

**ANALISIS PERFORMA *LIGHT GRADIENT BOOSTING MACHINE*
(LGBM) DAN KONTRIBUSI *FEATURE IMPORTANCE* PADA
KLASIFIKASI HEPATITIS C**

SKRIPSI



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

Disusun Oleh
THALIA ANINDYA ARDINE
09010222017

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

2026

PERNYATAAN KEASLIAN

Saya yang bertanda tangan di bawah ini,

Nama : Thalia Anindya Ardine
NIM : 09010222017
Program Studi : Matematika
Angkatan : 2022

Menyatakan bahwa saya tidak melakukan plagiat dalam penulisan skripsi saya yang berjudul ” Analisis Performa *Light Gradient Boosting Machine* (LGBM) dan Kontribusi *Feature Importance* Pada Klasifikasi Hepatitis C ”. 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, 17 Juni 2026

Yang menyatakan,



Thalia Anindya Ardine
NIM. 09010222017

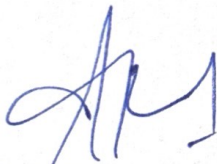
LEMBAR PERSETUJUAN PEMBIMBING

Skripsi oleh

Nama : Thalia Anindya Ardine
NIM : 09010222017
Judul proposal skripsi : Analisis Performa *Light Gradient Boosting Machine* (LGBM) pada Klasifikasi Penyakit Hepatitis C dan Analisis *Feature Importance*

telah diperiksa dan disetujui untuk diujikan.

Pembimbing I



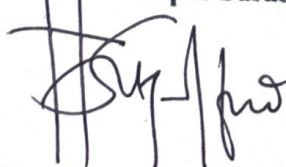
Aris Fanani, M. Kom
NIP. 198701272014031002

Pembimbing II



Dian Yuliati, M. Si
NIP. 198707142020122015

Mengetahui,
Ketua Program Studi Matematika
UIN Surjan Ampel Surabaya



Dr. Yuniar Farida, M.T
NIP. 197905272014032002

PENGESAHAN TIM PENGUJI SKRIPSI

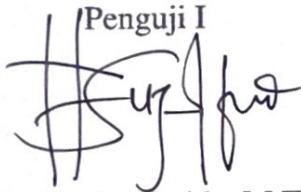
Skripsi oleh

Nama : Thalia Anindya Ardine
NIM : 09010222017
Judul Skripsi : Analisis Performa *Light Gradient Boosting Machine*
(LGBM) dan Kontribusi *Feature Importance* Pada
Klasifikasi Hepatitis C

Telah dipertahankan di depan Tim Penguji
pada tanggal 17 Juni 2026

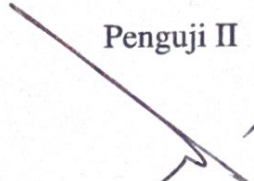
Mengesahkan,
Tim Penguji

Penguji I



Dr. Yuniar Farida, M.T
NIP. 1979052702014032002

Penguji II



Dr. Moh Hafiyusholeh, M.Si., M.PMat
NIP. 198002042014031001

Penguji III



Aris Fanani, M. Kom
NIP. 198701272014031002

Penguji IV



Dian Yuliati, M. Si
NIP. 198707142020122015

Mengetahui,

Dekan Fakultas Sains dan Teknologi
UIN Sunan Ampel Surabaya



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

LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI
KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

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

Nama : THALIA ANINDYA ARDINE
NIM : 09010222017
Fakultas/Jurusan : FAKULTAS SAINS DAN TEKNOLOGI / MATEMATIKA
E-mail address : thaliaardine10@gmail.com

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

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

ANALISIS PERFORMA LIGHT GRADIENT BOOSTING MACHINE
(LGBM) DAN KONTRIBUSI FEATURE IMPORTANCE PADA
KLASIFIKASI HEPATITIS C

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, 17 Juni 2026

Penulis



(THALIA ANINDYA-A)
nama terang dan tanda tangan

ABSTRAK

Analisis Performa *Light Gradient Boosting Machine* (LGBM) dan Kontribusi *Feature Importance* Pada Klasifikasi Hepatitis C

