
STUDI LITERATUR DATA MINING DALAM BIDANG KESEHATAN UNTUK ANALISIS DAN PREDIKSI PENYAKIT DALAM

Fira Yuniar¹, Hetty Rohayani²

^{1,2}Fakultas Sains dan Teknologi, Universitas Muhammadiyah Jambi, Jambi, Indonesia

Email: ¹firayuniar05@email.com, ²hettyrohayani@email.com,
Email Penulis Korespondensi: firayuniar05@gmail.com

Keyword :

Data Mining,
Prediksi Penyakit,
Algoritma,
Naïve Bayes,
Decision Tree,
Studi Literatur

ABSTRACT

Perkembangan teknologi informasi telah mendorong pertumbuhan data yang sangat pesat di bidang kesehatan, khususnya data rekam medis pasien yang tersimpan dalam jumlah besar di berbagai fasilitas kesehatan. Data tersebut menyimpan potensi pengetahuan yang sangat berharga jika dianalisis secara tepat. Salah satu pendekatan yang efektif untuk mengekstrak pengetahuan dari data tersebut adalah data mining, yaitu serangkaian metode untuk menemukan pola tersembunyi dan informasi penting dari kumpulan data yang besar, sehingga dapat digunakan dalam pengambilan keputusan medis secara lebih akurat dan efisien. Studi literatur ini membahas penerapan berbagai algoritma data mining, seperti Naïve Bayes, Decision Tree (C4.5 dan C5.0), K-Nearest Neighbor, dan Random Forest dalam analisis dan prediksi penyakit dalam. Hasil penelitian menunjukkan bahwa algoritma Decision Tree C4.5 yang dioptimasi dengan Particle Swarm Optimization (PSO) mampu mencapai akurasi hingga 99,67% untuk prediksi hepatitis C, sedangkan Naïve Bayes efektif dalam prediksi stroke dan penyebaran COVID-19 dengan akurasi tinggi. Selain itu, Random Forest terbukti unggul dalam prediksi penyakit stroke dan diabetes dengan akurasi di atas 90%. Studi ini memberikan gambaran komprehensif tentang efektivitas berbagai metode data mining dan menjadi referensi penting bagi pengembangan aplikasi prediksi penyakit di bidang kesehatan.

1. PENDAHULUAN

Perkembangan teknologi informasi telah menghasilkan pertumbuhan data yang sangat pesat di berbagai bidang, termasuk kesehatan. Data rekam medis pasien yang tersimpan dalam jumlah besar di fasilitas kesehatan menyimpan potensi pengetahuan yang sangat berharga apabila dianalisis dengan tepat. Salah satu pendekatan yang dapat digunakan untuk mengekstrak pengetahuan dari data tersebut adalah data mining, yaitu serangkaian metode untuk menemukan pola tersembunyi dan informasi penting dari kumpulan data yang besar. Data mining adalah sesuatu yang dapat di analisa pada suatu koleksi data (database) sehingga dapat ditemukan suatu pola yang menarik dengan tujuan mengfilter informasi dan pengetahuan yang akurat dan potensial,serta dapat dipahami dan berguna bagi pengambilan keputusan [1].

Penelitian oleh Haris Bugis “Metode Naïve Bayes Untuk Memprediksi Penyakit Stroke, Tahun 2022 dalam bidang kesehatan, data mining telah dimanfaatkan untuk berbagai keperluan, seperti prediksi penyakit, identifikasi pola penyebaran penyakit, serta evaluasi efektivitas pengobatan. Dengan memanfaatkan data mining, informasi seperti usia, jenis kelamin, tekanan darah, hingga riwayat medis pasien dapat digunakan

untuk memprediksi kemungkinan seseorang menderita penyakit tertentu dan membantu pengambilan keputusan medis secara lebih tepat dan efisien. Melihat pesatnya perkembangan di dunia Kesehatan, maka telah dilakukan penelitian untuk memprediksi penyakit stroke berdasarkan gaya hidup serta lingkungan masyarakat. Dengan tujuan mengklasifikasi dan memprediksi orang-orang tentang apakah orang tersebut akan berpotensi terkena penyakit stroke atau tidak. Peneliti menggunakan metode klasifikasi Naive Bayes yang merupakan salah-satu metode sederhana untuk mengklasifikasi dan memprediksi sebuah data[2].

