

**ANALISIS SMOTE DAN RANDOM SEARCH PADA
ALGORITMA MACHINE LEARNING UNTUK DIAGNOSIS
PENYAKIT STROKE**

SKRIPSI NON REGULER JALUR SCIENTIST

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



Disusun oleh :

UBAID KHOIR JULIO DN

21.11.3822

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

**ANALISIS SMOTE DAN RANDOM SEARCH PADA
ALGORITMA MACHINE LEARNING UNTUK DIAGNOSIS
PENYAKIT STROKE**

SKRIPSI NON REGULER JALUR SCIENTIST

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



Disusun oleh :

UBAID KHOIR JULIO DN

21.11.3822

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2020

HALAMAN PERSETUJUAN

JALUR SCIENTIST

ANALISIS SMOTE DAN RANDOM SEARCH PADA ALGORITMA MACHINE
LEARNING UNTUK DIAGNOSIS PENYAKIT STROKE

yang disusun dan diajukan oleh

Ubaid Khoir Julio Dn

21.11.3822

telah disetujui oleh Dosen Pembimbing
pada tanggal 19 Februari 2026

Dosen Pembimbing,



Majid Bahardi, S.Kom., M.Eng
NIK. 190302393

HALAMAN PENGESAHAN

JALUR SCIENTIST

ANALISIS SMOTE DAN RANDOM SEARCH PADA ALGORITMA MACHINE LEARNING PADA DIAGNOSIS PENYAKIT STROKE

yang disusun dan diajukan oleh

Ubaid Khoir Julio Da
21.11.3822

Telah dipertahankan di depan Dewan Penguji
pada tanggal 19 Februari 2026

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Bayu Setiaji, M.Kom.
NIK. 190302216

Anna Baiha, S.Kom., M.Kom.
NIK. 190302290

Majid Rahardi, S.Kom., M.Eng.
NIK. 190302393



Laporan ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 19 Februari 2026

DEKAN FAKULTAS ILMU KOMPUTER



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

HALAMAN PERNYATAAN KEASLIAN KARYA

Yang bertandatangan di bawah ini,

Nama mahasiswa : Ubaid Khoir Julio Dn

NIM : 21.11.3822

Menyatakan bahwa Laporan dengan judul berikut:

Analisis SMOTE dan Random Search pada Algoritma Machine Learning untuk Diagnosis Penyakit Stroke

Dosen Pembimbing : Majid Rahardi, S.Kom., M.Eng.

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan kegiatan SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
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 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 ketidak-benaran 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, 19 Februari 2026