Hepatitis C merupakan penyakit infeksi hati yang disebabkan oleh virus Hepatitis C (HCV) yang berpotensi berkembang menjadi kondisi kronis, sirosis, hingga kanker hati apabila tidak ditangani secara tepat. Oleh karena itu, klasifikasi kondisi pasien Hepatitis C perlu dilakukan secara akurat sebagai upaya deteksi dini untuk mendukung pengambilan keputusan medis. Berbagai indikator demografis dan hasil pemeriksaan laboratorium, seperti usia, jenis kelamin, albumin, bilirubin, enzim hati, dan protein total, menjadi faktor penting dalam menentukan kondisi pasien. Namun, banyaknya variabel yang digunakan menyebabkan proses analisis secara manual menjadi kurang efisien dan berpotensi menurunkan akurasi klasifikasi. Penelitian ini bertujuan membangun model klasifikasi kondisi pasien Hepatitis C menggunakan metode *Light Gradient Boosting Machine* (LGBM) serta menganalisis fitur yang paling berpengaruh melalui *feature importance*. Data yang digunakan merupakan dataset *Hepatitis C Prediction* yang terdiri atas 615 data pasien dengan lima kelas, yaitu *Blood Donor*, *Suspect Blood Donor*, *Hepatitis*, *Fibrosis*, dan *Cirrhosis*. Tahapan penelitian meliputi *preprocessing* data, penanganan *missing value* menggunakan median, pengelompokan data menggunakan *Stratified K-Fold Cross Validation* dengan nilai $k = 5$, pemodelan menggunakan LGBM, serta evaluasi menggunakan *confusion matrix* untuk memperoleh nilai *accuracy*, *weighted precision*, *weighted recall*, dan *weighted F1-score*. Hasil penelitian menunjukkan bahwa model LGBM mampu menghasilkan *accuracy* sebesar 93,98%, *weighted precision* sebesar 93,39%, *weighted recall* sebesar 93,98%, dan *weighted F1-score* sebesar 93,68%. Selain itu, hasil analisis *feature importance* menunjukkan bahwa fitur ALT (*Alanine Aminotransferase*) merupakan variabel yang paling berpengaruh dalam proses klasifikasi dengan nilai *feature importance* sebesar 1482. Hasil penelitian menunjukkan bahwa metode LGBM mampu memberikan performa klasifikasi yang sangat baik serta dapat digunakan sebagai pendukung dalam identifikasi kondisi klinis pasien Hepatitis C.

Kata kunci: *Feature Importance*, Hepatitis C, Klasifikasi, LGBM

ABSTRACT

Performance Analysis of *Light Gradient Boosting Machine* (LGBM) and *Feature Importance* Contribution in Hepatitis C Classification

Hepatitis C is a liver infectious disease caused by the Hepatitis C Virus (HCV) that can progress into chronic conditions, cirrhosis, and liver cancer if not treated properly. Therefore, accurate classification of Hepatitis C patient conditions is essential as an early detection effort to support medical decision-making. Various demographic indicators and laboratory examination results, such as age, sex, albumin, bilirubin, liver enzymes, and total protein, are important factors in determining patient conditions. However, the large number of variables involved makes manual analysis inefficient and may reduce classification accuracy. This study aims to develop a classification model for Hepatitis C patient conditions using the Light Gradient Boosting Machine (LGBM) method and to analyze the most influential features through *feature importance*. The dataset used in this study is the *Hepatitis C Prediction* dataset, consisting of 615 patient records with five classes, namely *Blood Donor*, *Suspect Blood Donor*, *Hepatitis*, *Fibrosis*, and *Cirrhosis*. The research stages include data preprocessing, handling missing values using the median method, data partitioning using *Stratified K-Fold Cross Validation* with $k = 5$, model development using LGBM, and performance evaluation using a *confusion matrix* to obtain *accuracy*, *weighted precision*, *weighted recall*, and *weighted F1-score*. The results show that the LGBM model achieved an *accuracy* of 93.98%, a *weighted precision* of 93.39%, a *weighted recall* of 93.98%, and a *weighted F1-score* of 93.68%. Furthermore, the *feature importance* analysis revealed that ALT (*Alanine Aminotransferase*) was the most influential feature in the classification process, with a feature importance value of 1482. These findings indicate that the LGBM method provides excellent classification performance and can be utilized as a supporting tool for identifying the clinical conditions of Hepatitis C patients.