Penelitian dari Fely Dany Prasetya, Handoyo Widi Nugroho, Joko Triloka, tentang “Analisa Data Mining Untuk Prediksi Penyakit Hepatitis C Menggunakan Algoritma Decision Tree C.45 tahun 2022, penelitian ini dengan Particle Swarm Optimization, Dalam penelitian ini metode yang digunakan adalah teknik klasifikasi dalam data mining dengan penggunaan metode C4.5 Decision Tree Algorithm hanya pada klasifikasi untuk mendapatkan nilai akurasi yang tinggi. yang paling optimal adalah 99,35%. Kemudian pengujian penggunaan algoritma Decision Tree C4.5 dengan optimasi Algoritma PSO menghasilkan nilai akurasi paling optimal sebesar 99,67% dengan demikian optimasi dengan Algoritma PSO dapat meningkatkan akurasi algoritma DecisionTree C4.5 sehingga menghasilkan hasil yang lebih akurasi optimal dengan selisih akurasi 0,32%[3].

Menurut penelitian dari Muhammad Sholeh, Dina Andayati, Rr. Yuliana Rachmawati dengan judul “Data Mining Model Klasifikasi Menggunakan Algoritma K-Nearest Neighbor dengan Normalisasi Untuk Prediksi Penyakit Diabetes” tahun 2022 penelitian ini tujuannya dengan Proses data mining menggunakan model klasifikasi dengan menggunakan algoritma k-NN. Datasheet yang digunakan merupakan datasheet public yaitu datasheet penyakit diabetes yang terdiri dari 768 record dan 8 atribut. Hasil dari pembuatan model ini menunjukkan proses normalisasi dapat memberikan nilai akurasi yang lebih baik. Model yang dikembangkan tanpa normalisasi menghasilkan nilai k=5 dengan akurasi 70%, normalisasi dengan metode Z-Score menghasilkan nilai k=21 dengan akurasi 72%, normalisasi dengan Min Max menghasilkan nilai k=3 dengan akurasi 74%. Model yang direkomendasi merupakan mode k-NN dengan nilai k=3[4].

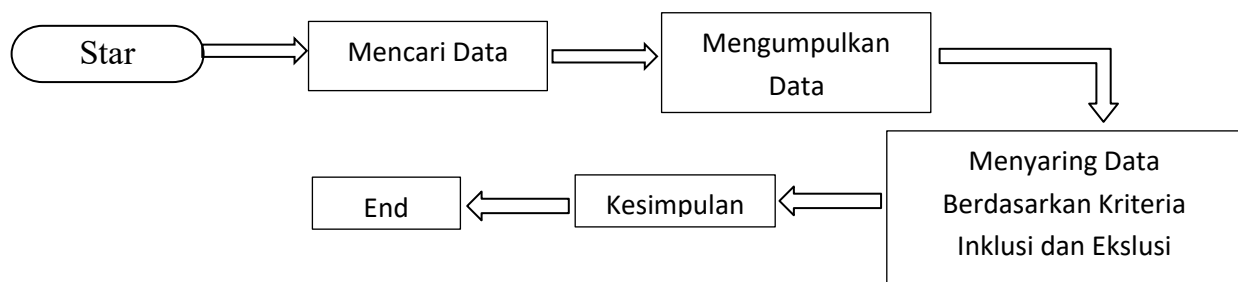
Menurut dari Hilih Firda, Rima Athiyah, Salsabila, Muhammad Ihsan Jambak, dengan judul “Perbandingan Algoritma Klasifikasi Data Mining Model C4.5 Dan Naive Bayes Untuk Prediksi Keganasan Kanker Payudara” tahun 2024. Penelitian ini bertujuan untuk membandingkan kinerja kedua algoritma tersebut dalam konteks medis ini, dengan referensi tambahan dari penelitian terdahulu. Hasil penelitian ini menunjukkan bahwa algoritma C4.5 mencapai akurasi sebesar 94,15%, sementara Naive Bayes mencapai 92,40%. Hal ini menegaskan bahwa algoritma C4.5 memiliki kinerja yang lebih unggul dibandingkan dengan Naive Bayes[5].

