

PERBANDINGAN KINERJA ALGORITMA NAIVE BAYES DAN C4.5 DALAM PREDIKSI PENYAKIT JANTUNG

(COMPARISON OF THE PERFORMANCE OF NAIVE BAYES AND C4.5 ALGORITHM IN HEART DISEASE PREDICTION)

Sri Wulandari¹⁾, Kusrini²⁾, Hanafi³⁾

^{1, 2,3)}Megiter Teknik Informatika Universitas Amikom Yogyakarta

e-mail: sriwulan@students.amikom.ac.id¹⁾, kusrini@amikom.ac.id²⁾, hanafi@amikom.ac.id³⁾

ABSTRAK

Teknologi Informasi adalah teknologi pemrosesan data dan berbagai cara untuk menghasilkan informasi berkualitas secara akurat, cepat, dan relevan dengan kebutuhan individu maupun bisnis. Informasi strategis dari teknologi ini sangat penting dalam pengambilan keputusan. Pengembangan teknologi informasi menjadi salah satu faktor utama dalam kemajuan zaman. Beberapa sektor penting yang terdampak oleh teknologi meliputi pendidikan, ekonomi, kesehatan, pemerintahan, dan sosial-budaya. Pada dasarnya, teknologi dikembangkan untuk membantu pekerjaan manusia dan kini menjadi kebutuhan pokok dalam kehidupan. Salah satu penerapan pentingnya adalah dalam bidang kesehatan, khususnya untuk memprediksi penyakit jantung. Prediksi yang akurat dapat membantu pengobatan pasien secara efisien sebelum serangan jantung terjadi. Tujuan ini dapat dicapai melalui model pembelajaran mesin yang optimal dan data kesehatan yang lengkap. Oleh karena itu, perbandingan kinerja algoritma Naïve Bayes dan C4.5 dalam prediksi penyakit jantung diperlukan untuk mendapatkan hasil yang lebih akurat. Sebelum pengujian, data perlu diperiksa kelayakannya dan dibagi menjadi data pelatihan serta pengujian. Penelitian ini menggunakan beberapa skenario pembagian data dengan confusion matrix. Terdapat enam skenario: masing-masing algoritma diuji dalam tiga variasi rasio data. Naïve Bayes menunjukkan akurasi 83% (70:30), 83% (80:20), dan 85% (90:10). Sementara itu, algoritma C4.5 menunjukkan performa lebih tinggi dengan akurasi 98% (70:30), 100% (80:20), dan 100% (90:10). Hasil ini menunjukkan bahwa C4.5 lebih unggul dalam prediksi penyakit jantung.

Kata Kunci: Perbandingan, Naïve Bayes, C4.5, Confusion Matrix

ABSTRACT

Information Technology refers to data processing technologies and various methods used to produce high-quality information accurately, quickly, and relevantly, tailored to the needs of individuals and businesses. Strategic information derived from this technology plays a crucial role in decision-making processes. The development of information technology is one of the key factors driving modern advancement. Several critical sectors are significantly influenced by technological progress, including education, economy, healthcare, government, and socio-cultural domains. Fundamentally, technology is developed to support human tasks and has now become an essential aspect of everyday life. One of the most important applications of information technology is in the healthcare sector, particularly in predicting heart disease. Accurate predictions are essential for providing effective treatment before a heart attack occurs. This goal can be achieved by employing optimal machine learning models combined with comprehensive heart disease datasets. Therefore, comparing the performance of the Naïve Bayes and C4.5 algorithms in heart disease prediction is necessary to obtain more accurate results. Before model evaluation, data must be assessed for suitability and then divided into training and testing sets. This study applies several data-splitting scenarios using a confusion matrix. A total of six experimental scenarios were conducted: each algorithm underwent three tests with varying training and testing ratios. Naïve Bayes achieved accuracy rates of 83% (70:30), 83% (80:20), and 85% (90:10). In contrast, the C4.5 algorithm demonstrated superior performance with accuracy rates of 98% (70:30), 100% (80:20), and 100% (90:10). These results indicate that the C4.5 algorithm outperforms Naïve Bayes in predicting heart disease.