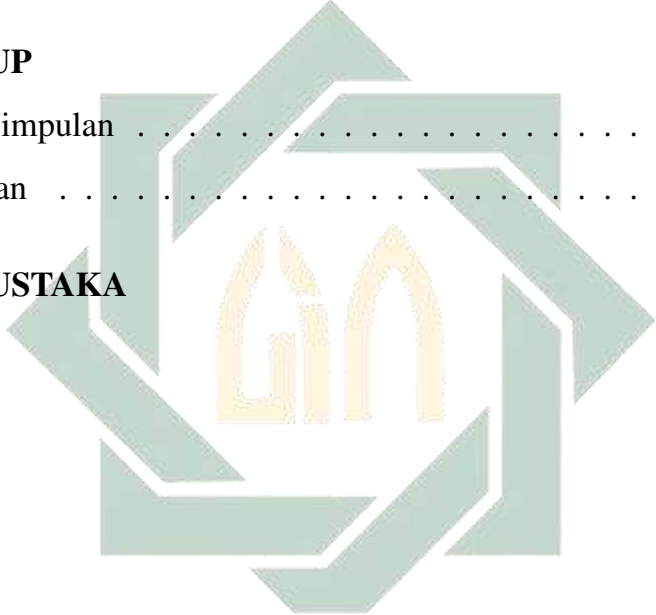
Keywords: Classification, Feature Importance, Hepatitis C, LGBM

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	xii
DAFTAR GAMBAR	xiii
ABSTRAK	xiv
ABSTRACT	xv
I PENDAHULUAN	1
1.1. Latar Belakang Masalah	1
1.2. Rumusan Masalah	9
1.3. Tujuan Penelitian	9
1.4. Batasan Masalah	9
1.5. Manfaat Teoritis	10
1.6. Manfaat Praktis	11

1.7. Sistematika Penulisan	11
II TINJAUAN PUSTAKA	13
2.1. Hepatitis	13
2.1.1. Hepatitis C	14
2.2. Faktor-Faktor yang Memengaruhi Kondisi Pasien Hepatitis C	15
2.3. Klasifikasi Tingkat Kondisi Penyakit Hepatitis C	18
2.4. <i>Stratified K-Fold Cross Validation</i> (SKCV)	20
2.5. <i>Light Gradient Boosting Machine</i> (LGBM)	21
2.6. <i>Confusion Matrix</i>	26
2.7. <i>Feature Importance</i> (FI)	30
2.8. Intergrasi Keilmuan	31
III METODE PENELITIAN	34
3.1. Jenis Penelitian	34
3.2. Sumber Data	34
3.3. Variabel Penelitian	35
3.3.1. Variabel Independen (X)	35
3.3.2. Variabel Dependen (Y)	36
3.4. Metode Analisis Data	37
3.5. Tahapan Penelitian	38
3.6. Skenario Uji coba	42
IV HASIL DAN PEMBAHASAN	44
4.1. Deskripsi Data	44
4.2. <i>Preprocessing</i> Data	45
4.2.1. Identifikasi dan Penanganan <i>Missing Value</i>	45
4.2.2. Analisis Ketidakseimbangan Kelas	46
4.3. <i>Stratified K-Fold Cross Validation</i> (SKCV)	47
4.4. <i>Light Gradient Boosting Machine</i> (LGBM)	49

4.4.1. Proses <i>Training</i>	49
4.4.2. Proses <i>Testing</i>	72
4.4.3. <i>Confusion Matrix</i>	73
4.4.4. Uji Coba	78
4.5. <i>Feature Importance (FI)</i>	80
4.6. Pembahasan	83
4.7. Integrasi Keilmuan	85
V PENUTUP	87
5.1. Kesimpulan	87
5.2. Saran	88
DAFTAR PUSTAKA	88



UIN SUNAN AMPEL
S U R A B A Y A

DAFTAR TABEL

3.1	Fitur Variabel Independen (X) pada Dataset Hepatitis C . . .	36
3.2	Variabel Dependen (Y) pada Dataset Hepatitis C	37
3.3	Skenario Uji Coba	43
4.1	Konversi Variabel Kategori ke Bentuk Numerik	44
4.2	Identifikasi dan Penanganan <i>Missing Value</i>	45
4.3	Distribusi Kelas pada Variabel Dependen	47
4.4	Pembagian Target pada Setiap <i>Fold</i>	48
4.5	Data Sampel Proses <i>Training</i>	50
4.6	Nilai Bobot Setiap Kelas	52
4.7	Perbandingan Nilai Information Gain pada Pohon Pertama LightGBM	65
4.8	Perbandingan Nilai Information Gain pada Node 2	68
4.9	Nilai <i>Leaf</i> Pohon Pertama untuk Setiap Kelas	70
4.10	Nilai <i>Leaf</i> untuk Setiap Iterasi dan Kelas	71
4.11	Data Proses Testing	72
4.12	Nilai <i>Leaf</i> Proses Testing	72
4.13	Nilai Logit Data Testing Pertama	73
4.14	Hasil Perhitungan Probabilitas Data Testing Pertama	73
4.15	Hasil Evaluasi Model pada Setiap Fold	78
4.16	Hasil Uji Coba Parameter LightGBM	79
4.17	Kombinasi Parameter Terbaik LightGBM	79
4.18	Hasil Feature Importance Model LGBM	81
4.19	Perbandingan Hasil Penelitian	85

