

**OPTIMALISASI AKURASI DETEKSI PENYAKIT LIVER DENGAN
ALGORITMA RANDOM FOREST PADA DATASET ILPD**

SKRIPSI

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



OLEH :

CINTA AZZARIA
NPM : 2113030060

FAKULTAS TEKNIK DAN ILMU KOMPUTER (FTIK)
UNIVERSITAS NUSANTARA PERSATUAN GURU REPUBLIK INDONESIA
UN PGRI KEDIRI
2025

Skripsi oleh:

CINTA AZZARIA
NPM : 2113030060

Judul:

**OPTIMALISASI AKURASI DETEKSI PENYAKIT LIVER DENGAN
ALGORITMA RANDOM FOREST PADA DATASET ILPD**

Telah disetujui untuk diajukan Kepada
Panitia Ujian/Sidang Skripsi Program Studi Sistem Informasi
FTIK UN PGRI Kediri

Tanggal : 4 Juli 2025

Pembimbing I



Erna Daniati, M.Kom
NIDN. 0723058501

Pembimbing II



Aidina Ristvawan, M.Kom
NIDN. 0721018801

Skripsi oleh :

CINTA AZZARIA
NPM : 2113030060

Judul :

**OPTIMALISASI AKURASI DETEKSI PENYAKIT LIVER DENGAN
ALGORITMA RANDOM FOREST PADA DATASET ILPD**

Telah dipertahankan di depan Panitia Ujian/Sidang Skripsi
Program Studi Sistem Informasi FTIK UN PGRI Kediri

Tanggal: 8 Juli 2025

Dan Dinyatakan telah Memenuhi Persyaratan

Panitia Penguji :

1. Ketua Penguji : Erna Daniati, M.Kom.

[.....]

2. Penguji 1 : Arie Nugroho, S.Kom, M.M, M.Kom

[.....]

3. Penguji 2 : Aidina Ristyawan, M.Kom.

[.....]



Mengetahui,
Dekan Fakultas Teknik dan Ilmu Komputer

Dr. Sulistiono, M.Si.
NIDN. 0007076801

PERNYATAAN

Yang bertanda tangan di bawah ini saya,

Nama : Cinta Azzaria
Jenis Kelamin : Perempuan
Tempat/tgl. lahir : Kediri, 29 Juni 2003
NPM : 2113030060
Fak/Prodi. : FTIK/ S1-Sistem Informasi

menyatakan dengan sebenarnya, 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, 4 Juli 2025
Yang Menyatakan



CINTA AZZARIA
NPM: 2113030060

HALAMAN MOTTO

“Kegagalan hanya terjadi jika kita menyerah.”

— *B.J. Habibie*

“Masih banyak makanan enak yg harus dicoba, serial yg harus ditamatkan, dan tempat indah yg harus dikunjungi. Mungkin bagi kita, hidup menyakitkan. Tapi bagi orang-orang di sekeliling kita, akan lebih menyakitkan kalau kita enggak hidup. Seberat apa pun keadaan, jangan menyerah ya..”

— *Fiersa Besari*

“Ketika semua orang menuntut bukti sebelum percaya, Ayah dan Ibu adalah dua manusia yang percaya bahkan sebelum aku tahu harus berbuat apa. Doa mereka menembus langit tanpa perlu diumbar, dan restu mereka menjadi penuntun setia dalam setiap keputusan dan perjalanan. Sesungguhnya, keberhasilan ini adalah milik mereka yang tak pernah berhenti berjuang dalam diam.”

— *Penulis*

ABSTARK

Cinta Azzaria: Optimalisasi Akurasi Deteksi Penyakit Liver dengan Algoritma Random Forest pada Dataset ILPD, Skripsi, Sistem Informasi, Fakultas Teknik dan Ilmu Komputer UN PGRI Kediri, 2025.

Kata Kunci: Penyakit Liver, *Random Forest*, *Data Mining*, *Hyperparameter Tuning*, SMOTE.

Penyakit liver menjadi salah satu penyebab kematian utama secara global. Kompleksitas gejala dan keterlambatan diagnosis membuat penanganannya sering kali kurang optimal. Oleh karena itu, dibutuhkan sistem deteksi dini yang akurat untuk membantu proses diagnosis. Pemanfaatan *data mining* dan algoritma *machine learning* seperti *Random Forest*, dinilai mampu mengatasi tantangan tersebut karena kemampuannya dalam mengelola data kompleks dan menghasilkan klasifikasi yang stabil.