Keywords: Comparison, Naïve Bayes, C4.5, Confusion Matrix.

I. PENDAHULUAN

Teknologi informasi adalah teknologi pemrosesan data dan berbagai metode untuk menghasilkan informasi berkualitas tinggi

informasi yang relevan, akurat, dan tepat waktu untuk keperluan pribadi dan bisnis. Informasi strategis untuk pengambilan keputusan. Perkembangan teknologi informasi merupakan salah satu faktor penting bagi kemajuan zaman. Terdapat

beberapa sektor yang penting bagi kemajuan teknologi dan mempengaruhi tingkat kemajuan suatu negara, seperti sektor pendidikan, sektor perekonomian, sektor kesehatan, sektor pemerintahan, dan sektor sosial budaya. Pada dasarnya teknologi dikembangkan untuk memudahkan pekerjaan manusia. Saat ini, teknologi sudah menjadi kebutuhan utama umat manusia. Faktanya, teknologi digunakan dalam setiap aspek kehidupan manusia dan menguraikan latar belakang permasalahan yang diselesaikan, isu-isu yang terkait dengan masalah yang diselesaikan, ulasan penelitian yang pernah dilakukan sebelumnya oleh peneliti lain yang relevan dengan penelitian yang dilakukan [1]. Penyakit jantung bisa mengancam jiwa jika pengobatannya tertunda. Masalah ini terjadi karena sulitnya mendeteksi dini pasien penyakit jantung karena selalu mengabaikan gejala pertama yang muncul. Memprediksi penyakit jantung secara akurat sangat penting untuk mengobati pasien jantung secara efisien sebelum serangan jantung terjadi. Tujuan ini dapat dicapai dengan menggunakan model pembelajaran mesin yang optimal dengan data kesehatan penyakit jantung yang lengkap [2]. Gagal jantung merupakan suatu sindrom linis kompleks yang ditandai dengan menurunnya kemampuan jantung dalam memompa darah guna memenuhi kebutuhan oksigen tubuh [3].

Berdasarkan latar belakang diatas, penelitian ini berfokus terhadap perbandingan algoritma Naive Bayes dan c4.5, pengujian menggunakan confusion matrix. Olah karena itu mengangkat penelitian tentang “Perbandingan Kinerja Algoritma Naive Bayes Dan C4.5 Dalam Prediksi Penyakit Jantung”

II. STUDI PUSTAKA

Pada penelitian membandingkan kinerja dari kedua algoritma yaitu algoritma naïve bayes dan C4.5. Algoritma naïve bayes adalah salah satu machine learning yang paling populer dan cocok untuk sebuah data atau analisis data [4]. Kemudian algoritma C4.5 adalah pengklasifikasi pohon pembuatan Keputusan dan sering digunakan untuk mengklasifikasikan data. Algoritma ini membuat model berdasarkan aturan yang dipilih dari data pelatihan [5].

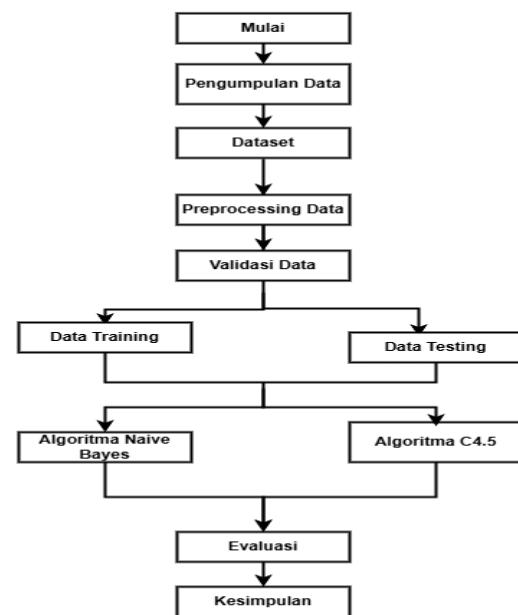
Berbagai penelitian sebelumnya telah dilakukan melakukan evaluasi algoritma C4.5 dan Naive Bayes. Berdasarkan hasil perbandingan kedua algoritma, tingkat akurasi algoritma Naive Bayes lebih baik dibandingkan C4.5, dengan selisih sebesar 11,77% [6]. Tetapi pada penelitian lain berdasarkan hasil pengolahan data, dirasa kurang persuasif

sehingga dilakukan pengolahan data juga dengan menggunakan metode klasifikasi lainnya yaitu Naive Bayes. Berdasarkan hasil perbandingan kedua algoritma, tingkat akurasi algoritma Naive Bayes lebih baik dibandingkan C4.5, dengan selisih sebesar 11,77% [7].

