

KLASIFIKASI PENYAKIT KANKER HATI DALAM ALGORITMA K-NEAREST NEIGHBORS (KNN)

Taniya Raisha Dwiva Putri, Tessa Arianvira, Revi Amelia Dwifa Futri,
Muhammad Farhan, Ken Ditha Tania, Allsela Meiriza

Sistem Informasi, Universitas Sriwijaya

Jl. Padang Selasa No. 524, Bukit Lama, Kota Palembang, Sumatera Selatan, Indonesia

taniyaraisha@gmail.com

ABSTRAK

Kanker hati (*liver cancer*) merupakan penyakit dengan tingkat mortalitas tinggi yang sering terdeteksi pada tahap lanjut, sehingga diperlukan pendekatan sistematis untuk mendukung diagnosis yang cepat dan akurat. Penelitian ini bertujuan menerapkan metode *data mining* menggunakan algoritma K-Nearest Neighbors (KNN) untuk mengklasifikasikan penyakit kanker hati serta mengevaluasi kinerja model. Data yang digunakan berasal dari Kaggle sebanyak 5.000 data pasien dengan 14 atribut klinis, mencakup faktor risiko seperti hepatitis, sirosis, dan diabetes, serta biomarker seperti skor fungsi hati dan kadar alpha fetoprotein. Tahap *preprocessing* dilakukan menggunakan *Z-Transformation Normalization* untuk menyetarakan skala atribut numerik, mengingat KNN berbasis perhitungan jarak. Selanjutnya, model dilatih dan diuji untuk mengukur performanya. Hasil penelitian menunjukkan bahwa model KNN mencapai akurasi sebesar 86,40% dengan nilai *class precision* pada prediksi positif sebesar 80,00%. Hasil ini menunjukkan bahwa algoritma KNN cukup efektif dalam mengolah data medis dan dapat dimanfaatkan sebagai alat bantu dalam proses identifikasi penyakit kanker hati secara lebih akurat.

Kata kunci : K-Nearest Neighbors (KNN), Kanker Hati, Data Mining, Klasifikasi, Confusion Matrix

1. PENDAHULUAN

Pemanfaatan data dalam jumlah besar telah menjadi bagian penting dalam berbagai bidang, termasuk bidang kesehatan. Data kesehatan yang berasal dari rekam medis pasien, hasil pemeriksaan laboratorium, serta data klinis lainnya menyimpan informasi berharga yang dapat digunakan untuk membantu proses analisis dan pengambilan keputusan. Namun, data tersebut belum sepenuhnya dimanfaatkan secara optimal untuk mendukung proses klasifikasi dan identifikasi penyakit secara akurat.

Salah satu penyakit yang memerlukan perhatian serius adalah kanker hati (*liver cancer*). Penyakit *liver* merupakan gangguan kesehatan yang masih menjadi perhatian serius karena memiliki angka kejadian dan tingkat kematian yang tinggi di berbagai wilayah [1].

Kanker hati merupakan penyakit dengan tingkat kematian yang tinggi dan sering kali baru terdeteksi pada tahap lanjut karena awal yang tidak spesifik. Kondisi ini menyebabkan proses penanganan menjadi kurang efektif dan meningkatkan risiko komplikasi yang lebih berat. Oleh karena itu, diperlukan suatu pendekatan yang mampu membantu mengklasifikasikan penyakit kanker hati secara lebih cepat dan tepat berdasarkan data yang tersedia.

Perkembangan ilmu komputer, khususnya bidang *machine learning* telah membuka peluang besar dalam membantu proses diagnosis penyakit secara otomatis [2].

Data mining merupakan pendekatan *machine learning* yang dapat digunakan untuk menggali pola dan informasi penting dari kumpulan data berukuran

besar. Salah satu teknik dalam *data mining* adalah klasifikasi, yaitu proses pengelompokan data ke dalam kelas tertentu berdasarkan karakteristik yang dimiliki. Beberapa algoritma pembelajaran mesin yang sering digunakan dalam proses klasifikasi antara lain *naïve bayes*, *support vector machine* (SVM), *k-nearest neighbors* (KNN), *decision tree*, dan *random forest* yang masing-masing memiliki kelebihan dan keterbatasan dalam akurasi, kecepatan komputasi serta ketahanan terhadap data yang bising [3].

