Terbit pada laman web jurnal : **http://ojsamik.amikmitragama.ac.id**

JURNAL JARINGAN SISTEM INFORMASI ROBOTIK (JSR)
Vol. 8 No. 1 TAHUN 2024 E - ISSN : 2579-373X

KOMPARASI ALGORITMA DECISION TREE, SVM, NAIVE BAYES DALAM PREDIKSI PENYAKIT LIVER

Diva Nabila Herisnan[1], Apriliani[2], Eric Dadynata[3], Rahmaddeni[4], Lusiana Efrizoni[5]
[12345]Teknik Informatika, STMIK Amik Riau, Pekanbaru, Indonesia
Email: [1*] divanabillapku123@gmail.com, [2] aprilia17tahun@gmail.com, [3]ericdadynata011@gmail.com, [4]rahmaddeni@sar.ac.id, [5]lusiana@stmik-amik-riau.ac.id

Abstrak
Penelitian ini membandingkan kinerja tiga algoritma klasifikasi dalam memprediksi penyakit hati: Decision Tree, Support Vector Machine (SVM), dan Naive Bayes. Prediksi akurat mengenai masalah kesehatan hati dapat membantu diagnosis dan pengobatan tepat waktu. Sebab, permasalahan ini mempunyai dampak yang cukup besar. Metode ini melatih dan mengevaluasi algoritma dengan kumpulan data klinis dan parameter biokimia. Meskipun Decision Tree memberikan interpretasi model yang lebih baik, SVM mengungguli algoritma lain dalam hal akurasi prediksi. Selain itu, Naive Bayes memiliki kinerja yang baik, terutama dalam menangani asumsi independensi fitur. Singkatnya, ketika seseorang memilih algoritma klasifikasi, mereka harus mempertimbangkan manfaat yang dihasilkan dari akurasi, interpretabilitas, dan asumsi model. Hasil penelitian ini semoga bermanfaat bagi para praktisi kesehatan dalam memilih metode untuk memprediksi penyakit liver dan tingkat akurasi terbaik terdapat pada algoritma Decision Tree menggunakan pembagian data Forward Selection 70:30 dengan tingkat akurasi 67,82%..

Kata Kunci: Hati, Prediksi, Decision Tree, Support Vector Machine (SVM). Naive Bayes.

Abstract
This study compared the performance of three classification algorithms in predicting liver disease: Decision Tree, Support Vector Machine (SVM), and Naive Bayes. Accurate prediction of liver health problems can help with timely diagnosis and treatment. Because this problem has quite a big impact. This method trains and evaluates the algorithm with clinical data sets and biochemical parameters. Although Decision Tree provides better model interpretation, SVM outperforms other algorithms in terms of prediction accuracy. Additionally, Naive Bayes performs well, especially in dealing with feature independence assumptions. In short, when one selects a classification algorithm, they should consider the benefits resulting from accuracy, interpretability, and model assumptions. The results of this research may be useful for health practitioners in choosing a method for predicting liver disease and the best level of accuracy is found in the Decision Tree algorithm using a 70:30 Forward Selection data division with an accuracy level of 67.82%.

Keywords: Liver, Prediction, Decision Tree, Support Vector Machine, Naive Bayes.

1. PENDAHULUAN

Kesehatan adalah salah satu aspek terpenting dalam kehidupan. Sejalan dengan itu, ada banyak penemuan logis sehubungan dengan obat-obatan, peralatan klinis, atau pengungkapan kesehatan baru. Selama bertahun-tahun banyak sekali penyakit yang bermunculan, baik karena infeksi, mikroorganisme, parasit, sel penyakit, atau sumbernya yang berbeda-beda. Salah satunya pada hati, organ tubuh. Penyakit hati digambarkan dengan kerusakan hati yang disebabkan oleh kontaminasi infeksi, zat beracun, atau mikroba, yang menyulitkan hati untuk bekerja secara normal.[1].