Penelitian dari Fazrin Meila Azzahra Sofyan, Affani Putri Riyandoro, Devi Fitriani Maulana, Jajam Haerul Jaman tahun 2023 tentang “Penerapan Data Mining dengan Algoritma C5.0 Untuk Prediksi Penyakit Stroke,” penelitian ini bertujuan Hasil Data yang digunakan dalam penelitian ini adalah dataset yang berisi informasi medis dan faktor risiko yang terkait dengan stroke. Hasil dari penelitian ini berupa Decision Tree (pohon keputusan) dengan nilai accuracy, recall, dan precision dengan melakukan split data 80% (data training) - 20% (data testing) hasil nilai Accuracy yang diperoleh sebesar 95%, Recall = 96%, dan Precision = 99%.[6]

Tujuan dari penelitian ini adalah untuk melakukan studi literatur yang komprehensif mengenai penerapan teknik data mining dalam bidang kesehatan, khususnya untuk analisis dan prediksi penyakit dalam. Penelitian ini bertujuan mengidentifikasi dan membandingkan berbagai algoritma data mining yang telah digunakan dalam penelitian sebelumnya untuk meningkatkan akurasi prediksi penyakit. Dengan demikian, penelitian ini diharapkan dapat memberikan gambaran yang jelas tentang metode data mining yang paling optimal dan aplikatif dalam diagnosis dan prediksi penyakit, sekaligus memberikan rekomendasi bagi pengembangan penelitian dan aplikasi di bidang kesehatan.

2. METODOLOGI PENELITIAN

Metode yang digunakan dalam penelitian ini adalah studi literatur. Data yang digunakan berupa data dari hasil penelitian yang memfokuskan pada 20 referensi jurnal untuk memperoleh hasil dari studi kasus setai jurnal tersebut. Teknik analisis data menggunakan analisis . Peneliti mengumpulkan, memeriksa, dan menganalisis informasi dari berbagai sumber untuk mendapatkan kesimpulan yang sesuai dengan tujuan penelitian



Dalam jurnal ini, metode penelitian yang digunakan adalah Systematic Literature Review (SLR), yaitu pendekatan terstruktur untuk mengidentifikasi, mengevaluasi, dan menganalisis literatur yang terkait dengan judul jurnal. Studi tinjauan pustaka ini diharapkan dapat memberikan wawasan mengenai metode data mining

1. Mencari Data,

Langkah pertama yang harus dilakukan pada tahap ini adalah mencari data yang diperlukan untuk analisis atau perbandingan dalam temuan yang ada dalam penelitian yang akan dikaji. Dalam Jurnal ini beberapa hasil penelitian terpilih yang diterbitkan sehingga dapat menjadi acuan bagi perkembangan pengetahuan beberapa hasil penelitian terpilih telah dipublikasikan dan dapat dijadikan sebagai referensi utama. Hasil-hasil penelitian tersebut tidak hanya memberikan gambaran mengenai perkembangan pengetahuan di bidang terkait, tetapi juga menjadi acuan penting dalam memahami tren, metodologi, serta temuan-temuan terbaru yang relevan. Dengan demikian, pemilihan dan analisis data dari penelitian-penelitian terpilih ini akan memperkaya landasan teori dan memperkuat argumen dalam penelitian yang sedang dilakukan[7][8].

2. Mengumpulkan Data

Setelah mencari data, langkah selanjutnya yang harus dilakukan adalah pada pengumpulan data untuk memilih jurnal yang akan menjadi data untuk studi literature ini, hal ini dapat menemukan data dari referensi jurnal untuk di analisis dan dibandingkan[8].

3. Menyaring Data Berdasarkan Kriteria Inklusi dan Ekskusi

Dari 20 jurnal dilakukan kelayakan Menyaring data berdasarkan kriteria inklusi dan ekskusi didapatkan yang ditemukan lalu dibandingkan kembali dan diuraikan pada tabel yang dianalisis secara mendalam sehingga dapat memperoleh hasil yang diinginkan [9].

4. Kesimpulan

Setelah dilakukan menyaring data berdasarkan kriteria inklusi dan eksklusi, dilakukan tahap kesimpulan. Kesimpulan yang ditemukan juga diverifikasi selama melakukan penelitian berlangsung, dengan cara menganalisis teori sehingga pada akhirnya kesimpulan akan muncul[10].