Algoritma *K-Nearest Neighbors* adalah metode dari *data mining* yang melakukan klasifikasi objek baru berdasarkan data *training* yang memiliki jarak *k* terdekat dengan objek tersebut (*nearest neighbors*) [4].

K-Nearest Neighbors (KNN) digunakan sebagai metode klasifikasi karena memiliki konsep yang sederhana, mudah diterapkan, serta mampu bekerja dengan baik pada data numerik, termasuk data medis. KNN adalah pendekatan yang populer dalam melakukan analisis data medis yang dapat mengklasifikasikan data berdasarkan kedekatannya dengan data yang lain [5].

Pada penelitian yang dilakukan oleh Depari, Tania, dan Sevtiyuni pada tahun 2025, dalam menerapkan metode Random Forest, XGBoost, KNN, dan SVM pada prediksi diabetes, menyatakan bahwa dari empat model pembandingan tersebut, KNN adalah model yang cukup baik dengan score 0,94 di semua aspek evaluasi [6].

Oleh karena itu, pada penelitian ini peneliti memutuskan untuk menggunakan model algoritma KNN dalam melakukan klasifikasi penyakit kanker hati. Model algoritma KNN juga dipilih karena

kesederhanaan dan kemudahan mengelompokan data dengan cara menilai kedekatan objek baru atau tetangga terdekat (k) [7].

Penelitian ini hanya menggunakan dataset *Liver Cancer* yang diambil dari platform *kaggle* yang berisi 5000 data pasien. Dataset ini terdiri dari 14 kolom dengan atribut seperti usia, jenis kelamin, *Body Mass Index*, kategori penggunaan alkohol, kategori perilaku merokok, hepatitis B, hepatitis C, skor fungsi hati, level alpha fetoprotein, riwayat sirosis, riwayat keluarga yang terkena kanker, level aktivitas fisik, dan diabetes. Setiap data memiliki label kelas *liver cancer* untuk menunjukkan status pasien penderita penyakit kanker hati, sehingga dapat digunakan untuk pengujian kinerja algoritma klasifikasi KNN.

Berdasarkan uraian latar belakang tersebut, penelitian ini dilakukan dengan tujuan untuk menerapkan metode data mining menggunakan algoritma *K-Nearest Neighbors* (KNN) dalam mengklasifikasikan dataset penyakit kanker hati serta mengevaluasi tingkat akurasi dan performanya.

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Penelitian yang bertujuan untuk mengevaluasi tingkat akurasi algoritma *K-Nearest Neighbors* (KNN) dalam mendiagnosis penyakit jantung pada populasi lansia [5].

Manfaat penelitian ini adalah untuk berkontribusi pada pengembangan sistem diagnosis penyakit jantung yang lebih cepat, efisien, dan akurat, khususnya untuk pasien lansia yang sering mengalami keterlambatan deteksi dini. Metode yang digunakan adalah kuantitatif dengan desain eksperimental, menggunakan dataset yang diperoleh dari *Kaggle* sebanyak 918 data pasien. Proses penelitian meliputi tahap pra-pemrosesan data, penerapan algoritma KNN dengan berbagai nilai k , dan evaluasi kinerja model menggunakan metrik akurasi, presisi, *recall*, dan *F1-score*. Hasil menunjukkan bahwa nilai $k = 9$ menghasilkan kinerja terbaik dengan nilai *recall* tertinggi 0,93 dan *F1-score* 0,81, meskipun akurasi yang diperoleh adalah 0,68. Temuan ini menunjukkan bahwa algoritma KNN sangat efektif untuk aplikasi medis yang memprioritaskan deteksi sebanyak mungkin kasus positif, sehingga dapat membantu proses diagnosis penyakit jantung secara lebih optimal pada lansia.

Penelitian yang sejenis yang bertujuan untuk menerapkan algoritma *K-Nearest Neighbors* (KNN) untuk mengklasifikasikan diabetes melitus pada penduduk Desa Jatitengah [4].