Tubuh dilindungi dari beberapa racun yang dapat menyerang kesehatan oleh organ hati. Fungsi hati termasuk penyimpanan glikogen. Faktor dalam perkembangan penyakit hati sedang memoles minuman keras, hati berminyak, sifat turun temurun dari wali, diabetes dan kegemukan dalam tubuh[2]. Dampak Peradangan, pembekuan darah, dan gagal hati merupakan cedera yang berhubungan dengan hati. Jika tidak ditangani segera, hal ini dapat membahayakan bagi tubuh. Oleh sebab itu, banyak temuan ilmiah mengenai obat, alat kesehatan, serta penemuan kesehatan baru. Sejak lama, banyak penyakit telah muncul, entah itu akibat virus, bakteri, parasite, sel kanker, atau sumber lainnya[3].

Hati adalah organ tubuh kita yang terbesar dan paling penting. Iritasi yang disebabkan oleh infeksi, mikroorganisme, atau zat berbahaya

sehingga hati tidak dapat bekerja sebagaimana mestinya dan mudah dideteksi pada tahap awal penyakit hati adalah penyakit hati. Pasien dengan penderita penyakit hati akan hidup lebih lama jika mereka ditangani lebih awal dari jadwal. Namun sayangnya, tidak semua ahli kesehatan memiliki keahlian luar biasa dalam mengambil kesimpulan klinis. Karena pembelajaran mesin banyak digunakan dalam bidang medis, sistem diagnosis otomatis secara medis mungkin sangat membantu.[4].

Penyakit hati yang sangat hebat mempengaruhi kemampuan hati, penyakit hati dapat dikenali dari adanya efek samping permasalahan klinis dan aktual yang muncul pada pasien, Efek samping klinis dilihat dari apa yang dirasakan pasien, sedangkan efek samping sebenarnya tidak sepenuhnya diatur dari keadaan tubuh pasien, Ada banyak efek samping dari penyakit hati dan kompleks, serta penyakit hati kedekatan efek samping dengan beberapa infeksi[5].

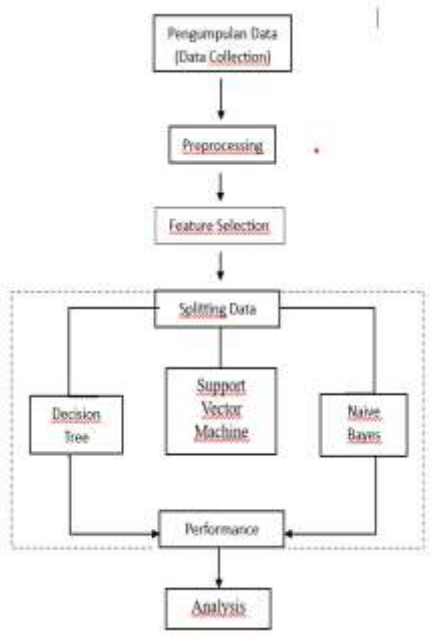
Karena kelemahan proses manual, peneliti menggunakan komputer sebagai alat bantu analisis dan mengembangkan proses yang bebas sepenuhnya pada manusia. Karena cara manual menggunakan bahasa yang hanya dimengerti manusia, Bahasa mesin adalah satu-satunya bahasa yang dikenal komputer, sehingga membuat hal ini sulit. Setelah data dikuantifikasi, mereka dapat dipisahkan dan dikelompokkan dengan menggunakan metode pembelajaran mesin[3].

Beberapa metode yang dapat digunakan untuk kasus prediksi termasuk Desicion tree, Support Vector Machine (SVM), dan naive bayes. Pada penelitian[6], menunjukkan akurasi sebesar 72.67% dengan menggunakan Decision Tree. Pada penelitian llainnya [7], mendapatkan akurasi tertinggi 62,3% menggunakan algoritma SVM dengan Forward Selection. Metode selanjutnya adalah Support Vector Machine (SVM) mampu bekerja menghindari masalah dimensionalitas dengan data berbagai dimensi. Selain itu, ada metode alternatif yang dapat digunakan, seperti naïve bayes. Keunggulan metode naive bayes adalah algoritmanya yang sederhana tetapi tetap mampu menghasilkan akurasi yang tinggi [8].

