

## Peningkatan Kinerja Deteksi Dini Diabetes Melalui Komparasi Algoritma Klasifikasi dengan Penerapan Hybrid Resampling SMOTE-ENN

Raihan Ariq Muzakki<sup>\*1</sup>, Bayu Aji Pradana<sup>2</sup>, Syahril Achamad Fahrezi<sup>3</sup>, Fazri Abdurrahman<sup>4</sup>,  
Awang Andriansyah Hermawan<sup>5</sup>, Advan Zharif Naja<sup>6</sup>, Dimas Hendri Pamugnkas<sup>7</sup>

<sup>1</sup>Informatika, Universitas Bhayangkara Jakarta Raya, Indonesia

Email: <sup>1</sup>202310715297@mhs.ubharajaya.ac.id

Received : Jun 9, 2025; Revised : Nov 20, 2025; Accepted : Nov 22, 2025; Published : Dec 11, 2025

### Abstract

Deteksi dini diabetes sangat penting untuk mengurangi komplikasi dan mendukung intervensi pencegahan. Namun, dataset indikator kesehatan yang umum digunakan untuk skrining diabetes sering kali memiliki ketidakseimbangan kelas yang tinggi, sehingga dapat membuat model klasifikasi bias terhadap kelas mayoritas dan menurunkan nilai recall untuk kasus positif. Penelitian ini mengkaji dampak hybrid resampling SMOTE-ENN terhadap deteksi dini diabetes melalui perbandingan tiga algoritma klasifikasi berbasis pohon: Decision Tree, Random Forest, dan XGBoost. Dataset yang digunakan berisi 253.680 data dengan 21 atribut prediktor terkait indikator kesehatan, serta distribusi target yang sangat tidak seimbang. Penelitian dilakukan dalam dua skenario eksperimen: (1) pelatihan baseline pada dataset asli yang tidak seimbang dan (2) pelatihan pada dataset hasil resampling SMOTE-ENN, kemudian dievaluasi menggunakan metrik accuracy, precision, recall, F1-score, dan ROC-AUC. Hasil penelitian menunjukkan bahwa SMOTE-ENN secara signifikan meningkatkan kemampuan deteksi kelas minoritas dengan menaikkan recall dan ROC-AUC pada seluruh model yang dievaluasi. Pendekatan dengan kinerja terbaik adalah Random Forest yang dikombinasikan dengan SMOTE-ENN, karena memberikan keseimbangan yang baik antara precision dan recall serta menghasilkan nilai ROC-AUC yang kuat. Temuan ini menunjukkan bahwa hybrid resampling dapat menjadi strategi yang efektif untuk meningkatkan model skrining diabetes ketika tujuan utamanya adalah meminimalkan kasus positif yang terlewat.

**Keywords :** class imbalance, diabetes prediction, early detection, machine learning, SMOTE-ENN, XGBoost

This work is an open access article licensed under a Creative Commons Attribution 4.0 International License.



## 1. INTRODUCTION

Diabetes mellitus (DM) merupakan penyakit kronis yang dapat menimbulkan komplikasi serius jika tidak terdeteksi dan ditangani lebih dini. Oleh karena itu, pengembangan sistem prediksi/skrining DM yang akurat menjadi penting untuk mendukung keputusan preventif dan mengurangi beban layanan kesehatan [1]-[3]. Studi terkini juga menunjukkan bahwa pendekatan berbasis machine learning (ML) semakin banyak digunakan untuk klasifikasi dan prediksi DM karena mampu mengolah data indikator kesehatan secara efisien [4], [5].

Salah satu tantangan utama dalam prediksi DM adalah ketidakseimbangan kelas (class imbalance), yaitu kondisi ketika sampel kelas positif (diabetes/prediabetes) jauh lebih sedikit dibanding kelas negatif (non-diabetes). Kondisi ini membuat model cenderung “aman” memilih mayoritas sehingga metrik seperti akurasi dapat terlihat tinggi, namun recall untuk kelas positif menjadi rendah (kasus diabetes banyak terlewat), yang justru berbahaya pada skenario deteksi dini [6], [7]. Literatur juga menegaskan bahwa ketidakseimbangan kelas dapat menurunkan kemampuan model dalam mempelajari pola kelas minoritas dan memerlukan penanganan khusus seperti resampling [7], [8].

Penelitian terkait telah banyak dilakukan. Chang dkk. mengevaluasi berbagai model ML untuk prediksi diabetes dari indikator kesehatan dan menunjukkan bahwa dataset indikator kesehatan memiliki

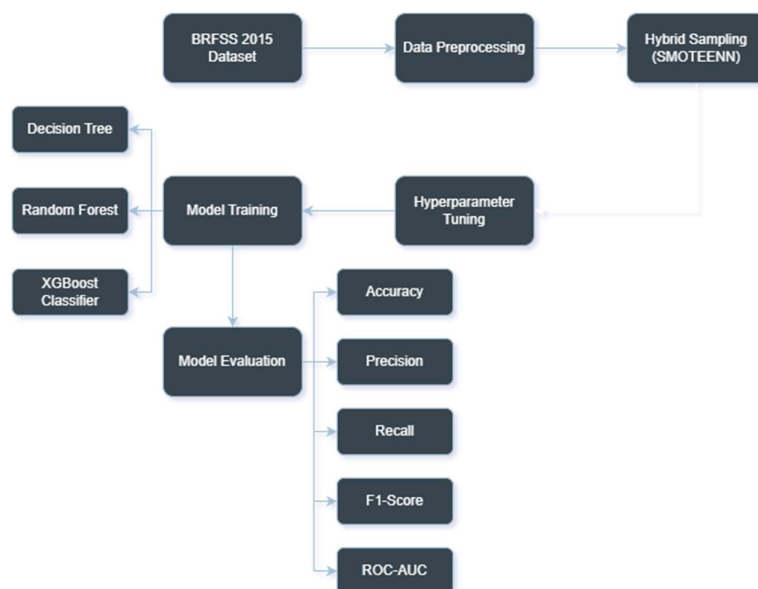
proporsi kelas positif yang relatif kecil dan membutuhkan strategi penyeimbangan agar performa meningkat [5]. Wantoro dkk. menerapkan feature selection serta class balancing berbasis SMOTE dan melaporkan bahwa SMOTE dapat meningkatkan performa beberapa algoritma (misalnya Random Forest dan Tree), walaupun tidak selalu meningkatkan semua model (SVM turun) [16]. Al Ghozali dkk. membandingkan beberapa model (XGBoost, Random Forest, ANN, SVM, Naive Bayes) dan memperoleh akurasi uji terbaik sekitar 0.8690 menggunakan XGBoost, namun studi tersebut masih membuka ruang untuk penguatan aspek resampling/imbalance dan eksplorasi metrik yang lebih berfokus pada kesalahan minoritas [17]. Sementara itu, Setiawan dkk. menunjukkan bahwa SMOTE-ENN dapat meningkatkan performa prediksi DM pada dataset imbalanced (BRFSS 2015 dan PIMA Indian), dan menekankan bahwa hybrid sampling mampu memperbaiki deteksi kelas minoritas [18].