Manfaat penelitian ini adalah untuk membantu deteksi dini diabetes, sehingga memberikan dasar pertimbangan bagi tenaga medis dan masyarakat dalam pencegahan dan pengobatan dini. Metode yang digunakan adalah pendekatan kuantitatif dengan desain eksperimental, di mana data pasien diproses melalui tahap pra-pemrosesan dan kemudian diklasifikasikan menggunakan algoritma KNN

berdasarkan atribut kesehatan pasien. Evaluasi dilakukan untuk mengukur tingkat keberhasilan klasifikasi. Hasil menunjukkan bahwa algoritma KNN mampu mengklasifikasikan diabetes melitus dengan tingkat akurasi sebesar 93% dengan $k = 9$ *precision* 100%, 60% *recall*, dan 75% *f1-score*.

2.2. Data Mining

Data mining atau penambangan data, adalah proses analisis data besar-besaran untuk menemukan pola, hubungan, atau pengetahuan baru yang berguna bagi pengambilan keputusan [8].

Penambangan data ini merupakan teknik pengolahan menggunakan dataset dengan jumlah besar untuk menemukan informasi penting yang sebelumnya tidak dapat diketahui jika diolah secara konvensional, sehingga hasil dari pengolahan dapat mendukung pengambilan keputusan penting dalam organisasi [9].

Metode yang ada pada data mining adalah *Classification*, *Association*, *Clustering*, *Regression*, *Forecasting*, *Sequencing*, dan *Descriptive*. Pada penelitian ini, metode *data mining* yang digunakan adalah klasifikasi. Klasifikasi adalah suatu metode dalam penentuan anggota di suatu kelas yang telah ditentukan sebelumnya [10].

Pada saat dilakukan pengklasifikasian, dataset dibagi menjadi dua yaitu data *training* dan data *testing* yang dimana keduanya berperan dalam menentukan hasil akurasi yang tepat dari penerapan metode klasifikasi yang digunakan.

2.3. K-Nearest Neighbors (KNN)

K-Nearest Neighbor (KNN) adalah salah satu teknik untuk menerapkan klasifikasi terhadap objek berdasarkan data pelatihan terdekat (*neighbors*) [11].

Dekat atau jauhnya tetangga biasanya dihitung berdasarkan jarak *Euclidean* [12].

Metode ini termasuk ke dalam teknik *lazy learning* dalam memprediksi atau mengklasifikasikan suatu data.

2.4. Confusion Matrix

Evaluasi dilakukan untuk menilai kemampuan dari model yang digunakan dalam mengklasifikasikan data uji secara lebih mendalam [13].

Hasil dari pengolahan dengan algoritma tersebut yang akan dihitung tingkat akurasi, presisi, dan *recall* [11]. Rumus yang digunakan pada evaluasi model adalah sebagai berikut [14]:

a. *Accuracy*

$$\frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (1)$$

b. *Precision*

$$\frac{TP}{TP + FP} \times 100\% \quad (2)$$

c. *Recall*

$$Recall = \frac{TP}{TP + FN} \times 100\% \quad (3)$$

2.5. Kanker Hati

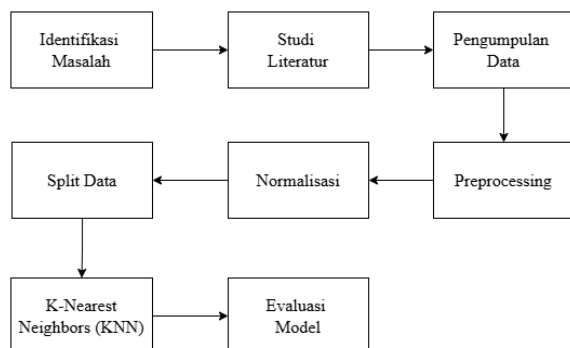
Kanker hati adalah jenis penyakit kanker yang lebih sering menyerang pria daripada wanita. Penyakit ini bisa disebabkan oleh beberapa masalah kesehatan, seperti infeksi hepatitis B atau C, serta kebiasaan buruk seperti minum alkohol berlebihan. Kanker hati sering sulit dideteksi karena gejalanya mirip dengan masalah hati lainnya. Untuk memudahkan deteksi penyakit *liver* diperlukan mesin yang dapat mendeteksi secara lebih efisien dengan mengklasifikasikan pasien yang menderita penyakit *liver* dan yang tidak menderita penyakit *liver* [15].

