

**OPTIMALISASI KINERJA ALGORITMA *RANDOM FOREST* UNTUK
DETEKSI DINI PENYAKIT STROKE MENGGUNAKAN DATA REKAM
MEDIS**

SKRIPSI

Diajukan Untuk Memenuhi Sebagian Syarat Guna
Memperoleh Gelar Sarjana Komputer (S. Kom)
Pada Program Studi Sistem Informasi



OLEH:

THEO KRISNA AMARYA

NPM: 2113030108

FAKULTAS TEKNIK DAN ILMU KOMPUTER (FTIK)
UNIVERSITAS NUSANTARA PERSATUAN GURU REPLUBIK INDONESIA
UN PGRI KEDIRI

2025

Skripsi Oleh:

THEO KRISNA AMARYA

2113030108

Judul :

**OPTIMALISASI KINERJA ALGORITMA *RANDOM FOREST* UNTUK
DETEKSI DINI PENYAKIT STROKE MENGGUNAKAN DATA REKAM
MEDIS**

Telah disetujui untuk diajukan Kepada

Panitia Ujian / Sidang Skripsi Prodi Sistem Informasi

FTIK UN PGRI Kediri

Tanggal : 2 Juli 2025

Pembimbing I



Rina Firliana, M. Kom
NIDN. 0731087703

Pembimbing II



Aidina Ristyawan, M. Kom
NIDN. 0721018801

Skripsi oleh:

THEO KRISNA AMARYA

NPM : 2113030108

Judul:

**OPTIMALISASI KINERJA ALGORITMA *RANDOM FOREST* UNTUK
DETEKSI DINI PENYAKIT STROKE MENGGUNAKAN DATA REKAM
MEDIS**

Telah dipertahankan di depan Panitia Ujian / Sidang Skripsi

Program Studi Sistem Informasi FTIK UN PGRI Kediri

Tanggal: 9 Juli 2025

Dan Dinyatakan telah Memenuhi Persyaratan

Panitia Penguji:

1. Ketua : Rina Firliana, M. Kom
2. Penguji I : Rini Indriati, S.Kom, M.Kom
3. Penguji II : Aidina Ristyawan, M. Kom

[.....]

[.....]

[.....]



Mengetahui,

Dekan Fakultas Teknik dan Ilmu Komputer

Dr. Sulistiono, M. Si

NIP. 196807071993031004

PERNYATAAN

Yang bertanda tangan dibawah ini saya,

Nama : Theo Krisna Amarya
Jenis Kelamin : Laki – laki
Tempat / Tgl Lahir : Kediri, 9 Maret 2001
NPM : 2113030108
Fak/ Jur/Prodi : Teknik dan Ilmu Komputer / Sistem Informasi

menyatakan dengan sebenar benarnya, bahwa dalam Skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu perguruan tinggi, dan sepanjang pengetahuan saya tidak terdapat karya tulis atau pendapat yang pernah diterbitkan oleh orang lain, kecuali yang secara sengaja dan tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Kediri, 2 Juli 2025

Yang menyatakan



THEO KRISNA AMARYA

NPM 2113030108

MOTTO

“Jadilah terang seperti cahaya lilin yang memberikan sinar terang dan hangat bagi orang di sekitarmu.”

(Matius 5 : 14)

“Ketika orang lain menganggap sesuatu mustahil, itulah saatnya bagi kita untuk membuktikan sebaliknya dengan keyakinan dan ketekunan”.

(Plato)

“Kepercayaan adalah fondasi dari segala hubungan manusia. Bangunlah kredibilitasmu dengan menjadi pribadi yang dapat diandalkan dan jujur”.

(Confucius)

“Ketika kuat, bertindaklah lemah; ketika lemah, bertindaklah kuat.”

(Sun Tzu dalam bukunya, Art of War.)

“Peliharalah hubungan baikmu dengan orang – orang, dan hal – hal baik akan datang kepadamu.”

(Malti Bhojwani.)

