

**KLASIFIKASI PENYAKIT STROKE MENGGUNAKAN *DECISION*
TREE C5.0 DENGAN *LOCAL OUTLIER FACTOR (LOF)* UNTUK
DETEKSI OUTLIER**

SKRIPSI



**UIN SUNAN AMPEL
S U R A B A Y A**

Disusun Oleh
MELINA AGUSTIYANINGSIH
09020222035

**PROGRAM STUDI MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SUNAN AMPEL
SURABAYA**

2026

PERNYATAAN KEASLIAN

Saya yang bertanda tangan di bawah ini,

Nama : MELINA AGUSTIYANINGSIH
NIM : 09020222035
Program Studi : Matematika
Angkatan : 2022

Menyatakan bahwa saya tidak melakukan plagiat dalam penulisan skripsi saya yang berjudul "KLASIFIKASI PENYAKIT STROKE MENGGUNAKAN *DECISION TREE C5.0* DENGAN *LOCAL OUTLIER FACTOR* (LOF) UNTUK DETEKSI OUTLIER". Apabila suatu saat nanti terbukti saya melakukan tindakan plagiat, maka saya bersedia menerima sanksi yang telah ditetapkan.

Demikian pernyataan keaslian ini saya buat dengan sebenar-benarnya.

Surabaya, 11 Januari 2026

Yang menyatakan,



MELINA AGUSTIYANINGSIH
NIM. 09020222035

LEMBAR PERSETUJUAN PEMBIMBING

Skripsi oleh

Nama : MELINA AGUSTIYANINGSIH
NIM : 09020222035
Judul proposal skripsi : KLASIFIKASI PENYAKIT STROKE
MENGUNAKAN *DECISION TREE C5.0*
DENGAN *LOCAL OUTLIER FACTOR (LOF)*
UNTUK DETEKSI OUTLIER

telah diperiksa dan disetujui untuk diujikan.

Pembimbing I



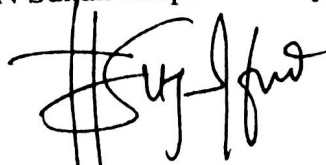
Dr. Yuniar Farida, M.T
NIP. 197905272014032002

Pembimbing II



Hani Khaulasari, M.Si
NIP. 199102092020122011

Mengetahui,
Ketua Program Studi Matematika
UIN Sunan Ampel Surabaya



Dr. Yuniar Farida, M.T
NIP. 197905272014032002

PENGESAHAN TIM PENGUJI SKRIPSI

Skripsi oleh:

Nama : MELINA AGUSTIYANINGSIH
NIM : 09020222035
Judul : KLASIFIKASI PENYAKIT STROKE
MENGUNAKAN *DECISION TREE C5.0* DENGAN
LOCAL OUTLIER FACTOR (LOF) UNTUK DETEKSI
OUTLIER

Telah dipertahankan di depan Tim Penguji pada tanggal 5 Januari 2026.

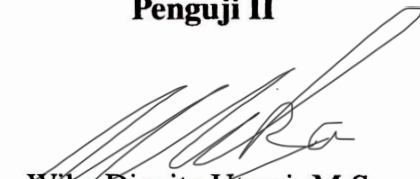
Mengesahkan,
Tim Penguji

Penguji I



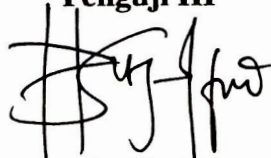
Dr. Dian Candra Rini Novitasari, M.Kom
NIP. 198511242014032001

Penguji II



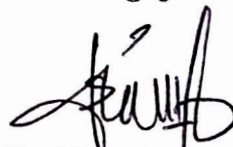
Wika Dianita Utami, M.Sc
NIP. 199206102018012003

Penguji III



Dr. Yuniar Farida, M.T
NIP. 197905272014032002

Penguji IV



Hani Khaulasari, M.Si
NIP. 199102092020122011

Mengetahui,

Dekan Fakultas Sains dan Teknologi
UTN Sunan Ampel Surabaya



Dr. A. Saepul Hamdani, M.Pd
NIP. 196507312000031002



UIN SUNAN AMPEL
SURABAYA

KEMENTERIAN AGAMA
UNIVERSITAS ISLAM NEGERI SUNAN AMPEL SURABAYA
PERPUSTAKAAN

Jl. Jend. A. Yani 117 Surabaya 60237 Telp. 031-8431972 Fax.031-8413300
E-Mail: perpustakaan@uinsby.ac.id

LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI
KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademika UIN Sunan Ampel Surabaya, yang bertanda tangan di bawah ini, saya:

Nama : MELINA AGUSTIYA NINGSIH
NIM : 09010222035
Fakultas/Jurusan : MATEMATIKA / SAINS DAN TEKNOLOGI
E-mail address : linaagustiya1@gmail.com

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Perpustakaan UIN Sunan Ampel Surabaya, Hak Bebas Royalti Non-Eksklusif atas karya ilmiah :

☒ Skripsi ☐ Tesis ☐ Desertasi ☐ Lain-lain (.....)
yang berjudul :

KLASIFIKASI PENYAKIT STROKE MENGGUNAKAN DECISION TREE C5.0
DENGAN LOCAL OUTLIER FACTOR (LOF) UNTUK DETEKSI OUTLIER

beserta perangkat yang diperlukan (bila ada). Dengan Hak Bebas Royalti Non-Eksklusif ini Perpustakaan UIN Sunan Ampel Surabaya berhak menyimpan, mengalih-media/format-kan, mengelolanya dalam bentuk pangkalan data (database), mendistribusikannya, dan menampilkan/mempublikasikannya di Internet atau media lain secara **fulltext** untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan atau penerbit yang bersangkutan.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak Perpustakaan UIN Sunan Ampel Surabaya, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini yang saya buat dengan sebenarnya.

Surabaya, 12 Januari 2026

Penulis

(Melina Agustyaningsih)
nama terang dan tanda tangan

ABSTRAK

KLASIFIKASI PENYAKIT STROKE MENGGUNAKAN *DECISION TREE C5.0* DENGAN *LOCAL OUTLIER FACTOR (LOF)* UNTUK DETEKSI OUTLIER

Stroke menempati urutan ketiga sebagai penyebab kematian terbanyak di dunia, berada setelah penyakit jantung dan kanker. Selain meningkatkan angka kematian, stroke juga menimbulkan beban kesehatan jangka panjang bagi penderitanya. Penelitian ini bertujuan untuk mengklasifikasikan penyakit stroke menggunakan metode *Decision Tree C5.0*. *Decision Tree* merupakan metode klasifikasi yang memanfaatkan struktur pohon keputusan untuk memetakan hubungan antar variabel. Sebelum melakukan proses klasifikasi data medis, sering ditemui kendala karena adanya data yang ekstrem atau keberadaan *outlier*. Untuk mendeteksi keberadaan *outlier*, penelitian ini menggunakan algoritma LOF (*Local Outlier Factor*) yang mengidentifikasi *outlier* berdasarkan kepadatan lingkungan di sekitarnya. Hasil penelitian menunjukkan penerapan LOF dengan penanganan *outlier* dihapus pada algoritma *Decision Tree C5.0* dengan pembagian data *5-fold cross validation* mampu menghasilkan kinerja klasifikasi akurasi sebesar 77,94%, presisi sebesar 14,69%, sensitivitas 71,91%, spesifisitas sebesar 78,25% dan F1-Score 24,33%. Hasil penelitian ini diharapkan dapat memberikan informasi mengenai risiko stroke berdasarkan hasil pola klasifikasi menggunakan *Decision Tree C5.0* dan deteksi *outlier* dengan LOF, sehingga pola distribusi risiko stroke dapat terlihat dengan jelas.

Kata kunci: *Decision Tree*, Klasifikasi, LOF, *Outlier*, Stroke

ABSTRACT

STROKE DISEASE CLASSIFICATION USING DECISION TREE C5.0 WITH LOCAL OUTLIER FACTOR FOR OUTLIER DETECTION

Stroke ranks as the third leading cause of death worldwide, following heart disease and cancer. In addition to increasing mortality rates, stroke also imposes a substantial long-term health burden on affected individuals. This study aims to classify stroke cases using the *Decision Tree C5.0* method. *Decision Tree* is a classification approach that employs a tree-based structure to represent relationships among variables. Prior to the classification of medical data, challenges often arise due to the presence of extreme values or *outliers*. To address this issue, this study applies the LOF (*Local Outlier Factor*) algorithm, which detects *outliers* by analyzing the local density of data points within their surrounding neighborhoods. The results indicate that the integration of LOF with the *Decision Tree C5.0* algorithm, evaluated using *5-fold cross validation*, produces reliable performance in classifying stroke risk, achieving an accuracy of 77,94%, precision of 14,69%, sensitivity of 71,91%, specificity of 78,25%, and an F1-score of 24,33%. This study is expected to provide meaningful insights into stroke risk through classification results obtained using *Decision Tree C5.0* combined with LOF-based *outlier* detection, thereby enabling clearer identification of stroke risk distribution patterns.