Penelitian ini bertujuan untuk mengimplementasikan algoritma *Random Forest* dalam mendeteksi penyakit liver dan mengoptimalkan akurasinya melalui teknik *hyperparameter tuning* menggunakan *Grid Search CV*. Metode penelitian yang digunakan adalah CRISP-DM, yang meliputi tahapan pemahaman bisnis, data, persiapan, pemodelan, evaluasi, dan deployment. Data yang digunakan berasal dari *Indian Liver Patient Dataset* (ILPD) sebanyak 570 data pasien, dengan penanganan ketidakseimbangan kelas menggunakan teknik SMOTE.

Hasil penelitian menunjukkan bahwa model *Random Forest* tanpa tuning menghasilkan akurasi sebesar 74%, namun masih menunjukkan ketimpangan dalam mengenali kelas minoritas. Setelah dilakukan *hyperparameter tuning*, akurasi meningkat menjadi 75% dan performa model terhadap kelas minoritas (pasien normal) juga membaik, ditunjukkan oleh peningkatan nilai *precision* dan *recall*.

Kesimpulannya, algoritma *Random Forest* terbukti efektif dalam mendeteksi penyakit liver. Penerapan *hyperparameter tuning* dan teknik penyeimbangan data seperti SMOTE mampu meningkatkan akurasi. Penelitian ini berkontribusi dalam pengembangan sistem pendukung keputusan di bidang medis, khususnya untuk diagnosis penyakit liver secara cepat dan akurat.

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 Akurasi Deteksi Penyakit Liver dengan Algoritma Random Forest pada Dataset ILPD ” 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 Fakultas Teknik dan Ilmu Komputer UN PGRI Kediri yang selalu memberikan dorongan motivasi kepada mahasiswa.
4. **Erna Daniati, M.Kom.** selaku Pembimbing skripsi I yang dengan sabar membimbing dan memotivasi dalam menyelesaikan skripsi ini.
5. **Aidina Ristyawan, M.Kom.** selaku Pembimbing skripsi II yang dengan sabar membimbing dan memotivasi dalam menyelesaikan skripsi ini.
6. Kedua orang tua dan adik-adik saya yang telah mendukung dalam do'a, dana dan motivasi dalam penyelesaian skripsi ini.
7. Teman-teman yang selalu membantu memberi semangat dalam penyelesaian skripsi ini.
8. Ucapan terima kasih juga disampaikan kepada pihak-pihak lain yang tidak dapat disebutkan satu persatu, yang telah banyak membantu menyelesaikan Skripsi ini.

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, 4 Juli 2025

CINTA AZZARIA
NPM : 2113030060

DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PENGESAHAN.....	ii
PERNYATAAN.....	iii
HALAMAN MOTTO	v
ABSTARK	vi
KATA PENGANTAR.....	vii
DAFTAR ISI	ix
DAFTAR TABEL.....	xii
DAFTAR GAMBAR	xiii
BAB I PENDAHULUAN.....	1
A. Latar Belakang Masalah.....	1
B. Batasan Masalah.....	4
C. Rumusan Masalah	5
D. Tujuan Penelitian.....	5
E. Manfaat Penelitian	6
BAB II KAJIAN PUSTAKA	Error! Bookmark not defined.
A. Kajian Teori.....	Error! Bookmark not defined.
1. Hati.....	Error! Bookmark not defined.
2. Penyakit Hati.....	Error! Bookmark not defined.
3. Optimalisasi	Error! Bookmark not defined.
4. <i>Data Mining</i>	Error! Bookmark not defined.
5. Klasifikasi	Error! Bookmark not defined.
6. Random Forest	Error! Bookmark not defined.
7. <i>Hyperparameter Tuning</i>	Error! Bookmark not defined.