III. METODE PENELITIAN

Jenis penelitian yang dilakukan adalah penelitian kuantitatif. Penelitian kuantitatif merupakan jenis penelitian yang digunakan untuk menganalisis data dengan cara mendeskripsikan atau mengilustrasikan data yang telah dikumpulkan sebelumnya sebagaimana berlaku. Penelitian deskriptif kuantitatif menggunakan metodologi penelitian korelasional. Tujuan analisis korelatif adalah untuk menentukan apakah ada hubungan yang benar atau salah, dan jika ada, seberapa kuat dan benar hubungan tersebut. Analisis korelasi berfokus pada dua variabel, atau lebih tepatnya, bagaimana variasi suatu variable berhubungan dengan variasi variabel lainnya.

Salah satu jenis penelitian Kuantitatif adalah penelitian empiris, data dalam bentuk angka dan penelitian kualitatif, dan data tidak dalam bentuk angka. Penelitian ini berfokus secara intens pada satu eksperimen tertentu yang dipelajari sebagai studi kasus [8]. Metode studi kasus memungkinkan peneliti melakukan analisis yang komprehensif dan bermakna. Alur penelitian dapat dilihat pada gambar 1 berikut ini:



Gambar 1. Alur Penelitian

Penelitian ini dilakukan secara sistematis dimulai dari pengumpulan data tahun 2020–2024, kemudian

membentuk dataset yang siap diolah. Data selanjutnya diproses pada tahap pre-processing untuk dibersihkan dan disesuaikan dengan kebutuhan analisis. Setelah itu, data divalidasi dan dibagi menjadi data training dan testing. Dua algoritma, yaitu Naïve Bayes dan C4.5, digunakan untuk membangun model prediksi. Naïve Bayes menggunakan pendekatan probabilistik, sedangkan C4.5 menggunakan metode pohon keputusan. Hasil dari kedua model kemudian dievaluasi, dan kesimpulan diambil berdasarkan akurasi yang diperoleh dari masing-masing algoritma.

IV. HASIL DAN PEMBAHASAN

Jenis penelitian yang dilakukan adalah penelitian kuantitatif. Penelitian kuantitatif adalah sejenis studi yang sistematis, terencana, terstruktur. Kemudian dapat diartikan penyelidikan masalah sosial berdasarkan tes teori, yang terdiri dari variabel berbasis numerik dan prosedur statistik, dan merupakan studi empiris di mana data dapat dihitung dengan cara tertentu, dengan prosedur statistik untuk menentukan apakah generalisasi prediksi teori adalah studi kuantitatif yang benar. Data yang akan diolah menghasilkan semua hasil akurasi atau numerik [9].

Pada Penelitian sebelumnya perbandingan komprehensif terhadap metode seleksi fitur dalam prediksi penyakit jantung, serta menunjukkan peran krusial seleksi fitur dalam mengoptimalkan kinerja model. Hasil ini memberikan panduan praktis dalam memilih teknik seleksi fitur yang sesuai dengan algoritma klasifikasi yang digunakan, sehingga berkontribusi pada pengembangan alat diagnostik yang lebih akurat dan efisien untuk pengambilan keputusan klinis di bidang kardiologi. Hasilnya, Mutual Information (MI) memberikan performa terbaik, terutama pada model neural network, dengan akurasi tertinggi 82,3% dan recall 0,94. Model logistic regression dan random forest juga menunjukkan peningkatan akurasi saat menggunakan MI. Sementara itu, model yang lebih sederhana seperti Naïve Bayes dan decision tree bekerja lebih baik dengan ANOVA dan Chi-Square, dengan akurasi masing-masing sekitar 76,45% dan 75,99%. Model k-NN dan SVM menunjukkan performa paling rendah, di bawah 55% akurasinya [10].