DAFTAR GAMBAR

2.1	Stratified K-Fold Cross Validation	20
2.2	<i>leaf wise growth</i>	22
2.3	Grafik fungsi <i>softmax</i>	26
2.4	Evaluasi <i>Confusion Matrix</i>	27
3.1	Flowchart Penelitian LightGBM	41
4.1	Visualisasi Variabel AST dengan Pemisah Data 31,050 . . .	61
4.2	Distribusi Jumlah Data pada Cabang Kiri dan Cabang Kanan Node 1 Berdasarkan Pemisah $AST \leq 31,050$	61
4.3	Visualisasi Variabel PROT dengan Pemisah Data 49,400 . .	66
4.4	Distribusi Data pada Cabang Kiri dan Cabang Kanan Node 2 Berdasarkan Pemisah $PROT \leq 49,400$	66
4.5	Evaluasi <i>Confusion Matrix</i>	74

UIN SUNAN AMPEL
SURABAYA

DAFTAR PUSTAKA

- Aditia, A. (2021). Covid-19 : Epidemiologi, virologi, penularan, gejala klinis, diagnosa, tatalaksana, faktor risiko dan pencegahan. *Jurnal Penelitian Perawat Profesional*, 3.
- Ainurrohmah and Wiyanti, D. T. (2023). Analisis performa algoritma decision tree, naïve bayes, k-nearest neighbor untuk klasifikasi zona daerah risiko covid-19 di indonesia. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, 10(1):115–122.
- Amalia, M., Hidayati, P. H., Yanti, A. K. E., Vitayani, S., and Gayatri, S. W. (2023). Karakteristik pasien sirosis hepatitis. *UMI Medical Journal*, 8(1):53–61.
- Amaliah, S., Nusrang, M., and Aswi, A. (2022). Penerapan metode random forest untuk klasifikasi varian minuman kopi di kedai kopi konijiwa bantaeng. *VARIANSI: Journal of Statistics and Its Application on Teaching and Research*, 4(3).
- Amani, D., Kristinawati, E., Jiwantoro, Y. A., and Inayati, N. (2024). Description of the enzyme gamma glutamyl transferase in the patient hepatitis. *Jurnal Analis Medika Biosains (JAMBS)*, 11(2):85–89.
- Aminuddin, M. F., Mahdasari, and Marsudi, L. O. (2023). Gambaran total protein dan albumin pada pasien di rsud abdul wahab sjahrane samarinda. *Jurnal Teknologi Laboratorium Medik Borneo*, 3(1):41–47.
- Andayani, D. E. and Mariani, C. (2024). Efektivitas pemberian probiotik dalam memperbaiki fungsi hati pada pasien dengan fatty liver non alkoholik. *Jurnal Sehat Indonesia (JUSINDO)*, 6(1):394–412.
- Bachtiar, R. R., Parewangi, M. L., Akil, F., Daud, N. A., Kusuma, S. H., and Rifai, A. (2022). Perubahan paradigma pada strategi

penanganan karsinoma hepatoseluler. *UMI Medical Journal*, 7(1):59–71.

Begum, F., Mazumder, M. W., Nahid, K. L., Jesmin, T., and Musabbir, N. (2023). Hepatitis c virus infection in children and adolescents: A management update. *Egyptian Pediatric Association Gazette*, 71(1):17.

Budianto, Y. and Akbar, M. A. (2022). Kenaikan kadar kolesterol ditinjau dari konsumsi gorengan. *Jurnal Kesehatan Abdurahman*, 11(2):8–13.

Dalimunthe, A. Y., Siregar, I. R., Rahmah, L., and Lubis, N. A. (2025). Gambaran hasil pemeriksaan hepatitis c virus (hcv) pada pendonor darah di udd pmi kota medan. *Jurnal Ilmu Kedokteran dan Kesehatan*, 12(11):2495–2501.