8. Grid Search CV	Error! Bookmark not defined.
9. Confusion Matrix	Error! Bookmark not defined.
B. Kajian Hasil Penelitian Terdahulu	Error! Bookmark not defined.
C. Kerangka Berfikir	Error! Bookmark not defined.
BAB III METODE PENELITIAN.....	Error! Bookmark not defined.
A. Bussines Understanding	Error! Bookmark not defined.
B. Data Understanding	Error! Bookmark not defined.
C. Data Preparation.....	Error! Bookmark not defined.
D. Modelling	Error! Bookmark not defined.
E. Evaluation.....	Error! Bookmark not defined.
F. Deployment	Error! Bookmark not defined.
BAB IV HASIL DAN PEMBAHASAN.....	Error! Bookmark not defined.
A. Bussines Understanding	Error! Bookmark not defined.
B. Data Understanding	Error! Bookmark not defined.
C. Data Preparation.....	Error! Bookmark not defined.
1. Penanganan Missing Value	Error! Bookmark not defined.
2. Penghapusan Data Duplikat	Error! Bookmark not defined.
3. Encoding Data Kategorikal	Error! Bookmark not defined.
4. Transformasi Label Target	Error! Bookmark not defined.
5. Pemisahan Fitur dan Label.....	Error! Bookmark not defined.
6. Pembagian Data Latih dan Data Uji.....	Error! Bookmark not defined.
7. Penanganan Ketidakseimbangan Kelas dengan SMOTE	Error! Bookmark not defined.
D. Modelling	Error! Bookmark not defined.
E. Evaluation.....	Error! Bookmark not defined.
BAB V PENUTUP.....	Error! Bookmark not defined.

A. Kesimpulan	Error! Bookmark not defined.
B. Saran	Error! Bookmark not defined.
DAFTAR PUSTAKA	7
LAMPIRAN	Error! Bookmark not defined.

DAFTAR TABEL

Tabel 2.1 Confusion Matrix	Error! Bookmark not defined.
Tabel 2.2 Penelitian Terdahulu.....	Error! Bookmark not defined.
Tabel 3.1 Deskripsi dataset ILPD.....	Error! Bookmark not defined.
Tabel 4.1 Classification Report dari Model Random Forest (Tanpa Tuning)	Error!
Bookmark not defined.	
Tabel 4.2 Classification Report dari Model Random Forest (Setelah Tuning)	
.....	Error! Bookmark not defined.

DAFTAR GAMBAR

Gambar 2.1 Algoritma Random Forest	Error! Bookmark not defined.
Gambar 2.2 Bagan Kerangka Berfikir	Error! Bookmark not defined.
Gambar 3.1 Model CRISP-DM	Error! Bookmark not defined.
Gambar 4.1 Tipe dataset ILPD.....	Error! Bookmark not defined.
Gambar 4.2 Temuan Missing Value	Error! Bookmark not defined.
Gambar 4.3 Setelah penanganan Missing Value	Error! Bookmark not defined.
Gambar 4.4 Data duplikasi.....	Error! Bookmark not defined.
Gambar 4.5 Distribusi Label sebelum SMOTE	Error! Bookmark not defined.
Gambar 4.6 Distribusi Label Setelah SMOTE.....	Error! Bookmark not defined.
Gambar 4.7 Confusion Matrix Model Random Forest (Tanpa Tuning).....	Error! Bookmark not defined.
Bookmark not defined.	
Gambar 4.8 Hasil Grid Search CV	Error! Bookmark not defined.
Gambar 4.9 Hasil Confusion Matrix Model Random Forest setelah tuning	Error! Bookmark not defined.
Bookmark not defined.	

BAB I

PENDAHULUAN

A. Latar Belakang Masalah

Penyakit liver menjadi salah satu masalah kesehatan yang signifikan secara global. Menurut data dari Organisasi Kesehatan Dunia (WHO), pada tahun 2020 tercatat sekitar 1,5 juta orang meninggal akibat penyakit liver, termasuk sirosis dan berbagai penyakit liver kronis lainnya. Penyakit liver sering kali disebabkan oleh infeksi virus seperti hepatitis B dan C, mengonsumsi alkohol secara berlebihan, serta penumpukan lemak di hati (Ginting et al., 2021). Mengingat beragamnya penyebab dan kompleksitas penyakit liver, dibutuhkan metode diagnosis dan klasifikasi untuk memastikan penanganan yang optimal bagi pasien.