Pada penelitian sebelumnya membandingkan algoritma C4.5 dan Naive Bayes untuk prediksi penyakit jantung menggunakan data dari UCI dan Kaggle. Hasilnya, C4.5 mencapai akurasi tertinggi 95,18% setelah seleksi fitur dengan PSO, sementara

Naive Bayes tanpa seleksi fitur mencapai akurasi maksimum 90,87% dan AUC 94%. Meski C4.5 unggul dalam akurasi total, Naive Bayes lebih unggul dalam beberapa metrik evaluasi seperti recall, F1-score, dan AUC, menunjukkan efisiensi yang baik tanpa proses seleksi fitur tambahan [11].

Pada penelitian sebelumnya membahas tentang tuberculosis yang disebabkan oleh bakteri *Mycobacterium Tuberculosis* dan merupakan salah satu penyakit menular yang masih tinggi kasusnya di Indonesia. Untuk membantu diagnosis dan deteksi dini, penelitian ini membandingkan algoritma klasifikasi Naïve Bayes dan C4.5. Dengan menggunakan model 10-fold cross validation dan confusion matrix, hasil menunjukkan bahwa C4.5 memiliki akurasi lebih tinggi (85%) dibandingkan Naïve Bayes (83%), sehingga lebih efektif untuk klasifikasi data tuberculosis [12].

Pada penelitian ini memperoleh data dari platform online yaitu Kaggle sebuah platform yang dapat diakses pada berbagai macam browser. Kaggle sangat membantu para peneliti atau data scientist untuk kebutuhan penelitian [13]. Data Kaggle dapat dilihat pada Gambar 2 berikut ini:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	target
1	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal		
2	52	1	0	125	212	0	1	168	0	1.0	2	2	3	0	
3	53	1	0	140	203	1	0	155	1	3.1	0	0	3	0	
4	70	1	0	145	223	0	1	174	0	2.6	0	0	3	0	
5	61	1	0	148	200	0	1	161	0	0	2	1	3	0	
6	62	0	0	138	294	1	1	108	0	1.9	1	3	2	0	
7	58	0	0	100	248	0	0	122	0	1	1	0	2	1	
8	58	1	0	114	318	0	2	140	0	4.4	0	3	1	0	
9	55	1	0	160	289	0	0	145	1	0.8	1	1	3	0	
10	46	1	0	120	249	0	0	144	0	0.6	2	0	3	0	
11	54	0	0	122	286	0	0	116	1	3.2	2	2	0		
12	71	0	0	112	149	0	1	125	0	1.6	1	0	2	1	
13	43	0	0	135	341	1	0	138	1	3	1	0	3	0	
14	34	0	0	118	210	0	1	192	0	0.7	2	0	2	1	
15	51	1	0	140	298	0	1	122	1	4.2	1	3	3	0	
16	52	1	0	128	204	1	1	158	1	1	1	0	0	0	
17	34	0	1	118	210	0	1	192	0	0.7	2	0	2	1	
18	51	0	2	140	308	0	0	142	0	1.5	2	1	2	1	
19	54	1	0	120	266	0	0	109	1	2.2	1	1	3	0	
20	50	1	0	120	244	0	1	162	0	1.1	2	0	2	1	
21	58	1	2	140	211	1	0	165	0	0	2	0	2	1	
22	60	1	2	140	185	0	0	155	0	3	1	0	2	0	
23	67	0	0	100	223	0	1	142	0	0.3	2	2	2	1	
24	AN	1	1	174	598	1	1	148	1	1	1	1	2	1	

Gambar 2. Mengumpulkan data

Mengolah dasaset pada google collaboratoty, perlu dilakukan mount pada gambar 3 dibawah ini:

[]	# Mount Google Drive													
from google.colab import drive	drive.mount('/content/drive')													
Mounted at /content/drive														
● # Baca dataset dari Google Drive (ganti path di bawah dengan path file kamu)														
path = '/content/drive/MyDrive/COLAB_NB_C45 PENYAKIT JANTUNG/TRAINING/DATASET/heart.csv' # sesuaikan lokasi file CSV														
df = pd.read_csv(path, sep=',')														
df														
age sex cp trestbps chol fbs restecg thalach exang oldpeak slope ca thal target														
0	52	1	0	125	212	0	1	168	0	1.0	2	2	3	0
1	53	1	0	140	203	1	0	155	1	3.1	0	0	3	0
2	70	1	0	145	174	0	1	125	1	2.6	0	0	3	0
3	61	1	0	148	203	0	1	161	0	0.0	2	1	3	0
4	62	0	0	138	294	1	1	106	0	1.9	1	3	2	0
...
1020	59	1	1	140	221	0	1	164	1	0.0	2	0	2	1
1021	60	1	0	125	258	0	0	141	1	2.8	1	1	3	0
1022	47	1	0	110	275	0	0	118	1	1.0	1	1	2	0
1023	50	0	0	110	254	0	0	159	0	0.0	2	0	2	1
1024	54	1	0	120	188	0	1	113	0	1.4	1	1	3	0
1025 rows x 14 columns														

Gambar 3. Melakukan mount pada dataset

Setelah menampilkan dataset, dilakukan perhitungan untuk mengetahui jumlah masing-masing kelas pada gambar 4 dibawah ini:

```
[73] # Menampilkan jumlah kelas pada keseluruhan data
print("Jumlah kelas pada seluruh data:")
print(y.value_counts())
print()

→ Jumlah kelas pada seluruh data:
target
1    526
0    499
Name: count, dtype: int64
```

Gambar 4 Perhitungan masing-masing kelas

a. Preprocessing adalah sebuah langkah penting saat menggunakan algoritma untuk pembelajaran mesin dan analisis teks. Tujuan utama preprocessing adalah untuk membersihkan dan membuat data dengan cara ini untuk membuatnya lebih mudah untuk dianalisis dan mendapatkan hasil yang lebih akurat [14]. Untuk memastikan dataset siap dan layak digunakan pada algoritma Naive Bayes dan C4.5 dalam melakukan klasifikasi penyakit jantung, perlu dilakukan preprocessing dengan langkah sebagai berikut:

1. Pembersihan Data: untuk memastikan tidak terdapat data yang hilang, perlu dilakukan pemeriksaan terlebih dahulu terhadap dataset yang disiapkan. proses pemeriksaan dataset dilakukan menggunakan fungsi `isnull` seperti pada gambar pada gambar 5 di bawah ini:

```
▶ # Cek nilai kosong
print(df.isnull().sum())

→ age      0
    sex      0
    cp      0
    trestbps 0
    chol      0
    fbs      0
    restecg  0
    thalach  0
    exang      0
    oldpeak  0
    slope      0
    ca      0
    thal      0
    target    0
    dtype: int64
```

Gambar 5. Validasi dataa

2. Pembagian Data: untuk dapat melakukan proses pelatihan (training) dan pengujian (testing) terhadap model, perlu dilakukan proses pembagian data. sebelum melakukan pembagian data training dan testing, terlebih dahulu dilakukan pemisahan antara fitur dan label seperti pada gambar 6 berikut ini:

```
[ ] # Pisahkan fitur (X) dan target/label (y)
X = df.drop('target', axis=1)
y = df['target']
```

```
[ ] # Bagi data menjadi train dan test
X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.3, random_state=42, stratify=y)
```

Gambar 6 Pembagian data

Menunjukkan proses pembagian dataset secara otomatis menggunakan fungsi `train_test_split`. dimana pada contoh ini 30% dataset dijadikan sebagai data *testing* dan 70% dijadikan sebagai data *training*. Variabel *X_train* akan menampung atribut untuk data *training* dan variabel *y_train* akan menampung label untuk data *training* demikian pula untuk variabel *X_test* dan *y_test* yang menampung atribut dan label untuk data *testing*. hasil pembagian dataset dapat dilihat pada tabel 1 beikut ini:

