



## Perbandingan Metode *Oversampling* SMOTE dan ADASYN pada Klasifikasi Diabetes Menggunakan Algoritma CatBoost

Zaenur Rozikin<sup>1\*</sup>, Jose Julian Hidayat<sup>2</sup>

<sup>1,2</sup>Program Studi Teknik Informatika, Fakultas Teknik, Universitas Pelita Bangsa, Bekasi, Indonesia

Email author: [zaenurrozikin@pelitabangsa.ac.id](mailto:zaenurrozikin@pelitabangsa.ac.id)<sup>1\*</sup>, [josejulianhidayat@gmail.com](mailto:josejulianhidayat@gmail.com)<sup>2</sup>

### Article Info

#### Article history:

Received April 15, 2026

Revised April 20, 2026

Accepted April 30, 2026

#### Keywords:

*Diabetes Classification, CatBoost Algorithm, SMOTE Oversampling, ADASYN Oversampling, Class Imbalance.*

### ABSTRACT

Class imbalance is a major challenge in diabetes classification, as it can lead models to become biased toward the majority class. Oversampling approaches such as the Synthetic Minority Oversampling Technique (SMOTE) and Adaptive Synthetic Sampling (ADASYN) are applied to address this issue by improving the representation of the minority class. This study compares the performance of both methods using the CatBoost algorithm on a diabetes classification dataset. The evaluation is carried out using accuracy, precision, recall, F1-score, and ROC-AUC metrics. The experimental results indicate that the baseline CatBoost model already achieves strong performance, with an accuracy of 0.9720 and a ROC-AUC of 0.9796; however, the recall for the minority class remains relatively low at 0.6935. The implementation of SMOTE yields the most optimal improvement, achieving an accuracy of 0.9727, precision of 0.9737, recall of 0.6971, and an F1-score of 0.8125, while maintaining a ROC-AUC of 0.9796. Meanwhile, ADASYN also improves performance compared to the baseline, but its results are slightly lower than SMOTE, with an accuracy of 0.9719 and recall of 0.6924. Overall, SMOTE proves to be more effective in enhancing the CatBoost model's ability to detect the minority class without compromising overall performance. Therefore, SMOTE is recommended as a more stable and optimal oversampling method for handling imbalanced data in diabetes classification tasks.

### Corresponding Author:

Zaenur Rozikin  
Universitas Pelita Bangsa  
Jl. Inspeksi Kalimalang No.9, Bekasi, 17530, Indonesia  
Email: [zaenurrozikin@pelitabangsa.ac.id](mailto:zaenurrozikin@pelitabangsa.ac.id)



## 1. INTRODUCTION

Salah satu masalah jangka panjang yang menjadi perhatian utama dalam bidang kesehatan di seluruh dunia adalah diabetes mellitus. Ketidakmampuan tubuh untuk menghasilkan atau menggunakan insulin secara efisien menyebabkan gangguan metabolisme glukosa, yang merupakan tanda penyakit ini (Dini Gladis Saputri et al., 2025). Kondisi tersebut menyebabkan kadar gula darah meningkat secara abnormal dalam jangka panjang (Mujabi & Yuniartika, 2018). Diabetes dapat

menyebabkan berbagai komplikasi serius, seperti penyakit kardiovaskular, neuropati, nefropati, hingga kebutaan, jika tidak ditangani dengan benar (Puspita Febriani et al., 2023). Akibatnya, upaya pencegahan dan deteksi dini sangat penting untuk mengurangi efek negatif yang ditimbulkan oleh penyakit ini.

Seiring dengan meningkatnya jumlah penderita diabetes di seluruh dunia, sistem deteksi yang efektif dan cepat semakin diperlukan. Metode konvensional yang bergantung pada pemeriksaan klinis biasanya membutuhkan banyak waktu dan biaya (Puspita Febriani et al., 2023). Selain itu, keterbatasan tenaga medis juga menjadi kendala dalam proses diagnosis yang optimal (Supiyan, 2025). Dalam konteks ini, pemanfaatan teknologi berbasis data menjadi solusi yang menjanjikan untuk membantu proses identifikasi risiko diabetes secara lebih efisien dan terukur (Tena, 2023).

Dengan kemajuan teknologi informasi, khususnya dalam bidang pengajaran mesin, ada banyak peluang baru untuk analisis data medis (Muhammad Idris, 2025). Memanfaatkan pembelajaran mesin, sistem dapat membuat prediksi dan mempelajari pola dari data historis yang akurat tanpa perlu diprogram secara eksplisit (Santoso et al., 2024). Dalam kasus klasifikasi diabetes, algoritma *machine learning* dapat digunakan untuk mengidentifikasi individu yang berisiko tinggi berdasarkan berbagai atribut kesehatan yaitu umur, keidealalan berat tubuh, serta riwayat medis lainnya (Hidayat et al., 2026).

Berbagai algoritma *machine learning* telah diterapkan dalam klasifikasi diabetes, seperti *Decision Tree*, *Random Forest*, *Support Vector Machine (SVM)*, hingga *Artificial Neural Network (ANN)* (Putri Pasaribu et al., 2025). Masing-masing algoritma memiliki keunggulan dan keterbatasan tersendiri dalam menangani data medis yang kompleks. Namun, dalam beberapa penelitian terbaru, algoritma berbasis *boosting* memperlihatkan kemampuan yang lebih ahli dalam hal ketepatan serta stabilitas model (Syaban & Mardawati, 2025).

Salah satu algoritma *boosting* yang semakin populer adalah *CatBoost*, yang dikembangkan untuk mengatasi kelemahan algoritma *boosting* tradisional. *CatBoost* memiliki kemampuan dalam menangani data kategorikal secara langsung tanpa memerlukan proses *encoding* yang kompleks (Putra et al., 2025). Selain itu, algoritma ini juga dirancang untuk mengurangi *overfitting* dan meningkatkan generalisasi model. Hal ini menjadikan *CatBoost* sebagai pilihan yang tepat untuk klasifikasi data medis yang memiliki karakteristik beragam (Zhou et al., 2021).

Meskipun algoritma yang digunakan sudah cukup canggih, tantangan utama dalam klasifikasi diabetes terletak pada ketidakseimbangan data (*class imbalance*) (Ernawati & Maulana, 2025). Dalam banyak dataset medis, jumlah data pasien tidak terjangkit diabetes jauh lebih besar dibandingkan pasien terdiagnosa diabetes (Sidiq et al., 2025). Kenyataan ini menyebabkan model bias terhadap kelas yang mendominasi, yang mengakibatkan performa yang kurang optimal untuk mengidentifikasi kelas minor (Samdro, 2026).

Dampak dari ketidakseimbangan data ini dapat dilihat pada rendahnya nilai *recall* untuk kelas minoritas (Johnson & Khoshgoftaar, 2019). Artinya, banyak kasus diabetes yang sebenarnya ada tidak berhasil terdeteksi oleh model. Hal ini tentu sangat berbahaya dalam konteks kesehatan, karena kesalahan dalam mendeteksi pasien diabetes dapat berakibat fatal. Maka dibutuhkan rencana terorganisir agar dapat menangani permasalahan yang disebabkan hal tersebut.

Sebagian dari banyaknya cara mengatasi data tidak seimbang adalah teknik *resampling*, khususnya *oversampling* (Putra et al., 2025). Dengan menghasilkan data sintesis, *oversampling* dilakukan dengan tujuan meningkatkan jumlah data yang tersedia untuk kelas minoritas (A. Rahim et al., 2023). Cara penyebaran data yang lebih merata, model diharapkan mampu belajar motif dari kedua kategori secara lebih adil serta menghasilkan prediksi yang lebih akurat (Rahman Wahid et al., 2023).