Maka Penelitian ini bertujuan untuk membandingkan metode Desicion tree, Support Vector Machine (SVM), dan Naive Bayes. Untuk mengetahui harapan mana yang paling dapat diandalkan untuk mendiagnosis infeksi hati membara. Tujuan studi ini adalah untuk membantu para profesional medis dalam mendiagnosis penyakit hati atau penyakit inflamasi hati.[3]

2. METODOLOGI PENELITIAN

Pada tahap ini dijelaskan tahapan dalam melakukan penelitian. Tahapan penelitian digambarkan pada Gambar 1.



Gambar 1. *Flowchat* Metodologi Penelitian

2.1 Pengumpulan Data (Data Collection)

Penelitian kali ini dimulai dengan pengumpulan data untuk proses analisis, data penting diperoleh melalui peninjauan data tentang penyakit liver yang ditemukan di situs *Kaggle.com.* Data akan diproses untuk membuat daftar pasien yang mungkin menderita penyakit liver.

2.2 Preprocessing

Sebelum melakukan pengujian model algoritma, proses ini adalah tahap awal. Saat ini, set data yang seharusnya akan diubah menjadi data bersih yang siap diuji sebelum digunakan [3]. Prosedur ini digunakan untuk memperbaiki kesalahan pada data mentah yang tidak lengkap dan diformat tidak teratur[9]. Penelitian ini adalah pemeriksaan terhadap data yang hilang juga dikenal sebagai nilai yang hilang dalam kumpulan data selama tahap prapemrosesan mendalam, Menghapus duplikat data, memeriksa ketidakkonsistenan data, dan memperbaiki kesalahan data tertentu, seperti kesalahan cetak, adalah beberapa contoh proses perbaikan data atau proses cleaning[10].

2.3 Feature Selection

Feature selection digunakan untuk menghilangkan fitur- fitur yang kurang berdampak signifikan pada proses prediksi. Keadaan di mana data latih yang akan digunakan bernilai "baik" sehingga nilai akurasi turun apabila data baru digunakan dapat menyebabkan masalah overfitting[11].

2.4 Spilitting Data

Setelah tahap *preprocessing*, sebuah data dibedakan menjadi data latih dan data uji. Data uji yaitu data yang mengukur sejauh mana memprediksi berlaku dalam mengkarakterisasi informasi secara akurat, sedangkan data latih adalah data yang telah ada sebelumnya sesuai dengan faktanya[12].Ini dilakukan dengan pemisahan 50:50, 70:30, dan 80:20. Dalam penelitian data pengujian adalah data yang tidak digunakan, namun berguna dalam mengevaluasi seberapa baik serta buruk nya penelitian. Data yang digunakan dalam penelitian dikenal sebagai data pelatihan[7].

Data yang telah diproses sebelumnya dari pasien penyakit hati akan digunakan untuk pemodelan pada saat ini. Berdasarkan nilai dataset untuk algoritma Decision Tree, SVM, dan Naive Bayes [13]. Dalam tahapan ini menggunakan metode:

2.4.1 Decision Tree

Decision tree seperti pohon dalam diagram air, pilihan pohon memiliki node internal yang menunjukkan variabel prediktor digunakan sebagai pemisah dan dihubungkan ke cabang, dan setiap node cabang menunjukkan kelas hasil klasifikasi. J Ross Quinlan mengembangkan algoritma ini pertama kali pada awal tahun 1980, mengembangkan jenis Decision Tree ID3[8].

2.4.2 Support Vector Machine(SVM)

Hipotesis fungsi-fungsi linear diterapkan pada sistem pembelajaran vector support. Algoritma pembelajaran berbasis teori optimasi digunakan untuk melatih sistem ini. SVM memaksimalkan fungsi pembatas batas (*hyperplane*) guna memisahkan dua buah dataset dari dua kelas yang berbeda[14]. Strategi SVM dibedakan menjadi dua jenis berdasarkan kualitasnya, yaitu SVM langsung dan tidak lurus. SVM langsung memisahkan informasi secara lurus, menggabungkan dua kelas pada hyperplane dengan tepi halus, sedangkan SVM tidak lurus menggunakan trik kernel untuk ruang berukuran besar[3].

