

**PERBANDINGAN ALGORITMA LGBM, XGBOOST, DAN
CATBOOST DENGAN BALANCING DAN HYPERPARAMETER
TUNING UNTUK PREDIKSI RISIKO HIPERTENSI PADA
DATASET KLINIS**

JALUR SCIENTIST

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Sistem Informasi



disusun oleh

DEWI AYU MURTININGSIH

22.12.2696

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

**PERBANDINGAN ALGORITMA LGBM, XGBOOST, DAN
CATBOOST DENGAN BALANCING DAN HYPERPARAMETER
TUNING UNTUK PREDIKSI RISIKO HIPERTENSI PADA
DATASET KLINIS**

JALUR NON REGULER - SCIENTIST

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Sistem Informasi



disusun oleh

DEWI AYU MURTININGSIH

22.12.269

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

HALAMAN PERSETUJUAN

JALUR NON REGULER - SCIENTIST

**PERBANDINGAN ALGORITMA LGBM, XGBOOST, DAN CATBOOST
DENGAN BALANCING DAN HYPERPARAMETER TUNING UNTUK
PREDIKSI RISIKO HIPERTENSI PADA DATASET KLINIS**

yang disusun dan diajukan oleh

Dewi Ayu Murtiningsih

22.12.2696

telah disetujui oleh Dosen Pembimbing
pada tanggal 2 Januari 2026

Dosen Pembimbing,



Bety Wulan Sari, S.Kom., M.Kom
NIK. 190302254

HALAMAN PENGESAHAN

JALUR SCIENTIST

**PERBANDINGAN ALGORITMA LGBM, XGBOOST, DAN CATBOOST
DENGAN BALANCING DAN HYPERPARAMETER TUNING UNTUK
PREDIKSI RISIKO HIPERTENSI PADA DATASET KLINIS**

yang disusun dan diajukan oleh

Dewi Ayu Murtiningsih

22.12.2696

Telah dipertahankan di depan Dewan Penguji
pada tanggal 20 Januari 2026

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Ika Nur Fajri, S.Kom., M.Kom.
NIK. 190302268

Acihmah Sidauruk, S.Kom., M.Kom.
NIK. 190302238

Bety Wulan Sari, S.Kom., M.Kom.
NIK. 190302254



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 20 Januari 2026

DEKAN FAKULTAS ILMU KOMPUTER



Prof. Dr. Kusriani, S.Kom., M.Kom.
NIK. 190302106

HALAMAN PERNYATAAN KONTRIBUSI KARYA

Yang bertandatangan di bawah ini,

Nama mahasiswa : Dewi Ayu Murtiningsih
NIM : 22.12.2696

Menyatakan bahwa karya dengan judul berikut:

Perbandingan Algoritma LGBM, XGBoost, Dan Catboost Dengan Balancing Dan Hyperparameter Tuning Untuk Prediksi Risiko Hipertensi Pada Dataset Klinis

Dosen Pembimbing : Bety Wulan Sari, S.Kom., M.Kom

1. Karya adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya merupakan gagasan, rumusan maupun penelitian yang orisinal dan SAYA memiliki KONTRIBUSI terhadap karya tersebut.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka atau Referensi pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 20 Januari 2026