Dua metode *oversampling* yang banyak digunakan adalah *Synthetic Minority Oversampling Technique (SMOTE)* dan *Adaptive Synthetic Sampling (ADASYN)*. *SMOTE* bekerja dengan cara membuat data sintesis berdasarkan interpolasi antara data minoritas yang berdekatan (Samudra et al., 2025). Pendekatan ini membantu memperluas ruang fitur kelas minoritas secara merata sehingga model dapat mengenali pola dengan lebih baik (Putranto et al., 2025).

Sementara itu, *ADASYN* merupakan pengembangan dari *SMOTE* yang bersifat lebih adaptif (Putra Sadewa & Kurniawan, 2026). Metode ini tidak hanya menghasilkan data sintesis, tetapi juga mempertimbangkan tingkat kesulitan dalam mengklasifikasikan data (Alwaliyanto et al., 2025). Dengan

demikian, lebih banyak data sintetis akan dihasilkan pada area yang sulit dipelajari oleh model, sehingga diharapkan dapat meningkatkan performa klasifikasi secara signifikan (Masruriyah et al., 2024).

Meskipun *SMOTE* dan *ADASYN* memiliki tujuan yang sama, yaitu mengatasi ketidakseimbangan data, keduanya memiliki pendekatan yang berbeda. *SMOTE* cenderung menghasilkan distribusi data yang lebih merata, sedangkan *ADASYN* lebih fokus pada area yang kompleks (Nurhayati & Rahardi, 2025). Perbedaan ini dapat memengaruhi kinerja model tergantung pada karakteristik dataset yang digunakan.

Dalam konteks penggunaan algoritma *CatBoost*, penerapan metode *oversampling* menjadi menarik untuk dikaji lebih lanjut. *CatBoost* yang memiliki kemampuan dalam menangani data kompleks diharapkan dapat bekerja lebih optimal ketika dikombinasikan dengan teknik penyeimbangan data yang tepat (Saputra et al., 2025). Oleh karena itu, perlu dilakukan analisis perbandingan antara *SMOTE* dan *ADASYN* dalam meningkatkan performa *CatBoost*.

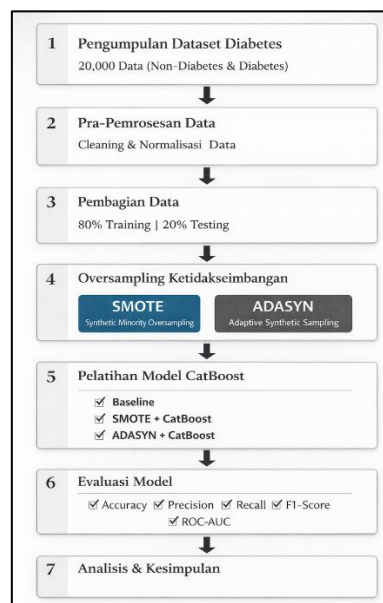
Penelitian ini tidak hanya menguji performa model setelah penerapan *oversampling*, tetapi juga membandingkannya dengan model *baseline* tanpa perlakuan khusus terhadap ketidakseimbangan data. Hal ini bertujuan untuk memberikan gambaran yang jelas mengenai dampak dari masing-masing metode terhadap peningkatan kinerja model.

Evaluasi kinerja dilakukan menggunakan berbagai metrik, seperti *accuracy*, *precision*, *recall*, *F1-score*, dan *ROC-AUC* (Hidayat et al., 2025). Penggunaan metrik yang beragam ini penting untuk memastikan bahwa model tidak hanya unggul secara keseluruhan, tetapi juga mampu mendeteksi kelas minoritas dengan baik (Anshor & Zy, 2024). Dengan demikian, hasil evaluasi dapat memberikan insight yang lebih komprehensif.

Novelty dalam penelitian ini terletak pada analisis komparatif antara metode *SMOTE* dan *ADASYN* yang diintegrasikan secara langsung dengan algoritma *CatBoost* pada dataset diabetes berskala besar, dengan fokus utama pada peningkatan kemampuan deteksi kelas minoritas. Selain itu, penelitian ini menekankan evaluasi yang komprehensif menggunakan berbagai metrik performa untuk mengidentifikasi metode *oversampling* yang paling efektif dan stabil. Pendekatan ini diharapkan dapat memberikan kontribusi baru dalam pengembangan model klasifikasi diabetes yang lebih akurat dan *robust*, khususnya dalam menghadapi permasalahan ketidakseimbangan data.

## 2. METHOD

Metode eksperimen dan pendekatan kuantitatif digunakan dalam penelitian ini untuk membandingkan kinerja metode *oversampling* *SMOTE* dan *ADASYN* pada algoritma *CatBoost* untuk klasifikasi diabetes. Proses pengumpulan data hingga evaluasi model dirancang secara sistematis untuk memungkinkan analisis menyeluruh dan tidak bias dari temuan penelitian.



Gambar 1. Alur dan Desain Penelitian

### 2.1. Dataset dan Sumber Data

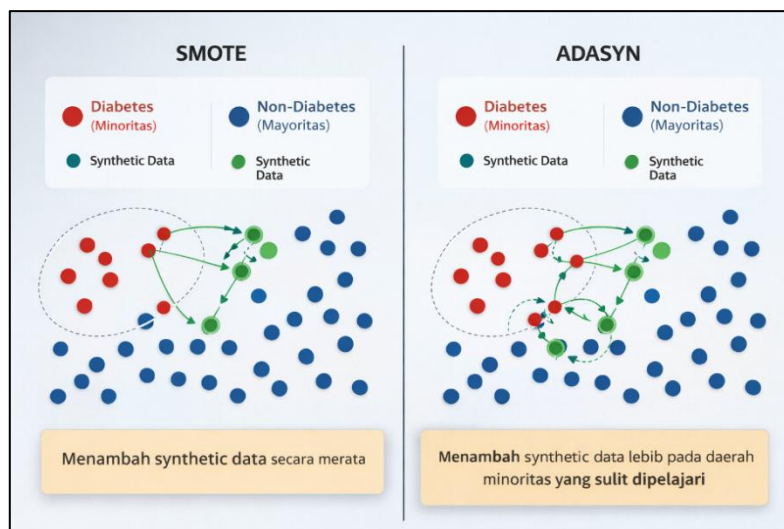
Dataset yang digunakan dalam penelitian ini merupakan dataset klasifikasi diabetes dari sumber data terbuka yaitu *Kaggle* yang dapat diakses dilaman berikut <https://www.kaggle.com/datasets/iammustafatz/diabetes-prediction-dataset>, yang terdiri dari 20.000 data dengan dua kelas, yaitu non-diabetes (kelas mayoritas) dan diabetes (kelas minoritas). Dataset ini memiliki ketidakseimbangan distribusi kelas, di mana jumlah data non-diabetes jauh lebih besar dibandingkan data diabetes. Setiap data terdiri dari beberapa fitur yang merepresentasikan kondisi kesehatan individu, seperti usia, indeks massa tubuh (BMI), tekanan darah, kadar glukosa, serta atribut lainnya yang relevan dalam prediksi diabetes.

### 2.2. Tahap Pra-pemrosesan Data

Sebelum dilakukan proses pemodelan, dataset terlebih dahulu melalui tahap pra-pemrosesan. Tahapan ini meliputi penanganan missing values menggunakan metode imputasi, normalisasi atau standarisasi data numerik, serta *encoding* pada fitur kategorikal jika diperlukan (Nugroho, Danny, et al., 2025). Selain itu, dilakukan juga pembagian dataset menjadi data latih (*training set*) dan data uji (*testing set*) dengan proporsi tertentu, misalnya 80:20, untuk memastikan evaluasi model yang adil (Kamalia et al., 2026).