Tingginya angka kematian akibat penyakit liver menunjukkan perlunya penelitian yang lebih mendalam. Rekam medis pasien, hasil tes laboratorium, dan survei epidemiologi adalah sumber data berharga yang dapat dimanfaatkan untuk mengenali pola penyakit, menentukan faktor resiko, dan merumuskan strategi pencegahan yang efektif. Data dalam rekam medis tidak hanya mencakup riwayat kesehatan pasien, tetapi juga berbagai hasil tes diagnostik yang memberikan gambaran menyeluruh tentang kondisi liver. Selain itu, survei epidemiologi berperan penting dalam memahami penyebaran dan prevalensi penyakit liver di berbagai populasi untuk pengembangan kebijakan kesehatan masyarakat yang efektif (Wulan Sari et al., 2023). Jika data ini dikelola dan dianalisis dengan benar, maka dapat memberikan pengetahuan yang lebih luas tentang penyakit liver.

Dengan pengelolaan dan analisis data yang efektif, *data mining* dapat berperan penting dalam penelitian penyakit liver. *Data mining* dimanfaatkan untuk menggali informasi penting dari kumpulan data yang besar dan kompleks yang mungkin tidak terlihat secara langsung (Triwardani et al., 2024). Misalnya, *data mining* dapat digunakan untuk menemukan faktor-faktor penting yang berkontribusi terhadap perkembangan penyakit liver, seperti faktor genetik,

lingkungan, dan gaya hidup. Selain itu, *data mining* juga dapat digunakan untuk membuat model prediktif yang membantu dalam diagnosis dan pencegahan dini penyakit liver (Iqbal, Destiana, et al., 2024). Proses ini melibatkan penggunaan berbagai teknik analisis untuk menjumpai pola dan hubungan dalam data yang dapat mendukung dalam pengambilan keputusan yang lebih baik dalam diagnosis dan perawatan.

Salah satu teknik yang dapat diintegrasikan dengan *data mining* dalam penelitian penyakit liver adalah *machine learning*. Dengan memanfaatkan *machine learning*, model dapat dilatih untuk mengenali pola kompleks dalam data penyakit liver (A'an Prasetyo et al., 2024). Model ini dapat digunakan untuk memprediksi kemungkinan seseorang mengidap penyakit liver, mengklasifikasikan jenis penyakit liver, atau bahkan memprediksi respons pasien terhadap pengobatan tertentu. Data medis seringkali mengandung *noise*, ketidakkonsistenan, dan variabel yang saling berkorelasi, maka *machine learning* mampu mengelola kompleksitas ini dan mengidentifikasi pola-pola yang tidak mudah terlihat oleh manusia. Kemampuan *machine learning* untuk belajar dari data besar dan kompleks menjadikannya alat yang sangat berharga dalam meningkatkan akurasi diagnosis dan personalisasi pengobatan (Syahputra & Hanifah, 2024).

Permasalahan utama dalam penanganan penyakit liver adalah keterlambatan diagnosis yang sering mengakibatkan pengobatan menjadi lebih kompleks. Banyak pasien baru menyadari kondisi mereka ketika penyakit sudah berada pada tahap lanjut, di mana kerusakan hati sudah parah dan sulit diperbaiki. Pendeteksian dan klasifikasi dini penyakit liver sangat penting untuk mencegah penyakit berkembang ke tahap yang lebih serius dan meningkatkan peluang kesembuhan pasien. Namun, gejala yang kompleks dan berbagai faktor risiko membuat proses diagnosis dini menjadi tantangan bagi tenaga medis. Oleh sebab itu, dibutuhkan sistem klasifikasi yang akurat dan efisien untuk menunjang profesional kesehatan mengidentifikasi serta mengkategorikan

penyakit liver sejak tahap awal, sehingga intervensi medis dapat dilakukan dengan tepat waktu dan efektif.