Persamaan *hyperplane* untuk kasus ini[8] adalah:

$$WX + b = 0 \quad (1)$$

dimana :

$W = \{w_1, w_{2,\ldots,} w_p\}$ adalah vektor pembobot dan P adalah banyak variabel X

b = bias atau suatu konstanta

Saat data dapat dipisahkan dengan *hyperplane* yang linier, maka fungsi 1 dapat berubah menjadi

$$f(x) = w^T x + b \quad (2)$$

Jika $f(x) \geq 0$ untuk $y_i = +1$ dan jika $f(x) < 0$ untuk $y_i = -1$

2.4.3 Naive Bayes

Naive Bayes merupakan pengklasifikasi prediksi sederhana penjumlahan frekuensi dari kumpulan data dan menggabungkan nilainya. Fakta bahwa metode ini hanya membutuhkan sebagian kecil data pelatihan guna menghitung estimasi yang dibutuhkan untuk proses prediksi adalah salah satu kelebihannya. Naive Bayes dapat melakukan pekerjaan yang lebih kompleks dan lebih baik di dunia nyata[15].

Perhitungan Naive Bayes dapat dilakukan dengan rumus [10]:

$$P(H \mid X) = \frac{P(X \mid H)\ P(H)}{P(X)} \quad (3)$$

dimana:

X : data dengan kelas yang belum diketahui

H : hipotesis data X merupakan suatu kelas spesifik

P(H | X) : probabilitas hipotesis H berdasaekan kondisi X

P(H) : probabilitas hipotesis H

P(X | H) : probabilitas X berdasarkan kondisi pada hipotesis H

P(X) : probabilitas hipotesis X

2.5 Analysis

Dalam tahapan analisis, performansi metode dinilai dengan mengevaluasi kinerja dari Decision Tree, SVM, dan Naive Bayes. Ini dilakukan dengan membandingkan akurasi yang dimiliki algoritma dengan splitting data 50:50, 70:30, dan 80:20.

3. HASIL DAN PEMBAHASAN

3.1. Pengumpulan Data (Data Collection)

Dataset yang digunakan pada studi kasus ini diperoleh dari dataset terbuka Kaggle[16], yang mencakup 579 dataset yang memiliki sebelas fitur data pasien yang menderita penyakit liver. Dataset yang digunakan disajikan dalam Tabel 1.

Tabel 1. Pengumpulan Data

NO	Age	Gender	Direct_Bilirubin	...	Dataset
1	65	0	0.1	...	1
2	62	1	5.5	...	1
3	62	1	4.1	...	1
4	58	1	0.4	...	1
...	...	...	...	...	...
579	38	1	0.3	...	2

3.2. Preprocessing

Proses *preprocessing* data melibatkan tahap cleaning data dan transformasi data. Sehingga proses pengolahan data dapat diubah menjadi bilangan binary. Data disajikan dalam tabel 2.

Tabel 2. Preprocessing Data

NO	Age	Gender	Direct_Bilirubin	...	Dataset
1	65	0	0.1	...	0
2	62	1	5.5	...	0
3	62	1	4.1	...	0
4	58	1	0.4	...	0
...	...	...	...	...	...
579	38	1	0.3	...	1

3.3. Feature Selection

Pada tahap ini menggunakan *forward* feature selection dengan metode *wrapper*. Sehingga hasil yang didapatkan yaitu menghapus variabel *Direct_Bilirubin* dan variabel *albumin*. Data disajikan dalam Tabel 3.