Berdasarkan gap tersebut, penelitian ini berfokus pada:

1. Komparasi algoritma klasifikasi (Decision Tree, Random Forest, XGBoost) pada deteksi dini diabetes,
2. Analisis dampak penerapan hybrid resampling SMOTE-ENN terhadap metrik yang relevan untuk skrining (precision, recall dan ROC-AUC), dan
3. Penyajian hasil sebelum vs sesudah resampling untuk menegaskan pengaruh penanganan class imbalance secara empiris.

Kontribusi utama penelitian ini adalah memberikan evaluasi terstruktur yang menekankan peningkatan kemampuan deteksi kasus positif diabetes (recall) melalui penerapan SMOTE-ENN serta menunjukkan model terbaik untuk skenario deteksi dini berbasis indikator kesehatan.

## 2. METHOD



Gambar 1 Desain dari Metode yang diusulkan

### 2.1. Dataset

Penelitian menggunakan dataset indikator kesehatan yang populer untuk prediksi diabetes berbasis survei BRFSS 2015 (tersedia sebagai Diabetes Health Indicators Dataset). Dataset berisi 253,680 data dengan total 22 kolom (21 fitur dan 1 target) dan target biner Diabetes\_binary [15]. Dataset ini juga digunakan di studi literatur sebagai basis prediksi DM berbasis indikator kesehatan [5], [18].

Tabel 1 Fitur Dataset BRFSS 2015

Fitur	Peran	Tipe
ID	ID	Integer
Diabetes_binary	Target	Binary
HighBP	Fitur	Binary
HighChol	Fitur	Binary
CholCheck	Fitur	Binary
BMI	Fitur	Integer
Smoker	Fitur	Binary
Stroke	Fitur	Binary
HeartDiseaseorAttack	Fitur	Binary
PhysActivity	Fitur	Binary
Fruits	Fitur	Binary
Veggies	Fitur	Binary
HvyAlcoholConsump	Fitur	Binary
AnyHealthcare	Fitur	Binary
NoDocbcCost	Fitur	Binary
GenHlth	Fitur	Integer
MentHlth	Fitur	Integer
PhysHlth	Fitur	Integer
DiffWalk	Fitur	Binary
Sex	Fitur	Binary
Age	Fitur	Integer
Education	Fitur	Integer
Income	Fitur	Integer

## 2.2. Data Preprocessing

Tahap preprocessing mencakup:

1. Pemisahan fitur dan target (X dan y),
2. Normalisasi Min-Max agar seluruh fitur berada pada rentang [0,1] sehingga kontribusi fitur lebih seimbang dan sesuai untuk model yang sensitif terhadap skala, dan
3. Pembagian data menjadi train-test split (80:20).

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

Normalisasi Min-Max dipilih karena sebagian besar fitur dalam dataset bersifat biner 0/1 sehingga rentang [0,1] selaras dengan distribusi fitur.

## 2.3. Hybrid Sampling (SMOTE-ENN)

SMOTE-ENN merupakan gabungan oversampling SMOTE (membuat sampel sintetis untuk kelas minoritas) dan undersampling ENN (membersihkan noise/instance yang “salah tetangga” berdasarkan nearest neighbors). Dengan demikian, SMOTE-ENN diharapkan tidak hanya menambah data minoritas, tetapi juga mengurangi data yang berpotensi mengganggu batas keputusan klasifikasi.

Penerapan SMOTE-ENN pada penelitian ini dilakukan menggunakan SMOTEENN(random\_state=42) dan menghasilkan perubahan distribusi kelas yang lebih seimbang dibanding kondisi awal (detail pada Bagian Hasil)

#### 2.4. Training Model Algoritma Klasifikasi

Penelitian membandingkan tiga algoritma klasifikasi:

- Decision Tree (DT)
- Random Forest (RF)
- XGBoost (XGB)

Pemilihan model ini didasarkan pada performa kuat model pohon keputusan dan ensemble pada data tabular serta relevansinya pada studi prediksi DM berbasis ML [4], [18].

#### 2.5. Skema Eksperimen

Eksperimen dilakukan pada dua skenario:

1. Baseline (tanpa resampling): model dilatih pada data train asli dan diuji pada test asli.
2. Dengan SMOTE-ENN: model dilatih pada data train hasil resampling dan diuji pada test hasil resampling.

Selain itu, dilakukan juga hyperparameter tuning menggunakan GridSearchCV pada masing-masing model untuk melihat dampak optimasi parameter terhadap performa. Pada notebook penelitian, tuning dilakukan dengan beberapa kandidat parameter penting seperti `n_estimators`, `max_depth`, `class_weight`, `scale_pos_weight`, dan `learning_rate`.

