menjadi beberapa *subset* atau *fold*. Dalam *10-Fold Cross-Validation*, dataset dibagi menjadi 10 bagian, model dilatih dengan 9 bagian dan diuji dengan 1 bagian, dan proses ini diulang hingga semua bagian digunakan sebagai data uji. Teknik ini membantu mencegah *overfitting* dan memberikan estimasi yang lebih stabil dan akurat. Tabel 2 menunjukkan hasil evaluasi tiga model: AdaBoost, Gradient Boosting, dan XGBoost berdasarkan akurasi, presisi, *recall*, *F1 score*, dan AUC.

Tabel 2. *10-Fold Cross-Validation*

Model	Akurasi	Presisi	<i>Recall</i>	<i>F1 Score</i>	AUC
AdaBoost	0.864450	0.833619	0.864450	0.835075	0.824833
Gradient Boosting	0.866081	0.835596	0.866081	0.832882	0.828730
XGBoost	0.864539	0.832660	0.864539	0.832229	0.824696

Model AdaBoost memiliki akurasi 86.45%, yang menunjukkan kemampuan klasifikasi yang kuat. Presisi sebesar 83.36% menunjukkan bahwa sebagian besar prediksi positifnya benar-benar positif, dan *recall* 86.45% menandakan kemampuannya mendeteksi sampel positif. *F1 Score* 83.51% menunjukkan keseimbangan presisi dan *recall*, sementara AUC 82.48% mencerminkan kemampuan model membedakan kelas.

Model Gradient Boosting memiliki performa unggul pada hampir semua metrik. Akurasinya 86.61%, sedikit lebih tinggi dibanding AdaBoost dan XGBoost, serta presisi 83.56%, yang menunjukkan akurasi tinggi dalam memprediksi kelas positif. *Recall* juga sebesar 86.61%, menunjukkan keunggulan dalam mendeteksi sampel positif. AUC tertinggi sebesar 82.87% menunjukkan bahwa Gradient Boosting unggul dalam membedakan kelas, meskipun *F1 Score* 83.29% sedikit lebih rendah dibandingkan AdaBoost.

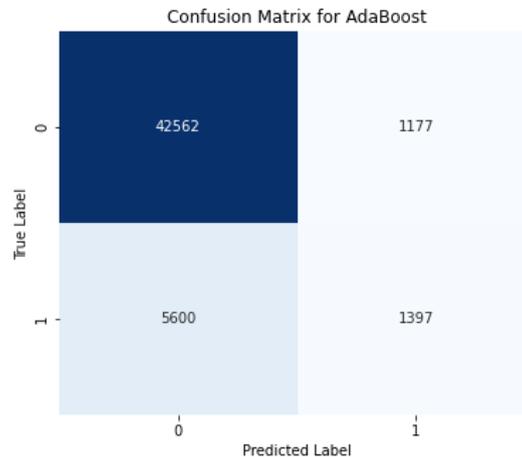
XGBoost mencatat akurasi serupa dengan AdaBoost, 86.45%, namun presisi 83.27% sedikit lebih rendah dibandingkan kedua model lainnya. *Recall* yang sama dengan akurasinya mencerminkan kesetaraan dalam mendeteksi sampel positif. *F1 Score* 83.22% dan AUC 82.47% menunjukkan bahwa XGBoost memiliki performa yang hampir setara, meski sedikit lebih rendah dalam presisi dan *F1 Score*.

Secara keseluruhan, Gradient Boosting memberikan performa terbaik dalam akurasi, presisi, dan AUC, ideal untuk akurasi tinggi dan deteksi positif. AdaBoost menonjol pada *F1 Score*, cocok jika ada keseimbangan presisi dan *recall*, sementara XGBoost hampir sebanding, namun dapat ditingkatkan lebih lanjut.

3.4. Confussion Matrix

3.4.1. Adaboost

Confussion Matrix ini adalah alat evaluasi yang membantu memahami performa model klasifikasi dengan memberikan informasi detail tentang prediksi yang benar dan salah pada setiap kelas. Gambar 3 menunjukkan confusion matrix untuk algoritma klasifikasi AdaBoost.