### 2.3. Penanganan Ketidakseimbangan Data

Terdapat dua metode oversampling yang digunakan dalam penelitian ini: Teknik Oversampling Minority Synthetic (SMOTE) dan Adaptive Synthetic Oversampling (ADASYN). Metode-metode ini digunakan untuk mengatasi masalah ketidakseimbangan kelas pada dataset. Untuk menghindari kehilangan data, proses oversampling hanya digunakan pada data latihan (training data). Ini dilakukan untuk menghindari bias dalam hasil evaluasi model (Dn & Rahardi, 2026). Metode ini membuat distribusi kelas dalam data latih lebih seimbang, yang memungkinkan model untuk mempelajari lebih baik karakteristik kelas minor.



Gambar 2. Cara Kerja SMOTE vs ADASYN

*SMOTE* bekerja dengan menghasilkan data sintesis berdasarkan kedekatan antar sampel pada kelas minoritas melalui teknik interpolasi, sehingga distribusi data menjadi lebih merata (Nugroho, Wiyanto, et al., 2025). Sementara itu, *ADASYN* merupakan pengembangan dari *SMOTE* yang menghasilkan data sintesis secara adaptif dengan memberikan fokus lebih pada sampel yang sulit dipelajari oleh model (Nurhopipah & Magnolia, 2023). Kombinasi kedua metode ini digunakan untuk menganalisis efektivitas masing-masing dalam meningkatkan performa model klasifikasi, khususnya dalam meningkatkan kemampuan deteksi terhadap kelas minoritas.

## 2.4 Pembangunan Model

Model yang digunakan dalam penelitian ini adalah *CatBoost*, yaitu algoritma berbasis *gradient boosting* yang dirancang untuk menangani data kategorikal secara efisien serta mampu mengurangi *overfitting* melalui teknik *ordered boosting* (Febriansyah Istianto et al., 2024). *CatBoost* dipilih karena kemampuannya dalam menghasilkan performa yang stabil pada data dengan karakteristik kompleks, seperti dataset medis (Suryaputri & Rahardi, 2026). Dalam penelitian ini, model dibangun menggunakan data latih yang telah melalui tahap pra-pemrosesan, sehingga siap digunakan dalam proses pembelajaran mesin.

Untuk memastikan perbandingan yang adil, model *CatBoost* dikonfigurasi dengan parameter yang sama untuk setiap skenario selama proses pelatihan. Untuk menjaga konsistensi hasil, digunakan parameter berikut: *iterations* = 1000, *learning\_rate* = 0.05, kedalaman = 6, *loss\_function* = "Logloss", *eval\_metric* = "Accuracy", dan *random\_state* = 42. Selain itu, *l2\_leaf\_reg* = 3 digunakan sebagai parameter regularisasi untuk meningkatkan generalisasi model dan mengurangi *overfitting*. Selain itu, *border\_count* = 254 digunakan untuk menghitung jumlah pembagian (*binning*) pada fitur numerik, yang menghasilkan representasi data yang lebih baik.

Jika data validasi tidak menunjukkan peningkatan dalam kinerja, mekanisme *early\_stopping\_rounds* = 50 digunakan untuk menghentikan pelatihan lebih awal selama proses pelatihan. Dengan menetapkan *task\_type* = 'GPU', implementasi model memanfaatkan akselerasi komputasi, yang menghasilkan proses pelatihan yang lebih cepat dan efektif, terutama untuk dataset berukuran besar.

Tiga skenario pemodelan diterapkan untuk menganalisis pengaruh metode *oversampling* terhadap kinerja model, yaitu *Baseline CatBoost* tanpa perlakuan *oversampling*, SMOTE dengan *CatBoost*, dan *ADASYN* dengan *CatBoost*. Setiap model dilatih menggunakan konfigurasi yang sama untuk memastikan perbandingan yang adil. Hasil dari ketiga skenario tersebut kemudian dievaluasi untuk mengetahui metode mana yang paling efektif dalam meningkatkan kemampuan model, khususnya dalam mendeteksi kelas minoritas pada kasus klasifikasi diabetes.

## 2.6 Evaluasi dan Visualisasi

Evaluasi kinerja model dalam penelitian ini dilakukan secara komprehensif menggunakan beberapa metrik utama, yaitu *accuracy*, *precision*, *recall*, *F1-score*, dan *ROC-AUC* (Hidayat & Hasanudin, 2026). *Accuracy* digunakan untuk mengukur tingkat ketepatan prediksi secara keseluruhan, sementara *precision* menunjukkan ketepatan model dalam mengklasifikasikan kelas positif (Amin et al., n.d.). *Recall* menjadi metrik penting karena mengukur kemampuan model dalam mendeteksi kelas minoritas, yang dalam konteks ini adalah pasien diabetes (Nurriqfi Fakhri Fikrillah et al., 2023). *ROC-AUC* menunjukkan kemampuan model untuk membedakan kelas positif dan negatif secara keseluruhan, sedangkan *F1-score* mengukur keseimbangan antara *precision* dan *recall* (Rifai et al., 2024). Selain itu, matriks confusion juga digunakan untuk memberikan gambaran rinci tentang distribusi prediksi yang benar dan salah pada masing-masing kelas (Reynaldi Valerian et al., 2025).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

Selain evaluasi numerik, penelitian ini juga dilengkapi dengan visualisasi untuk mendukung interpretasi hasil model (Setyowati et al., n.d.). Visualisasi yang digunakan meliputi kurva *ROC (Receiver Operating Characteristic)* untuk melihat performa klasifikasi pada berbagai *threshold*, serta kurva *Precision-Recall* untuk mengevaluasi kinerja model pada data tidak seimbang (Werdana, n.d.). Selain itu, grafik *confusion matrix* ditampilkan untuk mempermudah analisis kesalahan klasifikasi, serta perbandingan metrik antar model disajikan dalam bentuk grafik batang (*bar chart*). Dengan adanya visualisasi ini, hasil evaluasi menjadi lebih mudah dipahami dan memberikan *insight* yang lebih mendalam terhadap performa masing-masing model.

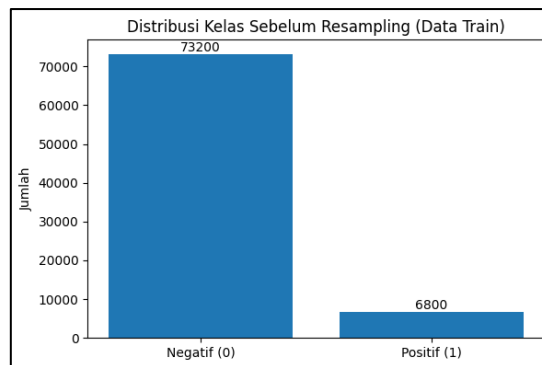
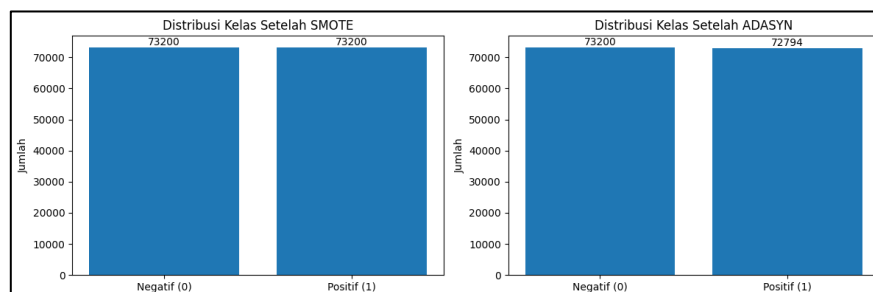
### 3. HASIL DAN PEMBAHASAN

Dataset yang digunakan dalam penelitian ini terdiri dari 20.000 data dengan dua kelas, yaitu non-diabetes sebagai kelas mayoritas dan diabetes sebagai kelas minoritas. Data ini memiliki ketidakseimbangan distribusi kelas yang cukup signifikan, di mana jumlah data non-diabetes jauh lebih banyak dibandingkan data diabetes. Setiap data memuat berbagai atribut kesehatan yang digunakan sebagai fitur dalam proses klasifikasi.