Keywords: Classification, *Decision Tree*, LOF, *Outlier*, Stroke

DAFTAR ISI

HALAMAN JUDUL	i
LEMBAR PERSETUJUAN PEMBIMBING	ii
PENGESAHAN TIM PENGUJI SKRIPSI	iii
HALAMAN PERNYATAAN KEASLIAN	iv
MOTTO	v
PERSEMBAHAN	vi
KATA PENGANTAR	vii
DAFTAR ISI	ix
DAFTAR TABEL	xi
DAFTAR GAMBAR	xii
ABSTRAK	xiii
ABSTRACT	xiv
I PENDAHULUAN	1
1.1. Latar Belakang Masalah	1
1.2. Rumusan Masalah	8
1.3. Tujuan Penelitian	8
1.4. Batasan Masalah	9
II TINJAUAN PUSTAKA	10
2.1. Stroke	10
2.2. Klasifikasi	13
2.3. <i>Preprocessing</i>	14
2.4. <i>One-Hot Encoding</i>	16
2.5. Deteksi Outlier (<i>Local Outlier Factor</i>)	16
2.6. Distribusi Normal (Shapiro-Wilk)	19
2.7. <i>K-fold Cross Validation</i>	20
2.8. <i>Decision Tree</i>	21
2.9. Algoritma C5.0	23

2.10. <i>Confusion Matrix</i>	25
2.11. Integrasi Keilmuan	27
III METODE PENELITIAN	30
3.1. Jenis Penelitian	30
3.2. Sumber Data	30
3.3. Variabel Penelitian	31
3.4. Metode Analisis Data	32
3.5. Tahapan Penelitian	32
IV HASIL DAN PEMBAHASAN	37
4.1. Deskripsi Data	37
4.2. <i>Preprocessing</i>	42
4.2.1. Imputasi Data <i>Missing Value</i>	42
4.2.2. Deteksi <i>outlier</i> (<i>Local Outlier Factor</i>)	45
4.2.3. Membagi Data <i>Training</i> dan <i>Testing</i> (<i>5-Fold Cross Validation</i>)	52
4.3. Klasifikasi Risiko Stroke dengan <i>Decision Tree C5.0</i>	53
4.3.1. Melatih Model Pada Data <i>Training</i>	53
4.3.2. Prediksi Model Pada Data <i>Testing</i>	61
4.4. Evaluasi Hasil	61
4.4.1. Evaluasi Hasil Setiap Fold	61
4.4.2. Perbandingan Evaluasi Hasil Klasifikasi C5.0 Tanpa LOF dan dengan LOF	64
4.5. Integrasi Keilmuan	67
V PENUTUP	69
5.1. Kesimpulan	69
5.2. Saran	69
DAFTAR PUSTAKA	69

DAFTAR TABEL

2.1	<i>Confusion Matrix</i>	26
3.1	Deskripsi Variabel Data	31
3.2	Skenario Uji Coba Fold	34
4.1	Analisis Deskripsi Data Kategorik	37
4.2	Analisis Deskripsi Data Numerik	42
4.3	Analisis Deskripsi Data Numerik	42
4.4	10 Data Awal Sebelum Preprocessing	43
4.5	10 Data Awal Setelah Imputasi	44
4.6	Data One-Hot Encoding	44
4.7	15 Indeks Tetangga Terdekat	46
4.8	Skor <i>Local Outlier Factor</i> (LOF) yang Terindikasi Outlier	49
4.9	Hasil Uji Normalitas Sebelum dan Sesudah Penghapusan Outlier	50
4.10	Hasil Variansi Sebelum dan Sesudah Penghapusan Outlier	50
4.11	Hasil Uji Normalitas Sebelum dan Sesudah Imputasi Outlier	51
4.12	Hasil Variansi Sebelum dan Sesudah Imputasi Outlier	51
4.13	Pembagian Data Training dan Testing per Fold	52
4.14	Threshold Optimal per Fold (1–5)	61
4.15	Probabilitas prediksi stroke untuk 10 sampel	62
4.16	Tabel Confusion Matrix per Fold (1–5)	63
4.17	Perbandingan Hasil Evaluasi Model	65