“Ketika datang kesempatan, belajarlah untuk mengambilnya dengan bijaksana”

(Confucius.)

“Kesuksesan besar selalu diawali dengan keberanian untuk memulai.”

(John F. Kennedy.)

ABSTRAK

Theo Krisna Amarya: Optimalisasi Kinerja Algoritma *Random Forest* Untuk Deteksi Dini Penyakit Stroke Menggunakan Data Rekam Medis, Skripsi, Sistem Informasi, Fakultas Teknik dan Ilmu Komputer UN PGRI Kediri, 2025.

Kata kunci: *Adaboost, Hybrid Sampling, Random Forest, Smote-Tomek, Stroke Predict.*

Stroke merupakan kondisi darurat medis yang terjadi ketika pasokan darah ke otak terganggu, baik karena penyumbatan (*stroke iskemik*) maupun pecahnya pembuluh darah (*stroke hemoragik*). Kondisi ini tidak hanya menjadi penyebab kematian kedua terbesar di dunia tetapi juga penyebab utama kecacatan jangka panjang. Faktor risiko seperti hipertensi, diabetes, gaya hidup tidak sehat, dan usia semakin meningkatkan prevalensi stroke. Deteksi dini stroke sangat krusial untuk mengurangi dampak buruknya, namun tantangan seperti ketidakseimbangan data dan akurasi model prediksi sering menghambat efektivitas sistem diagnosis. Oleh karena itu, penelitian ini berfokus pada pengembangan model prediksi stroke yang lebih akurat dengan mengatasi masalah ketidakseimbangan data dan optimasi parameter algoritma.

Penelitian ini bertujuan meningkatkan kinerja algoritma *Random Forest* sebagai model deteksi dini stroke dengan mengombinasikan metode *hybrid sampling SMOTETomek* untuk menangani ketidakseimbangan data. SMOTETomek menggabungkan *oversampling (SMOTE)* dan *undersampling (Tomek Links)* guna menyeimbangkan distribusi kelas. Selain itu, dilakukan *hyperparameter tuning* pada *Random Forest* untuk mengoptimalkan performanya. Algoritma ini juga diintegrasikan dengan *AdaBoost* sebagai *boosting technique* guna meningkatkan stabilitas dan akurasi prediksi. Dataset yang digunakan mencakup variabel seperti tekanan darah, kadar gula, riwayat kesehatan, dan gaya hidup pasien.

Hasil penelitian menunjukkan peningkatan signifikan dibandingkan studi sebelumnya, dengan akurasi mencapai 96% (*standar deviasi* 0%) dan nilai *Area Under Curve (AUC) Receiver Operating Characteristic (ROC)* sebesar 0.96 (96%). Kombinasi SMOTETomek dan optimasi parameter berhasil mengurangi bias pada data minoritas, sementara integrasi *AdaBoost* memperkuat generalisasi model. Hasil ini mengungguli penelitian sebelumnya yang hanya mencapai akurasi 94% dengan *standar deviasi* 2%, membuktikan efektivitas pendekatan *hybrid* dan *tuning* parameter.

Penelitian ini berhasil mengoptimalkan algoritma *Random Forest* untuk deteksi dini stroke dengan akurasi tinggi dan konsistensi yang baik. Namun, limitasi seperti ketergantungan pada data historis dan kebutuhan validasi eksternal masih perlu diatasi. Untuk pengembangan selanjutnya, disarankan eksplorasi *deep learning* atau *ensemble methods* lain, serta uji coba pada dataset yang lebih besar dan beragam untuk memastikan *robustnes* model. Implementasi *real-time* dengan integrasi sistem kesehatan juga dapat menjadi langkah strategis berikutnya.