3. HASIL DAN PEMBAHASAN

No	Judul	Tahun Terbit dan Penerbit	Data	Metode	Hasil
1.	Analisa Data Mining Untuk Prediksi Penyakit Hepatitis C Menggunakan Algoritma Decision Tree C.45 Dengan Particle Swarm Optimization[3]	Fely Dany Prasetyala, Handoyo Widi Nugroho, Joko Triloka 2022	Memprediksi pasien yang terinfeksi hepatitis C	Klasifikasi data mining C4.5 Decision Tree	Pengujian penggunaan algoritma Decision tree C4.5 dengan optimasi algoritma PSO menghasilkan nilai akurasi paling optimal sebesar 99,67%
2.	Metode Naïve Bayes Untuk Memprediksi Penyakit Stroke [2]	Haris, 2022	Memprediksi penyakit stroke	Klasifikasi data mining Naïve Bayes	Metode naïve bayes tingkat akurasi sangat tinggi 93,94%
3.	Prediksi Penyebaran Penyakit TBC dengan Metode K-Means Clustering Menggunakan Aplikasi Rapidminer [10]	Muhammad Farid Iqbal Al-Rizki, Ida Widaningrum,, Ghulam Asrofi Buntoro, 2020	menentukan pola penyebaran penyakit Tuberculosis (TBC)	algoritma K-Means Clustering dengan Cross-Industry Standard Process for Data Mining (CRISP-DM)	accuracy dan performa Area under the Curve (AUC) dari K-Means sebesar 84,13% dan 0,837
4.	Analisis Metode Decision Tree Dan Naïve Bayes Pada Pasien Penyakit Liver[11]	Damar Adji Sodikin, El Thariq Is'ad, Rangga Prayoga, Ahmad Nur Ihsan Purwanto, 2024	Data 416 catatan pasien liver/hati dan 167 catatan pasien tanpa penyakit liver/hati	Klasifikasi data mining metode naïve bayes	nilai akurasi optimal sebesar 70.29%, 67.05%. salah satu metode yang dapat memecahkan masalah penentuan penyakit liver.
5.	Implementasi Algoritma Decision Tree C4.5 dan Support Vector Regression untuk Prediksi Penyakit Stroke[12]	Firman Akbar, Hanif Wira Saputra, Adhitya Karel Maulaya, Muhammad Fikri Hidayat, Rahmaddeni, 2022	Mengidentifikasi pengidap Penyakit Stroke	Algoritma Decision Tree C4.5 dan Support Vector Regression	algoritma Decision Tree C4.5 terhadap rasio 70 : 30 bernilai 0.235 dan algoritma Support Vector Regression terhadap rasio 70 : 30 bernilai 39,09%

6.	Analisa Data Mining Untuk Prediksi Penyakit Ginjal Kronik Dengan Algoritma Regresi Linier[13]	Angga Kurniadi Hermawan, Agung Nugroho, Edora, 2023	Memprediksi penyakit ginjal kronis mengguna	Algoritma regresi linier	algoritma regresi linear mampu memberikan hasil yang baik akurasi 81,00%.
7.	Komparasi Algoritma Data Mining dan Perancangan Aplikasi Prediksi Harapan Hidup Pasien Gagal Jantung[14]	Agustiena Merdekawati, 2022	memprediksi harapan hidup penyakit gagal jantung	algoritma C4.5 dengan PSO (Particle Swarm Optimization)	algoritma C4.5 dan PSO memiliki akurasi grafik ROC sebesar 1.00%,
8.	Implementasi Data Mining untuk Memprediksi Kelulusan Mahasiswa Menggunakan Algoritma K-Nearest Neighbor di Universitas Muhammadiyah Jambi[15]	Shandy Amandha, Hetty Rohayani, Kevin Kurniawansyah, 2024	memprediksi kelulusan mahasiswa	Algoritma K-Nearest Neighbor (K-NN)	pendekatan K-NN dalam penelitian ini menghasilkan nilai akurasi sebesar 93,33%.
9.	Penerapan Algoritma K-Means untuk Mengidentifikasi Pola Penjualan Frozen Food yang Paling Populer [16]	Dwi Nopriyani, Hetty Rohayani, Zulfikri Akbar Fakultas, 2024	Mengidentifikasi Pola Penjualan Frozen Food yang Paling Populer	Algoritma K-Means	jumlah cluster positif (C1) 79,5%. Sedangkan persentase jumlah cluster negatif (C2) 19,7%
10.	Komparasi Metode Klasifikasi Data Mining Untuk Prediksi Penyakit Stroke [16]	Fitri Adha Hariyati Airi, Tati Suprapti, Agus Bahtiar, 2023	Prediksi Penyakit Stroke	algoritma Naive Bayes, K-Nearest Neighbor dan Random Forest. algoritma K-NN	Hasil yang diperoleh pada penelitian ini untuk algoritma Naive Bayes sebesar 71.9% accuracy, 71.7% precision, 71.9% recall. Sedangkan untuk algoritma K-NN mendapatkan nilai accuracy sebesar 73.6%, precision sebesar 73%, recall 73.6% dan untuk algoritma Random Forest mendapatkan nilai accuracy sebesar 92.5%, precision 92.5%, recall 92.5%.
11.	Perbandingan Akurasi Algoritma Naïve Bayes Dan Algoritma Xgboost Pada Klasifikasi Penyakit Diabetes [17]	Muhammad Kaddafi Nasution, Rd. Rohmat Saedudin, Vandha	Prediksi penyakit Diabetes	Algoritma Naïve Bayes dan Algoritma XGBoost	Algoritma Naïve Bayes dengan nilai hasil akurasi model sebesar 79.68% dan Algoritma XGBoost memiliki performa yang lebih