3. METODE PENELITIAN

Penelitian ini termasuk ke dalam penelitian deskriptif kuantitatif dan studi literatur untuk menggambarkan dan menjelaskan hasil penelitian secara objektif [16].

Hasil dari penelitian ini dapat menarik kesimpulan berupa data berbentuk angka-angka yang telah disajikan dalam menganalisis data terkait dengan klasifikasi penyakit kanker hati.

Metode yang digunakan adalah metode *data mining* algoritma *K-Nearest Neighbors* (KNN) untuk mengklasifikasikan penyakit kanker hati. Adapun tahapan dalam penelitian ini antara lain studi literatur, pengambilan data, *pre-processing*, *split* data, pelatihan data dengan algoritma KNN, evaluasi model, dan kesimpulan. Berikut gambar dari tahapan penelitian ini:



Gambar 1. Tahapan Penelitian

Tahapan yang akan dilakukan seperti pada gambar 1 di atas adalah sebagai berikut :

3.1. Pengumpulan Data

Data yang digunakan adalah dataset yang diambil dari situs *Kaggle* dengan judul dataset "*Predict Liver Cancer from and Clinical Features*". Data sintesis ini mencakup 14 atribut yang menunjukkan variabel klinis, seperti usia, jenis kelamin, *Body Mass Index*, kategori penggunaan alkohol, kategori perilaku merokok, hepatitis B, hepatitis C, skor fungsi hati, level alpha fetoprotein, riwayat sirosis, riwayat keluarga yang terkena kanker, level aktivitas fisik, dan diabetes. Variabel yang menjadi atribut target atau label pada penelitian ini adalah variabel *liver_cancer*. Dataset ini memiliki

5000 data yang bisa digunakan untuk melakukan klasifikasi penyakit kanker hati.

Tabel 1. Dataset

A	B	C	D	E	F	G	H	I	J	K	L	M	N
68	1	18	1	1	0	0	51	16	0	0	1	0	0
81	1	19	3	2	0	0	41	8	0	0	2	1	0
58	1	25	2	2	0	0	76	0	0	0	2	0	0
44	2	16	2	1	0	0	50	19	0	0	1	1	0
72	2	21	3	1	0	0	39	4	1	0	1	1	1
37	1	23	1	2	0	0	50	0	0	0	2	0	0
50	2	19	1	3	0	0	68	0	1	0	2	0	0
68	2	15	1	1	0	0	70	40	0	0	3	0	0
48	2	27	3	1	0	0	70	4	0	0	1	0	0
52	2	26	3	2	1	0	48	8	0	1	3	0	0

Keterangan.

A : <i>age</i>	H : <i>liver_function_score</i>
B : <i>gender</i>	I : <i>alpha_fetoprotein_level</i>
C : <i>BMI</i>	J : <i>cirrhosis_history</i>
D : <i>alcohol_consumption</i>	K : <i>family_history_cancer</i>
E : <i>smoking_status</i>	L : <i>physical_activity_level</i>
F : <i>hepatitis_b</i>	M : <i>diabetes</i>
G : <i>hepatitis_c</i>	N : <i>liver_cancer</i> (label)

Pada tabel 1 di atas, tabel tersebut menampilkan sepuluh data pertama dari dataset penyakit kanker hati yang digunakan dalam penelitian ini. Terlihat pada tabel, bahwa dataset terdiri dari 14 kolom atribut, termasuk atribut yang dinyatakan sebagai label klasifikasi yaitu *liver_cancer* yang ada pada kolom N. Label klasifikasi ini menunjukkan status diagnosis pasien, yaitu terindikasi penyakit kanker hati atau tidak terindikasi penyakit kanker hati. Dataset tersebut merupakan sampel awal yang digunakan untuk proses *pre-processing* dan implementasi ke dalam model algoritma KNN.

3.2. Pre-processing Data

Pada tahapan ini bertujuan untuk mengubah data mentah yang belum terstruktur (*unstructured*) menjadi data text yang lebih terstruktur sehingga siap untuk di analisis oleh model algoritma. Proses yang dilakukan yaitu menangani data yang kosong atau tidak lengkap agar tidak menyebabkan eror dan bias pada proses pemodelan.