KATA PENGANTAR

Puji Syukur Kami panjatkan Kehadirat Allah Tuhan Yang Maha Kuasa, karena hanya atas perkenan-Nya penyusunan skripsi ini dapat diselesaikan. Skripsi dengan judul “Optimalisasi Kinerja Algoritma *Random Forest* Untuk Deteksi Dini Penyakit Stroke Menggunakan Data Rekam Medis” ini ditulis guna memenuhi sebagian syarat untuk memperoleh gelar Sarjana Komputer, pada Program Studi Sistem Informasi Fakultas Teknik dan Ilmu Komputer UN PGRI Kediri. Pada kesempatan ini diucapkan terima kasih dan penghargaan yang setulus-tulusnya kepada:

1. **Dr. Zainal Afandi, M. Pd.** selaku Rektor UN PGRI Kediri yang selalu memberikan dorongan motivasi kepada mahasiswa.
2. **Dr. Sulistiono, M.Si.** selaku Dekan Fakultas Teknik dan Ilmu Komputer UN PGRI Kediri yang selalu memberikan dorongan motivasi kepada mahasiswa.
3. **Sucipto, M. Kom.** selaku Ka Prodi Sistem Informasi UN PGRI Kediri yang selalu memberikan dorongan motivasi kepada mahasiswa.
4. **Rina Firliana, M. Kom** dan **Aidina Ristyan, M. Kom** selaku Pembimbing skripsi yang dengan sabar membimbing dan memotivasi dalam menyelesaikan skripsi ini.
5. Ayah, Kakak, Adik, Alm. Ibu, dan seorang perempuan spesial dalam hidup saya, serta teman – teman Prodi Sistem Informasi.
6. Bpk & Ibu Dosen Sistem Informasi yang selalu memberikan dorongan dan motivasi.

Disadari bahwa skripsi ini masih banyak kekurangan, maka diharapkan tegur sapa, kritik dan saran-saran, dari berbagai pihak sangat diharapkan. Akhirnya, disertai harapan semoga skripsi ini ada manfaatnya bagi kita semua, khususnya bagi dunia pendidikan, meskipun hanya ibarat setitik air bagi samudra yang luas.

Kediri, 2 Juli 2025

THEO KRISNA AMARYA

NPM : 2113030108

DAFTAR ISI

Halaman Sampul	i
Halaman Persetujuan.....	ii
Halaman Pengesahan	iii
Halaman Pernyataan.....	iv
Halaman Motto.....	v
ABSTRAK	vi
KATA PENGANTAR.....	vii
DAFTAR ISI.....	viii
DAFTAR TABEL.....	x
DAFTAR GAMBAR.....	xi
DAFTAR LAMPIRAN	xii
BAB I PENDAHULUAN.....	1
A. Latar Belakang.....	1
B. Batasan Masalah.....	2
C. Rumusan Masalah	2
D. Tujuan Penelitian.....	3
E. Manfaat Penelitian.....	3
F. Sistematika Penulisan.....	3
BAB II LANDASAN TEORI	5
A. Kajian Teori.....	5
1. Penyakit Stroke	5
2. Machine Learning	5
3. Data Mining	6
4. <i>Classification</i> atau Klasifikasi.....	6
5. <i>Random Forest</i>	6
B. Kajian Hasil Penelitian Terdahulu.....	9
C. Kerangka Berpikir	11

BAB III METODE PENELITIAN	13
A. Gambaran Objek Penelitian.....	13
B. Metode Penelitian.....	13
C. Metode Pengumpulan Data	14
D. Metode Analisis Data	14
E. Alur Penelitian.....	14
BAB IV HASIL DAN PEMBAHASAN.....	18
BAB V PENUTUP	28
A. Kesimpulan.....	28
B. Saran	28
DAFTAR PUSTAKA	29
LAMPIRAN.....	34

DAFTAR TABEL

Tabel 4. 1 Atribut Dataset Stroke	18
Tabel 4. 2 Nilai Hilang Pada Dataset Stroke	18
Tabel 4. 3 Hasil Impitasi Missing Value	19
Tabel 4. 4 Hasil Transformasi Data	20
Tabel 4. 5 Skema Parameter <i>Random Forest</i>	21
Tabel 4. 6 Hasil Klasifikasi <i>WRF</i>	21
Tabel 4. 7 Hasil Klasifikasi <i>RFWFS</i>	22
Tabel 4. 8 Hasil Klasifikasi <i>RFWSS</i>	23
Tabel 4. 9 Hasil Klasifikasi <i>ABRF</i>	24
Tabel 4. 10 Hasil Klasifikasi <i>RFWP</i>	25
Tabel 4. 11 Hasil Perbandingan Kinerja Tekni Pengoptimalan.....	26