DAFTAR GAMBAR

2.1	<i>Cross Validation</i>	21
2.2	<i>Struktur Decision Tree</i>	23
3.1	Diagram Penelitian	36
4.1	Distribusi Stroke Berdasarkan Jenis Kelamin	38
4.2	Distribusi Stroke Berdasarkan Riwayat Jantung	39
4.3	Distribusi Stroke Berdasarkan Riwayat Hipertensi	39
4.4	Distribusi Stroke Berdasarkan Status Menikah	40
4.5	Distribusi Stroke Berdasarkan Status Merokok	40
4.6	Distribusi Stroke Berdasarkan Status Pekerjaan	41
4.7	<i>Persebaran Data Stroke</i>	48
4.8	<i>Cabang Awal Pohon Keputusan C5.0</i>	56
4.9	<i>Cabang Kedua Pohon Keputusan C5.0</i>	58
4.10	<i>Pohon Keputusan C5.0</i>	59
4.11	<i>Grafik Perbandingan Hasil Evaluasi</i>	65

DAFTAR PUSTAKA

- Ahadi, G. D., Nur, N., and Ersela, L. (2023). The Simulation Study of Normality Test Using Kolmogorov-Smirnov ., *Eigen Mathematics Journal*, 6(1).
- Altaf, I. and Ahmad, M. (2025). Advances in Density-Based Outlier Detection Algorithms: Exploration of LOF with Experimental Analysis. *Procedia Computer Science*, 258:1833–1843.
- Asgedom, S. G., Gebrewahd, D. T., Abdi, M. G., Wondie, Y. A., Mekonnen, B. A., and Belete, K. T. (2025). Treatment outcome and risk factors of ischemic stroke among adult patients in two selected hospital in Addis Ababa, Ethiopia. *Brain Disorders*, 19(May):100242.
- Aulia, Y., Andriyansyah, A., Suharjito, S., and Nensi, S. W. (2024). Analisis Prediksi Stroke dengan Membandingkan Tiga Metode Klasifikasi Decision Tree, Naïve Bayes, dan Random Forest. *Jurnal Ilmu Komputer dan Informatika*, 3(2):89–98.
- Banafshah Shafa, Hanny Hikmayanti Handayani, Santi Arum Puspita Lestari, and Yana Cahyana (2024). Prediksi Kanker Paru dengan Normalisasi menggunakan Perbandingan Algoritma Random Forest, Decision Tree dan Naïve Bayes. *Decode: Jurnal Pendidikan Teknologi Informasi*, 4(3):1057–1070.
- Biyantoro, A. S. and Prasetyo, B. (2024). Penerapan Decision Tree untuk Klasifikasi Status Kesehatan dengan Perbandingan KNN dan Naive Bayes. *Indonesian Journal of Informatic Research and Software Engineering*, 4(1):47–55.
- Bugis, H. (2022). Metode Naïve Bayes Untuk Memprediksi Penyakit Stroke. *Jurnal SISKOM-KB (Sistem Komputer dan Kecerdasan Buatan)*, 6(1):8–14.
- DocXian (2022). Stroke prediction. Kaggle Notebook. Diakses: akses 23 Agustus 2025, dari kaggle.com/code/docxian/stroke-prediction.