Beragam penelitian dan implementasi telah dilakukan untuk meningkatkan akurasi dan efisiensi algoritma, termasuk penelitian sebelumnya oleh (Ayudhitama & Pujianto, 2020) melaporkan akurasi tertinggi dari *Decision Tree* sebesar 72,74%, diikuti *Neural Network* 69,64%, KNN 66,03%, dan *Naïve Bayes* 55,42%. Penelitian oleh (Rahman, 2020) mencatat akurasi *Decision Tree* sebesar 70,29% dan *Naïve Bayes* 67,05%. Kemudian, penelitian oleh (Lubis et al., 2021) menunjukkan *Random Forest* unggul dengan akurasi 70,60% dibandingkan *Naïve Bayes* 55,80%. Pada penelitian (Aldana & Sasongko Wibowo, 2024) mencatat akurasi K-NN sebesar 70%, dan (Afni Alviola et al., 2023) memperoleh akurasi tertinggi 71,79% dengan ANN. Masing-masing algoritma memiliki kelemahan yaitu *Naïve Bayes* kurang optimal dengan data kelas yang tidak seimbang dan fitur non-kategorikal, sedangkan K-NN sensitif terhadap outlier dan menurun performanya pada data berdimensi tinggi. *Decision Tree* rentan terhadap *overfitting*, dengan hasil yang sangat bervariasi akibat perubahan kecil dalam data. Di sisi lain, *Neural Network* membutuhkan waktu pelatihan yang lama dan sulit untuk diinterpretasikan. Oleh karena itu, penelitian ini memilih *Random Forest* sebagai alternatif yang lebih stabil, mampu menangani banyak fitur, tahan terhadap *overfitting*, dan memberikan interpretasi yang lebih baik, dengan harapan dapat meningkatkan akurasi klasifikasi secara optimal (Santoso et al., 2024).

Dalam konteks penyakit liver, *Random Forest* efektif dalam menangani data medis yang kompleks berkat kemampuannya mengelola variabel numerik dan kategorikal secara bersamaan, seperti hasil tes laboratorium (Christian & Yani, 2025). Salah satu kekuatan utama *Random Forest* adalah kemampuannya untuk menghasilkan model yang lebih generalisasi dan akurat, berkat penggunaan teknik *bagging* yang secara efektif mengatasi masalah *overfitting*, di mana sampel data diambil secara acak dengan penggantian untuk membangun setiap pohon keputusan (Andriyani et al., 2024). Metode ini juga

menyediakan peringkat kepentingan fitur yang membantu peneliti mengidentifikasi faktor-faktor paling berpengaruh dalam diagnosis penyakit liver. Selain itu, *Random Forest* mampu menyelesaikan nilai yang hilang dan outlier, mengingat data medis seringkali tidak lengkap atau mengandung nilai-nilai ekstrem (Ratnawati & Sulistyaningrum, 2020).

Untuk meningkatkan akurasi dalam proses klasifikasi, penyesuaian parameter atau *hyperparameter tuning* merupakan langkah yang sangat penting. Pada algoritma *Random Forest*, proses ini mencakup pengaturan beberapa parameter utama, seperti jumlah pohon yang digunakan (*n_estimators*), kedalaman maksimum pohon (*max_depth*), serta jumlah minimum sampel pada setiap daun (*min_samples_leaf*) (Hidayat et al., 2023). Penyesuaian parameter ini terbukti mampu meningkatkan performa model secara nyata. Hal ini dibuktikan melalui penelitian oleh (Ryan Dana et al., 2024) di mana akurasi model *Random Forest* berhasil meningkat 1% setelah dilakukan proses *hyperparameter tuning*. Dalam data medis seperti diagnosis penyakit liver, *hyperparameter tuning* membantu model menyesuaikan diri dengan kompleksitas data pasien untuk menghasilkan prediksi yang lebih akurat.

Dengan mengimplementasikan algoritma *Random Forest*, penelitian ini diharapkan dapat meningkatkan akurasi dari penelitian sebelumnya serta menghasilkan model klasifikasi yang tepat dalam mendeteksi penyakit liver pada tahap awal. Diharapkan penelitian ini dapat memberikan kontribusi pada pengembangan sistem pendukung keputusan klinis yang dapat membantu profesional kesehatan membuat diagnosis dengan lebih cepat dan tepat, sehingga meningkatkan efektivitas penanganan pasien penyakit liver.

B. Batasan Masalah

Penelitian ini memiliki beberapa batasan sebagai berikut :

1. Penelitian ini memanfaatkan dataset pasien liver India (*Indian Liver Patient Dataset* atau ILPD). Dataset ini merupakan sumber data terbuka yang dapat diakses secara bebas melalui *UCI Machine Learning Repository*.