		Pradwiyasma Widartha			baik dengan nilai hasil akurasi yang didapat sebesar 90.10%.
12	Prediksi Penyebaran Virus COVID-19 Dari Hasil PCR Menggunakan Metode Naïve Bayes [18]	Hetty Rohayani,Sitti Nur Alam, Muhammad Fauzi, Rico	Prediksi Penyebaran Virus COVID-19	Algoritma Naïve Bayes	Tingkat akurasi yang tinggi dalam mengklasifikasikan status terkonfirmasi Covid-19 yaitu 97%
13.	Komparasi Metode Deep Learning, Naïve Bayes dan Random Forest untuk Prediksi Penyakit Jantung[19]	Ivana Alhabib , Ahmad Faqih ,Fatihanursari Dikananda	Prediksi Penyakit Jantung	Deep Learning, Naives Bayes dan Random Forest	Menghasilkan akurasi sebesar 83,49%.
14.	Prediksi Kanker Paru-Paru menggunakan Algoritme Random Forest Decision Tree [20]	Rafly Dwi Marzuq, Satrio Agung Wicaksono, Nanang Yudi Setiawan	Prediksi Kanker Paru-Paru	Algoritma Random Forest Decision Tree	Rata rata akurasi yaitu 88,09%
15.	Deteksi Dini Penyakit Diabetes dengan Menggunakan Algoritma Random Forest[20]	Duwi Cahya Putri Buani	Deteksi Dini Penyakit Diabetes	Algoritma Random Forest	Nilai akurasi dari Random Forest sebesar 98,78%

4. KESIMPULAN

Dari 15 artikel yang dipilih ada beberapa penelitian yang sangat akurat untuk judul ini Studi Literatur Data Mining Dalam Bidang Kesehatan Analisis Dan Prediksi Penyakit Dalam Penelitian yang dilakukan oleh beberapa peneliti yaitu, Fely Dany Prasetya, Handoyo Widi Nugroho, dan Joko Triloka (2022) menunjukkan bahwa penggunaan algoritma Decision Tree C4.5 yang dikombinasikan dengan optimasi Particle Swarm Optimization (PSO) sangat efektif dalam memprediksi pasien yang terinfeksi hepatitis C. Metode ini berhasil mencapai tingkat akurasi yang sangat tinggi, yaitu 99,67%, yang menandakan kemampuan model dalam mengklasifikasikan data pasien dengan sangat tepat. Dengan demikian, kombinasi Decision Tree C4.5 dan PSO dapat dijadikan solusi andal dalam sistem prediksi penyakit hepatitis C berbasis data mining.

Penelitian yang dilakukan oleh Haris (2022) menunjukkan bahwa Metode Naïve Bayes sangat efektif untuk memprediksi penyakit stroke. Algoritma klasifikasi data mining Naïve Bayes yang diterapkan dalam studi ini mampu menghasilkan tingkat akurasi yang sangat tinggi, yaitu sebesar 93,94%. Hasil ini mengindikasikan bahwa Naïve Bayes merupakan pendekatan yang kuat dan dapat diandalkan untuk identifikasi risiko atau prediksi penyakit stroke.