DAFTAR GAMBAR

Gambar 2. 1 Diagram <i>Random Forest</i>	7
Gambar 2. 2 Kerangka Berpikir	11
Gambar 3. 1 Alur Penelitian.....	15
Gambar 4. 1 Sampel Nilai Hilang pada Kolom Bmi	19
Gambar 4. 2 Hasil Penyeimbangan Data	21
Gambar 4. 3 Grafik <i>ROC</i> dan <i>AUC WRF</i>	22
Gambar 4. 4 Grafik <i>ROC</i> dan <i>AUC RFWFS</i>	23s
Gambar 4. 5 Grafik <i>ROC</i> dan <i>AUC RFWSS</i>	24
Gambar 4. 6 Grafik <i>ROC</i> dan <i>AUC ABRF</i>	25
Gambar 4. 7 Grafik <i>ROC</i> dan <i>AUC RFWP</i>	26

DAFTAR LAMPIRAN

Lampiran 1. Kajian Literatur	35
Lampiran 2. Isi Dataset Stroke.....	36
Lampiran 3. <i>Type Data</i> Dari Dataset Stroke.....	37
Lampiran 4. Cek Data yang Kosong.....	37
Lampiran 5. Transformasi Data	38
Lampiran 6. Menangani Data Yang Tidak Seimbang	38
Lampiran 7. Bimbingan Skripsi.....	39
Lampiran 8. Surat Keterangan Bebas Similarity PPL.....	40
Lampiran 9. Bukti Halaman Awal cek similarity	41
Lampiran 10. Upload Jurnal JPIT.....	42
Lampiran 11. Upload Jurnal Kedua Semnas Inotek	42

BAB I PENDAHULUAN

A. Latar Belakang

Stroke merupakan sesuatu keadaan medis yang terjadi pada saat aliran darah ke otak tersendat, mengakibatkan kerusakan pada jaringan otak (Riyadina & Rahajeng, 2013). Stroke ialah salah satu pemicu utama kematian serta kecacatan nomor 2 didunia, penyakit ini bisa melanda orang dari seluruh umur serta dipengaruhi oleh bermacam aspek resiko, seperti pola hidup tidak sehat, tekanan darah tinggi, serta rata-rata kandungan gula, dan resiko lainnya.

Sangat penting untuk mendeteksi stroke pada pasien segera mungkin untuk mencegahnya. Teknologi machine learning (Ramadhanu et al., 2023) telah digunakan secara luas dalam banyak bidang, termasuk bidang perawatan kesehatan. Klasifikasi adalah teknik yang sering digunakan untuk diagnosis dan prognosis penyakit. Klasifikasi menggunakan sejumlah algoritma untuk memprediksi risiko stroke berdasarkan data pasien. Dengan menemukan algoritma terbaik yang dapat memberikan hasil yang paling akurat dan konsisten, mengingat jumlah algoritma yang tersedia dan dataset yang besar.

Algoritma seperti *Random Forest*, *K-Nearest Neighbors* (KNN), *Naive Bayes*, *Decision Trees*, *Support Vector Machines* (SVM), *Neural Networks*, dan *Logistic Regression* sering digunakan untuk klasifikasi (Wibawa, Guntur, Purnama, Fathony Akbar, & Dwiyanto, 2018). Namun, kinerja setiap algoritma dapat berbeda tergantung pada karakteristik dataset dan teknik *preprocessing* yang digunakan, seperti *balancing data* dan *cross-validation*. Penyebaran data yang tidak merata, seperti ketidakseimbangan data pada dataset, dapat menyebabkan bias (kesalahan) atau deviasi sistematis yang terjadi dalam model, yang bisa memengaruhi hasil klasifikasi serta mengurangi akurasi algoritma (Bradter et al., 2022). Untuk membuat model klasifikasi lebih akurat dalam memprediksi kasus stroke, penerapan metode *balancing data* dapat digunakan untuk membantu meningkatkan kinerja dari algoritma klasifikasi. Selain itu, teknik *cross-validation* juga digunakan untuk menilai kinerja model.