Yang Menyatakan,



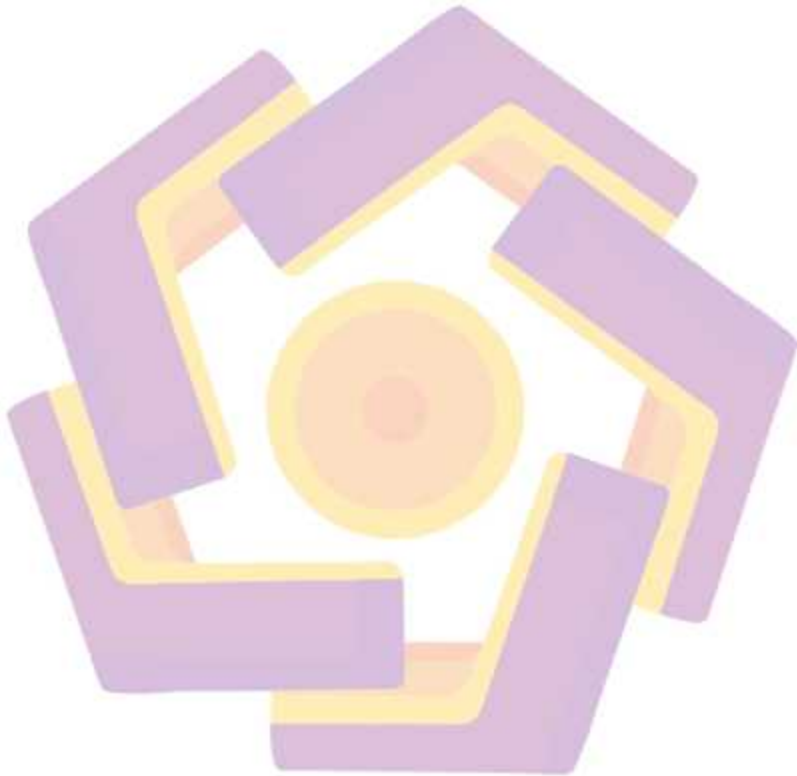
Dewi Ayu Murtiningsih

HALAMAN PERSEMBAHAN

Alhamdulillahirabbil'alamin, puji syukur kehadiran Allah SWT atas rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi ini dengan lancar. Sebagai ungkapan terima kasih penulis untuk karya ilmiah ini, penulis persembahkan rasa syukur dan penghargaan setinggi-tingginya kepada :

1. Kedua orang tua tercinta, terutama ibu, yang senantiasa memberikan doa, kasih sayang, perhatian, serta dukungan yang tiada henti dalam setiap langkah kehidupan penulis. Segala perjuangan, pengorbanan, dan ketulusan yang telah diberikan menjadi sumber kekuatan, semangat, dan motivasi utama bagi penulis dalam menempuh pendidikan hingga tahap penyusunan skripsi ini.
2. Kakak yang selalu memberikan dukungan, perhatian, serta semangat kepada penulis. Kehadiran dan perhatian menjadi penguat tersendiri bagi penulis dalam menjalani proses perkuliahan dan menyelesaikan skripsi ini.
3. Seluruh keluarga terutama kakek dan nenek, yang senantiasa memberikan doa, nasihat, dan dukungan penuh kepada penulis. Doa dan kasih sayang yang diberikan menjadi penyemangat dan penguat bagi penulis dalam menghadapi setiap proses dan tantangan selama penyusunan skripsi ini.
4. Seluruh teman saya Tina, Pinky, Fatma, Yunita, Armel, Asti dan teman-teman yang pernah hadir dalam perjalanan penulis dan tidak bisa penulis sebutkan satu persatu. Semua dukungan, bantuan dan perhatian yang diberikan selama masa perkuliahan dan proses penyusunan skripsi ini menjadi salah satu sumber motivasi untuk tetap berjuang.
5. Blackpink yang telah menjadi salah satu kebahagiaan yang mengiringi proses penulis dan menjadi inspirasi untuk terus berjuang dan bekerja keras. Kerja keras dan pencapaian mereka yang membuat penulis terus termotivasi untuk terus melangkah kedepan.
6. Ibu Beti selaku dosen pembimbing, yang telah meluangkan waktu dan tenaga untuk memberikan arahan serta masukan yang sangat berharga selama proses penyusunan skripsi ini.

7. Terakhir dan yang paling terkepada penulis yaitu saya sendiri, yang telah berjuang, bertahan, dan tidak menyerah dalam menghadapi berbagai rintangan dan tantangan selama ini hingga tahap akhir penyusunan skripsi.