Penelitian oleh Fitri Adha Hariyati Airi, Tati Suprpti, dan Agus Bahtiar (2023) melakukan komparasi tiga metode klasifikasi data mining untuk prediksi penyakit stroke, yaitu Naive Bayes, K-Nearest Neighbor (K-NN), dan Random Forest. Hasil penelitian menunjukkan bahwa, Naive Bayes memiliki akurasi sebesar 71,9%, precision 71,7%, dan recall 71,9%, K-Nearest Neighbor (K-NN) sedikit lebih baik dengan akurasi 73,6%, precision 73%, dan recall 73,6%, Random Forest memberikan performa terbaik dengan akurasi, precision, dan recall yang sama-sama tinggi yaitu 92,5%.

Dengan demikian, algoritma Random Forest terbukti paling unggul dalam memprediksi penyakit stroke dibandingkan Naive Bayes dan K-NN, sehingga lebih direkomendasikan untuk digunakan dalam aplikasi prediksi penyakit stroke.

Penelitian oleh Hetty Rohayani, Sitti Nur Alam, Muhammad Fauzi, dan Rico menunjukkan bahwa algoritma Naive Bayes efektif digunakan untuk prediksi penyebaran virus COVID-19 berdasarkan hasil PCR. Metode ini mampu mengklasifikasikan status terkonfirmasi COVID-19 dengan tingkat akurasi yang tinggi, yaitu sebesar 97%. Hal ini menegaskan bahwa Naive Bayes merupakan pendekatan yang andal dan efisien dalam membantu deteksi dan pengendalian penyebaran COVID-19.

Penelitian Duwi Cahya Putri Buani, menunjukkan bahwa algoritma Random Forest sangat efektif dalam memprediksi risiko diabetes dengan tingkat akurasi yang tinggi, yaitu sekitar 98,78%

Hal ini bisa dilihat dalam analisis dan memprediksi penyakit dalam penelitian yang banyak digunakan adalah algoritma metode naive bayes.

REFERENCES

- [1] L. Setiyani, M. Wahidin, D. Awaludin, and S. Purwani, "Analisis Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Data Mining Naive Bayes : Systematic Review," *Fakt. Exacta*, vol. 13, no. 1, p. 35, 2020, doi: 10.30998/faktorexacta.v13i1.5548.
- [2] H. Bugis, "Metode Naive Bayes Untuk Memprediksi Penyakit Stroke," *J. SISKOM-KB (Sistem Komput. dan Kecerdasan Buatan)*, vol. 6, no. 1, pp. 8–14, 2022, doi: 10.47970/siskom-kb.v6i1.317.
- [3] F. D. Prasetya, H. W. Nugroho, and J. Triloka, "Analisa Data Mining Untuk Prediksi Penyakit Hepatitis C Menggunakan Algoritma Decision Tree C.45 Dengan Particle Swarm Optimization," *Semin. Nas. Has. Penelit. dan Pengabd. Masy.*, no. April 1989, pp. 199–209, 2022, [Online]. Available: <http://archive.ic>
- [4] M. Sholeh, D. Andayati, and R. Y. Rachmawati, "Data Mining Model Klasifikasi Menggunakan Algoritma K-Nearest Neighbor Dengan Normalisasi Untuk Prediksi Penyakit Diabetes," *TeKa*, vol. 12, no. 02, pp. 77–87, 2022, doi: 10.36342/teika.v12i02.2911.
- [5] H. Firda, R. Athiyah, and M. Ihsan Jambak, "Perbandingan Algoritma Klasifikasi Data Mining Model C4.5 Dan Naive Bayes Untuk Prediksi Keganasan Kanker Payudara," *Just-IT*, vol. 14, no. 3, pp. 150–233, 2024, [Online]. Available: <https://jurnal.umj.ac.id/index.php/just-it/index>
- [6] F. M. A. Sofyan, A. P. Riyandoro, D. F. Maulana, and J. H. Jaman, "Penerapan Data Mining dengan Algoritma C5.0 Untuk Prediksi Penyakit Stroke," *J-SISKO TECH (Jurnal Teknol. Sist. Inf. dan Sist. Komput. TGD)*, vol. 6, no. 2, p. 619, 2023, doi: 10.53513/jsk.v6i2.8578.
- [7] L. Belakang *et al.*, "BAB I," pp. 1–5, 2011.
- [8] N. F. Hadi and N. K. Afandi, "Literature Review is A Part of Research," *Sultra Educ. J.*, vol. 1, no. 3, pp. 64–71, 2021, doi: 10.54297/seduj.v1i3.203.
- [9] M. K. Murtiningsih, K. Pandelaki, and B. P. Sedli, "Gaya Hidup sebagai Faktor Risiko Diabetes Melitus Tipe 2," *e-CliniC*, vol. 9, no. 2, p. 328, 2021, doi: 10.35790/ecl.v9i2.32852.
- [10] M. F. I. Al-Rizki, I. Widaningrum, and G. A. Buntoro, "Prediksi Penyebaran Penyakit TBC dengan Metode K-Means Clustering Menggunakan Aplikasi Rapidminer," *JTERA (Jurnal Teknol. Rekayasa)*, vol. 5, no. 1, p. 1, 2020, doi: 10.31544/jtera.v5.i1.2019.1-10.
- [11] D. Adji Sodikin, E. Thaariq Is'ad, R. Prayoga, and A. Nur Ihsan Purwanto, "Analisis Metode Decision