Damayanti, A. and Testiana, G. (2023). Penerapan data mining untuk prediksi penyakit hepatitis c menggunakan algoritma naïve bayes. *Jurnal Manajemen Informatika Jayakarta*, 3(2):177–186.

Feliansyah, A. W. and Purwanto, E. (2024). Analisis faktor yang berhubungan dengan penyakit hepatitis di indonesia. *Holistik Jurnal Kesehatan*, 18(9):1131–1138.

Ferdiansyah, A., Umbara, F. R., and Kasyidi, F. (2025). Prediksi penyakit diabetes menggunakan teknik imputasi missforest dan klasifikasi lightgbm. *MIND (Multimedia Artificial Intelligent Networking Database) Journal*, 10(2):221–234.

Hafid, H. (2023). Penerapan k-fold cross validation untuk menganalisis kinerja algoritma k-nearest neighbor pada data kasus covid-19 di indonesia. *Journal of Mathematics Computations and Statistics*, 6(2).

Hutagaol, N. M. and Tarigan, C. (2024). Pendekatan diagnosis sirosis hepatis: Laporan kasus.

- Indreswara, D. R., Maimunah, U., and Retnowati, E. (2022). Serum afp (alpha fetoprotein) levels profile of hepatocellular carcinoma patients in dr. soetomo general academic hospital, surabaya, indonesia. *Majalah Biomorfologi (Biomorphology Journal)*, 32(1):6–12.
- Iparraguirre-Villanueva, O., Flores-Castañeda, R. O., Chero-Valdivieso, H., and Sierra-Liñan, F. (2024). Predicting hepatitis c infection with machine learning algorithms: a prospective study. *IAES International Journal of Artificial Intelligence*, 13(4):4403–4413.
- Jakaria, F., Nikma, Lewa, I. B. H., and Husen, S. H. (2024). Analisis kadar alp (alkaline phospatase) pada ibu hamil trimester iii di wilayah kerja puskesmas gambesi kota ternate selatan. *Jurnal Analis Kesehatan Kendari*, 7(1):30–36.
- Kamal, M. F. F. and Darmadi (2024). Hiperbilirubinemia. *GALENICAL: Jurnal Kedokteran dan Kesehatan Mahasiswa Malikussaleh*, 3(6):90–101.
- Kementerian Kesehatan Republik Indonesia (2024). Angka hepatitis b dan c di indonesia turun. <https://kemkes.go.id/id/angka-hepatitis-b-dan-c-di-indonesia-turun>. Diakses pada 2 April 2026.
- Kurniadi, F. I. and Larasati, P. D. (2022). Light gradient boosting machine untuk deteksi penyakit stroke. *Jurnal Sistem Komputer dan Kecerdasan Buatan*, 6(1):67–72.
- Lestari, A. and Tyas, T. A. W. (2023). Profil pemeriksaan hematologi dan fungsi hati pada lansia dengan sirosis hepatis. *Muhammadiyah Journal of Geriatric*, 4(1):65–72.
- Mardhiyatillah, N., Akbar, T. I. S., and Utariningsih, W. (2024). Gambaran hasil skrining infeksi menular lewat transfusi darah (imltd) pendonor di unit transfusi darah (utd) pmi kabupaten aceh utara periode 2017-2021. *Galenical: Jurnal Kedokteran dan Kesehatan Mahasiswa Malikussaleh*, 3(2).

- Mayaserli, D. P., Rosita, B., and Remadhani, E. (2022). Pengaruh waktu paparan pestisida organofosfat terhadap kadar kolinesterase dalam darah dengan metode komperator. *Jurnal Kesehatan Perintis*, 9(1):31–38.
- Microsoft LightGBM Developers (2023). Lightgbm documentation: Parameters. Accessed: 2026-01-10.
- Multazam, S., Cholissodin, I., and Adinugroho, S. (2020). Implementasi metode extreme learning machine pada klasifikasi jenis penyakit hepatitis berdasarkan faktor gejala. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 4(3):789–797.
- Nurwananda, S. S. and Sulaiman, R. (2022). Aplikasi himpunan fuzzy intuisisionistik dalam diagnosa penyakit hepatitis menggunakan extended hausdorff distance. *MATHunesa: Jurnal Ilmiah Matematika*, 10(1):41–49.
- Paramitha, A. F., Arimbi, Y. D., Riyanto, S., Apriani, N. F., and Siagian, A. H. A. M. (2024). Klasifikasi transaksi keuangan mencurigakan dengan metode light gradient boosting machine (lgbm) berdasarkan indikator social network analysis (sna). *SISTEMASI: Jurnal Sistem Informasi*, 12(2):572–582.
- Pratama, N. A. and Utomo, D. W. (2024). Komparasi metode xgboost dan lightgbm untuk deteksi diabetes mellitus menggunakan smote-tomek. *Jurnal Informatika Sunan Kalijaga (JISKa)*, 9(2):101–110.
- Putra, I. G. N. P. S., Saputra, I. P. G. H., Eryana, I. M., and Wijaya, I. G. H. A. (2023). Penanda biologis kerusakan ginjal pada penderita sirosis hepatis. *Intisari Sains Medis*, 14(3):1032–1039.
- Putri, T. (2024). Tantangan dan solusi terkini dalam penanganan penyakit menular. *SURABAYA BIOMEDICAL JOURNAL*, 4.
- Rizky, P. S., Hirzi, R. H., and Hidayaturrohman, U. (2022). Perbandingan metode lightgbm dan xgboost dalam menangani data