- Elmukhsinur and Kusumarini, N. (2021). The Correlation of Modifiable Risk Factors with Stroke Incidence Hubungan Faktor Risiko yang Dapat dirubah dengan Kejadian Stroke. *Jurnal Proteksi Kesehatan*, 10(2):89–95.
- Fauziningrum, E. and Suryaningsih, E. I. (2021). *Penerapan Data Mining Metode Decision Tree*.
- Feigin, V. L. and Abate, M. D. (2024). Global, regional, and national burden of stroke and its risk factors, 1990–2021: a systematic analysis for the Global Burden of Disease Study 2021. *The Lancet Neurology*, 23(10):973–1003.
- Feigin, V. L., Brainin, M., Norrving, B., Martins, S. O., Pandian, J., Lindsay, P., Grupper, M. F., and Rautalin, I. (2025). World stroke organization: Global stroke fact sheet 2025. *International Journal of Stroke*, 20(2):132–144.
- Fuadah, Y. N., Ubaidullah, I. D., Ibrahim, N., Taliningsing, F. F., Sy, N. K., and Pramuditho, M. A. (2022). Optimasi Convolutional Neural Network dan K-Fold Cross Validation pada Sistem Klasifikasi Glaukoma. *ELKOMIKA: Jurnal Teknik Energi Elektrik, Teknik Telekomunikasi, dan Teknik Elektronika*, 10(3):728.
- Geneva, R. and Usman, S. (2023). Gambaran Karakteristik Individu Dengan Kejadian Stroke Pada Pasien Poliklinik Penyakit Saraf Description of Individual Characteristics With Stroke Events in Patients in Nerve Diseases Polychnic. *Jurnal Kedokteran STM (Sains dan Teknologi Medik)*, 6(2):159–167.
- Guo, Z., Shi, Y., Huang, F., Fan, X., and Huang, J. (2021). Landslide susceptibility zonation method based on C5.0 decision tree and K-means cluster algorithms to improve the efficiency of risk management. *Geoscience Frontiers*, 12(6):101249.
- Haiga, Y., Prima Putri Salman, I., and Wahyuni, S. (2022). Perbedaan Diagnosis Stroke Iskemik dan Stroke Hemoragik dengan Hasil Transcranial Doppler di RSUP Dr. M. Djamil Padang. *Scientific Journal*, 1(5):391–400.

- Hapsari, R. K., Sumarni, T., Hidayatulloh, T., Herianto, F., Elfaladonna, F., Nurfitriana, N. A., Kurniawan, H., Yusuf, K. L., and Widiatama, Y. (2024). *Data Mining*. INDIE PRESS, Bandung.
- Hartono (2016). *MACHINE LEARNING Panduan Praktis dan Sistematis Belajar Machine Learning*. pages 1–23.
- Hermawan, S. M. and Wihardja, H. (2020). Hubungan Karakteristik Individu dan Riwayat Stroke dengan Kekuatan Otot Pada Pasien Stroke. *Dunia Keperawatan: Jurnal Keperawatan dan Kesehatan*, 8(3):406.
- Ishiwata, T. and Seyama, K. (2019). The relationship between smoking and copd. *Systematic Review and Meta-Analysis*, 98(12):1–8.
- Isnaini, M., Afgani, M. W., Haqqi, A., and Azhari, I. (2025). Teknik Analisis Data Uji Normalitas. *J-CEKI: Jurnal Cendekia Ilmiah*, 4(2):1377–1384.
- Jauhari, W. (2018). *Kaidah Fikih: Adh-Dhararu Yuzal*. Rumah Fiqih Publishing, Jakarta Selatan, cetakan pe edition.
- Kamaladevi, M., Sekar, K. R., Venkataraman, V., and Kannan, K. (2019). Outlier Detection in Imbalanced Data Classification. *International Journal of Recent Technology and Engineering (IJRTE)*, 7(6):972–975.
- Karisik, A., Moelgg, K., Buergi, L., Scherer, L., Dejakum, B., Felicetti, S., Boehme, C., Toell, T., Pechlaner, R., Sollereeder, S., Rossi, S., Eller, M. T., Schoenherr, G., Lang, W., Kiechl, S., Knoflach, M., and Mayer-Suess, L. (2025). Impact of dysphagia on early psychosocial consequences after acute ischemic stroke. *Journal of the Neurological Sciences*, 476(June).
- Kuhn, M. and Johnson, K. (2013). *Applied Predictive Modeling*. Springer International Publishing.
- Laili, E. F., Alawi, Z., Rohmah, R., Barata, M. A., Informatika, T., Nahdlatul, U., Sunan, U., Informasi, S., Nahdlatul, U., Sunan, U., Komputer, S., Nahdlatul, U., Sunan, U., and Bojonegoro, K. (2025). Komparasi Algoritma Decision Tree Dan Support Vector Machine (Svm) Dalam. *Jurnal Sistem Informasi dan Informatika (Simika)*, 8(1):67–76.