Tabel 1. Dataset Penelitian

Age	Gender	BMI	HbA1c_level	...	Diabetes
80	Female	25.19	6.6	...	0
54	Male	27.32	6.6	...	0
28	Female	27.32	5.7	...	0
36	Male	23.45	5.0	...	0
76	Female	20.14	4.8	...	0
44	Male	19.31	6.5	...	1
67	Female	32.45	7.2	...	1
29	Male	24.89	5.4	...	0
58	Female	30.11	6.8	...	1
....	...	...	...	...	...
62	Male	28.76	6.9	...	1

Kondisi ketidakseimbangan ini menjadi tantangan utama karena dapat menyebabkan model cenderung bias terhadap kelas mayoritas, sehingga diperlukan teknik penanganan khusus seperti *oversampling* untuk meningkatkan representasi kelas minoritas. Setelah *oversampling*, Gambar 4 menunjukkan distribusi kelas menjadi lebih seimbang.

Gambar 3. Dataset Sebelum *Resampling*Gambar 4. Dataset Setelah *Resampling*

Hasil pengujian menunjukkan bahwa algoritma *CatBoost* memiliki performa yang sangat baik dalam klasifikasi diabetes, bahkan tanpa penerapan teknik penyeimbangan data. Model baseline *CatBoost* memperoleh nilai *accuracy* sebesar 0,9720 dan *ROC-AUC* sebesar 0,9796, yang menunjukkan kemampuan model dalam membedakan kelas secara keseluruhan sudah sangat tinggi. Selain itu, nilai *precision* yang mencapai 0,9680 mengindikasikan bahwa prediksi terhadap kelas positif cukup akurat. Namun demikian, nilai *recall* pada kelas minoritas masih berada pada angka 0,6935, yang menunjukkan bahwa masih terdapat sejumlah kasus diabetes yang tidak terdeteksi oleh model.

Tabel 2. *Classification Report* dari *Baseline CatBoost*

Kelas	Precision	Recall	F1-Score	Support
0	0.97	1.00	0.98	18300
1	0.97	0.69	0.81	1700
Accuracy	-	-	0.97	20000
Macro Avg	0.97	0.85	0.90	20000
Weighted Avg	0.97	0.97	0.97	20000

Penerapan metode *SMOTE* memberikan peningkatan performa yang paling signifikan dibandingkan metode lainnya. Model *SMOTE* dan *CatBoost* menghasilkan *accuracy* sebesar 0,9727, *precision* sebesar 0,9737, *recall* sebesar 0,6971, dan *F1-score* sebesar 0,8125, dengan *ROC-AUC* tetap tinggi di angka 0,9796. Peningkatan recall ini menunjukkan bahwa model menjadi lebih sensitif dalam mendeteksi kasus diabetes setelah dilakukan penyeimbangan data menggunakan *SMOTE*. Selain itu, peningkatan *F1-score* menunjukkan adanya keseimbangan yang lebih baik antara *precision* dan *recall*, sehingga model menjadi lebih optimal dalam klasifikasi.

Tabel 3. *Classification Report CatBoost with SMOTE*

Kelas	Precision	Recall	F1-Score	Support
0	0.97	1.00	0.99	18300
1	0.97	0.70	0.81	1700
Accuracy			0.97	20000
Macro Avg	0.97	0.85	0.90	20000
Weighted Avg	0.97	0.97	0.97	20000

Sementara itu, penerapan *ADASYN* juga menunjukkan adanya peningkatan performa dibandingkan model baseline, meskipun tidak seoptimal *SMOTE*. Model *ADASYN* dan *CatBoost* memperoleh *accuracy* sebesar 0,9719, *precision* sebesar 0,9687, *recall* sebesar 0,6924, dan *F1-score* sebesar 0,8075, dengan *ROC-AUC* sebesar 0,9792. Meskipun *ADASYN* dirancang untuk fokus pada data yang sulit dipelajari, hasil yang diperoleh menunjukkan bahwa pendekatan ini belum mampu memberikan peningkatan yang signifikan dalam konteks dataset yang digunakan.

Tabel 4. *Classification Report CatBoost with ADASYN*

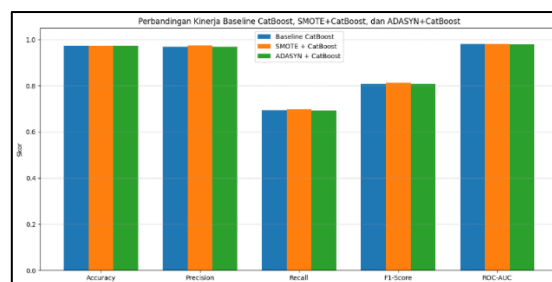
Kelas	Precision	Recall	F1-Score	Support
0	0.97	1.00	0.98	18300
1	0.97	0.69	0.81	1700
Accuracy			0.97	20000
Macro Avg	0.97	0.85	0.90	20000
Weighted Avg	0.97	0.97	0.97	20000

Jika dibandingkan secara keseluruhan, *SMOTE* terbukti menjadi metode *oversampling* yang paling efektif dalam penelitian ini. Hal ini terlihat dari nilai *accuracy*, *precision*, *recall*, dan *F1-score* yang

lebih tinggi dibandingkan metode lainnya, serta kestabilan nilai *ROC-AUC* yang tetap tinggi. *SMOTE* mampu meningkatkan representasi kelas minoritas secara merata tanpa menimbulkan noise yang berlebihan, sehingga model dapat belajar dengan lebih optimal.

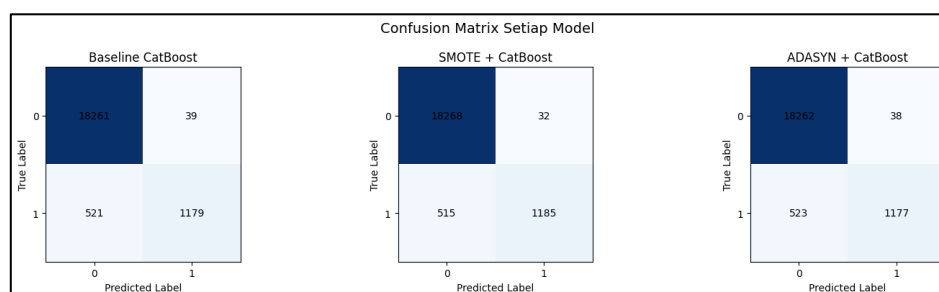
Peningkatan *recall* dari 0.69 menjadi 0.70 pada skenario *SMOTE* menunjukkan bahwa sensitivitas model terhadap kelas minoritas telah meningkat. Kesalahan tipe II (kesalahan negatif palsu) memiliki konsekuensi yang jauh lebih besar dibandingkan dengan kesalahan tipe I (kesalahan positif palsu) dalam kerangka pembelajaran yang sensitif terhadap biaya secara statistik, meskipun peningkatan tersebut tampaknya kecil. Metode *oversampling* seperti *SMOTE* secara teoritis bertujuan untuk meningkatkan representasi ruang karakteristik kelas minoritas dengan membuat sampel sintesis yang didasarkan pada interpolasi antar tetangga terdekat. Ini membantu model belajar tentang batas keputusan yang lebih representatif dan tidak bias terhadap kelas mayoritas. Oleh karena itu, peningkatan *recall* yang dihasilkan tidak sekadar fluktuasi acak; itu menunjukkan perbaikan dalam generalisasi model terhadap distribusi data minoritas.