Tabel 3. Feature Selection

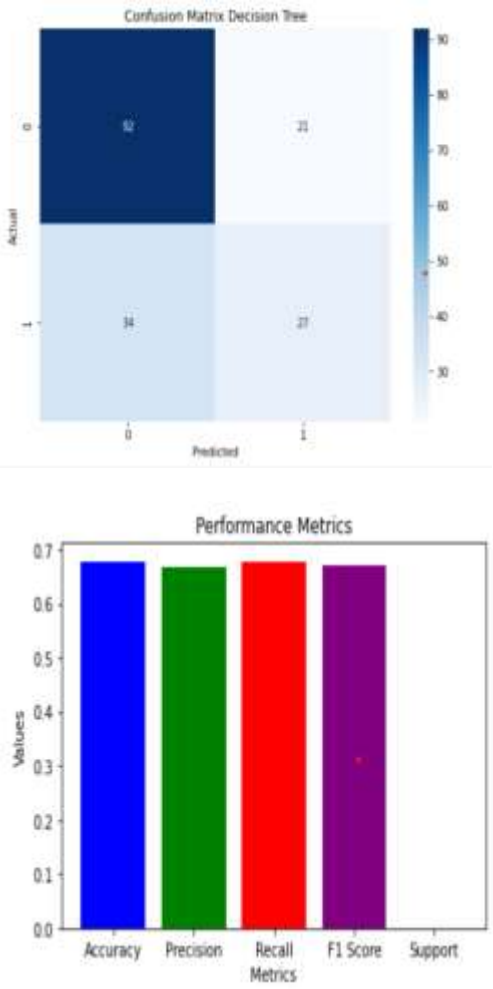
NO	Age	Gender	Total_Bilirubin	...	Dataset
1	65	0	0.7	...	0
2	62	1	10.9	...	0
3	62	1	7.3	...	0
4	58	1	1	...	0
...	...	...	...	...	...
579	38	1	1	...	1

3.4. Splitting Data

Pada tahap ini dilakukan proses membagi data yaitu data training dan data testing menjadi tiga tahap dengan rasio 50 : 50, 70: 30 dan 90 : 10. Pada tahap 50:50, 50 persen 50 persen datanya digunakan sebagai data latih, dan sisanya digunakan sebagai data uji. Pada tahap berikutnya, 70% dari jumlah keseluruhan informasi digunakan untuk menyiapkan informasi, dan 30 persen digunakan sebagai data uji, dan 90 persen dari total data digunakan sebagai data latih, dan 10 persen digunakan sebagai data uji. Metode ini memungkinkan untuk mengeksplorasi kinerja model dalam berbagai skenario pembagian data latih dan uji. Data yang telah diikuti ketiga tahap ini kemudian diuji menggunakan algoritma Decision Tree, Support Vector Machine dan Naive Bayes untuk menilai kinerja model pada setiap skenario pembagian data

3.4.1 Decision Tree

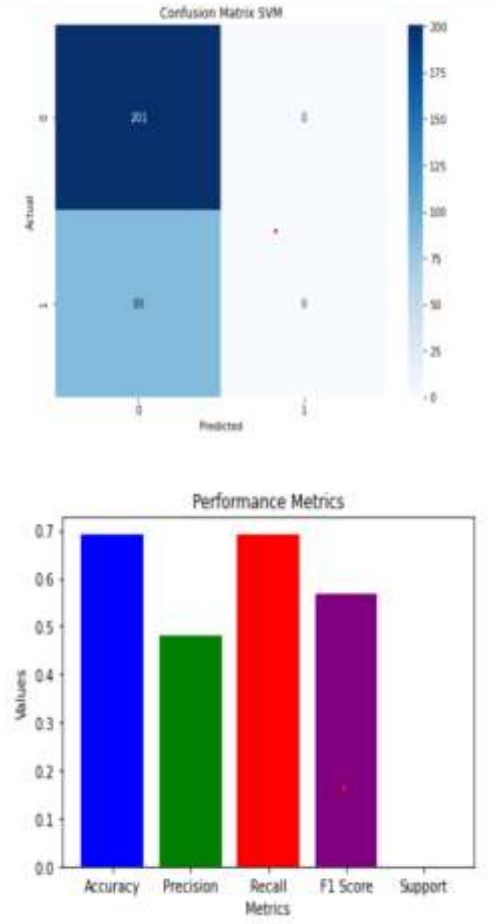
Memodelkan algoritma Decision Tree dengan splitting data 70:30, prediksi Decision Tree menunjukkan tingkat akurasi tertinggi 67,82%.



Gambar 2. Performance Metrics Decision Tree

3.4.2 Support Vector Machine(SVM)

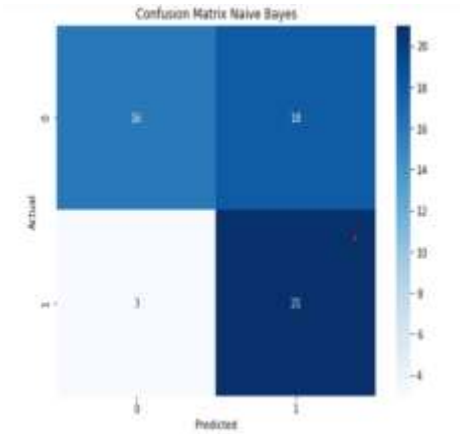
Memodelkan algoritma dengan splitting data 50 : 50, prediksi Support Vector Machine menunjukkan tingkat akurasi terbaik dengan 69,31%.