Gambar 3 *Confusion Matrix* AdaBoost

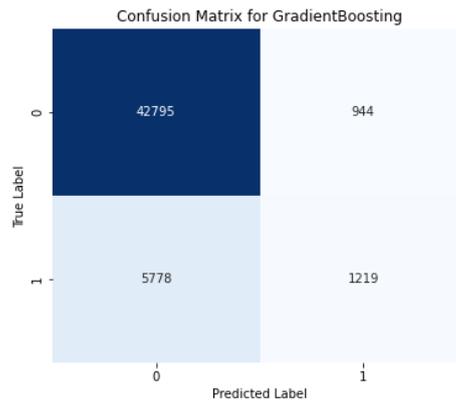
Confusion Matrix berbentuk persegi, dengan dua baris dan dua kolom yang mewakili dua kelas dalam dataset ini, yaitu kelas 0 (negatif) dan kelas 1 (positif). Sumbu y pada matriks ini menunjukkan "*True Label*" atau label asli dari data, sementara sumbu x menunjukkan "*Predicted Label*" atau label yang diprediksi oleh model. Setiap elemen di dalam matriks menunjukkan jumlah sampel yang diklasifikasikan dalam kategori tertentu, berdasarkan kombinasi dari label asli dan prediksi model. Berikut adalah penjelasan lebih lanjut tentang setiap elemen dalam matriks ini:

1. *True Negatives* (TN): Elemen pada baris pertama dan kolom pertama, yaitu nilai 42,562, menunjukkan jumlah sampel dengan label asli 0 (negatif) yang diprediksi sebagai 0 oleh model. Ini berarti ada 42,562 sampel yang diklasifikasikan dengan benar sebagai kelas negatif oleh AdaBoost.
2. *False Positives* (FP): Elemen pada baris pertama dan kolom kedua, yaitu nilai 1,177, menunjukkan jumlah sampel dengan label asli 0 tetapi diprediksi sebagai 1 oleh model. Ini berarti ada 1,177 sampel negatif yang salah diklasifikasikan sebagai positif oleh AdaBoost.
3. *False Negatives* (FN): Elemen pada baris kedua dan kolom pertama, yaitu nilai 5,600, menunjukkan jumlah sampel dengan label asli 1 tetapi diprediksi sebagai 0 oleh model. Ini berarti ada 5,600 sampel positif yang tidak dikenali dengan benar oleh AdaBoost dan justru diklasifikasikan sebagai negatif.
4. *True Positives* (TP): Elemen pada baris kedua dan kolom kedua, yaitu nilai 1,397, menunjukkan jumlah sampel dengan label asli 1 yang diprediksi dengan benar sebagai 1 oleh model. Ini berarti AdaBoost mampu mengklasifikasikan 1,397 sampel positif dengan benar.

Model AdaBoost menunjukkan performa baik dalam mengidentifikasi kelas negatif dengan 42.562 *true negatives*, namun mengalami kesalahan lebih tinggi pada kelas positif dengan 5.600 *false negatives*. Ini mengindikasikan ketidakseimbangan terhadap kelas negatif. Ketidakseimbangan antara *false positives* (1.177) dan *false negatives* dapat berisiko, terutama dalam konteks di mana kegagalan mendeteksi kelas positif berakibat merugikan. Oleh karena itu, model ini perlu penyesuaian lebih lanjut untuk meningkatkan *recall* pada kelas positif.

3.4.2. Gradient Boosting

Gambar 4 menunjukkan *confusion matrix* yang digunakan untuk mengevaluasi performa model klasifikasi Gradient Boosting. Matriks ini membantu dalam memahami sejauh mana model Gradient Boosting mampu mengklasifikasikan sampel dengan benar, serta seberapa besar tingkat kesalahan klasifikasinya.



Gambar 4. Confusion Matrix Gradient Boosting

Berikut adalah rincian penjelasan untuk setiap elemen dalam matriks ini:

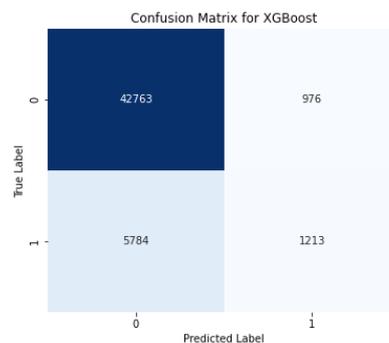
1. *True Negatives* (TN): Elemen pada baris pertama dan kolom pertama, yaitu nilai 42,795, menunjukkan jumlah sampel dengan label asli 0 (negatif) yang diprediksi sebagai 0 oleh model. Ini menunjukkan bahwa Gradient Boosting mampu mengklasifikasikan 42,795 sampel negatif dengan benar.
2. *False Positives* (FP): Elemen pada baris pertama dan kolom kedua, yaitu nilai 944, menunjukkan jumlah sampel dengan label asli 0 tetapi diprediksi sebagai 1 oleh model. Dengan kata lain, ada 944 sampel negatif yang salah diklasifikasikan sebagai positif. *False positives* ini merupakan kesalahan klasifikasi yang terjadi saat model mengira sampel negatif sebagai positif.
3. *False Negatives* (FN): Elemen pada baris kedua dan kolom pertama, yaitu nilai 5,778, menunjukkan jumlah sampel dengan label asli 1 tetapi diprediksi sebagai 0 oleh model. Ini berarti Gradient Boosting gagal mengenali 5,778 sampel positif, yang justru diklasifikasikan sebagai negatif. Kesalahan ini bisa bermasalah tergantung pada konteks aplikasi, terutama jika kegagalan mendeteksi sampel positif memiliki konsekuensi serius.
4. *True Positives* (TP): Elemen pada baris kedua dan kolom kedua, yaitu nilai 1,219, menunjukkan jumlah sampel dengan label asli 1 yang diprediksi dengan benar sebagai 1 oleh model. Dengan kata lain, model Gradient Boosting berhasil mengenali 1,219 sampel positif.

Gradient Boosting menunjukkan performa yang baik dalam mengklasifikasikan sampel negatif (42,795 *true negatives*), sementara jumlah *false positives* (944) relatif kecil. Namun, seperti pada model AdaBoost, Gradient Boosting juga memiliki kesulitan dalam mendeteksi sampel positif, yang terlihat dari jumlah *false negatives* yang cukup tinggi (5,778), dibandingkan dengan *true positives* (1,219). Ini menunjukkan bahwa model ini mungkin memiliki bias terhadap kelas negatif, atau cenderung konservatif dalam mengklasifikasikan sampel sebagai positif.

Kondisi ini dapat memiliki konsekuensi yang berbeda tergantung pada konteks aplikasinya. Untuk konteks seperti ini, model mungkin perlu ditingkatkan lebih lanjut untuk mengurangi *false negatives*, terutama jika deteksi sampel positif sangat penting.

3.4.3. XGBoost

Dalam penelitian ini, confusion matrix digunakan untuk mengevaluasi model XGBoost yang telah dilatih untuk memprediksi risiko penyakit jantung, seperti terlihat pada gambar 5.



Gambar 5. Confusion Matrix XGBoost

Gambar 5 menunjukkan *confusion matrix* dari model XGBoost, yang berfungsi untuk mengevaluasi performa model dalam mengklasifikasikan data ke dalam dua kelas: kelas 0 (negatif) dan kelas 1 (positif). Dari matriks ini, kita bisa menganalisis kekuatan model dalam mengenali sampel negatif dan positif, serta mengidentifikasi area yang masih memerlukan peningkatan agar performa model lebih optimal. Berikut adalah rincian penjelasan setiap elemen dalam matriks:

1. *True Negatives* (TN): Angka di kuadran kiri atas, yaitu 42,763, adalah jumlah kasus di mana model berhasil memprediksi kelas 0 dengan benar, dan kelas sebenarnya juga 0. Ini berarti model secara akurat mengklasifikasikan 42,763 sampel negatif.
2. *False Positives* (FP): Angka di kuadran kanan atas, yaitu 976, menunjukkan jumlah kasus di mana model salah memprediksi kelas 1, padahal kelas sebenarnya adalah 0. Ini disebut sebagai kesalahan positif palsu, yang dalam beberapa konteks, dapat mengindikasikan bahwa model mungkin terlalu sering memberi prediksi positif saat tidak seharusnya.
3. *False Negatives* (FN): Angka di kuadran kiri bawah, yaitu 5,784, menunjukkan jumlah kasus di mana model salah memprediksi kelas 0, padahal kelas sebenarnya adalah 1. Ini menunjukkan bahwa model cenderung melewati beberapa kasus positif, yang dapat menjadi hal yang penting jika kelas positif memiliki arti signifikan (misalnya, jika positif menunjukkan kondisi penyakit atau fraud).
4. *True Positives* (TP): Angka di kuadran kanan bawah, yaitu 1,213, menunjukkan jumlah kasus di mana model berhasil memprediksi kelas 1 dengan benar, dan kelas sebenarnya juga 1. Ini adalah kasus di mana model secara akurat mengidentifikasi sampel positif.