Selain itu, ada stabilitas nilai presisi yang tetap tinggi (0.97) pada seluruh skenario, yang menunjukkan bahwa tidak ada penurunan dalam performa dalam hal nilai prediktor positif meskipun ada peningkatan jumlah ingatan. Ini menunjukkan bahwa model tetap dapat mempertahankan keseimbangan antara sensitivitas dan presisi, yang merupakan karakteristik penting dalam sistem yang diklasifikasikan sebagai berisiko tinggi. Meskipun dalam skala kecil, peningkatan *recall* memiliki dampak yang signifikan dari perspektif praktis dan klinis. Dalam sistem deteksi penyakit, setiap peningkatan kemampuan untuk menemukan kasus positif menurunkan jumlah kasus yang tidak terdeteksi. Oleh karena itu, meskipun peningkatan dalam kinerja secara numerik terlihat kecil, secara konseptual dan aplikatif masih penting dan relevan.



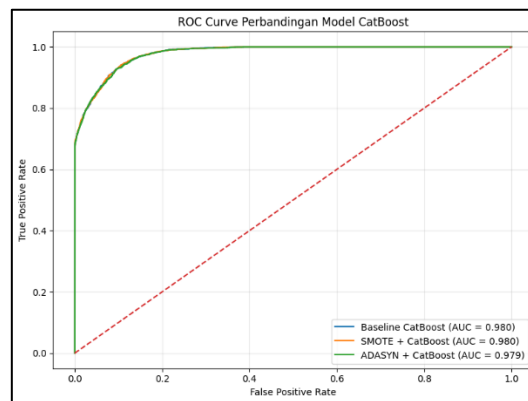
Gambar 5. Perbandingan Kinerja Model

Analisis *confusion matrix* juga menunjukkan bahwa penerapan *SMOTE* berhasil mengurangi jumlah kesalahan klasifikasi pada kelas minoritas dibandingkan dengan *baseline* dan *ADASYN*. Hal ini mengindikasikan bahwa model menjadi lebih baik dalam mengidentifikasi pasien diabetes yang sebelumnya sulit terdeteksi. Sebaliknya, pada metode *ADASYN*, peningkatan jumlah data sintesis yang berfokus pada area tertentu cenderung tidak memberikan dampak signifikan terhadap peningkatan performa model.



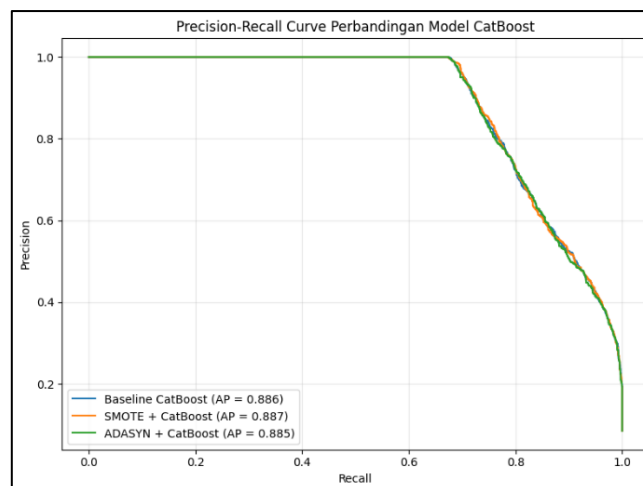
Gambar 6. Perbandingan Confusion Matrix pada Tiap Model

Grafik *ROC* menunjukkan bahwa ketiga model, yaitu *Baseline CatBoost*, *SMOTE* dikombinasikan *CatBoost*, dan *ADASYN* dikombinasikan *CatBoost*, memiliki performa yang sangat baik dengan nilai *AUC* yang hampir sama, yaitu sekitar 0,98. Kurva dari ketiga model berada jauh di atas garis diagonal, yang menandakan kemampuan klasifikasi yang sangat tinggi dalam membedakan kelas diabetes dan non-diabetes. Meskipun terdapat sedikit perbedaan, model dengan *SMOTE* dan *baseline* menunjukkan performa yang sedikit lebih baik dibandingkan *ADASYN*, namun secara keseluruhan ketiganya memiliki kinerja yang sangat optimal dan stabil.



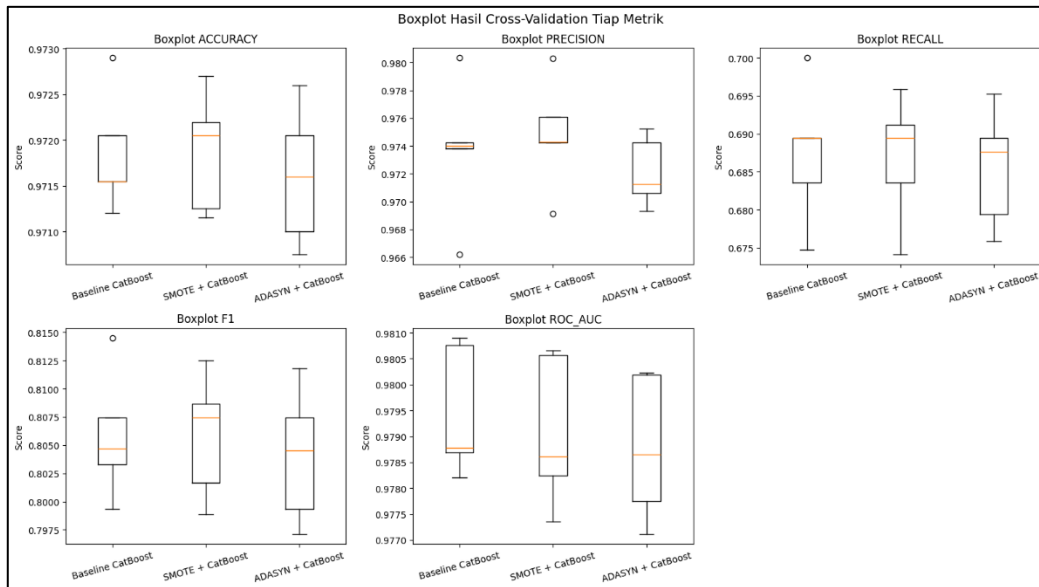
Gambar 7. Perbandingan *ROC Curve* tiap model

Grafik *Precision-Recall* menunjukkan bahwa ketiga model memiliki performa yang sangat mirip dengan nilai *Average Precision (AP)* yang tinggi, yaitu sekitar 0,88. Kurva ketiga model hampir saling tumpang tindih, yang menandakan bahwa kemampuan model dalam menjaga keseimbangan antara *precision* dan *recall* sudah sangat baik. Model *SMOTE* dipadukan *CatBoost* sedikit unggul dibandingkan lainnya, diikuti oleh *baseline* dan *ADASYN*, namun perbedaannya sangat kecil sehingga secara keseluruhan ketiga model menunjukkan performa yang stabil dalam menangani data yang tidak seimbang.