2. Algoritma yang digunakan dalam deteksi penyakit liver adalah *Random Forest*.
3. Penelitian terbatas pada algoritma *Random Forest* sebagai metode klasifikasi utama tanpa membandingkan secara langsung dengan algoritma lain.
4. Penelitian memanfaatkan teknik *hyperparameter tuning* dengan metode *Grid Search CV* dan teknik penyeimbangan data SMOTE.
5. Klasifikasi difokuskan pada dua kelas (biner):
 - 1 = Pasien terdiagnosis penyakit liver
 - 0 = Pasien normal (tidak terdiagnosis penyakit liver)
6. *Python* dipilih sebagai bahasa pemrograman untuk menjalankan penelitian ini.
7. Evaluasi model hanya dilakukan berdasarkan metrik klasifikasi seperti accuracy, precision, recall, F1-score, dan *Confusion Matrix* tanpa mempertimbangkan aspek waktu komputasi atau efisiensi sistem.

C. Rumusan Masalah

Mengacu pada batasan masalah yang telah dijelaskan diatas, rumusan masalah yang diperoleh adalah sebagai berikut :

1. Seberapa akurat algoritma *Random Forest* dalam mendeteksi penyakit liver ?
2. Bagaimana pengaruh penyetelan *hyperparameter tuning* pada algoritma *Random Forest* terhadap kinerja model dalam mendeteksi penyakit liver?

D. Tujuan Penelitian

Berdasarkan latar belakang dan rumusan masalah yang telah disebutkan, maka penelitian ini bertujuan untuk mengimplementasikan algoritma *Random Forest* guna meningkatkan performa deteksi dan akurasi diagnosis penyakit liver, serta mengoptimalkan performa model melalui teknik *hyperparameter tuning* menggunakan metode *Grid Search CV* untuk memperoleh kombinasi parameter terbaik yang dapat memaksimalkan akurasi deteksi.

E. Manfaat Penelitian

Berikut adalah manfaat dari penelitian ini :

1. Meningkatkan akurasi diagnosis penyakit liver melalui penerapan algoritma *Random Forest*.
2. Bagi pusat tenaga medis dalam membantu pasien dalam mendiagnosa penyakit liver berdasarkan gejala dan data yang tersedia.

DAFTAR PUSTAKA

- A'an Prasetyo, M., Khanif Zyen, A., & Kusumodestoni, R. H. (2024). OPTIMASI ALGORITMA NAIVE BAYES BERBASIS KERNEL UNTUK KLASIFIKASI PENYAKIT HATI. *Jurnal Informatika Teknologi Dan Sains*. <https://www.kaggle.com/datasets/johnsmith88/heart-disease-dataset?resource>.
- Afni Alviola, N., Fathurrahman, Z., Nur Rifai, R., Shabrina Afrah, A., Studi Teknik Informatika, P., Sains dan Teknologi, F., Maulana Malik Ibrahim, U., Gajayana No, J., Lowokwaru, K., Malang, K., & Timur, J. (2023). *Sistem Diagnosa Penyakit Liver Menggunakan Metode Artificial Neural Network: Studi Berdasarkan Dataset Indian Liver Patient Dataset*. 8(3).
- Agustiani, S., Tajul Arifin, Y., Junaidi, A., Khotimatul Wildah, S., & Mustopa, A. (2022). *Klasifikasi Penyakit Daun Padi menggunakan Random Forest dan Color Histogram 1, ** (Vol. 10, Issue 1). <https://www.kaggle.com/vbookshelf/rice-leaf->
- Aldana, S., & Sasongko Wibowo, J. (2024). *Penerapan Data Mining Terhadap Klasifikasi Pasien Penderita Penyakit Liver Menggunakan Metode K-Nearest Neighbor*. <https://doi.org/10.35889/progresif.v20i1.1376>
- Andini, E., Reza Faisal, M., Herteno, R., Adi Nugroho, R., & Abadi, F. (2022). PENINGKATAN KINERJA PREDIKSI CACAT SOFTWARE DENGAN HYPERPARAMETER TUNING PADA ALGORITMA KLASIFIKASI DEEP FOREST. In *Jurnal MNEMONIC* (Vol. 5, Issue 2). <https://github.com/bharlow058/AEEEM-and-other->
- Andriyani, W., Natsir, F., Lubis, H., Hidayatul Yulianing Tyas, S., Meidelfi, D., Faizah, S., Kurniawan, H., Wahyuningtyas, I., Noor Hasan, F., Ricky Afandi, I., & Hikmawati, E. (2024). *PERANGKAT LUNAK DATA MINING*. www.freepik.com