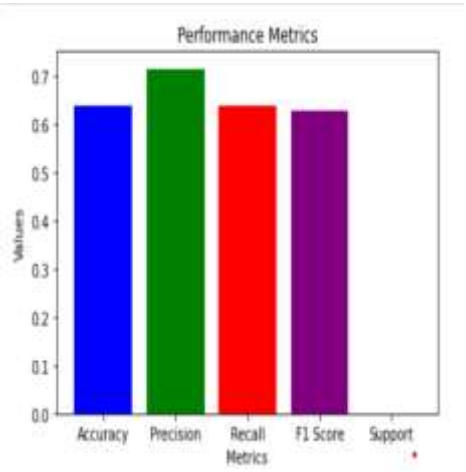


Gambar 3. Performance Metrics Support Vector Machine

3.4.3 Naive Bayes

Memodelkan algoritma dengan splitting data 90 : 10, terbukti bahwa prediksi Naive Bayes memiliki tingkat akurasi terbaik dengan 66%.





Gambar 4. Performance Metrics Naïve Bayes

3.5. Analysis

Dalam penelitian ini hasil perbandingan dari metode algoritma Decision Tree, Decision Tree + *Forward Selection*, Support Vector Machine (SVM), SVM + *Forward Selection*, Naive Bayes dan Naive Bayes + *Forward Selection* menggunakan rasio 50:50, 70:30 dan, 90:10. Dapat disajikan pada tabel 4.

Tabel 4. Hasil Perbandingan

Splitting Data	Decision Tree	Decision Tree + Forward Selection
90:10	65,52%	53,45%
70:30	64,95%	67,82%
50:50	64,14%	65,52%

Splitting Data	SVM	SVM + Forward Selection
90:10	59%	58,62%
70:30	65%	64,94%
50:50	69%	69,31%

Splitting Data	Naive Bayes	Naive Bayes + Forward Selection
90:10	66%	66%
70:30	61%	60%
50:50	58%	57%

4. KESIMPULAN

Berdasarkan hasil dari pengolahan serta analisis pada data pasien penyakit liver yang terdiri dari 579 dataset yang memiliki sebelas fitur dengan membandingkan tiga algoritma, yaitu Decision Tree, Support Vector Machine(SVM) dan Naive Bayes. Dengan splitting data 70:30 dan tingkat akurasi 67,82%, algoritma Decision Tree mencapai tingkat akurasi terbaik dan nilai *Precision* tertinggi yaitu 66% serta *F1 Score* 66% dengan penggunaan *Forward Selection*. Dengan splitting data 50 : 50 , Hasil *Recall* terbaik pada algoritma Support Vector Machine (SVM) yaitu 69%.

5. REFERENCES

[1] A. P. Sampurna, I. G. Santi Astawa, N. A. Sanjaya ER, A. A. I. Ngurah Eka Karyawati, I. W. Santiyasa, and I. M. Widiartha, "Seleksi Atribut Pada Diagnosis Penyakit Liver Menggunakan Decision Tree Dengan Algoritma Genetika," *JELIKU (Jurnal Elektron. Ilmu Komput. Udayana)*, vol. 11, no. 2, p. 329, 2022, doi: 10.24843/jlk.2022.v11.i02.p12.

[2] N. T. Rahman, "Analisa Algoritma Decision Tree Dan Naïve Bayes Pada Pasien Penyakit Liver," *J. Fasilkom*, vol. 10, no. 2, pp. 144–151, 2020, doi: 10.37859/jf.v10i2.2087.

[3] M. Nurkholifah, Jasmarizal, Y. Umar, and Rahmaddeni, "Analisa Performa Algoritma Machine Learning Dalam Prediksi Penyakit Liver," *J. Indones. Manaj. Inform. dan Komun.*, vol. 4, no. 1, pp. 164–172, 2023, doi: 10.35870/jimik.v4i1.149.