KATA PENGANTAR

Puji syukur penulis panjatkan kehadiran Tuhan Yang Maha Esa, yang telah melimpahkan rahmat, hidayah, serta karunia-Nya, sehingga penulis dapat menyelesaikan laporan tugas akhir ini dengan lancar. Tugas akhir ini disusun untuk memenuhi salah satu syarat kelulusan Program Sarjana pada Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.

Selain itu penulis dengan segala kerendahan hati ingin menyampaikan rasa terima kasih kepada semua pihak yang telah berjasa memberikan dukungan dan bantuan untuk menyelesaikan laporan tugas akhir ini. Untuk itu penulis mengucapkan terima kasih kepada :

1. *Bapak Prof. Dr. M. Suyanto, M.M.* selaku Rektor Universitas Amikom Yogyakarta
2. *Ibu Prof. Dr. Kusri, S.Kom., M.Kom.* selaku Dekan Program Fakultas Ilmu Komputer
3. *Bapak Anggit Dwi Hartanto, M.Kom* selaku Ketua Program Studi Sistem Informasi
4. *Ibu Bety Wulan Sari, S.Kom., M.Kom* selaku dosen pembimbing yang memberikan arahan, saran, dan motivasi kepada penulis
5. Seluruh dosen Universitas Amikom Yogyakarta, yang telah memberikan ilmu, pengetahuan, dan dukungan yang sangat berarti selama penulis menempuh pendidikan di Universitas Amikom Yogyakarta.
6. Kepada kedua orang tua, keluarga besar, dan teman-teman tercinta yang memberikan semangat dan doa kepada penulis.

Yogyakarta, 20 Januari 2026

Penulis

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KONTRIBUSI.....	iii
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vii
DAFTAR ISI.....	viii
INTISARI	xii
ABSTRACT	xiii
BAB I IDENTITAS PUBLIKASI	1
BAB II PROSES SUBMIT	2
2.1 Lembar Review	2
2.2.1 Review Round 1	4
2.2 Lembar Persetujuan (LoA)	9
BAB III ISI KARYA ILMIAH.....	10
3.1 Intisari	10
3.2 Pendahuluan.....	11
3.3 Metode.....	14
3.4 Hasil dan Pembahasan	21
3.5 Kesimpulan.....	33
DAFTAR PUSTAKA	34
LAMPIRAN.....	37

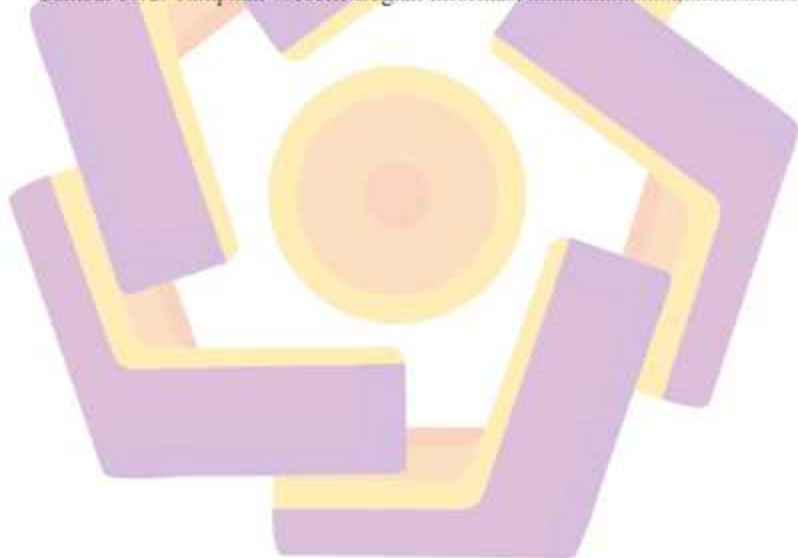
DAFTAR TABEL

Tabel 1.1. Identitas Publikasi.....	1
Tabel 2.1. Lembar Review.....	2
Tabel 2.2. Jawaban Review.....	4
Tabel 3.1. Informasi Dataset.....	15
Tabel 3.2. Confusion Matrix.....	20
Tabel 3.3. Hasil Seleksi Fitur.....	22
Tabel 3.4. Perbandingan Distribusi Data.....	24
Tabel 3.5. Perbandingan Akurasi SMOTE.....	25
Tabel 3.6. Perbandingan Akurasi ADASYN.....	25
Tabel 3.7. Parameter Grid Search.....	26
Tabel 3.8. Evaluasi Model Setelah Tuning.....	27
Tabel 3.9. Classification Report.....	29