Penulis ingin melanjutkan penelitian sebelumnya yang dikerjakan oleh Theo Krisna, dkk yang berjudul "Analisa Perbandingan Algoritma Classification Berdasarkan Komposisi Label", pada Tahun 2024 . Penelitian tersebut bertujuan untuk menemukan algoritma klasifikasi dengan kinerja terbaik. Hasil dari penelitian tersebut mengatakan bahwa algoritma yang terbaik untuk klasifikasi penyakit stroke yaitu algoritma *Random Forest* dengan tingkat akurasi sebesar 94% dengan standar deviasi 2% (Amarya, Galuh, Achmad, Daniati, & Ristyawan, 2024) Pada penelitian lain yang menggunakan dataset sama seperti yang dilakukan oleh (Guhdar, Ismail Melhum, & Luqman Ibrahim, 2023) membahas prediksi penyakit stroke menggunakan algoritma *Logistik Regression* menghasilkan kinerja sebesar 86%, dan penelitian lain yang pernah dilakukan oleh (Wulandari et al., 2024) membahas klasifikasi penyakit stroke menggunakan algoritma SVM (*Support Vector Machine*) menghasilkan kinerja sebesar 85.45% dengan perbandingan data 80:20 dan sebesar 85.24% dengan perbandingan data 70:30. Penelitian (Wang et al., 2020) menerapkan algoritma *Random Forest* sebagai algoritma dalam memprediksi stroke, kemudian penelitian (Huang et al., 2022) membuktikan bahwa algoritma *Random Forest* adalah algoritma yang cocok dalam memprediksi penyakit stroke, hal ini yang membuat penelitian ini berfokus dalam mengoptimalkan algoritma *Random Forest* agar menghasilkan standar deviasi seminimal mungkin dan mengembangkan model prediksi berdasarkan algoritma *Random Forest* sebagai sarana deteksi dini penyakit stroke dan untuk mengetahui potensi penyakit stroke secara dini, dengan menggunakan data rekam medis.

B. Batasan Masalah

Penelitian ini dibatasi pada klasifikasi penyakit stroke menggunakan dataset rekam medis *WHO* (Pathan, Jianbiao, John, Nag, & Dev, 2020) dari website *kaggle*. Algoritma yang diuji dalam klasifikasi ini adalah *Random Forest*.

C. Rumusan Masalah

Dari paparan pada latar belakang, rumusan masalah yang ada yaitu bagaimana mengoptimalkan kinerja algoritma *Random Forest* untuk deteksi dini penyakit stroke berdasarkan data rekam medis?

D. Tujuan Penelitian

- 1) Mengembangkan model deteksi berbasis algoritma *Random Forest* sebagai alat deteksi dini penyakit stroke.
- 2) Untuk mengetahui potensi terjadinya penyakit stroke secara dini, menggunakan data rekam medis.
- 3) Untuk meminimalisir tingkat bias (kesalahan) pada model yang diterapkan.

E. Manfaat Penelitian

- 1) Terbentuknya model deteksi dini penyakit stroke berbasis algoritma *Random Forest* dengan kinerja yang baik.
- 2) Mengetahui secara dini potensi terjadinya penyakit stroke pada pasien, berdasarkan data rekam medis.
- 3) Mengetahui teknik meminimalisir tingkat bias (kesalahan) pada model.