- Ayudhitama, A. P., & Pujiyanto, U. (2020). *JIP (Jurnal Informatika Polinema) ANALISA 4 ALGORITMA DALAM KLASIFIKASI PENYAKIT LIVER MENGGUNAKAN RAPIDMINER*.
<https://doi.org/https://doi.org/10.33795/jip.v6i2.274>
- Christian, A., & Yani, A. (2025). *ANALISIS MACHINE LEARNING UNTUK PREDIKSI PENYAKIT PARU-PARU MENGGUNAKAN RANDOM FOREST* (Vol. 7). <https://www.kaggle.com/>
- Desiani, A. (2022). Perbandingan Implementasi Algoritma Naïve Bayes dan K-Nearest Neighbor Pada Klasifikasi Penyakit Hati. *SIMKOM*, 7(2), 104–110.
<https://doi.org/10.51717/simkom.v7i2.96>
- Faruqziddan, M., Herdika Septa Aulia, E., Dini Azzahra, S., Ristyawan, A., & Daniati, E. (2024). Klasifikasi Risiko Kambuhnya Kanker Tiroid Menggunakan Algoritma Random Forest. In *Agustus* (Vol. 8). Online.
- Ginting, L. B., Laila Sekawan, & Imelda Juwita. (2021). *PENYULUHAN TENTANG PERAWATAN PASIEN PENYAKIT HEPATITIS (LIVER) DI DESA PALUH SIBAJI*.
- Hidayat, Sunyoto, A., & Al Fatta, H. (2023). *Klasifikasi Penyakit Jantung Menggunakan Random Forest Clasifier*.
- Iqbal, M., Destiana, H., & Hamid, A. (2024). *Data Mining Berbasis Machine Learning Untuk Analitik Prediktif Dalam Kelulusan: Vols. x, No.x*.
<https://doi.org/https://doi.org/10.55679/semantik.v10i2.67>
- Iqbal, M., Jauhar, I., Bayu, A. A., Ristyawan, A., & Daniati, E. (2024). Klasifikasi Penggunaan Video Game Dengan Menggunakan Metode Algoritma Naive Bayes. In *Agustus* (Vol. 8). Online.
- Junus, C. Z. V., Tarno, T., & Kartikasari, P. (2023). *KLASIFIKASI MENGGUNAKAN METODE SUPPORT VECTOR MACHINE DAN RANDOM FOREST UNTUK DETEKSI AWAL RISIKO DIABETES*

MELITUS. *Jurnal Gaussian*, 11(3), 386–396.

<https://doi.org/10.14710/j.gauss.11.3.386-396>

Khatib Sulaiman, J., Tegar Kusuma, S., Bayu Sasongko, T., & Amikom Yogyakarta, U. (2023). Optimasi K-Nearest Neighbor dengan Grid Search CV pada Prediksi Kanker Paru-Paru. *Indonesian Journal of Computer Science Attribution*, 12(4), 2162.

Kudus, U. M., Ganesha, J., & Kudus, P. (2020). *Klasifikasi Penderita Penyakit Diabetes Menggunakan Algoritma Decision Tree C4.5*.

Kurniawan, M., Wijaya, S., Satria, D., Daniati, E., & Ristyawan Aidina. (2024). *Analisis Data Mining Untuk Klasifikasi Kasus Covid 19*.

Lestari, M. J. (2023). *HUBUNGAN LAMA KONSUMSI MIRAS TRADISIONAL KAMEKO TERHADAP PENINGKATAN KADAR SGPT DI DESA PARIDA KABUPATEN MUNA*.

Lubis, A. I., Erdiansyah, U., & Siregar, R. (2021). *Komparasi Akurasi Pada Naïve Bayes Dan Random Forest Dalam Klasifikasi Penyakit Liver*.
<https://doi.org/https://doi.org/10.24114/cess.v7i1.28888>

Ma'rifah, H., Prasetya Wibawa, A., & Akbar, M. I. (2020). *Sains, Aplikasi, Komputasi dan Teknologi Informasi Klasifikasi artikel ilmiah dengan berbagai skenario preprocessing*. 2(2), 70.