Tabel 2 Algoritma klasifikasi dan hyperparameter untuk tuning

Algoritma	Nama Parameter	Nilai
Random Forest	<code>n_estimators</code>	[50, 100, 200]
	<code>max_depth</code>	[None, 10, 20]
	<code>class_weight</code>	['balanced', 'balanced_subsample']
XGBoost	<code>n_estimators</code>	[50, 100, 200]
	<code>max_depth</code>	[3, 10, 20]
	<code>scale_pos_weight</code>	[1, ratio]
	<code>learning_rate</code>	[0.01, 0.1, 0.2]
Decision Tree	<code>max_depth</code>	[None, 10, 20]
	<code>min_samples_split</code>	[2, 5, 10]
	<code>class_weight</code>	['balanced']

Tabel 3 Nilai hyperparameter terbaik setelah tuning

Model	Nama Parameter	Best Parameter	
		GridSearch	GridSearch+SMOTE-ENN
Random Forest	<code>n_estimators</code>	200	200
	<code>max_depth</code>	None	None
	<code>class_weight</code>	'balanced_subsample'	'balanced_subsample'
XGBoost	<code>n_estimators</code>	200	200
	<code>max_depth</code>	3	20

	scale_pos_weight	1	1
	learning_rate	0.2	0.1
	max_depth	None	20
Decision Tree	min_samples_split	2	10
	class_weight	balanced'	balanced'

## 2.6. Metrik Evaluasi

Penelitian ini mengevaluasi performa model klasifikasi menggunakan Confusion Matrix untuk menurunkan nilai metrik accuracy, precision, recall, F1-score, serta ROC-AUC. Confusion matrix merepresentasikan jumlah prediksi benar dan salah pada tiap kelas, yaitu True Positive (TP), False Positive (FP), False Negative (FN), dan True Negative (TN) [16].

Secara klinis, kesalahan FN (pasien diabetes diprediksi tidak diabetes) berisiko menunda diagnosis dan penanganan, sehingga dalam konteks skrining dini, metrik seperti recall menjadi krusial, disamping menjaga FP tetap wajar [18].

Tabel 4 Confusion Matrix (CM)

Aktual \ Prediksi	Negatif (Tidak Diabetes)	Positif (Diabetes)
Negatif (Tidak Diabetes)	True Negative	False Negative
Positif (Diabetes)	False Positive	True Positive

TP (True Positive), TN (True Negative), FP (False Positive), FN (True Negative) didefinisikan seperti pada Tabel 2, maka metrik yang digunakan adalah:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

$$F1 = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \quad (5)$$

Untuk mengukur kemampuan model membedakan kelas positif vs negatif pada berbagai ambang keputusan [18], digunakan kurva ROC dan nilai AUC (luas area di bawah kurva ROC):

$$AUC = \int_0^1 TPR(FPR) d(FPR) \quad (6)$$

## 3. RESULT

### 3.1. Karakteristik Data dan Ketidakseimbangan Kelas

Dataset indikator kesehatan BRFSS 2015 yang digunakan terdiri dari **253,680** baris data dan **21** variabel prediktor dengan target biner **Diabetes\_binary** (0 = tidak diabetes, 1 = prediabetes/diabetes). Distribusi kelas pada data asli menunjukkan kondisi **imbalanced**, yaitu kelas 0 sebanyak **218,334** dan kelas 1 sebanyak **35,346**.

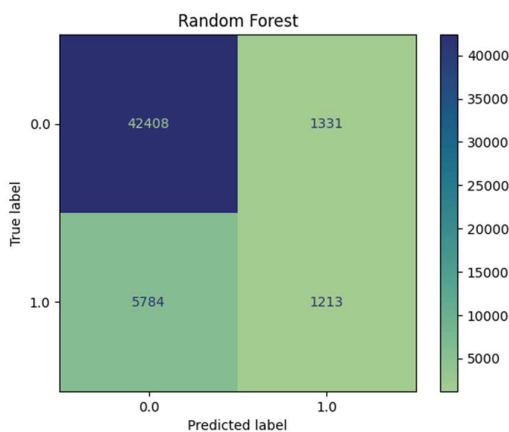
Tabel 5 Distribusi kelas sebelum dan sesudah resampling SMOTE-ENN

Kondisi Data	Kelas 0 (Non-diabetes)	Kelas 1 (Diabetes/Prediabetes)
Data asli (imbalanced)	218,334	35,346
Setelah SMOTE-ENN	145,088	185,263

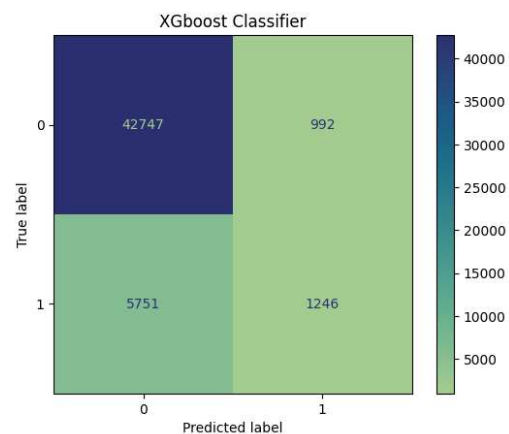
Setelah penerapan **SMOTE-ENN**, distribusi kelas menjadi jauh lebih seimbang. Perubahan ini menunjukkan bahwa SMOTE-ENN tidak hanya menambah sampel minoritas (SMOTE), tetapi juga melakukan pembersihan (ENN) sehingga sebagian sampel yang dianggap noise dapat terhapus.