F. Sistematika Penulisan

BAB I PENDAHULUAN

Bab ini terdiri dari latar belakang, identifikasi masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

BAB II LANDASAN TEORI

Bab ini berisikan teori – teori yang digunakan untuk memahami permasalahan yang dibahas pada penelitian ini yaitu teori yang berkaitan dengan penyakit stroke, *machine learning*, *Data Mining*, *Classification*, *Random Forest*, *Weighted Random Forest*, *Random Forest With Feature Selection*, *Random Forest With Stratified Selection*, *AdaBoost Random Forest*, dan *Random Forest With Parallelization*.

BAB III METODE PENELITIAN

Bab ini berisi tentang tahapan yang akan dilakukan pada penelitian dimana tahapan dideskripsikan dalam bentuk diagram alir serta masing – masing tahapan berisi penjelasan rinci yang terkait.

BAB IV HASIL DAN PEMBAHASAN

Bab ini berisi tentang proses yang terdapat pada tahap penelitian yang telah dibuat serta hasil temuan – temuan yang didapat pada saat penelitian.

BAB V PENUTUP

Bab ini berisi tentang kesimpulan dari seluruh uraian pada bab – bab sebelumnya serta saran yang diharapkan berguna untuk pengembangan selanjutnya.

DAFTAR PUSTAKA

- Abdullah, K., Jannah, M., Aiman, U., Hasda, S., Fadilla, Z., Taqwin, T., ... Sari, M. E. (2022). Metodologi Penelitian Kuantitatif. In N. Saputra (Ed.), *Yayasan Penerbit Muhammad Zaini*. Yayasan Penerbit Muhammad Zaini. Retrieved from <https://penerbitzaini.com/>
- Adha, F., Airi, H., Suprati, T., & Bahtiar, A. (2023). KOMPARASI METODE KLASIFIKASI DATA MINING UNTUK PREDIKSI PENYAKIT STROKE. *Jurnal Teknik Elektro dan Informatika*, 18(1), 73–79.
- Aini, E. D. N., Khasanah, R. A., Ristyawan, A., & Daniati, E. (2024). Penggunaan Data Mining untuk Prediksi tingkat Obesitas di Meksiko Menggunakan Metode Random Forest. *Prosiding SEMNAS INOTEK (Seminar Nasional Inovasi Teknologi)*, 8, 1256–1265. Retrieved from <https://proceeding.unpkediri.ac.id/index.php/inotek/>
- Amarya, T. K., Galuh, A. C. A., Achmad, R., Daniati, E., & Ristyawan, A. (2024). Analisa Perbandingan Algoritma Classification Berdasarkan Komposisi Label. *Prosiding SEMNAS INOTEK*, 8(1), 32–40. <https://doi.org/https://doi.org/10.29407/inotek.v8i1.4906>
- Bradter, U., Altringham, J. D., Kunin, W. E., Thom, T. J., O'Connell, J., & Benton, T. G. (2022). Variable ranking and selection with random forest for unbalanced data. *Environmental Data Science*, 1, 1–23. <https://doi.org/10.1017/eds.2022.34>
- Daniati, E., Firliana, R., & Wardani, A. S. (2024a). *Dukungan Keputusan Dengan Pendekatan Machine Learning*. Kediri: Universitas Nusantara PGRI Kediri. <https://doi.org/https://doi.org/10.29407/punp.vi.68>
- Daniati, E., Firliana, R., & Wardani, A. S. (2024b). *Dukungan Keputusan Dengan Pendekatan Machine Learning*. Kediri: Universitas Nusantara PGRI Kediri. <https://doi.org/https://doi.org/10.29407/punp.vi.68>
- Elwirehardja, G. N., Suparyanto, T., & Pardamean, B. (2023). *MACHINE LEARNING UNTUK PEMULA* (1st ed.; F. A. Munawar, ed.). Yogyakarta:

INSTIPER PRESS (IKAPI & APPTI).