Tabel 1. Pembagian data

NO	RASIO	JUMLAH DATA TRAINING			JUMLAH DATA TESTING		
		JUMLAH	JUMLAH LABEL 0	JUMLAH LABEL 1	JUMLAH	JUMLAH LABEL 0	JUMLAH LABEL 1
1	70:30	717	349	368	308	150	158
2	80:20	820	399	421	205	100	105
3	90:10	922	449	473	103	50	53

b. Skenario Percobaan Untuk mendapatkan model klasifikasi yang diharapkan, skenario percobaan yang akan dibuat dengan membandingkan dua algoritma. Dalam penelitian ini akan melakukan 6 (enam) skenario percobaan yang dapat dilihat pada Tabel 2 berikut ini:

Tabel 2. Skenario percobaan

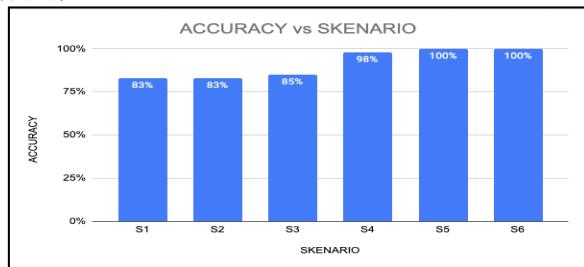
NO	SKENARIO	ALGORITMA	DATASET
1	S1	NB	70:30
2	S2	NB	80:20
3	S3	NB	90:10
4	S4	C4.5	70:30
5	S5	C4.5	80:20
6	S6	C4.5	90:10

c. Analisis Hasil Percobaan Dalam penelitian ini, algoritma Naive Bayes dan C4.5 diterapkan untuk melakukan klasifikasi penyakit jantung. Evaluasi terhadap kinerja kedua algoritma tersebut menggunakan confusion matrix. Confusioan matrix adalah sebuah tabel yang membandingkan hasil prediksi dengan data asli dan memberikan informasi tentang jumlah prediksi yang benar (benar kesalahan prediksi positif dan benar) dan (alse positif dan false negative) [15]. Berikut hasil analisis masing-masing skenario dapat dilihat pada table 3 berikut ini:

Tabel 3. Hasil percobaan

NO	SKENARIO	ACCURACY	DATA TRAINING	DATA TESTING	
1	S1	83%	717	308	
2	S2	83%	820	205	
3	S3	85%	922	103	[4]
4	S4	98%	717	308	
5	S5	100%	820	205	
6	S6	100%	922	103	

Dari tabel 3 diketahui skenario lima dan enam merupakan skenario dengan akurasi tertinggi yang mencapai 100% kedua skenario terbaru menggunakan algoritma C4.5. Algoritma Naive Bayes yang diuji pada skenario satu, dua dan tiga, tidak menghasilkan akurasi yang lebih baik dari semua pengujian pada algoritma C4.5. sebagai perbandingan dapat dilihat grafik pada gambar 7 berikut ini:



Gambar 7. Grafik Akurasi

V. KESIMPULAN

Perbandingan kinerja algoritma sangat di pengaruhi oleh pembagian data pelatihan dan data pengujian yang digunakan. Kemudian dengan adanya perbandingan algoritma dapat menjadi perbandingan pada saat menggunakan algoritma atau pembagian data. Pada skenario lima dan enam merupakan skenario dengan akurasi tertinggi yang mencapai 100% kedua skenario terbaru menggunakan algoritma C4. algoritma Naive Bayes yang diuji pada skenario satu, dua dan tiga, tidak menghasilkan akurasi yang lebih baik dari semua pengujian pada algoritma C4.5