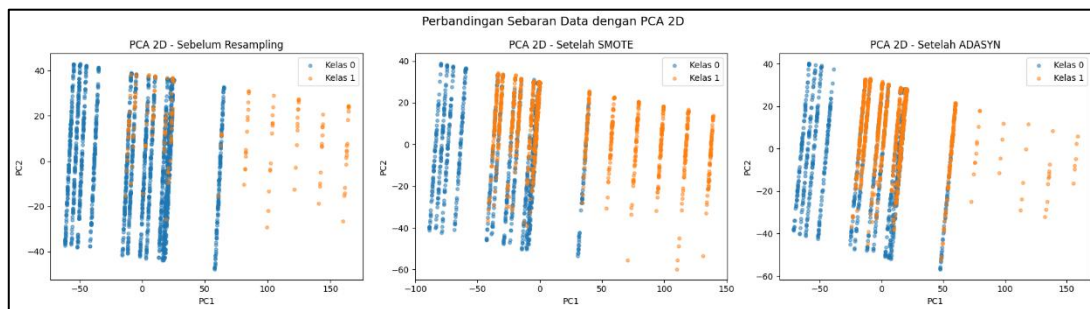
Gambar 8. *Precision-Recall Curve* Perbandingan Model *CatBoost*

*Boxplot* hasil *cross-validation* menunjukkan bahwa ketiga model memiliki performa yang stabil pada seluruh metrik evaluasi. Model *SMOTE* dikolaborasikan *CatBoost* cenderung memiliki median tertinggi pada hampir semua metrik, terutama pada *recall* dan *F1-score*, yang menunjukkan peningkatan kemampuan dalam mendeteksi kelas minoritas. Sementara itu, *baseline CatBoost* juga menunjukkan performa yang konsisten dengan variasi yang relatif kecil, sedangkan *ADASYN* dikolaborasikan *CatBoost* memiliki sebaran nilai yang sedikit lebih luas pada beberapa metrik. Secara keseluruhan, *SMOTE* memberikan hasil yang paling optimal dan stabil dibandingkan metode lainnya.



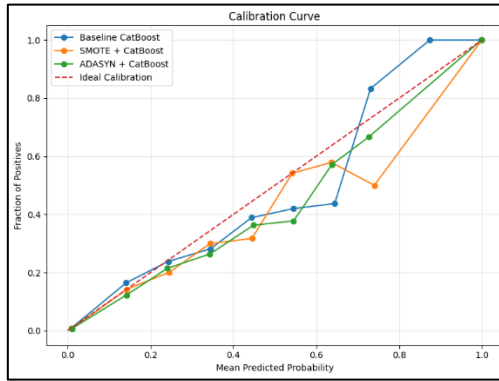
Gambar 9. Boxplot

Visualisasi *PCA 2D* menunjukkan perbedaan distribusi data sebelum dan setelah penerapan metode *oversampling*. Sebelum resampling, terlihat bahwa data kelas minoritas (kelas 1) jauh lebih sedikit dan tersebar tidak merata dibandingkan kelas mayoritas (kelas 0). Setelah penerapan *SMOTE*, jumlah data kelas minoritas meningkat secara signifikan dan distribusinya menjadi lebih merata di ruang fitur, sehingga pola antar kelas menjadi lebih seimbang. Sementara itu, pada *ADASYN*, data sintesis juga bertambah namun lebih terkonsentrasi pada area tertentu yang sulit dipelajari, sehingga distribusinya terlihat lebih terfokus dibandingkan *SMOTE*. Secara keseluruhan, *SMOTE* memberikan penyebaran data yang lebih merata, sedangkan *ADASYN* lebih adaptif terhadap area kompleks.



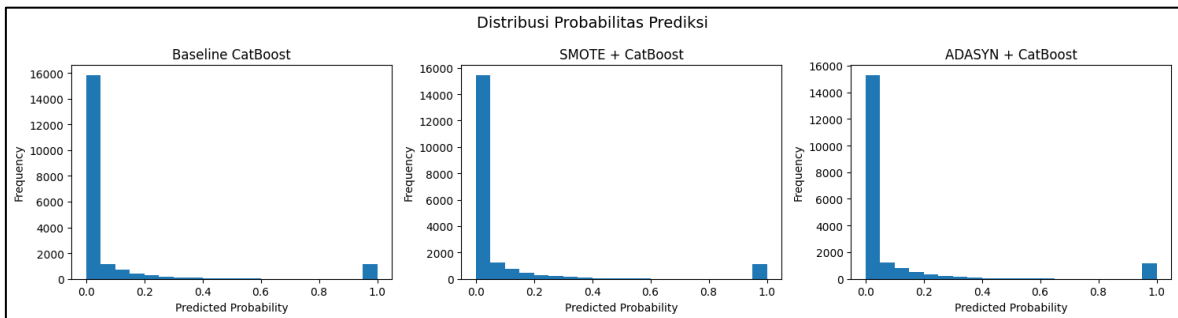
Gambar 10. Persebaran Data Tiap Model

Grafik *calibration curve* menunjukkan bahwa ketiga model memiliki kemampuan kalibrasi yang cukup baik, ditunjukkan dengan kurva yang relatif mendekati garis ideal. Model *baseline CatBoost* terlihat paling mendekati garis diagonal pada beberapa interval probabilitas, yang menandakan prediksi probabilitasnya cukup akurat. Sementara itu, *SMOTE* dan *ADASYN* juga menunjukkan performa yang baik, meskipun terdapat sedikit deviasi pada beberapa titik, terutama pada rentang probabilitas menengah. Secara keseluruhan, ketiga model memiliki kualitas probabilitas prediksi yang cukup reliabel, dengan baseline sedikit lebih unggul dalam aspek kalibrasi.



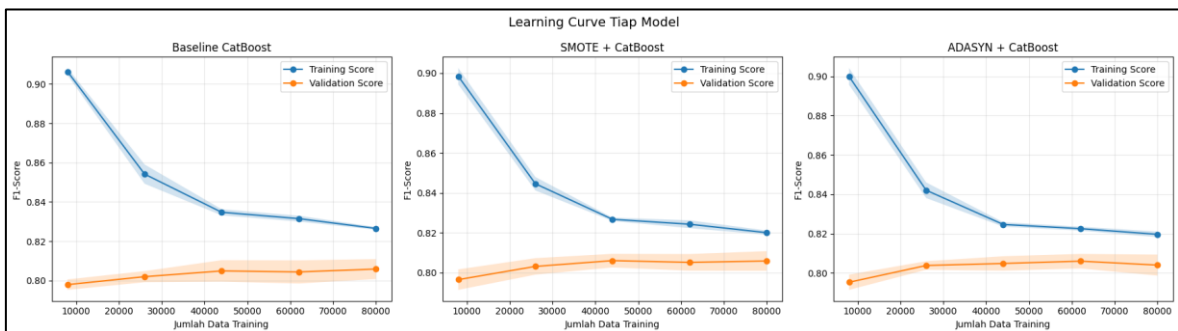
Gambar 11. Calibration Curve

Distribusi probabilitas prediksi menunjukkan bahwa ketiga model cenderung menghasilkan probabilitas yang terkonsentrasi pada nilai rendah (mendekati 0), yang mencerminkan dominasi kelas mayoritas dalam dataset. Namun, terlihat juga adanya distribusi pada nilai tinggi (mendekati 1) yang merepresentasikan prediksi untuk kelas minoritas. Model SMOTE + CatBoost dan ADASYN + CatBoost menunjukkan penyebaran probabilitas yang sedikit lebih luas dibandingkan baseline, menandakan peningkatan sensitivitas terhadap kelas minoritas. Secara keseluruhan, ketiga model mampu memisahkan kelas dengan cukup baik, dengan model berbasis oversampling memberikan distribusi prediksi yang lebih seimbang.



Gambar 12. Distribusi Probabilitas Prediksi

Grafik learning curve menunjukkan bahwa ketiga model mengalami peningkatan performa pada data validasi seiring bertambahnya jumlah data training, sementara skor training cenderung menurun dan stabil. Hal ini menunjukkan bahwa model tidak mengalami overfitting yang signifikan dan memiliki kemampuan generalisasi yang baik. Model SMOTE + CatBoost dan ADASYN + CatBoost menunjukkan kurva yang lebih stabil dengan gap yang lebih kecil antara training dan validation dibandingkan baseline, menandakan performa yang lebih seimbang. Secara keseluruhan, penambahan data training memberikan dampak positif terhadap performa model, dengan metode oversampling membantu meningkatkan stabilitas pembelajaran.