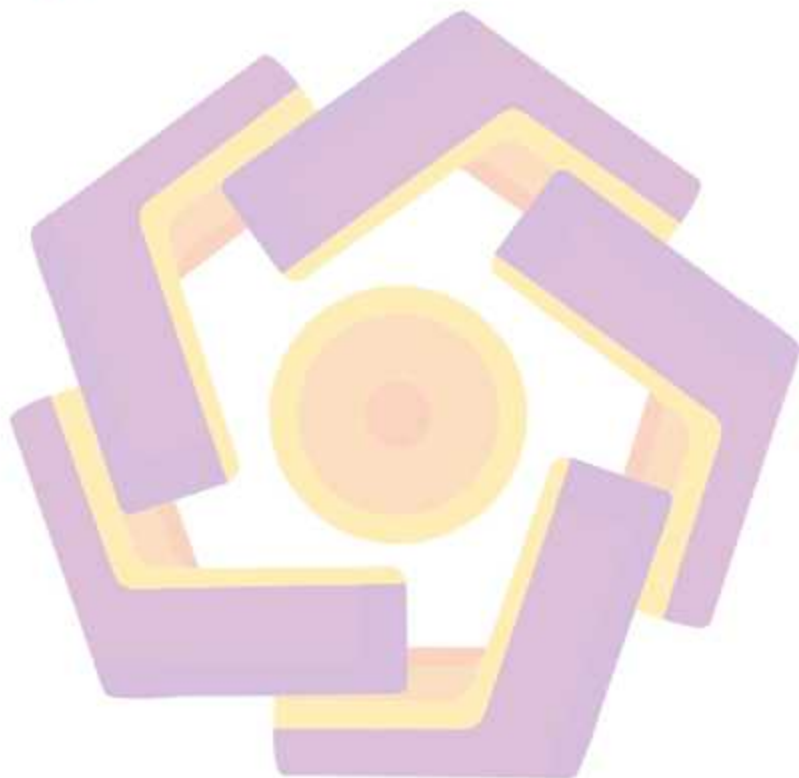
DAFTAR GAMBAR

Gambar 2.1. Bukti Email	3
Gambar 2.2. Lembar Persetujuan.....	9
Gambar 3.1. Alur Penelitian	14
Gambar 3.2. Bar Chart Distribusi Kelas Target.....	15
Gambar 3.3. Korelasi Fitur dengan Kelas Target	16
Gambar 3.4. Hasil Voting Fitur	23
Gambar 3.5. Confusion Matrix Algoritma XGBoost.....	27
Gambar 3.6. Confusion Matrix Algoritma CatBoost.....	28
Gambar 3.7. Confusion Matrix Algoritma LGBM	28
Gambar 3.8. Curva ROC-AUC Algoritma CatBoost.....	30
Gambar 3.9. SHAP Summary Plot	31
Gambar 3.10. Tampilan Website Bagian Formulir.....	32
Gambar 3.11. Tampilan Website Bagian Visualisasi	32
Gambar 3.12. Tampilan Website Bagian Informasi	32



DAFTAR LAMPIRAN

Lampiran 1. Proses EDA Kelas Target.....	39
Lampiran 2. Split Data.....	39
Lampiran 3. Feature Selection	38
Lampiran 4. Hasil Voting Feature Selection.....	39
Lampiran 5. Naskah Publikasi	41