XGBoost menunjukkan performa yang baik dalam mengklasifikasikan sampel negatif, dengan 42,763 *true negatives* (TN), yang berarti model secara konsisten berhasil mengenali sampel negatif. Jumlah *false positives* (976) relatif kecil, menunjukkan bahwa model jarang salah dalam mengklasifikasikan sampel negatif sebagai positif.

XGBoost juga memiliki kesulitan dalam mendeteksi sampel positif. Hal ini terlihat dari jumlah *false negatives* (5,784) yang cukup tinggi dibandingkan dengan *true positives* (1,213). Kesalahan ini menunjukkan bahwa model mungkin memiliki kecenderungan bias terhadap kelas negatif atau bersifat konservatif dalam mengklasifikasikan sampel sebagai positif. Kondisi ini dapat berdampak signifikan, terutama jika deteksi sampel positif sangat penting.

Penelitian ini menunjukkan bahwa Gradient Boosting, AdaBoost, dan XGBoost memiliki keunggulan berbeda dalam tugas klasifikasi dua kelas. Gradient Boosting memberikan performa terbaik pada akurasi (0.866081), presisi, *recall*, dan AUC (0.828730), menunjukkan kemampuan tinggi dalam membedakan kelas positif dan negatif dengan tingkat kesalahan yang rendah. Model ini sangat cocok untuk aplikasi yang memerlukan akurasi dan deteksi kuat terhadap kelas positif, meskipun keseimbangan antara presisi dan *recall* bisa ditingkatkan untuk kasus yang memerlukan sensitivitas tinggi.

AdaBoost, meski akurasinya sedikit di bawah Gradient Boosting, unggul pada *F1 Score* (0.835075), memberikan keseimbangan optimal antara presisi dan *recall*. Ini menjadikan AdaBoost pilihan ideal dalam situasi di mana kesalahan positif dan negatif perlu dikendalikan secara proporsional. AUC-nya (0.824833) yang sedikit lebih rendah dari Gradient Boosting tetap menunjukkan bahwa AdaBoost cukup andal dalam membedakan kelas.

XGBoost menampilkan performa yang mendekati AdaBoost dengan sedikit penurunan pada presisi dan *F1 Score*. Kelebihan utama XGBoost terletak pada efisiensi pemrosesan dan kemampuan menangani dataset besar, menjadikannya pilihan baik untuk aplikasi produksi yang menuntut kecepatan dan skalabilitas.

Pemilihan model bergantung pada beberapa konteks, yaitu Gradient Boosting untuk akurasi dan deteksi kelas yang kuat, AdaBoost untuk keseimbangan antara presisi dan *recall*, dan XGBoost untuk efisiensi pemrosesan.

Untuk penelitian di masa yang akan datang, penelitian dapat menggabungkan ketiga model ini melalui teknik *ensemble* seperti *stacking* atau *blending*, memanfaatkan keunggulan masing-masing model. Misalnya, mengombinasikan Gradient Boosting dan AdaBoost dapat menghasilkan model gabungan yang lebih tangguh, dengan memanfaatkan kekuatan akurasi dari Gradient Boosting dan keseimbangan presisi, *recall* dari AdaBoost, ditambah logistic regression atau SVM untuk menyempurnakan prediksi akhir.

4. KESIMPULAN

Model AdaBoost, Gradient Boosting, dan XGBoost memiliki kekuatan dan kelemahan masing-masing dalam tugas klasifikasi dua kelas. Gradient Boosting terbukti unggul dalam akurasi, presisi, *recall*, dan AUC, membuatnya ideal untuk aplikasi yang memerlukan deteksi akurat sampel positif, terutama di bidang yang menuntut keandalan tinggi. Meski memiliki akurasi sedikit lebih rendah, AdaBoost menonjol dalam *F1 Score*, memberikan keseimbangan baik antara presisi dan *recall*, cocok untuk aplikasi yang membutuhkan pengendalian proporsional atas kesalahan pada kedua kelas. XGBoost, meskipun sedikit kurang presisi dibandingkan dua model lainnya, menawarkan efisiensi dan stabilitas yang membuatnya ideal dalam pemrosesan data besar atau *real-time*, dengan potensi performa tinggi melalui *tuning parameter* yang lebih lanjut. Penelitian ini menunjukkan bahwa pemilihan model bergantung pada konteks aplikasi, yaitu Gradient Boosting untuk akurasi tinggi, AdaBoost untuk keseimbangan performa, dan XGBoost untuk efisiensi pemrosesan. Kombinasi dan tuning lebih lanjut dari ketiga model ini dapat meningkatkan performa klasifikasi yang lebih kompleks di masa depan.