Gambar 13. Perbandingan Learning Curve

#### 4. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian, dapat disimpulkan bahwa algoritma CatBoost mampu memberikan performa klasifikasi yang sangat baik pada kasus diabetes, namun masih memiliki keterbatasan dalam mendeteksi kelas minoritas. Penerapan metode oversampling, khususnya SMOTE, terbukti paling efektif dalam meningkatkan recall dan F1-score tanpa menurunkan performa keseluruhan model, sehingga menghasilkan model yang lebih seimbang dan optimal dibandingkan baseline maupun ADASYN. Oleh karena itu, SMOTE direkomendasikan sebagai metode penanganan ketidakseimbangan data pada kasus serupa. Sebagai saran, penelitian selanjutnya dapat mengeksplorasi kombinasi metode lain seperti hybrid sampling, tuning hyperparameter yang lebih mendalam, atau penggunaan algoritma lain untuk memperoleh performa yang lebih baik serta meningkatkan kemampuan deteksi kelas minoritas secara lebih optimal.

#### REFERENCES

- A. Rahim, A. M., Ingrid Yanuar Risca Pratiwi, & Muhammad Ainul Fikri. (2023). Klasifikasi Penyakit Jantung Menggunakan Metode Synthetic Minority Over-Sampling Technique Dan Random Forest Classifier. *Indonesian Journal of Computer Science*, 12(5). <https://doi.org/10.33022/ijcs.v12i5.3413>
- Alwaliyanto, Siska Kurnia Gusti, Iis Afrianty, & Fadhilah Syafria. (2025). Penerapan Metode ADASYN Dalam Mengatasi Imbalanced Data Untuk Klasifikasi Penyakit Stroke Menggunakan Support Vector Machine. *Bulletin of Computer Science Research*, 5(4), 532–541. <https://doi.org/10.47065/bulletincsr.v5i4.612>
- Amin, M. D. I., Hidayat, J. J., Setyowati, C., Fitri, E. K., Anggraini, A. N., & Werdana, A. P. (n.d.). Implementasi Model LSTM Untuk Peramalan Curah Hujan Di Bekasi Dengan Pemanfaatan Data Cuaca BMKG. *Jurnal Teknologi Informasi Digital*, 1(2), 90–99. Retrieved <https://jurnal.ipdig.id/index.php/jtid/article/view/200>
- Anshor, A. H., & Zy, A. T. (2024). Implementasi Metode Decision Tree pada Sistem Prediksi Status Kualitas Produk Minuman A. *Jurnal Ilmiah Informatika Global*, 15(1), 17–22. <https://doi.org/10.36982/jiig.v15i1.3778>
- Dini Gladis Saputri, Chika Amelia Putri, Cyntia Ramandani, Depi Sugesti, Ayuni Dwi Andini, Carmia Alysia Dina Haqiqi, & Liss Dyah Dwi Arini. (2025). Diabetes Melitus sebagai Gangguan Endokrin: Tinjauan Patofisiologi dan Pendekatan Diagnosis. *Jejak Digital: Jurnal Ilmiah Multidisiplin*, 1(6 SE-Articles), 4382–4387. <https://doi.org/10.63822/ggpkeg19>
- Dn, U. K. J., & Rahardi, M. (2026). Analysis of SMOTE and Random Search on Machine Learning Algorithms for Stroke Disease Diagnosis. *Journal of Applied Informatics and Computing*, 10(1), 847–855. <https://doi.org/10.30871/jaic.v10i1.12046>
- Ernawati, S., & Maulana, I. (2025). Meningkatkan Klasifikasi Penyakit Diabetes Menggunakan Metode Ensemble Softvoting Dengan SMOTE-ENN dan Optimasi Bayesian. *Evolusi: Jurnal Sains Dan Manajemen*, 13(1), 71–86. <https://doi.org/10.31294/evolusi.v13i1.8267>
- Febriansyah Istianto, A., Id Hadiana, A., & Rakhmat Umbara, F. (2024). PREDIKSI CURAH HUJAN MENGGUNAKAN METODE CATEGORICAL BOOSTING (CATBOOST). *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(4), 2930–2937. <https://doi.org/10.36040/jati.v7i4.7304>
- Hidayat, J. J., Amin, M. D. I., Fitri, E. K., Anggraini, A. N., Werdana, A. P., Setyowati, C., & Sasongko, A. T. (2026). Prediksi Diabetes Menggunakan Deep Neural Network dengan Penyesuaian Hiperparameter Berbasis Bayesian Optimization. *Journal of Practical Computer Science*, 5(2), 130–143. <https://doi.org/10.37366/jpcs.v5i2.6419>
- Hidayat, J. J., & Hasanudin, S. (2026). Prediksi Volatilitas IHSG Dengan Hybrid Model GARCH–Random Forest Berbasis Machine Learning. *Jurnal Manajemen Informatika & Teknologi*, 6(1 SE-Articles), 130–140. <https://doi.org/10.51903/mifortekh.v6i1.1134>
- Hidayat, J. J., Setyowati, C., & Werdana, A. P. (2025). Perancangan Sistem Prediksi Penyakit pada Tanaman Padi Berbasis Image Processing Menggunakan Algoritma VGG-16 Transfer Learning dan K-Means Segmentation. *Journal of Practical Computer Science*, 5(1), 1–15. <https://doi.org/10.37366/jpcs.v5i1.5759>