INTISARI

Hipertensi adalah kondisi jangka panjang yang sangat umum dan secara signifikan berkontribusi terhadap masalah kardiovaskular, sehingga identifikasi dini menjadi tindakan pencegahan yang penting. Penelitian ini mengevaluasi efektivitas tiga algoritma boosting, eXtreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LGBM), dan CatBoost dalam memprediksi risiko hipertensi. Dataset yang tersedia secara publik yang terdiri dari 4.363 sampel digunakan, diikuti dengan preprocessing data, seleksi fitur melalui metode voting yang mengintegrasikan Boruta, Recursive Feature Elimination (RFE) dan SelectKBest, serta penanganan ketidakseimbangan kelas menggunakan Synthetic Minority Over-sampling Technique (SMOTE) dan Adaptive Synthetic Sampling Approach (ADASYN). Model-model tersebut juga disesuaikan lebih lanjut melalui optimisasi hyperparameter dengan GridSearchCV dan Repeated Stratified K- Fold Cross Validation. Hasil evaluasi menunjukkan ketiga algoritma memiliki kemampuan prediksi yang kuat dengan CatBoost menjadi terdepan, dengan akurasi 0.992, precision 0.992, recall 0.992, F1 score 0.992 dan nilai ROC AUC 0.9987. Analisis confusion matrix juga memvalidasi CatBoost memiliki jumlah misclassification paling rendah dibandingkan XGBoost dan LGBM. Selain itu penggunaan SHaply Additive exPlanations (SHAP) untuk interpretabilitas model menyoroti bahwa faktor utama yang memengaruhi prediksi risiko hipertensi adalah tekanan darah, indeks massa tubuh, aktivitas fisik keseluruhan, lingkar pinggang, kadar trigliserida, usia dan kadar kolesterol, yang selaras dengan pengetahuan medis yang telah mapan. Untuk memfasilitasi penggunaan di dunia nyata, model dengan performa terbaik diimplementasikan ke dalam antarmuka website yang ramah pengguna, memungkinkan pengguna memprediksi risiko hipertensi secara interaktif. Temuan ini menggambarkan bahwa algoritma boosting terutama CatBoost yang menawarkan metode machine learning dengan akurat, andal dan dapat diinterpretasikan untuk membangun sistem prediksi risiko hipertensi.

Kata kunci : Hipertensi, Prediksi, LGBM, SMOTE, Seleksi Fitur

ABSTRACT

Hypertension is a long-lasting condition that is highly prevalent and significantly contributes to cardiovascular issues, making early identification a crucial preventive action. This research evaluates the efficacy of three boosting algorithms, eXtreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LGBM), and of 4,363 samples was employed, followed by data preprocessing, feature selection through a voting method that integrates Boruta, Recursive Feature Elimination (RFE), and SelectKBest, as well as addressing class imbalance using the Synthetic Minority Over-sampling Technique (SMOTE) and ADASYN (Adaptive Synthetic Sampling Approach). The models were additionally fine-tuned through hyperparameter optimization using GridSearchCV and Repeated Stratified K-Fold Cross Validation. The evaluation results demonstrate that all three algorithms exhibited strong predictive capabilities, with CatBoost leading the way, achieving an accuracy of 0.992, precision of 0.992, recall of 0.992, F1-score of 0.992, and ROC- AUC of 0.9987. Analyzing the confusion matrix further validated that CatBoost had the lowest number of misclassifications when compared to XGBoost and LGBM. Additionally, the use of SHapley Additive exPlanations (SHAP) for model interpretability highlighted that the key factors influencing the prediction of hypertension risk are blood pressure, body mass index (BMI), overall physical activity, waist circumference, triglyceride levels, age, and LDL cholesterol levels, aligning with established medical knowledge. To facilitate real-world use, the top- performing model was implemented into a user-friendly website interface, allowing users to predict their hypertension risk interactively. These findings illustrate that boosting algorithms, especially CatBoost, offer an accurate, dependable, and interpretable machine learning method for creating hypertension risk prediction systems.

Keywords : Hypertension, Prediction, LGBM, SMOTE, Feature selection