### 3.2. Baseline Performance (Tanpa Resampling)

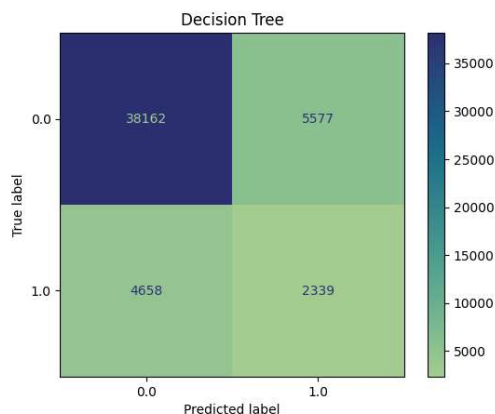
Evaluasi baseline dilakukan pada data asli (imbalanced). Hasilnya menunjukkan bahwa akurasi relatif tinggi, namun recall kelas positif rendah, terutama pada Random Forest dan XGBoost. Kondisi ini mengindikasikan banyak kasus positif (diabetes/prediabetes) yang terklasifikasi sebagai negatif.



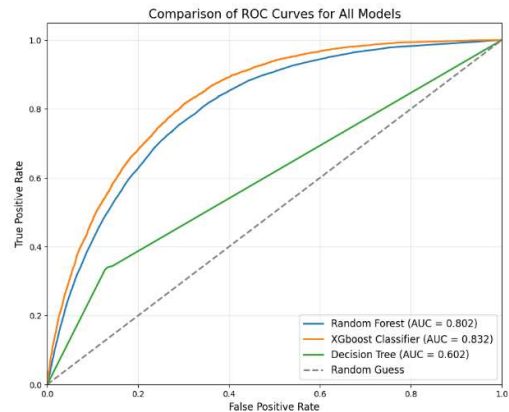
Gambar 2 CM Baseline Random Forest



Gambar 3 CM Baseline XGBoost



Gambar 4 CM Baseline Decision Tree



Gambar 5 Kurva ROC Baseline Model

Tabel 6 Baseline performance pada data imbalanced

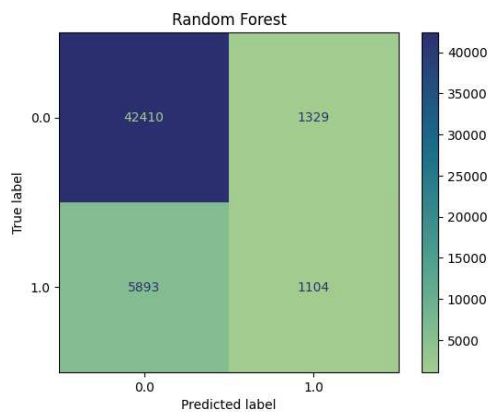
Model	Accuracy	Precision	Recall	F1-score	ROC-AUC
Random Forest	0.859764	0.476808	0.173360	0.254271	0.801842
XGBoost	0.867096	0.556747	0.178076	0.269843	0.831897
Decision Tree	0.798269	0.295478	0.334286	0.313686	0.602269

Pada baseline, XGBoost memberikan ROC-AUC tertinggi (0.831897), namun recall tetap rendah (0.178076). Random Forest juga menunjukkan pola serupa (recall 0.173360). Decision Tree memiliki recall lebih tinggi (0.334286), tetapi AUC rendah (0.602269) sehingga pemisahan kelas kurang baik.

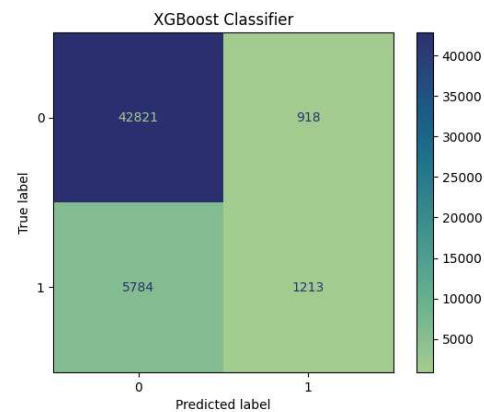
Untuk memperjelas dampak false negative pada baseline, confusion matrix Random Forest menunjukkan FN jauh lebih tinggi daripada TP. Dari confusion matrix tersebut terdapat FN = 5,784 dan TP = 1,213, artinya sekitar 82.7% kasus positif pada data uji masih terlewat pada skenario baseline (4).

### 3.3. Baseline + Hyperparameter Tuning (GridSearchCV)

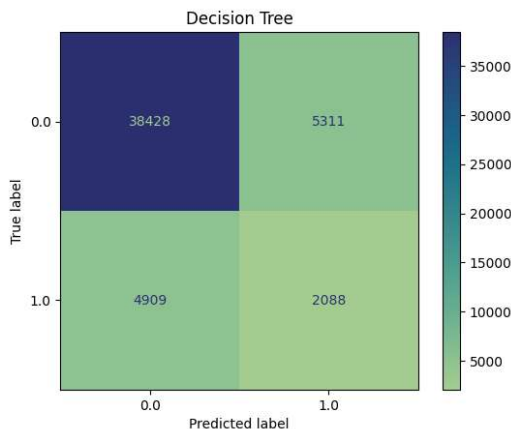
Optimasi hyperparameter dengan GridSearchCV pada data imbalanced memberikan perubahan performa yang relatif kecil dan belum menyelesaikan masalah utama, yaitu recall kelas positif yang tetap rendah.