VI. DAFTAR PUSTAKA

- [1] Cecep Abdul Cholik, “Teknologi Informasi, ICT,” *J. Fak. Tek.*, vol. 2, no. 2, pp. 39–46, 2021.
- [2] A. Sepharni, I. E. Hendrawan, and C. Rozikin, “Klasifikasi Penyakit Jantung dengan Menggunakan Algoritma C4.5,” *STRING (Satuan Tulisan Ris. dan Inov. Teknol.)*, vol. 7, no. 2, p. 117, 2022, doi: 10.30998/string.v7i2.12012.
- [3] T. A. Afifah, R. Martiansah, M. Alviyoni, and A. Arifin, “Perbandingan Algoritma Klasifikasi C4.5 Dan Naive Bayes untuk Memprediksi Gagal Jantung,” *SENTIMAS Semin. Nas. Penelit. dan Pengabd. Masy.*, pp. 78–85, 2023, [Online]. Available: <https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction>,
- [4] C. L. Rohmat, R. Tito Aprilianto, F. Fathurrohman, and I. Iin, “Analisis Sentimen Pada Ulasan Produk Pakaian Zirdigo Dengan Menggunakan Algoritma Naive Bayes,” *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 1, pp. 668–674, 2024, doi: 10.36040/jati.v8i1.8697.
- [5] N. Leniawati and S. Wijayanto, “Klasifikasi Desa Wisata di Provinsi Jawa Tengah Menggunakan Algoritma C4.5,” *J. Sains Komput. Inform. (J-SAKTI)*, vol. 8, no. 1, pp. 171–184, 2024.
- [6] F. Fatmawati and N. Narti, “Perbandingan Algoritma C4.5 dan Naive Bayes Dalam Klasifikasi Tingkat Kepuasan Mahasiswa Terhadap Pembelajaran Daring,” *JTIM J. Teknol. Inf. dan Multimed.*, vol. 4, no. 1, pp. 1–12, 2022, doi: 10.35746/jtim.v4i1.196.
- [7] D. Septiani, “Dan Naive Bayes Untuk Prediksi Penyakit Hepatitis,” *J. Pilar Nusa Mandiri*, vol. 13, no. 1, pp. 76–84, 2017, [Online]. Available: <http://archive.ics.uci.edu/ml/>.
- [8] H. Syahrizal and M. S. Jailani, “Jenis-Jenis Penelitian Dalam Penelitian Kuantitatif dan Kualitatif,” *J. QOSIM J. Pendidik. Sos. Hum.*, vol. 1, no. 1, pp. 13–23, 2023, doi: 10.61104/jq.v1i1.49.
- [9] M. M. Ali, T. Hariyati, M. Y. Pratiwi, and S. Afifah, “Metodologi Penelitian Kuantitatif dan Penerapannya dalam Penelitian,” *Educ. Journal*. 2022, vol. 2, no. 2, pp. 1–6, 2022.
- [10] B. Ahmad, Jinfu Chen, and H. Chen, “Feature selection strategies for optimized heart disease diagnosis using ML and DL models,” no. Cvd, pp. 1–19, 2025, [Online]. Available: <https://arxiv.org/pdf/2503.16577>
- [11] P. A. Sihotang and D. Sitanggang, “Penerapan Metode Algoritma C4.5 Dan Naive Bayes Untuk Prediksi Penyakit Jantung,” *J. Tek. Inf. dan Komput.*, vol. 7, no. 2, p. 899, 2024, doi: 10.37600/tekinkom.v7i2.1535.
- [12] P. N. Isnaeni, H. Rakhamawati, F. A. Tyas, S. Muhammadiyah, and P. Brebes, “Perbandingan Algoritma Naïve Bayes Dan C4.5 Pada Klasifikasi Penyakit Tuberculosis,” *J. Informatics Comput.*, vol. 3, no. 1, pp. 41–48, 2024.
- [13] A. S. Setyanegara and E. Zuliarso, “Menerapkan Data Mining Dengan Algoritma Fp-Growth Pada Analisis Pola Pembelian Konsumen Pada Donat Bolong Semarang,” *INTECOMS J. Inf. Technol. Comput. Sci.*, vol. 6, no. 2, pp. 866–872, 2023,

doi: 10.31539/intecoms.v6i2.7106.

- [14] M. Z. Haq, C. S. Octiva, A. Ayuliana, U. W. Nuryanto, and D. Suryadi, “Algoritma Naïve Bayes untuk Mengidentifikasi Hoaks di Media Sosial,” *J. Minfo Polgan*, vol. 13, no. 1, pp. 1079–1084, 2024, doi: 10.33395/jmp.v13i1.13937.
- [15] F. R. Valerian *et al.*, “Klasifikasi tingkat obesitas menggunakan metode gbm dan confusion matrix,” vol. 9, no. 2, pp. 2242–2249, 2025.