- Lee, C. S., Cheang, P. Y. S., and Moslehpour, M. (2022). Predictive Analytics in Business Analytics: Decision Tree. *Advances in Decision Sciences*, 26(1):1–29.
- Lembang, S. T., Vonnisye, M. L., and Yuan Andinny, K. (2023). *Statistika Pendidikan*. Penerbit Media Sans Indonesia, Bandung.
- Lin, C.-H., Hsu, K.-C., Johnson, K. R., Luby, M., and Fann, Y. C. (2020). Datasets for Validation of Stroke Clinical Outcomes. *Int J Med Inform*, 13(2):1–23.
- Marcot, B. G. and Hanea, A. M. (2021). What is an optimal value of k in k-fold cross-validation in discrete Bayesian network analysis. *Computational Statistics*, 36(3):1–23.
- Maskuri, M. N., Sukerti, K., and Herdian Bhakti, R. M. (2022). Penerapan Algoritma K-Nearest Neighbor (KNN) untuk Memprediksi Penyakit Stroke Stroke Disease Predict Using KNN Algorithm. *Jurnal Ilmiah Intech : Information Technology Journal of UMUS*, 4(1):130–140.
- Mienye, I. D. and Jere, N. (2024). A Survey of Decision Trees : Concepts , Algorithms , and Applications. *IEEE Access*, 12:86716–86727.
- Mizuno, T., Mizuta, I., Watanabe-Hosomi, A., Mukai, M., and Koizumi, T. (2020). Clinical and Genetic Aspects of CADASIL. *Frontiers in Aging Neuroscience*, 12(May):1–10.
- Musa, D. M., Sakti, D., Shantiony, K. A., Zega, S. K. P., Hamzah, S., Zega, Y. J., and Lubis, B. O. (2024). Penerapan Data Mining Untuk Klasifikasi Data Penjualan Pakan Ternak Terlaris Dengan Algoritma C4.5. *Jurnal Teknologi Informatika dan Komputer*, 10(1):168–182.
- Mutmainah, M. and Yustanti, W. (2024). Studi Komparasi Local Outlier Factor (LOF) dan Isolation Forest (IF) pada Analisis Anomali Kinerja Dosen. *Journal of Informatics and Computer Science (JINACS)*, 6(02):532–540.
- Nugraha, K. W. (2024). *Profil Kesehatan Indonesia Tahun 2023*. Kementerian Kesehatan Republik Indonesia, Jakarta.

- Nurliansyah, F. H. (2023). Penelitian mengenai klasifikasi dan user education koleksi perpustakaan: Sebuah narative literatur review pada Google Scholar. *BIBLIOTIKA : Jurnal Kajian Perpustakaan dan Informasi*, 7(2):252–259.
- Oktavia, L. D. and Hendrati, L. Y. (2023). Implementasi Upaya Pencegahan Stroke Melalui Program GATROK SEDI (Cegah Stroke Sejak Dini) pada Siswa SMP Al-Huda Surabaya. *Media Publikasi Promosi Kesehatan Indonesia (MPPKI)*, 6(11):2217–2223.
- Pakpahan, J. E. S. and Hartati, B. (2022). Hubungan dislipidemia dengan kejadian stroke. *Holistik Jurnal Kesehatan*, 16(6):542–551.
- Pinzon, R. T. (2016). *Awas Stroke*, volume 1.
- Pradana, D., Luthfi Alghifari, M., Farhan Juna, M., and Palaguna, D. (2022). Klasifikasi Penyakit Jantung Menggunakan Metode Artificial Neural Network. *Indonesian Journal of Data and Science*, 3(2):55–60.
- Primadi, O. (2020). *Profil Kesehatan Indonesia Tahun 2019*. Kementerian Kesehatan Republik Indonesia, Jakarta.
- Quinlan, J. R. (1986). Induction of Decision Trees. *Machine Learning*, 1:81–106.
- Raharjo, B. (2021). *Pembelajaran Mesin*.
- Rahayu, S., Masrurah, U., Slamet, Murtiyasa, B., and Sumardi (2024). Analisis Kesulitan dalam Pembelajaran Konsep Mean, Median, dan Modus Pada Peserta Didik Sekolah Dasar. *Jurnal Ilmiah Pendidikan Dasar*, 09(3):106–119.
- Rahayu, S. and Yamasari, Y. (2024). Klasifikasi Penyakit Stroke dengan Metode Support Vector Machine (SVM). 05(3):440–446.
- Rahayu, T. G. (2023). The Analysis of Stroke Risk Factors and Stroke Types. *Faletehan Health Journal*, 10(01):48–53.
- Rohmah, S. (2025). Kajian pengeluaran biaya kesehatan pada pasien stroke. *Jurnal Pharma Saintika*, 8:1–18.