3.3. Normalisasi

Tahapan ini penting dilakukan ketika menggunakan pemodelan dengan metode *K-Nearest Neighbors* (KNN), karena algoritma KNN bergantung pada perhitungan jarak *euclidean*. Jarak ini memiliki fitur dengan skala nilai yang besar akan mendominasi perhitungan jarak dan mengurangi akurasi model. Metode normalisasi yang digunakan pada penelitian ini adalah *Z-Transformation*.

3.4. K-Nearest Neighbors (KNN)

Metode KNN bekerja dengan menentukan dan mencari nilai *k* terdekat. Nilai *k* terbaik dapat mengurangi efek kesalahan pada proses klasifikasi, namun membuat model menjadi tidak maksimal.

Diperlukan suatu sistem klasifikasi sebagai sistem yang dapat mencari informasi. Adapun rumus jarak atau *Euclidean* seperti yang ditunjukkan persamaan berikut:

$$euc = \sqrt{(a_1 - b_1)^2 + \dots + (a_n - b_n)^2} \quad (4)$$

Dimana $a = a_1, a_2, \dots, a_n$ dan $b = b_1, b_2, \dots, b_n$ menyatakan nilai n atribut dari dua register. Untuk atribut dengan nilai kelas. Skor diprediksi untuk suatu jenis berdasarkan peringkat tertinggi dari tetangga sekitarnya.

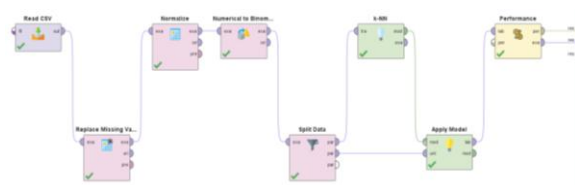
Pada penelitian ini, nilai k yang digunakan dalam pemodelan adalah $k = 3$. Nilai ini merupakan nilai terbaik dari dataset penyakit kanker hati. Sehingga, dengan pemilihan nilai k terbaik, maka hasil akurasi dari evaluasi model *confusion matrix* juga akan meningkat.

3.5. Evaluasi Model

Pada tahapan evaluasi model, proses dilakukan setelah pengolahan data dilakukan menggunakan algoritma *K-Nearest Neighbors* (KNN). Tahapan ini penting dalam proses machine learning karena dapat menilai performa model klasifikasi, termasuk mendeteksi *overfitting* atau *underfitting* pada model. Evaluasi model dilakukan dengan mencocokkan hasil prediksi yang dihasilkan oleh model terhadap data yang sudah memiliki label sebenarnya, kemudian dinilai kinerjanya melalui metrik evaluasi [17].

4. HASIL DAN PEMBAHASAN

Penelitian ini menggunakan 5000 data penyakit kanker hati. Data tersebut akan dilakukan klasifikasi menggunakan bantuan *RapidMiner*.



Gambar 1. Model *RapidMiner*

Pada gambar di atas, tahapan awal pada model *RapidMiner* adalah penggunaan operator *Read CSV* untuk membaca dataset penyakit kanker hati. Atribut yang digunakan sebagai label (*liver_cancer*) harus di atur terlebih dahulu *role* nya. Lalu, untuk 13 atribut lainnya dibiarkan menjadi atribut reguler. Atribut dari label harus diubah menggunakan operator *Numerical to Binomial* agar tipe data berbentuk binomial (0/1 atau 'Yes/No'). Setelah pemberian atribut, tahap selanjutnya adalah normalisasi, dan pembersihan data.

Kemudian, dari total keseluruhan data yang sudah dilakukan tahap pembersihan akan di bagi menjadi 70% yang digunakan untuk data *training*, dan 30% untuk data *testing*. Hasilnya berupa tabel *confusion matrix* yang menguji performa dari model

algoritma *K-Nearest Neighbors* dengan nilai $k = 3$. Berikut tabel hasil dari pengujian dan evaluasi model.