- Johnson, J. M., & Khoshgoftaar, T. M. (2019). Survey on deep learning with class imbalance. *Journal of Big Data*, 6(1), 27. <https://doi.org/10.1186/s40537-019-0192-5>
- Kamalia, A. Z., Choiriyatun Nisa Latansa, & Zaenur Rozikin. (2026). Klasifikasi Kondisi Pasar Harga Emas ANTAM Indonesia Menggunakan Algoritma Decision Tree. *Jurnal Komputer Teknologi Informasi Sistem Informasi (JUKTISI)*, 4(3), 2087–2098. <https://doi.org/10.62712/juktisi.v4i3.800>
- Masruriyah, A., Novita, H., Sukmawati, C., Ramadhan, A., Arif, S., & Dermawan, B. (2024). Pengukuran Kinerja Model Klasifikasi dengan Data Oversampling pada Algoritma Supervised Learning untuk Penyakit Jantung. *Computer Science (CO-SCIENCE)*, 4(1), 62–70. <https://doi.org/10.31294/coscience.v4i1.2389>
- Muhammad Idris. (2025). Pemanfaatan Machine Learning untuk Optimasi Big Data dalam Sistem Informatika Modern. *Journal of Information Systems and Technology*, 1(1), 33–39. <https://doi.org/10.64845/jistech.v1i1.41>
- Mujabi, M. F., & Yuniartika, W. (2018). Hubungan Kadar Gula Darah Dengan Tingkat Depresi Dan Aktifitas Fisik Pada Penderita Diabetes Mellitus. *Jurnal Berita Ilmu Keperawatan*, 11(2), 73–83.
- Nugroho, A., Danny, M., & Nawangsih, I. (2025). Ensemble Learning for Robust Anomaly Detection in Banking Fraud. *2025 8th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, 649–654. <https://doi.org/10.1109/ISRITI68345.2025.11393116>
- Nugroho, A., Wiyanto, & Maulana, D. (2025). COMPARATIVE ANALYSIS OF CLASSIFICATION ALGORITHMS IN HANDLING IMBALANCED DATA WITH SMOTE OVERSAMPLING APPROACH. *JITK (Jurnal Ilmu Pengetahuan Dan Teknologi Komputer)*, 11(2), 487–495. <https://doi.org/10.33480/jitk.v11i2.6956>
- Nurhayati, L. D., & Rahardi, M. (2025). Impact of SMOTE and ADASYN on Class Imbalance in Metabolic Syndrome Classification Using Random Forest Algorithm. *Journal of Applied Informatics and Computing*, 9(5), 2807–2813. <https://doi.org/10.30871/jaic.v9i5.10657>
- Nurhopipah, A., & Magnolia, C. (2023). Perbandingan metode Resampling pada Imbalanced Dataset untuk klasifikasi komentar program MBKM. *Jurnal Publikasi Ilmu Komputer Dan Multimedia*, 2(1), 9–22.
- Nurrifqi Fakhri Fikrillah, H., Hudawiguna, S., & Juliane, C. (2023). Klasifikasi Penerima Bansos Menggunakan Algoritma Naive Bayes. *Jurnal Teknik Informatika Dan Sistem Informasi*, 10(1), 683–695. <http://jurnal.mdp.ac.id>
- Puspita Febriani, R., Agung Prabowo, N., Puspa Putri, D., Setyawan, S., & Benedictus. (2023). PANDUAN DIABETES BAGI PASIEN DAN KELUARGA. Penerbit Tahta Media, (SE-Katalog Buku). <https://tahtamedia.co.id/index.php/issj/article/view/433>
- Putra, M. R. P., Juwariyah, S., Ridwan, M., & Marco, R. (2025). Optimasi Prediksi Kelayakan Pinjaman dengan Teknik Resampling dan Algoritma Boosting. *Komputika : Jurnal Sistem Komputer*, 14(2). <https://doi.org/10.34010/komputika.v14i2.15485>
- Putra Sadewa, F., & Kurniawan, D. (2026). Application of ADASYN and Optuna in the XGBoost Algorithm for Stunting Detection. *Journal of Applied Informatics and Computing*, 10(1), 1006–1014. <https://doi.org/10.30871/jaic.v10i1.12035>
- Putranto, A. F., -, A. S., -, D. B. P., -, R. E. D. M., & -, R. F. A. (2025). Optimasi Ulasan Palsu Menggunakan ADASYN Dan SMOTE. *Journal of Informatics and Interactive Technology*, 2(3), 413–418. <https://doi.org/10.63547/jiite.v2i3.81>
- Putri Pasaribu, J., Indra, Z., Iskandar Al Idrus, S., Nasution, H., & Yandra Niska, D. (2025). STUDI KOMPARATIF: EVALUASI PERFORMA ALGORITMA ARTIFICIAL NEURAL NETWORK DENGAN ALGORITMA MACHINE LEARNING DALAM KLASIFIKASI PENYAKIT DIABETES. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 9(5), 8245–8253. <https://doi.org/10.36040/jati.v9i5.15071>
- Rahman Wahid, M. A., Nugroho, A., & Halim Anshor, A. (2023). Prediksi Penyakit Kanker Paru-Paru Dengan Algoritma Regresi Linier. *Bulletin of Information Technology (BIT)*, 4(1), 63–74. <https://doi.org/10.47065/bit.v4i1.501>
- Reynaldi Valerian, F., Syarief, M., & Abdul Fatah, D. (2025). KLASIFIKASI TINGKAT OBESITAS MENGGUNAKAN METODE GBM DAN CONFUSION MATRIX. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 9(2), 2242–2249. <https://doi.org/10.36040/jati.v9i2.13062>
- Rifai, A. M., Raharjo, S., Utami, E., & Ariatmanto, D. (2024). Analysis for diagnosis of pneumonia symptoms using chest X-ray based on MobileNetV2 models with image enhancement using white balance and contrast limited adaptive histogram equalization (CLAHE). *Biomedical Signal Processing and Control*, 90, 105857. <https://doi.org/10.1016/j.bspc.2023.105857>
- Samodro, M. (2026). Analisis Pengaruh Ketidakseimbangan Data terhadap Kinerja Model Klasifikasi Penyakit Jantung. *Journal of Software Engineering and Information System (SEIS)*, 6(1 SE-Articles), 56–62. <https://ejournal.umri.ac.id/index.php/SEIS/article/view/11050>

- Samudra, G., Anggraeni, A. W., Jamroni, A. R. B., Sarif, A., & Wiyanto, W. (2025). Efektivitas Teknik SMOTE Dalam Meningkatkan Performa Naïve Bayes Deteksi Gangguan Kecemasan Mahasiswa. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, 12(3). <https://doi.org/10.35957/jatisi.v12i3.12197>
- Santoso, J. T., Manongga, D., Setyawan, I., Purnomo, H. D., & Hendry. (2024). Exploring Data Analytics in Attendance Systems: Unveiling Machine Learning Techniques, Patterns, Practices, and Emerging Trends. *Scientific Journal of Informatics*, 11(2), 325–340. <https://doi.org/10.15294/sji.v11i2.3438>
- Saputra, G. E., Swari, M. H. P., & Nurlaili, A. L. (2025). Implementasi Algoritma XGBoost, CatBoost, dan LGBM untuk Klasifikasi Pencemaran Udara. *JIIP - Jurnal Ilmiah Ilmu Pendidikan*, 8(12), 14135–14139. <https://doi.org/10.54371/jiip.v8i12.10102>
- Setyowati, C., Anggraini, A. N., & Fitri, E. K. (n.d.). Analisis Sentimen dan Karakteristik Linguistik Komentar Publik terhadap Kebijakan Militer Menggunakan Model RoBERTa. *Jurnal Teknologi Informasi Digital*, 2(1), 39–46. Retrieved <https://jurnal.ipdig.id/index.php/jtid/article/view/219>
- Sidiq, S., Alfian, A., & Mabrur, N. S. (2025). Pengembangan Model Prediksi Risiko Diabetes Menggunakan Pendekatan AdaBoost dan Teknik Oversampling SMOTE. *Jurnal Ilmiah Informatika Dan Ilmu Komputer (JIMA-ILKOM)*, 4(1 SE-Articles), 13–23. <https://doi.org/10.58602/jima-ilkom.v4i1.41>
- Supiyana, D. (2025). Pengembangan Sistem Pakar Untuk Diagnosa Penyakit Diabetes Melitus Menggunakan Metode Forward Chaining. *Bit-Tech*, 7(3), 918–927. <https://doi.org/10.32877/bt.v7i3.2244>
- Suryaputri, C. O., & Rahardi, M. (2026). Analysis of Gradient Boosted Trees Algorithm in Breast Cancer Classification. *Journal of Applied Informatics and Computing*, 10(1), 605–618. <https://doi.org/10.30871/jaic.v10i1.11875>
- Syaban, K., & Mardiaty. (2025). Evaluasi Model Ensemble Learning pada Identifikasi Faktor Risiko Diabetes Mellitus. *Jurnal Teknologi Dan Informasi*, 15(2), 121–130. <https://doi.org/10.34010/jati.v15i2.16238>
- Tena, A. (2023). *Penggunaan teknologi berbasis e-health sebagai upaya dalam mengontrol glikemik pasien diabetes mellitus: A Scoping Review*. Universitas Hasanuddin.
- Werdana, A. P. (n.d.). Pemodelan Klasifikasi Efisiensi Kalori Berbasis Data Aktivitas dan Kondisi Fisiologis Menggunakan Random Forest dan SMOTE. *Jurnal Teknologi Informasi Digital*, 2(1), 54–62. Retrieved <https://jurnal.ipdig.id/index.php/jtid/article/view/222>
- Zhou, F., Pan, H., Gao, Z., Huang, X., Qian, G., Zhu, Y., & Xiao, F. (2021). Fire prediction based on catboost algorithm. *Mathematical Problems in Engineering*, 2021(1), 1929137.