- Samariya, D., Ma, J., Aryal, S., and Zhao, X. (2023). Detection and explanation of anomalies in healthcare data. *Health Information Science and Systems*, 11(1):1–23.
- Santini, A., Man, A., and Voidāzan, S. (2020). Accuracy of Diagnostic Tests. *The Journal of Critical Care Medicine*, 7(3):241–248.
- Saputri, A. F., Hoyyi, A., and Sugito (2017). Prediksi jumlah penumpang kereta api menggunakan model variasi kalender dengan deteksi outlier. *Jurnal Gaussian*, 6(3):281–289.
- Saputro, I. W. and Sari, B. W. (2019). Uji Performa Algoritma Naïve Bayes untuk Prediksi Masa Studi Mahasiswa. *Citec Journal*, 6(1):1–11.
- Soesana, A., Subakti, H., Karwanto, Fitri, A., Kuswandi, S., Sastri, L., Falani, I., Aswan, N., Hasibuan, F. A., and Lestari, H. (2023). *Metodologi Penelitian Kuantitatif*. Yayasan Kita Menulis, cetakan 1 edition.
- Sofyan, F. M. A., Riyandoro, A. P., Maulana, D. F., and Jaman, J. H. (2023). Penerapan Data Mining dengan Algoritma C5.0 Untuk Prediksi Penyakit Stroke. *J-SISKO TECH (Jurnal Teknologi Sistem Informasi dan Sistem Komputer TGD)*, 6(2):619.
- Sulistijo, E. (2025). *Profil Kesehatan Indonesia Tahun 2024*. Kementerian Kesehatan Republik Indoensia, Jakarta.
- Sultradewi Kesuma, N. M. T., Krismashogi Dharmawan, D., and Fatmawati, H. (2019). Gambaran faktor risiko dan tingkat risiko stroke iskemik berdasarkan stroke risk scorecard di RSUD Klungkung. *Intisari Sains Medis*, 10(3):720–729.
- Syukriyah, S. R. Z. M. D. and Al Fauzi, A. (2024). Hypertension as a risk factor in stroke : An overview. *World Journal of Advanced Research and Reviews*, 21(01):2370–2372.
- Tabassum, M., Mahmood, S., Bukhari, A., Alshemaimri, B., and Daud, A. (2024). Anomaly - based threat detection in smart health using machine learning. *BMC Medical Informatics and Decision Making*, 4(1):24–347.
- Tukey, J. W. (1977). *Exploratory Data Analysis*.

- Utomo, T. Y. (2024). *Buku Ajar Stroke*.
- Vences, M. A., Barboza, M. A., Carbonera, L. A., and Failoc-Rojas (2025). Case-fatality of acute ischemic stroke in stroke units of Latin American hospitals. *Journal of Stroke and Cerebrovascular Diseases*, 34(9).
- Wahono, R. S. (2023). *Data Mining*, volume 2.
- World Health Organization (2025). Hypertension. Online. Diakses: 10 November 2025, dari who.int/news-room/fact-sheets/detail/hypertension.
- Wororomi, J. K., Reba, F., Mandowen, S. A., Sroyer, A. M., Manurung, H. E., Koibur, M. E., Pujiarini, E. H., and Marwan (2024). *Data Mining Memahami Pola Di Balik Angka*.
- Xu, J., Liu, X., Gu, Z., and Xiao, G. (2025). A rapid cross-validation computing for three-way decisions in imbalanced data. *Information Sciences*, 707(February):122016.
- Yousufuddin, M. and Young, N. (2019). Aging and ischemic stroke Mohammed. *Aging*, 11(9):2542–2544.
- Zou, D., Xiang, Y., Zhou, T., Peng, Q., Dai, W., Hong, Z., Shi, Y., Wang, S., Yin, J., and Quan, H. (2023). Outlier detection and data filling based on KNN and LOF for power transformer operation data classification. *Energy Reports*, 9:698–711.