DAFTAR PUSTAKA

- [1] K. Ogurtsova *et al.*, “IDF diabetes Atlas: Global estimates of undiagnosed diabetes in adults for 2021,” *Diabetes Res. Clin. Pract.*, vol. 183, 2022, doi: 10.1016/j.diabres.2021.109118.
- [2] U. e. Laila, K. Mahboob, A. W. Khan, F. Khan, dan W. Taekeun, “An Ensemble Approach to Predict Early-Stage Diabetes Risk Using Machine Learning: An Empirical Study,” *Sensors*, vol. 22, no. 14, hal. 1–15, 2022, doi: 10.3390/s22145247.
- [3] M. H. D. M. Ribeiro dan L. dos Santos Coelho, “Ensemble approach based on bagging, boosting and stacking for short-term prediction in agribusiness time series,” *Appl. Soft Comput. J.*, vol. 86, hal. 105837, 2020, doi: 10.1016/j.asoc.2019.105837.
- [4] S. Wu dan S. Meng, “A Modern Communication Path for Traditional Chinese Cultural Design Concepts Based on AdaBoost Model,” *Appl. Math. Nonlinear Sci.*, vol. 9, no. 1, hal. 90781403, 2024, doi: 10.2478/amns.2023.2.00068.
- [5] P. Chen dan C. Pan, “Diabetes classification model based on boosting algorithms,” *BMC Bioinformatics*, vol. 19, no. 1, hal. 1–9, 2018, doi: 10.1186/s12859-018-2090-9.
- [6] P. K. Anand, D. R. Shin, dan M. L. Memon, “Adaptive boosting based personalized glucose monitoring system (PGMS) for non-invasive blood glucose prediction with improved accuracy,” *Diagnostics*, vol. 10, no. 5, hal. 1–22, 2020, doi: 10.3390/diagnostics10050285.
- [7] F. NUSRAT, B. UZBAŞ, dan Ö. K. BAYKAN, “Gradient Boosting Classification kullanakar Diabetes Mellitus Tahmini,” *Eur. J. Sci. Technol.*, no. September, hal. 268–272, 2020, doi: 10.31590/ejosat.803504.
- [8] S. Deepa dan B. Booba, “Predict Diabetes Healthcare Analytics Using Hybrid Gradient Boosting Machine Learning Model,” vol. 30, no. 5, hal. 2928–2945, 2024, doi: 10.53555/kuvey.v30i5.3371.
- [9] S. P. Nainggolan dan A. Sinaga, “Comparative Analysis of Accuracy of Random Forest and Gradient Boosting Classifier Algorithm for Diabetes Classification,” *Sebatik*, vol. 27, no. 1, hal. 97–102, 2023, doi: 10.46984/sebatik.v27i1.2157.
- [10] Kartina Diah Kusuma Wardani dan Memen Akbar, “Diabetes Risk Prediction using Feature Importance Extreme Gradient Boosting (XGBoost),” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 7, no. 4, hal. 824–831, 2023, doi: 10.29207/resti.v7i4.4651.
- [11] S. Gündoğdu, “Efficient prediction of early-stage diabetes using XGBoost classifier with random forest feature selection technique,” *Multimed. Tools Appl.*, vol. 82, no. 22, hal. 34163–34181, 2023, doi: 10.1007/s11042-023-15165-8.
- [12] F. Mazhar, W. Akbar, M. Sajid, N. Aslam, M. Imran, dan H. Ahmad, “Boosting Early Diabetes Detection: An Ensemble Learning Approach with XGBoost and LightGBM,” *J. Comput. & Biomed. Informatics*, vol. 6, no. 02, hal. 127–138, 2024.
- [13] A. Maulana *et al.*, “Machine Learning Approach for Diabetes Detection Using Fine-Tuned XGBoost Algorithm,” *Infolitika J. Data Sci.*, vol. 1, no. 1, hal. 1–7, 2023, doi: 10.60084/ijds.v1i1.72.