- Fadellia Azzahra, Suarna, N., & Arie Wijaya, Y. (2024). Penerapan Algoritma Random Forest Dan Cross Validation Untuk Prediksi Data Stunting. *Kopertip : Jurnal Ilmiah Manajemen Informatika Dan Komputer*, 8(1), 1–6. <https://doi.org/10.32485/kopertip.v8i1.238>
- Fedesoriano. (2020). Stroke Prediction Dataset. Retrieved October 4, 2024, from <https://www.kaggle.com/fedesoriano> website:
<https://www.kaggle.com/datasets/fedesoriano/stroke-prediction-dataset>
- Fci, H., Fan, Z., Wang, C., Zhang, N., Wang, T., Chen, R., & Bai, T. (2022). Cotton Classification Method at the County Scale Based on Multi-Features and Random Forest Feature Selection Algorithm and Classifier. *Remote Sensing*, 14(4), 1–28. <https://doi.org/10.3390/rs14040829>
- Ginantra, N. L. W. S. R., Arifah, F. N., Wijaya, H. A., Septarini, R. S., Ahmad, N., Ardiana, D. P. Y., ... Negara, E. S. (2021). *Data Mining dan Penerapan Algoritma* (R. Watrionthos & J. Simarmata, Eds.). Medan: Yayasan Kita Menulis.
- Guhdar, M., Ismail Melhum, A., & Luqman Ibrahim, A. (2023). Optimizing Accuracy of Stroke Prediction Using Logistic Regression. *Journal of Technology and Informatics (JoTI)*, 4(2), 41–47. <https://doi.org/10.37802/joti.v4i2.278>
- Gultom, S. I. (2020). Implementasi Data Mining Menentukan Pola Hidup Sehat Bagi Pengguna KB Menggunakan Algoritma Adaboost (Studi Kasus :Dinas Serdang Bedagai). *Jurnal Informasi Dan Teknologi Ilmiah (INTI)*, 7(3), 298–304.
- Huang, X., Cao, T., Chen, L., Li, J., Tan, Z., Xu, B., ... Liu, L. (2022). Novel Insights on Establishing Machine Learning-Based Stroke Prediction Models Among Hypertensive Adults. *Frontiers in Cardiovascular Medicine*, 9(May), 1–11. <https://doi.org/10.3389/fcvm.2022.901240>
- Ismafillah, D., Rohana, T., & Cahyana, Y. (2023). Implementasi Model Support

- Vector Machine dan Logistic Regression Untuk Memprediksi Penyakit Stroke. *Jurnal Riset Komputer*, 10(1), 2407–389.
<https://doi.org/10.30865/jurikom.v10i1.5478>
- Krisna Amarya, T., Candra Andy, A. G., Achmad, R., Daniati, E., & Ristyawan, A. (2024). Analisa Perbandingan Algoritma Classification Berdasarkan Komposisi Label. *Prosiding SEMNAS INOTEK*, 8, 2549–7952. Retrieved from <https://proceeding.unpkediri.ac.id/index.php/inotek>
- Louppe, G. (2014). *Understanding Random Forests: From Theory to Practice*. Retrieved from <http://arxiv.org/abs/1407.7502>
- Lugianti, D., Nugroho, B. I., & Surejo, S. (2022). Penerapan Sistem Interaksi Manusia Dan Komputer Pada Website E Learning Di STMIK Tegal. *Jurnal Manajemen Informatika Politeknik Ganesa (Jurnal Minfo Polgan)*, 11(2), 36–44.
- Muvida, & Amar, F. (2024). The Features of Comorbidity of Stroke in The Indonesian Population: Findings from The Indonesian Family Life Survey (IFLS-5). *Magna Neurologica*, 2(2), 42–47.
<https://doi.org/10.20961/magnaneurologica.v2i2.948>
- Nugroho, A., Fanani, A. Z., & Shidik, G. F. (2021). *Evaluation of Feature Selection Using Wrapper For Numeric Dataset With Random Forest Algorithm*. IEEE.
<https://doi.org/https://doi.org/10.1109/iSemantic52711.2021.9573249>
- Pathan, M. S., Jianbiao, Z., John, D., Nag, A., & Dev, S. (2020). Identifying Stroke Indicators Using Rough Sets. *IEEE Access*, 8, 210318–210327.
<https://doi.org/10.1109/ACCESS.2020.3039439>
- Putri, M., & Erlin. (2024). Prediksi Penyakit Stroke Menggunakan Machine Learning Dengan Algoritma Random Forest. *Jurnal Infomedia: Teknik Informatika*, 9(2), 16–21.
- Raharjo, B. (2021). *Pembelajaran Mesin (Machine Learning)* (1st ed.; M. C. Wibowo, Ed.). Semarang: Yayasan Prima Agus Teknik.