Nurrofikoh, M., Fatima, A., Hastuti, H., Fauziyah, O., Nursiswati, N., & Pebrianti, S. (2023). Cegah dan Kenali Kondisi Hati (CEK SI HATI) sebagai Upaya Pendidikan Kesehatan terkait Sirosis Hati Kepada Masyarakat. *Jurnal Kreativitas Pengabdian Kepada Masyarakat (PKM)*, 6(7), 2984–3008.
<https://doi.org/10.33024/jkpm.v6i7.10160>

Pane, R., Nasution Marnis, & Irmayani, D. (2020). *PENERAPAN METODE DEMPSTER SHAFER UNTUK MENDIAGNOSA PENYAKIT LIVER*.

- Putri, R. W., Ristyawan, A., & Muzaki, M. N. (2022). Comparison Performance of K-NN and NBC Algorithm for Classification of Heart Disease. *JTECS : Jurnal Sistem Telekomunikasi Elektronika Sistem Kontrol Power Sistem Dan Komputer*, 2(2), 143. <https://doi.org/10.32503/jtecs.v2i2.2708>
- Rahman, N. (2020). *ANALISA ALGORITMA DECISION TREE DAN NAÏVE BAYES PADA PASIEN PENYAKIT LIVER*.
<https://doi.org/https://doi.org/10.37859/jf.v10i2.2087>
- Ratnawati, L., & Sulistyaningrum, D. (2020). *Penerapan Random Forest untuk Mengukur Tingkat Keparahannya Penyakit pada Daun Apel*.
<https://doi.org/10.12962/j23373520.v8i2.48517>
- Rianti, A., Wachid, N., Majid, A., Fauzi, A., Sistem, P., Informasi, T., Telekomunikasi, S., Upi Di Purwakarta, K., & Edu, A. (2023). *CRISP-DM: Metodologi Proyek Data Science*.
- Ristyawan, A., Nugroho, A., & Amarya, T. K. (2025). *Optimasi Preprocessing Model Random Forest Untuk Prediksi Stroke*. 12(1), 29–44.
- Ryan Dana, A., Valentino Kristananda, R., Bagas Satrio Wibowo, M., & Arman Prasetya, D. (2024). Perbandingan Algoritma Decision Tree dan Random Forest dengan Hyperparameter Tuning dalam Mendeteksi Penyakit Stroke. In *Seminar Nasional Informatika Bela Negara (SANTIKA)* (Vol. 4).
- Santoso, H. T., Felmidi, F. A., Fadhila, A. N., Ristyawan, A., & Daniati, E. (2024). Analisis Kinerja Algoritma Data Mining pada Klasifikasi Tingkat Obesitas dengan K-Fold Cross Validation dan AUC. In *Agustus* (Vol. 8). Online.
- Saripah, I., Tasawuf, J., Psikoterapi, D., Uin, U., Gunung, S., & Bandung, D. (2023). Pandangan Tokoh-Tokoh Muslim mengenai Sabar sebagai Metode Penyembuhan Penyakit Hati. *Gunung Djati Conference Series*, 19.
- Syahputra, R. A., & Hanifah, M. R. (2024). *Metode Analisis Kesehatan Dengan Menggunakan Machine Learning Atau Artificial Inteligenci Atau Data Mining Literature Review*. <http://jurnal.utu.ac.id/invasi/>

- Tarigan, V. (2023). *Pembuatan Aplikasi Data Mining Untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma Naive Bayes*. 11(1).
<https://doi.org/https://doi.org/10.36987/informatika.v11i1.3847>
- Triwardani, A. A., Ulfiah, M. H., Ristyawan, A., & Daniati, E. (2024). Prosiding SEMNAS INOTEK (Seminar Nasional Inovasi Teknologi) 1218 Prediksi Kematian Akibat Gagal Jantung Menggunakan Algoritma K-Nearest Neighbors. In *Agustus* (Vol. 8). Online.
- Wulan Sari, N., Hairil Akbar, Mk., Ika Nirmala Masliah, Me., Sartika, Mk., Mudyawati Kamaruddin, Mk., Susanti Sinaga, E., Elmi Nuryati, M., & Saraswati Haylian Chiani, Me. (2023). *TEORI DAN APLIKASI EPIDEMIOLOGI KESEHATAN*.