Ubaid Khoir Julio Dn

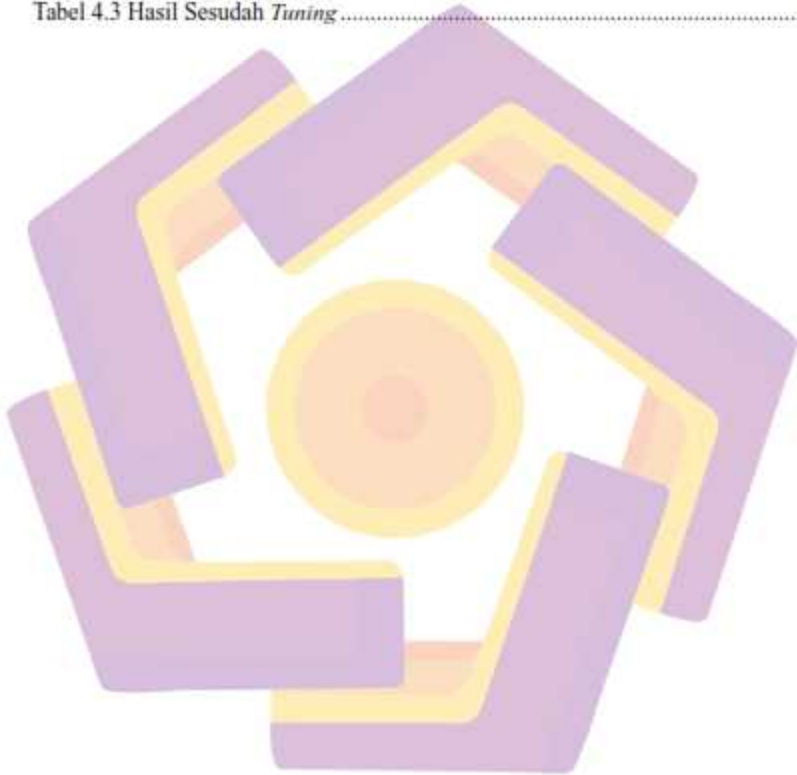
DAFTAR ISI

Halaman Judul.....	i
Halaman Persetujuan.....	ii
Halaman Pengesahan	iii
Halaman Pernyataan Keaslian Karya.....	iv
Halaman Persembahan	v
Kata Pengantar	vi
Daftar Isi.....	vii
Daftar Tabel.....	ix
Daftar Gambar.....	x
Intisari	xi
<i>Abstract</i>	xii
Bab I Pendahuluan	1
1.1. Gambaran Umum	1
1.2. Rumusan Masalah	3
1.3. Batasan Masalah.....	3
1.4. Tujuan.....	4
Bab II Tinjauan Pustaka	5
2.1. Studi Literatur	5
2.2. Landasan Teori	6
BAB III Metode Penelitian	8
3.1. Metode.....	8
3.1.1. Dataset Penelitian.....	9
3.1.2. Preprocessing Data	11
3.1.3. Pembuatan Model.....	11
3.1.4. Hyperparameter	14
3.1.5. Evaluasi	15
BAB IV Pembahasan	19
4.1. Sub Pembahasan	19
4.1.1 Dataset.....	19
4.1.2 Hasil Preprocessing Data	20

4.1.3	Analisis Pengaruh <i>SMOTE</i>	22
4.1.4	Hasil Evaluasi Model dengan <i>Hyperparameter Tuning</i>	24
4.1.5	Analisis Pengaruh <i>Hyperparameter Tuning</i>	24
4.1.6	Analisis <i>Overfitting</i> dan <i>Interpretabilitas</i> Model.....	27
4.1.7	Keterbatasan Penelitian.....	27
BAB V Kesimpulan		28
5.1.	Kesimpulan	28
5.2.	Saran.....	29
Referensi		30
Curriculum Vitae		32
Lampiran dan Bukti Pendukung.....		33
a.	Letter of Acceptance (LOA).....	33
b.	Lembar Review	34
c.	Bukti Terbit/Terindex	36
d.	Bukti pembayaran	37

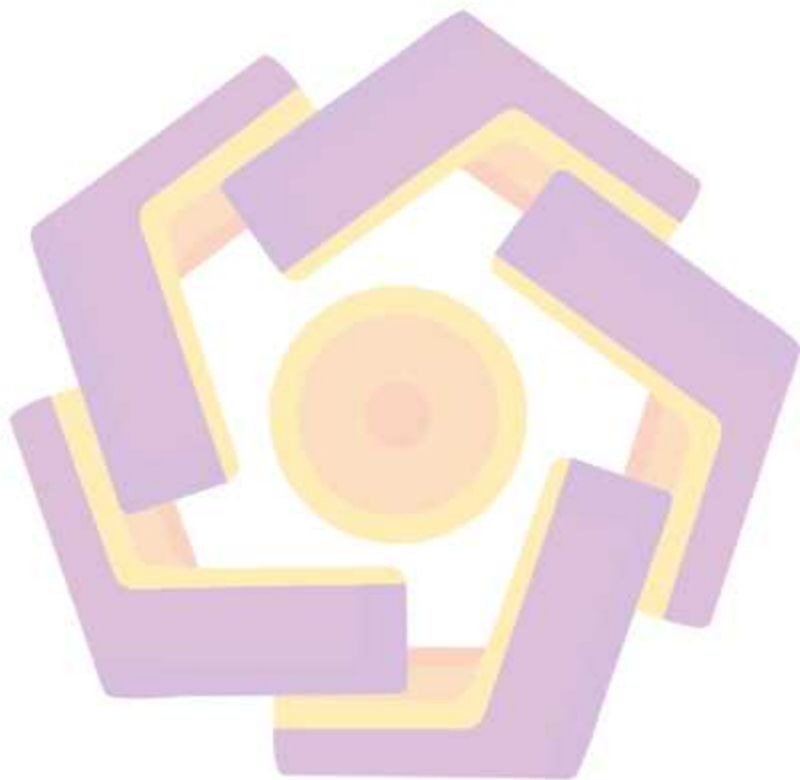
DAFTAR TABEL

Tabel 3.1 Deskripsi Atribut Dataset Stroke.....	10
Tabel 4.1 Tabel <i>Missing Value</i> Sebelum dan Sesudah <i>Imputasi</i>	21
Tabel 4.2 Hasil Sebelum <i>Tuning</i>	25
Tabel 4.3 Hasil Sesudah <i>Tuning</i>	25