Tree Dan Naïve Bayes Pada Pasien Penyakit Liver,” *Pros. TAU SNARS-TEK Semin. Nas. Rekayasa dan Teknol.*, vol. 3, no. 1, pp. 12–18, 2024, doi: 10.47970/snarstek.v2i1.545.

[12] R. A. Saputra *et al.*, “Detecting Alzheimer’s Disease by the Decision Tree Methods Based on Particle Swarm Optimization,” *J. Phys. Conf. Ser.*, vol. 1641, no. 1, pp. 61–67, 2020, doi: 10.1088/1742-6596/1641/1/012025.

[13] A. Kurniadi Hermawan, A. Nugroho, and Edora, “Analisa Data Mining Untuk Prediksi Penyakit Ginjal Kronik Dengan Algoritma Regresi Linier,” *Bull. Inf. Technol.*, vol. 4, no. 1, pp. 37–48, 2023, doi: 10.47065/bit.v4i1.475.

[14] A. Merdekawati, “Komparasi Algoritma Data Mining dan Perancangan Aplikasi Prediksi Harapan Hidup Pasien Gagal Jantung,” *CSRID J.*, vol. 14, no. 3, pp. 188–202, 2022.

[15] S. Amandha, H. Rohayani, and K. Kurniawansyah, “Implementation of Data Mining for Predicting Student Graduation Using the K-Nearest Neighbor Algorithm at Jambi Muhammadiyah University,” *Indones. J. Artif. Intell. Data Min.*, vol. 7, no. 1, p. 134, 2024, doi: 10.24014/ijaidm.v7i1.26150.

[16] D. Nopriyani, H. Rohayani, and Z. Akbar, “Penerapan Algoritma K-Means untuk Mengidentifikasi Pola Penjualan Frozen Food yang Paling Populer,” *Bull. Comput. Sci. Res.*, vol. 5, no. 1, pp. 98–104, 2024, doi: 10.47065/bulletincsr.v5i1.445.

[17] M. F. Rofie *et al.*, “Perbandingan Akurasi Algoritma Naïve Bayes Dan Algoritma Xgboost Pada Klasifikasi Penyakit Diabetes,” *e-Proceeding Eng.*, vol. 8, no. 5, pp. 9765–9772, 2021, [Online]. Available: <https://journal.ubpkarawang.ac.id/mahasiswa/index.php/ssj/article/view/424/338%0Ahttps://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/15759>

[18] H. Rohayani, S. N. Alam, M. Fauzi, and R. Rico, “Prediksi Penyebaran Virus COVID-19 Dari Hasil PCR Menggunakan Metode Naïve Bayes,” *J. Comput. Syst. Informatics*, vol. 4, no. 1, pp. 109–115, 2022, doi: 10.47065/josyc.v4i1.2577.

[19] I. Alhabib, “Komparasi Metode Deep Learning, Naïve Bayes Dan Random Forest Untuk Prediksi Penyakit Jantung,” *INFORMATICS Educ. Prof. J. Informatics*, vol. 6, no. 2, p. 176, 2022, doi: 10.51211/itbi.v6i2.1881.

[20] R. D. Marzuq, S. A. Wicaksono, and N. Y. Setiawan, “Prediksi Kanker Paru-Paru menggunakan Algoritme Random Forest Decision Tree,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 7, pp. 3448–3456, 2023.