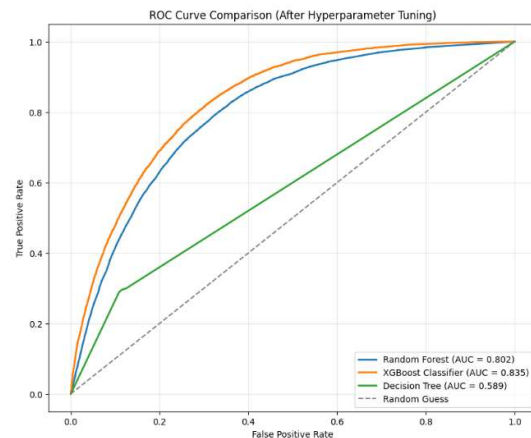
Gambar 6  
CM Baseline + Hyperparameter Tuning  
(GridSearchCV) Random Forest



Gambar 7  
CM Baseline + Hyperparameter Tuning  
(GridSearchCV) XGBoost



Gambar 8  
CM Baseline + Hyperparameter Tuning  
(GridSearchCV) Decision Tree



Gambar 9  
Kurva ROC Baseline + Hyperparameter Tuning  
(GridSearchCV) Model

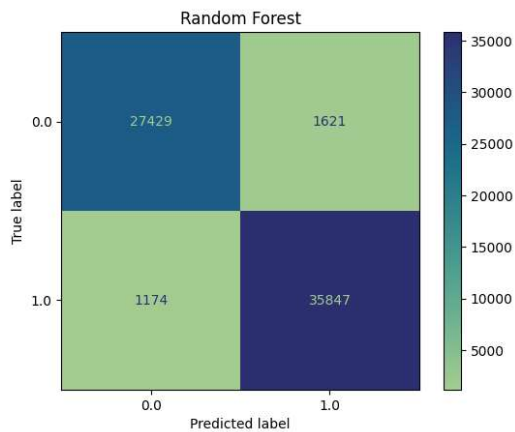
Table 1 Baseline + GridSearchCV pada data imbalanced

Model	Accuracy	Precision	Recall	F1-score	ROC-AUC
Random Forest	0.857655	0.453761	0.157782	0.234146	0.802090
XGBoost	0.867904	0.569216	0.173360	0.265776	0.834632
Decision Tree	0.798565	0.282200	0.298414	0.290081	0.589112

Dapat dilihat bahwa tuning tidak meningkatkan recall secara signifikan; bahkan pada Random Forest recall menurun menjadi 0.157782. Hal ini menegaskan bahwa kendala utama pada baseline berasal dari ketidakseimbangan kelas, bukan sekadar konfigurasi parameter.

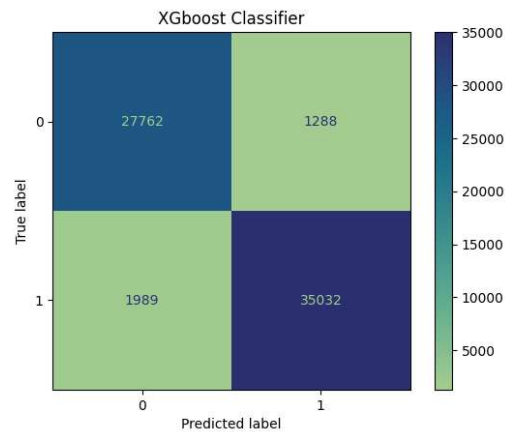
### 3.4. Performance After Hybrid Resampling SMOTE-ENN

Setelah dilakukan resampling SMOTE-ENN, seluruh model mengalami peningkatan yang sangat besar, khususnya pada recall, F1-score, dan ROC-AUC. Ini menunjukkan bahwa model menjadi jauh lebih sensitif terhadap kelas positif.



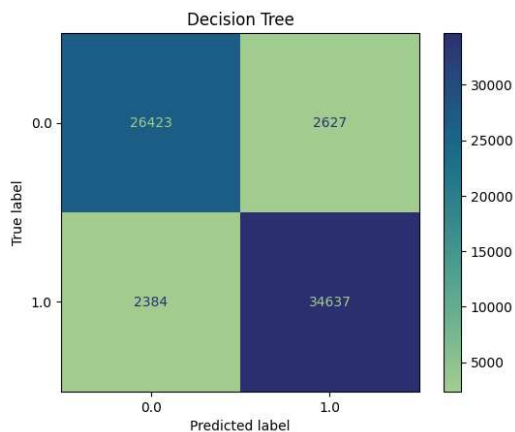
Gambar 10

CM After Hybrid Resampling SMOTE-ENN  
Random Forest



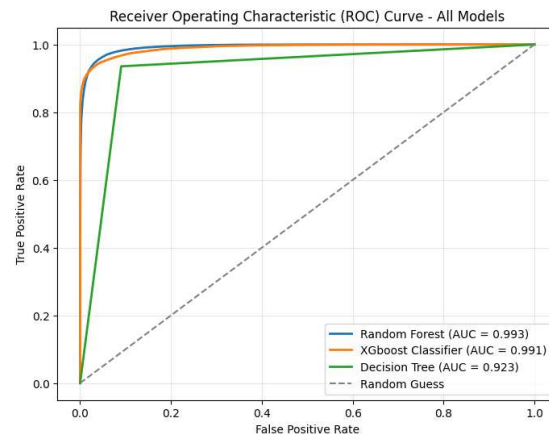
Gambar 11

CM After Hybrid Resampling SMOTE-ENN  
XGBoost



Gambar 12

CM After Hybrid Resampling SMOTE-ENN  
Decision Tree



Gambar 13

Kurva ROC After Hybrid Resampling SMOTE-ENN Model

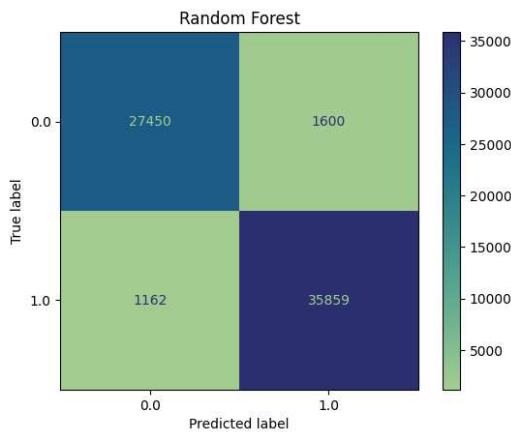
Table 2 Performa model setelah SMOTE-ENN

Model	Accuracy	Precision	Recall	F1-score	ROC-AUC
Random Forest	0.957697	0.956736	0.968288	0.962478	0.992809
XGBoost	0.950402	0.964537	0.946274	0.955318	0.990583
Decision Tree	0.924157	0.929503	0.935604	0.932544	0.922587

Random Forest memberikan performa terbaik dengan recall **0.968288** dan ROC-AUC **0.992809**. Peningkatan recall pada Random Forest dari baseline (0.173360 → 0.968288) menunjukkan kenaikan sekitar **5.6×**, yang sangat relevan untuk kebutuhan skrining dini.