- dengan kelas tidak seimbang. *J Statistika: Jurnal Ilmiah Teori dan Aplikasi Statistika*, 15(2):228–236.
- Saputra, D. D. and Munir, M. S. (2025). Klasifikasi kualitas udara menggunakan metode light gradient boosting machine (lightgbm). *Prosiding Seminar Nasional Teknologi Informasi dan Komunikasi (SNATIF)*, 5(1):1–8.
- Saraswati, A., Larasati, T. A., and Suharmanto (2022). Faktor risiko terjadinya penyakit hepatitis c. *Jurnal Penelitian Perawat Profesional*, 4(2):649–654.
- Selfitriana, Djuwita, R., Hikmah, K., Hermanto, J., and Nurdiansyah, I. (2023). Analisis deskripsi faktor risiko infeksi hepatitis c. *Media Publikasi Promosi Kesehatan Indonesia*, 6(6):1103–1109.
- Sitorus, D. S. and Desiani, A. (2024). Hepatitis disease diagnosis expert system using certainty factor method. *JAMASTIKA*, 3(1):212–219.
- Susetyoko, R., Yuwono, W., Purwantini, E., and Ramadijanti, N. (2022). Perbandingan metode random forest, regresi logistik, naïve bayes, dan multilayer perceptron pada klasifikasi uang kuliah tunggal (ukt). *Jurnal Infomedia: Teknik Informatika, Multimedia & Jaringan*, 7(1):8–16.
- Sutrisno, S. and Jupron, J. (2024). Analisa klasifikasi penyakit diabetes dengan algoritma neural network. *bit-Tech*, 6(3):303–310.
- Syafei, R. M. and Efrilianda, D. A. (2023). Machine learning model using extreme gradient boosting (xgboost) feature importance and light gradient boosting machine (lightgbm) to improve accurate prediction of bankruptcy. *Recursive Journal of Informatics*, 1(2):64–72.
- Wantoro, A., Zulkifli, Yulia, A. F., Ayu, D. Y., and Mustofa, S. (2025). Evaluasi kinerja algoritma machine learning (ml) menggunakan seleksi fitur pada klasifikasi diabetes. *JIP (Jurnal Informatika Polinema)*, 11(3):311–316.

- Widodo, S., Brawijaya, H., and Samudi, S. (2022). Stratified k-fold cross validation optimization on machine learning for prediction. *Sinkron: Jurnal dan Penelitian Teknik Informatika*, 6(4):2407–2414.
- Wijiyanto, Pradana, A. I., Sopingi, and Atina, V. (2024). Teknik k-fold cross validation untuk mengevaluasi kinerja mahasiswa. *Jurnal Algoritma*, 21(1):239–248.
- Yamaoka, N. A. S. P. U. M. R. H. A. I. F. M. M. Y. (2022). Quality of life of chronic hepatitis b patients consuming nucleoside analog: A case-control clinical study in indonesia. *Biomolecular and Health Science Journal*, 5:6–10.
- Yusfila, F. Q., Khotimah, B. K., Anamisa, D. R., and Ni'mah, A. T. (2025). Klasifikasi penyakit hepatitis c dengan menggunakan k-nearest neighbor. *Sains Data: Jurnal Studi Matematika dan Teknologi*, 3(1):33–42.
- Zignego, A. L., Monti, M., and Gragnani, L. (2024). The epidemiology of chronic hepatitis c: Where we are now. *Diseases*, 4(2):13.

UIN SUNAN AMPEL
S U R A B A Y A