[4] N. Musyaffa and B. Rifai, "Model Support Vector Machine Berbasis Particle Swarm Optimization Untuk Prediksi Penyakit Liver," *JITK (Jurnal Ilmu Pengetah. Dan Teknol. Komputer)*, vol. 3, no. 2, pp. 189–194, 2018, doi: https://doi.org/10.33480/jitk.v3i2.

[5] F. L. D. Cahyanti, F. Sarasati, W. Astuti, and E. Firasari, "Klasifikasi Data Mining Dengan Algoritma Machine Larning Untuk Prediksi Penyakit Liver," *Technol. J. Ilm.*, vol. 14, no. 2, p. 134, 2023, doi: 10.31602/tji.v14i2.10093.

[6] I. Setiawati, A. P. Wibowo, and A. Hermawan, "Pendahuluan Tinjauan Pustaka Penelitian Sebelumnya Klasifikasi," *J. Inf. Syst. Manag.*, vol. 1, no. 1, pp. 13–17, 2019.

[7] D. Septhya *et al.*, "Implementasi Algoritma Decision Tree dan Support Vector Machine untuk Klasifikasi Penyakit Kanker Paru," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 3, no. 1, pp. 15–19, 2023, doi: 10.57152/malcom.v3i1.591.

[8] E. Pusporani, S. Qomariyah, and I. Irhamah, "Klasifikasi Pasien Penderita Penyakit Liver dengan Pendekatan Machine Learning," *Inferensi*, vol. 2, no. 1, p. 25, 2019, doi: 10.12962/j27213862.v2i1.6810.

[9] M. Kesuma, "Prediksi Penyakit Liver Menggunakan Algoritma Random Forest," *J. Inf. dan Komput.*, vol. 11, no. 2, p. 2023, 2023.

[10] A. Desiani, "Perbandingan Implementasi Algoritma Naïve Bayes dan K-Nearest Neighbor Pada Klasifikasi Penyakit Hati," *Simkom*, vol. 7, no. 2, pp. 104–110, 2022, doi: 10.51717/simkom.v7i2.96.

[11] L. Dwi Yulianto, E. Heni Hermaliani, and L. Kurniawati, "RESOLUSI : Rekayasa Teknik Informatika dan Informasi Penerapan Machine Learning Dalam Analisis Stadium Penyakit Hati Untuk Proses Diagnosis dan Perawatan," *Media Online*, vol. 3, no. 4, pp. 303–313, 2023, [Online]. Available: https://djournals.com/resolusi

[12] Baiq Nurul Azmi, Arief Hermawan, and Donny Avianto, "Analisis Pengaruh Komposisi Data Training dan Data Testing pada Penggunaan PCA dan Algoritma Decision Tree untuk Klasifikasi Penderita Penyakit Liver," *JTIM J. Teknol. Inf. dan Multimed.*, vol. 4, no. 4, pp. 281–290, 2023, doi: 10.35746/jtim.v4i4.298.

[13] A. D. Adhi Putra, "Analisis Sentimen pada Ulasan pengguna Aplikasi Bibit Dan Bareksa dengan Algoritma KNN," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 8, no. 2, pp. 636–646, 2021, doi: 10.35957/jatisi.v8i2.962.

[14] R. Dasmasela, B. P. Tomasouw, and Z. A. Leleury, "Penerapan Metode Support Vector Machine (SVM) untuk Mendeteksi Penyalahgunaan Narkoba," *J. Param.*, vol. 01, no. 02, pp. 111–122, 2022.

[15] M. Fatchan, M. Ir. Nanang Tedi K., Alfiyan, and Kurniawan, "Perbandingan Dalam Memprediksi Penyakit Liver Menggunakan Algoritma Naïve Bayes Dan K-Nearest Neighbor," *J. Pelita Teknol.*, vol. 16, no. 1, pp. 15–21, 2021.

[16] SAMAELFUKADA, "Data Pasien penyakit liver," Kaggle. Accessed: Dec. 23, 2023. [Online]. Available: https://www.kaggle.com/datasets/ahmadreginald/data-pasien-penyakit-liver