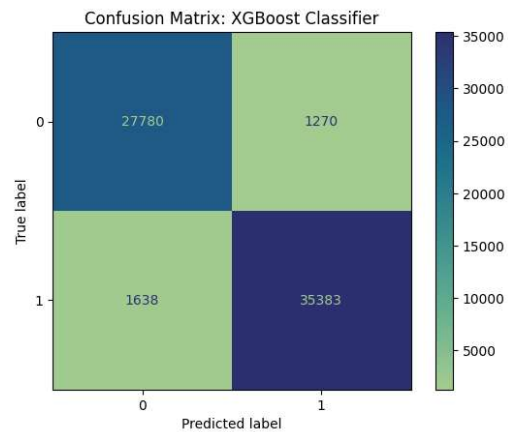
### 3.5. SMOTE-ENN + Hyperparameter Tuning (GridSearchCV)

Eksperimen tuning setelah SMOTE-ENN menunjukkan bahwa peningkatan performa tidak selalu besar. Random Forest mengalami peningkatan kecil, sedangkan XGBoost dan Decision Tree sedikit menurun.



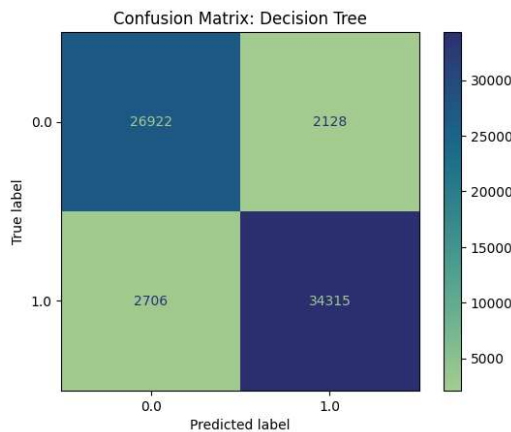
Gambar 14

CM After SMOTE-ENN + Hyperparameter Tuning (GridSearchCV) Random Forest



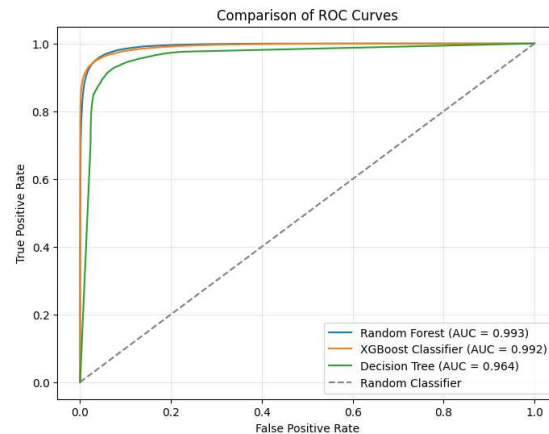
Gambar 15

CM After SMOTE-ENN + Hyperparameter Tuning (GridSearchCV) XGBoost



Gambar 16

CM After SMOTE-ENN + Hyperparameter Tuning (GridSearchCV) Decision Tree



Gambar 17

Kurva ROC After SMOTE-ENN + Hyperparameter Tuning (GridSearchCV) Model

Table 3 Performa model setelah SMOTE-ENN + GridSearchCV

Model	Accuracy	Precision	Recall	F1-score	ROC-AUC
Random Forest (tuned)	0.958196	0.957287	0.968612	0.962916	0.993097
XGBoost (tuned)	0.955987	0.965531	0.955755	0.960529	0.992462
Decision Tree (tuned)	0.926836	0.941607	0.926906	0.934199	0.963515

Dengan demikian, faktor yang paling dominan menaikkan performa adalah **resampling SMOTE-ENN**, sedangkan tuning cenderung berperan sebagai penyempurna dan tidak terlalu menaikkan semua model.

## 4. DISCUSSIONS

### 4.1. Dampak Imbalanced Data terhadap Model Baseline

Hasil baseline menunjukkan fenomena umum pada data imbalanced: akurasi terlihat baik, tetapi recall kelas minoritas rendah. Pada konteks deteksi dini diabetes, rendahnya recall berarti banyak individu berisiko (kelas positif) tidak terdeteksi sehingga berpotensi menunda intervensi. Studi Chang dkk. juga menegaskan bahwa dataset indikator kesehatan seperti BRFSS 2015 bersifat sangat bias sehingga diperlukan teknik penyeimbangan seperti SMOTE agar model tidak berat ke kelas mayoritas [5].

### 4.2. Mengapa SMOTE-ENN Meningkatkan Kinerja

Peningkatan performa setelah SMOTE-ENN dapat dijelaskan melalui dua mekanisme utama:

1. SMOTE meningkatkan representasi kelas minoritas sehingga model memperoleh lebih banyak contoh pola positif, dan
2. ENN menghapus sampel yang ambigu/noise sehingga batas keputusan antar kelas lebih “bersih”.

Temuan ini sejalan dengan Setiawan dkk. yang menyatakan bahwa resampling (khususnya SMOTE-ENN) meningkatkan performa berbagai algoritma, dan pada dataset tidak seimbang, pengaruh resampling dapat lebih signifikan daripada pemilihan algoritma. Pola tersebut juga terlihat pada hasil kamu: setelah SMOTE-ENN, gap performa antar model mengecil (akurasi selisih ~3–4% dan recall selisih ~3%) [18].