Tabel 2. Hasil evaluasi model *K-Nearest Neighbors*

	True False	True True	Class Precision
Pred. false	1132	163	87,41%
Pred. true	41	164	80%
Class recall	96,50%	50,15%	
Accuracy			86,40%

Berdasarkan Tabel 2 di atas, hasil pengujian model menggunakan algoritma *K-Nearest Neighbors* (KNN) menunjukkan performa klasifikasi yang tergolong baik dengan tingkat akurasi keseluruhan mencapai 86,40%. Nilai akurasi ini mengindikasikan bahwa sebagian besar data uji berhasil diklasifikasikan dengan benar oleh model. Secara lebih rinci, model mampu memprediksi 1.132 data sebagai kategori negatif (false) dengan tepat, serta 164 data sebagai kategori positif (true) secara akurat. Hal ini menunjukkan bahwa model memiliki kemampuan yang cukup baik dalam membedakan antara pasien yang tidak terindikasi penyakit dan yang terindikasi.

Nilai Class Recall sebesar 96,50% pada label false menunjukkan bahwa model sangat efektif dalam mengidentifikasi pasien yang benar-benar tidak memiliki penyakit, sehingga kemungkinan terjadinya kesalahan berupa false negative pada kelas ini relatif kecil. Di sisi lain, nilai recall pada kelas true sebesar 50,15% mengindikasikan bahwa masih terdapat sejumlah kasus positif yang belum berhasil terdeteksi dengan optimal oleh model. Sementara itu, nilai Class Precision sebesar 87,41% pada prediksi negatif dan 80% pada prediksi positif menunjukkan bahwa prediksi yang dihasilkan model cukup dapat dipercaya, khususnya dalam mengonfirmasi pasien yang benar-benar terindikasi penyakit liver cancer.

Secara keseluruhan, performa model KNN ini sudah cukup baik terutama dalam mendeteksi kasus negatif, namun masih perlu dilakukan peningkatan pada kemampuan model dalam mengidentifikasi kasus positif agar risiko kesalahan diagnosis, khususnya pada pasien yang sebenarnya sakit tetapi terklasifikasi sebagai sehat, dapat diminimalkan.

5. KESIMPULAN DAN SARAN

Kesimpulan dari hasil penelitian yang sudah dilakukan terhadap klasifikasi penyakit kanker hati menggunakan model algoritma KNN mendapatkan hasil akurasi sebesar 86,40%. Ketepatan memprediksi (*precision*) sebesar 80% menunjukkan potensi model sebagai alat bantu diagnosis awal, dan nilai *class recall* sebesar 96,50% dalam mengidentifikasi subjek sehat. Hal ini menunjukkan bahwa model cukup baik dalam melakukan klasifikasi untuk 14 atribut yang mengindikasikan adanya penyakit kanker hati.

Berdasarkan hasil penelitian, terdapat beberapa poin yang masih bisa dipertimbangkan untuk penelitian selanjutnya. Peneliti menyarankan untuk

melakukan komparasi model algoritma selain KNN untuk melihat potensi peningkatan akurasi pada dataset yang sama. Penambahan jumlah data juga dapat dilakukan untuk penelitian selanjutnya agar model dapat mempelajari pola klinik secara mendalam.