DAFTAR GAMBAR

Gambar 3.1 Alur Penelitian.....	9
Gambar 4.1 Dataset Stroke	20
Gambar 4.2 Sebelum <i>SMOTE</i>	22
Gambar 4.3 Sesudah <i>SMOTE</i>	23



INTISARI

Stroke adalah kondisi medis kritis di mana prediksi *negatif* palsu dapat menyebabkan keterlambatan pengobatan dan peningkatan angka kematian. Oleh karena itu, model prediktif di bidang medis harus memprioritaskan sensitivitas (*recall*) selain akurasi keseluruhan. Studi ini menganalisis dampak dari *Synthetic Minority Over-sampling Technique (SMOTE)* dan optimasi *hyperparameter Random Search* pada lima algoritma pembelajaran mesin—*Random Forest*, *XGBoost*, *Support Vector Machine (SVM)*, *Logistic Regression*, dan *CatBoost*—untuk diagnosis penyakit stroke. Dua skenario eksperimental dilakukan, yaitu model yang dilatih tanpa *SMOTE* dan model yang dilatih dengan *SMOTE* yang hanya diterapkan pada data pelatihan untuk mencegah kebocoran data. Kinerja model dievaluasi menggunakan *akurasi*, *presisi*, *recall*, dan *F1-score*, dengan penekanan khusus pada *recall* karena relevansinya secara klinis. Dalam praktik klinis, *recall* yang rendah dapat menyebabkan prediksi negatif palsu, di mana pasien stroke berisiko tinggi tidak teridentifikasi oleh sistem, yang berpotensi mengakibatkan keterlambatan intervensi medis. Oleh karena itu, *recall* ditekankan sebagai metrik kinerja utama dalam studi ini. Hasil eksperimen menunjukkan bahwa *SMOTE* secara konsisten meningkatkan *recall* di semua model, sementara *Random Search* lebih meningkatkan kinerja. *CatBoost* mencapai kinerja terbaik dengan akurasi 96,61%, *recall* 97%, dan *F1-score* 97%. Terlepas dari kinerjanya yang unggul, potensi risiko *overfitting* dibahas secara kritis. Temuan ini menunjukkan bahwa pendekatan yang diusulkan menghasilkan model pendukung keputusan yang relevan secara klinis untuk prediksi risiko stroke.

Kata kunci: *CatBoost*, *Machine Learning*, *Random Search*, *SMOTE*, Prediksi Stroke.

ABSTRACT

Stroke is a critical medical condition in which false negative predictions may lead to delayed treatment and increased mortality. Therefore, predictive models in the medical domain should prioritize sensitivity (recall) in addition to overall accuracy. This study analyzes the impact of the Synthetic Minority Over-sampling Technique (SMOTE) and Random Search hyperparameter optimization on five machine learning algorithms—Random Forest, XGBoost, Support Vector Machine (SVM), Logistic Regression, and CatBoost—for stroke disease diagnosis. Two experimental scenarios were conducted, namely models trained without SMOTE and models trained with SMOTE applied only to the training data to prevent data leakage. Model performance was evaluated using accuracy, precision, recall, and F1-score, with particular emphasis on recall due to its clinical relevance. In clinical practice, low recall may lead to false negative predictions, where high-risk stroke patients are not identified by the system, potentially resulting in delayed medical intervention. Therefore, recall is emphasized as the primary performance metric in this study. Experimental results demonstrate that SMOTE consistently improves recall across all models, while Random Search further enhances performance. CatBoost achieved the best performance with an accuracy of 96.61%, recall of 97%, and F1-score of 97%. Despite its superior performance, potential overfitting risks are critically discussed. These findings indicate that the proposed approach produces a clinically relevant decision-support model for stroke risk prediction.

Keyword: CatBoost, Machine Learning, Random Search, SMOTE, Stroke Prediction.