### 4.3. Komparasi terhadap Penelitian Terdahulu

Tabel 7 Komparasi penelitian terdahulu dengan penelitian ini (diabetes detection)

Studi	Dataset	Model Terbaik	Accuracy	Precision	Recall	F1-Score	ROC-AUC
Chang dkk. [5]	Subset BRFSS (70,692 respon); merujuk dataset BRFSS2015 (253,680 record, 21 fitur)	Random Forest + SMOTE	0.8226	0.8347	0.8045	0.8226	0.8226

Setiawan dkk. [18]	BRFSS Diabetes 2015	KNN + SMOTE-ENN	0.97	0.96	0.99	0.98	0.99
A. Wantoro dkk. [16]	BRFSS Diabetes 2015	XGBoost + NearMiss	0.869	0.88	0.87	0.87	0.8682
<b>Penelitian Ini</b>	<b>Diabetes Health Indicators (BRFSS2015) 253,680 record; 21 fitur</b>	<b>Random Forest + SMOTE-ENN + GridSearchCV</b>	<b>0.9582</b>	<b>0.9573</b>	<b>0.9686</b>	<b>0.9630</b>	<b>0.9931</b>

#### 4.4. Threats to Validity dan Rekomendasi Pengetesan Lanjutan

Walaupun peningkatan setelah SMOTE-ENN sangat tinggi, penelitian lanjutan sebaiknya menambahkan protokol evaluasi yang lebih ketat untuk menghindari bias, misalnya:

- Melakukan resampling hanya pada data training (misalnya melalui pipeline + cross-validation stratified), sedangkan data test dibiarkan merepresentasikan distribusi asli; dan/atau
- Menambahkan evaluasi tambahan seperti PR-AUC atau analisis cost-sensitive jika fokus utama adalah meminimalkan false negative.

Langkah ini akan membuat klaim peningkatan performa menjadi lebih kuat dan lebih aman saat direview.

## 5. CONCLUSION

Penelitian ini menunjukkan bahwa dataset indikator kesehatan untuk prediksi diabetes memiliki ketidakseimbangan kelas yang signifikan, dan kondisi ini menyebabkan model baseline cenderung bias terhadap kelas mayoritas sehingga recall untuk kasus diabetes rendah. Penerapan hybrid resampling SMOTE-ENN mampu memperbaiki masalah tersebut dan meningkatkan performa secara konsisten pada tiga model yang diuji (Decision Tree, Random Forest, dan XGBoost).

Berdasarkan evaluasi akurasi, precision, recall, F1-score, dan ROC-AUC, model terbaik adalah Random Forest + SMOTE-ENN, dengan performa tinggi dan keseimbangan metrik yang kuat untuk kebutuhan deteksi dini (meminimalkan kasus diabetes yang terlewat).

Untuk pengembangan penelitian selanjutnya, disarankan:

1. menerapkan resampling secara pipeline di dalam cross-validation untuk menghindari bias evaluasi,
2. menambahkan analisis feature importance dan/atau feature selection, serta
3. menguji generalisasi model pada dataset lain atau skenario multi-class (misal pemisahan prediabetes vs diabetes).

## REFERENCES

- [1] "Diabetes Complications | International Diabetes Federation." Accessed: Dec. 19, 2025. [Online]. Available: <https://idf.org/about-diabetes/diabetes-complications/>

- 
- [2] A. D. Association, "2. Classification and Diagnosis of Diabetes: Standards of Medical Care in Diabetes—2020," *Diabetes Care*, vol. 43, no. Supplement\_1, pp. S14–S31, Dec. 2019, doi: 10.2337/dc20-S002.
- [3] A. A. Yameny, "Diabetes Mellitus Overview 2024," *Journal of Bioscience and Applied Research*, vol. 10, no. 3, pp. 641–645, 2024, doi: 10.21608/jbaar.2024.382794.
- [4] N. P. Tigga and S. Garg, "Prediction of Type 2 Diabetes using Machine Learning Classification Methods," *Procedia Comput Sci*, vol. 167, pp. 706–716, 2020, doi: <https://doi.org/10.1016/j.procs.2020.03.336>.
- [5] V. Chang, M. A. Ganatra, K. Hall, L. Golightly, and Q. A. Xu, "An assessment of machine learning models and algorithms for early prediction and diagnosis of diabetes using health indicators," *Healthcare Analytics*, vol. 2, p. 100118, 2022, doi: <https://doi.org/10.1016/j.health.2022.100118>.
- [6] U. M. Butt, S. Letchmunan, M. Ali, F. H. Hassan, A. Baqir, and H. H. R. Sherazi, "Machine Learning Based Diabetes Classification and Prediction for Healthcare Applications," *J Healthc Eng*, vol. 2021, no. 1, p. 9930985, Jan. 2021, doi: <https://doi.org/10.1155/2021/9930985>.
- [7] S. Rezvani and X. Wang, "A broad review on class imbalance learning techniques," *Appl Soft Comput*, vol. 143, p. 110415, 2023, doi: <https://doi.org/10.1016/j.asoc.2023.110415>.
- [8] G. Husain *et al.*, "SMOTE vs. SMOTEENN: A Study on the Performance of Resampling Algorithms for Addressing Class Imbalance in Regression Models," *Algorithms*, vol. 18, no. 1, Jan. 2025, doi: 10.3390/a18010037.
- [9] L. C. M. Liaw, S. C. Tan, P. Y. Goh, and C. P. Lim, "A histogram SMOTE-based sampling algorithm with incremental learning for imbalanced data classification," *Inf Sci (N Y)*, vol. 686, p. 121193, 2025, doi: <https://doi.org/10.1016/j.ins.2024.121193>.
- [10] C. C. Olisah, L. Smith, and M. Smith, "Diabetes mellitus prediction and diagnosis from a data preprocessing and machine learning perspective," *Comput Methods Programs Biomed*, vol. 220, p. 106773, 2022, doi: <https://doi.org/10.1016/j.cmpb.2022.106773>.
- [11] M. O. Edeh *et al.*, "A Classification Algorithm-Based Hybrid Diabetes Prediction Model," *Front Public Health*, vol. Volume 10-2022, 2022, [Online]. Available: <https://www.frontiersin.org/journals/public-health/articles/10.3389/fpubh.2022.829519>
- [12] K. Kangra and J. Singh, "Comparative analysis of predictive machine learning algorithms for diabetes mellitus," *Bulletin of Electrical Engineering and Informatics*, vol. 12, pp. 1728–1737, Jun. 2023, doi: 10.11591/eei.v12i3.4412.
- [13] S. Nematzadeh, F. Kiani, M. Torkamanian-Afshar, and N. Aydin, "Tuning hyperparameters of machine learning algorithms and deep neural networks using metaheuristics: A bioinformatics study on biomedical and biological cases," *Comput Biol Chem*, vol. 97, p. 107619, 2022, doi: <https://doi.org/10.1016/j.compbiolchem.2021.107619>.
- [14] P. Charilaou and R. Battat, "Machine learning models and over-fitting considerations," *World J Gastroenterol*, vol. 28, no. 5, pp. 605–607, Feb. 2022, doi: 10.3748/wjg.v28.i5.605.
- [15] "Diabetes Health Indicators Dataset." Accessed: Dec. 19, 2025. [Online]. Available: <https://www.kaggle.com/datasets/alexteboul/diabetes-health-indicators-dataset>
- [16] A. Wantoro, A. F. Yuliana, D. Y. A. Andini, I. Awaliyani, and W. Caesarendra, "Optimizing Type 2 Diabetes Classification with Feature Selection and Class Balancing in Machine
-