DAFTAR PUSTAKA

- [1] I. Fauzi And C. Harmon, "Analisis Kinerja Algoritma K-Nearest Neighbor Dalam Klasifikasi Penyakit Liver," *J. Tek. Inform. Unika St. Thomas*, Pp. 357–364, 2025.
- [2] A. E. Ginting, A. F. Fadililah, M. Kusumah, Y. Nuryamin, And A. Priyatna, "Analisis Prediksi Penyakit Hepatitis Menggunakan Algoritma Decision Tree Dan Naive Bayes," *J. Nas. Teknol. Komput.*, Vol. 5, No. 4, Pp. 1083–1091, 2025.
- [3] M. Al Fachrozi And K. D. Tania, "Comparison Of Naive Bayes, Svm, K-Nn, Decision Tree, And Random Forest In Sentiment Analysis Based On Seabank Application Aspects," *Jurteksi (Jurnal Teknol. Dan Sist. Informasi)*, Vol. 12, No. 1 Se-Articles, Pp. 69–80, Dec. 2025, Doi: 10.33330/Jurteksi.V12i1.4189.
- [4] H. A. D. Fasnuari, H. Yuana, And M. T. Chulkamdi, "Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Penyakit Diabetes Melitus: Studi Kasus: Warga Desa Jatitengah," *Antivirus J. Ilm. Tek. Inform.*, Vol. 16, No. 2, Pp. 133–142, 2022.
- [5] D. S. Sipayung And S. Atika, "Analisis Akurasi Algoritma K-Nearest Neighbor Untuk Diagnosis Penyakit Jantung Pada Lansia," *J. Sist. Inf. Dan Ilmu Komput.*, Vol. 8, No. 2, 2025.
- [6] A. D. Sembiring Depari, K. D. Tania, And P. E. Sevtiyuni, "Penerapan Metode Machine Learning Dan Teknik Smote Untuk Prediksi Diabetes," *J. Sist. Komput. Dan Inform.*, Vol. 7, No. 2 Se-Articles, Pp. 436–447, Dec. 2025, Doi: 10.30865/Json.V7i2.9032.
- [7] Z. Fatah, "Klasifikasi Penyakit Tbc Menggunakan Metode K-Nearest Neighbors (Knn)," *J. Mhs. Tek. Inform.*, Vol. 4, No. 2, Pp. 114–118, 2025.
- [8] E. Saputri, "Teknik Dan Aplikasi Data Mining Di Indonesia: Tinjauan Literatur Satu Dekade (2015-2024)," *It-Explore J. Penerapan Teknol. Inf. Dan Komun.*, Vol. 4, No. 2, Pp. 138–149, 2025.
- [9] M. Y. Zidane, B. N. Sari, I. Maulana, A. Primaya, And G. Garno, "Penerapan Data Mining Dalam Klasifikasi Data Transaksi Produk Koperasi Di Smk Pgri 2 Karawang," *Jati (Jurnal Mhs. Tek. Inform.*, Vol. 9, No. 1, Pp. 263–269, 2025.
- [10] M. A. Shinami And S. Bahri, "Klasifikasi Penyakit Kanker Payudara Menggunakan Metode K-Nearest Neighbors (Knn)," *J. Fourier*, Vol. 12, No. 2, Pp. 79–85, 2023.
- [11] F. A. H. Airi, T. Suprpti, And A. Bahtiar, "Komparasi Metode Klasifikasi Data Mining Untuk Prediksi Penyakit Stroke," *E-Link J. Tek. Elektro Dan Inform.*, Vol. 18, No. 1, Pp. 73–79, 2023.
- [12] I. N. Abrar, A. Abdullah, And S. Sucipto, "Klasifikasi Penyakit Liver Menggunakan Metode Elbow Untuk Menentukan K Optimal Pada Algoritma K-Nearest Neighbor (K-Nn)," *Vol*, Vol. 12, Pp. 218–228, 2023.
- [13] N. Ikhsan, "Klasifikasi Penyakit Diabetes Menggunakan Metode K-Nearest Neighbors (Knn) Berdasarkan Data Klinis," *Infotech J. Inform. Teknol.*, Vol. 7, No. 1, Pp. 19–26, 2026.
- [14] Z. Y. Nabila, K. P. A. Simaremare, And M. F. Billyahsan, "Analisis Akurasi Prediksi Tuberkulosis Menggunakan Algoritma Knn," In *Seminar Nasional Informatika (Senatika)*, 2026, Pp. 199–204.
- [15] K. W. Kayohana, "Klasifikasi Penyakit Hati Menggunakan Random Forest Dan Knn," *Jati (Jurnal Mhs. Tek. Inform.*, Vol. 8, No. 4, Pp. 7924–7929, 2024.
- [16] A. Homaidi And Z. Fatah, "Implementasi Metode K-Nearest Neighbors (Knn) Untuk Klasifikasi Penyakit Jantung," *G-Tech J. Teknol. Terap.*, Vol. 8, No. 3, Pp. 1720–1728, 2024.
- [17] M. A. Devira, N. Mintarsih, And Y. M. Arianti, "Klasifikasi Gigi Berlubang Dan Gigi Tidak Berlubang Menggunakan Convolutional Neural Network (Cnn) Dengan Pendekatan Transfer Learning Nasnetmobilearia," *Portal Ris. Dan Inov. Sist. Perangkat Lunak*, Vol. 4, No. 1, Pp. 52–59, 2026.