- Ramadhanu, A., Ayu Mahessya, R., Raihan Zaky, M., Isra, M., Informasi, S., & Putra Indonesia YPTK Padang, U. (2023). PENERAPAN TEKNOLOGI MACHINE LEARNING DENGAN METODE VADER PADA APLIKASI SENTIMEN TAMU DI HOTEL DYMENS. *JOISIE Journal Of Information System And Informatics Engineering*, 7(1), 165–173.
- Riyadina, W., & Rahajeng, E. (2013). Determinan Penyakit Stroke. *Kesmas: National Public Health Journal*, 7(7), 324.
<https://doi.org/10.21109/kesmas.v7i7.31>
- Sadiq, B. H., & Zeebarec, S. R. (2024). Parallel Processing Impact on Random Forest Classifier Performance: A CIFAR-10 Dataset Study. *Indonesian Journal of Computer Science Attribution*, 13(2), 1833–1846.
- Saputro, M. R. (2023). *Implementasi Metode Adaptive Boosting Untuk Prediksi Hasil Pertandingan Sepak Bola*. Universitas Nusantara PGRI Kediri.
- Shahhosseini, M., & Hu, G. (n.d.). *Improved Weighted Random Forest for Classification Problems*.
- Susanto, S., & Suryadi, D. (2010). *Pengantar Data Mining Menggali Pengetahuan Dari Bongkahan Data* (N. WK, Ed.). Yogyakarta: CV.ANDYOFFSET.
- Utomo, T. Y. (2024). *BUKU AJAR STROKE* (1st ed.; N. Khairusyifa & A. A. Hayuwaskita, Eds.). Purbalingga: EUREKA MEDIA AKSARA, JUNI 2024
ANGGOTA IKAPI JAWA TENGAH NO. 225/JTE/2021.
- Wang, W., Kiik, M., Peek, N., Curcin, V., Marshall, I. J., Rudd, A. G., ... Bray, B. (2020). A systematic review of machine learning models for predicting outcomes of stroke with structured data. *PLoS ONE*, 15(6), 1–16.
<https://doi.org/10.1371/journal.pone.0234722>
- Wibawa, A. P., Guntur, M., Purnama, A., Fathony Akbar, M., & Dwiyanto, F. A. (2018). Metode-metode Klasifikasi. *Prosiding Seminar Ilmu Komputer Dan Teknologi Informasi*, 3(1), 134–138.
- Wulandari, E., Witanti, A., Mercu, U., Yogyakarta, B., Sleman, K., Informatika,

- J., & Informasi, F. T. (2024). Classification Of Stroke Prediction Using The Support Vector Machine (SVM). *Jurnal Teknik Informatika Dan Sistem Informasi*, 11(3), 17–29.
- Wyner, A. J., Olson, M., Bleich, J., & Mease, D. (2017). Explaining the Success of AdaBoost and Random Forests as Interpolating Classifiers. In *Journal of Machine Learning Research* (Vol. 18). Retrieved from <http://jmlr.org/papers/v18/15-240.html>.
- Ye, Y., Wu, Q., Zhexue Huang, J., Ng, M. K., & Li, X. (2013). Stratified sampling for feature subspace selection in random forests for high dimensional data. *Pattern Recognition*, 46(3), 769–787. <https://doi.org/10.1016/J.PATCOG.2012.09.005>