- Learning,” *Jurnal Teknik Informatika (Jutif)*, vol. 6, no. 4, pp. 2625–2637, Aug. 2025, doi: 10.52436/1.jutif.2025.6.4.5166.
- [17] I. H. Al Ghozali, M. A. Fathin, and A. R. Handoko, “A Comparative Study of Machine Learning with Statistical Feature Selection for Risk Detection of Diabetic,” *Jurnal Ilmiah FIFO*, vol. 17, no. 2, p. 102, Nov. 2025, doi: 10.22441/fifo.2025.v17i2.001.
- [18] A. E. Setiawan, S. Rustad, A. Syukur, M. A. Soeleman, M. Akrom, and A. W. Setiawan, “SMOTE-ENN Resampling to Optimize Diabetes Prediction in Imbalanced Data,” *Ingenierie des Systemes d’Information*, vol. 30, no. 5, pp. 1163–1176, May 2025, doi: 10.18280/isi.300505.
- [19] I. Popchev and D. Orozova, “Algorithms for Machine Learning with Orange System,” *International Journal of Online and Biomedical Engineering (iJOE)*, vol. 19, no. 04, pp. 109–123, Apr. 2023, doi: 10.3991/ijoe.v19i04.36897.
- [20] T. Yan, S.-L. Shen, A. Zhou, and X. Chen, “Prediction of geological characteristics from shield operational parameters by integrating grid search and K-fold cross validation into stacking classification algorithm,” *Journal of Rock Mechanics and Geotechnical Engineering*, vol. 14, no. 4, pp. 1292–1303, 2022, doi: <https://doi.org/10.1016/j.jrmge.2022.03.002>.
- [21] R. D. Joshi and C. K. Dhakal, “Predicting type 2 diabetes using logistic regression and machine learning approaches,” *Int J Environ Res Public Health*, vol. 18, no. 14, Jul. 2021, doi: 10.3390/ijerph18147346.
- [22] R. Islam, A. Sultana, Md. N. Tuhin, Md. S. H. Saikat, and M. R. Islam, “Clinical Decision Support System for Diabetic Patients by Predicting Type 2 Diabetes Using Machine Learning Algorithms,” *J Healthc Eng*, vol. 2023, no. 1, p. 6992441, Jan. 2023, doi: <https://doi.org/10.1155/2023/6992441>.
- [23] Md. M. Islam *et al.*, “Identification of the risk factors of type 2 diabetes and its prediction using machine learning techniques,” *Health Systems*, vol. 12, no. 2, pp. 243–254, Apr. 2023, doi: 10.1080/20476965.2022.2141141.
- [24] O. Iparraguirre-Villanueva, K. Espinola-Linares, R. O. Flores Castañeda, and M. Cabanillas-Carbonell, “Application of Machine Learning Models for Early Detection and Accurate Classification of Type 2 Diabetes,” *Diagnostics*, vol. 13, no. 14, Jul. 2023, doi: 10.3390/diagnostics13142383.
- [25] A. Agliata, D. Giordano, F. Bardozzo, S. Bottiglieri, A. Facchiano, and R. Tagliaferri, “Machine Learning as a Support for the Diagnosis of Type 2 Diabetes,” *Int J Mol Sci*, vol. 24, no. 7, Apr. 2023, doi: 10.3390/ijms24076775.
- [26] M. Lugner, A. Rawshani, E. Helleryd, and B. Eliasson, “Identifying top ten predictors of type 2 diabetes through machine learning analysis of UK Biobank data,” *Sci Rep*, vol. 14, no. 1, p. 2102, 2024, doi: 10.1038/s41598-024-52023-5.
- [27] H. Guan *et al.*, “The role of machine learning in advancing diabetic foot: a review,” *Front Endocrinol (Lausanne)*, vol. Volume 15-2024, 2024, [Online]. Available: <https://www.frontiersin.org/journals/endocrinology/articles/10.3389/fendo.2024.1325434>