

ANALISIS PREDIKSI PENYAKIT HEPATITIS MENGGUNAKAN ALGORITMA DECISION TREE DAN NAIVE BAYES

Akbar Fajar Fadillah¹, Axel Elyas Ginting², Mayang Kusumah³, Yamin Nuryamin, S.Kom, M.Kom⁴, Ade Priyatna, S.Kom, MM, M.Kom⁵

^{1,2,3,4,5}Program Studi S1 Teknologi Informasi, Universitas Bina Sarana Informatika

¹akbarfadillahkc@gmail.com, ²axeleyas3@gmail.com*, ³Kusumahmayang544@gmail.com,

⁴yamin.yny@bsi.ac.id, ⁵ade.aeq@bsi.ac.id

Corresponding Author: Axel Elyas Ginting

ABSTRACT

Hepatitis is a critical global health issue and continues to be a major challenge in the medical world because it attacks the liver and is difficult to detect in its early stages. This condition requires faster, more effective diagnostic methods that can provide accurate predictions to assist medical personnel in early identification. With the development of technology, machine learning-based approaches are increasingly being adopted to support automatic medical data analysis. This study aims to analyze and compare the performance of two classification algorithms, namely Decision Tree and Naive Bayes, in predicting hepatitis disease status based on blood biochemistry data. The dataset used was sourced from the Kaggle platform, consisting of 615 patient data and 14 laboratory attributes. The research stages included data pre-processing, normalization, data division into 70% training data and 30% test data, model training, and evaluation using accuracy, precision, recall, F1-score, and confusion matrix metrics. The results showed that the Decision Tree algorithm achieved an accuracy of 88.65%, slightly higher than the Naive Bayes algorithm, which achieved an accuracy of 87.03%. This finding indicates that Decision Tree is more effective in handling data variation and non-linear relationships between biochemical features. Overall, this study confirms that both models can be used as decision support systems in detecting hepatitis, but Decision Tree is a more promising model for implementation in the early stages of diagnosis.

Keywords: Hepatitis, Machine Learning, Decision Tree, Naive Bayes, Disease Prediction

ABSTRAK

Penyakit hepatitis merupakan salah satu masalah kesehatan global yang kritis dan terus menjadi tantangan besar dalam dunia medis karena menyerang organ hati dan sulit terdeteksi pada tahap awal. Kondisi ini menuntut adanya metode diagnosis yang lebih cepat, efektif, dan mampu memberikan prediksi akurat untuk membantu tenaga medis dalam proses identifikasi dini. Seiring perkembangan teknologi, pendekatan berbasis machine learning mulai banyak diadopsi untuk mendukung analisis data medis secara otomatis. Penelitian ini bertujuan untuk menganalisis dan membandingkan performa dua algoritma klasifikasi, yaitu Decision Tree dan Naive Bayes, dalam memprediksi status penyakit hepatitis berdasarkan data biokimia darah. Dataset yang digunakan bersumber dari platform Kaggle, terdiri atas 615 data pasien dan 14 atribut laboratorium. Tahapan penelitian meliputi pra-pemrosesan data, normalisasi, pembagian data menjadi 70% data latih dan 30% data uji, pelatihan model, serta evaluasi menggunakan metrik akurasi, presisi, recall, F1-score, dan confusion matrix. Hasil penelitian menunjukkan bahwa algoritma Decision Tree memperoleh akurasi sebesar 88,65%, sedikit lebih tinggi dibandingkan algoritma Naive Bayes yang mencapai akurasi 87,03%. Temuan ini mengindikasikan bahwa Decision Tree lebih efektif dalam menangani variasi data dan hubungan non-linear antar fitur biokimia. Secara keseluruhan, penelitian ini menegaskan bahwa kedua model mampu digunakan sebagai sistem pendukung keputusan dalam mendeteksi hepatitis, namun Decision Tree menjadi model yang lebih potensial untuk diimplementasikan pada tahap diagnosis awal.

Kata Kunci: Hepatitis, Machine Learning, Decision Tree, Naive Bayes, Prediksi Penyakit

1. Pendahuluan

Penyakit hepatitis adalah salah satu masalah kesehatan global yang dipicu oleh infeksi virus yang menyerang organ hati. Hepatitis adalah kondisi yang menyebabkan peradangan di hati ataupun liver [1]. World Health Organization (WHO) menyatakan bahwa hepatitis adalah penyakit menular pembunuh terbesar setelah tuberculosis (TBC) [2]. Hal ini menunjukkan



Lisensi

Lisensi Internasional Creative Commons Attribution-ShareAlike 4.0.

pentingnya deteksi dini untuk mengidentifikasi individu yang terinfeksi sejak tahap awal agar penanganan medis dapat dilakukan dengan cepat dan efektif. Namun, diagnosis hepatitis umumnya masih memerlukan pemeriksaan laboratorium yang kompleks dan biaya yang cukup tinggi, sehingga dibutuhkan pendekatan berbasis data yang lebih efisien dan terjangkau untuk membantu proses identifikasi penyakit tersebut.

Perkembangan ilmu komputer, khususnya di bidang machine learning, telah membuka peluang besar dalam membantu proses diagnosis penyakit secara otomatis. Data merupakan komponen krusial dalam Machine Learning karena berfungsi sebagai bahan bakar utama yang memungkinkan model membangun metode penyelesaian masalah [3]. Melalui penerapan algoritma klasifikasi, sistem komputer dapat memprediksi status kesehatan pasien berdasarkan parameter biokimia darah seperti kadar enzim hati (ALT, AST, ALP), protein, bilirubin, kolesterol, dan komponen darah lainnya. Permasalahan utama yang dihadapi adalah bagaimana menentukan model klasifikasi yang memiliki tingkat akurasi tinggi dan mampu mengidentifikasi kategori pasien secara tepat berdasarkan data laboratorium tersebut. Dalam konteks ini, algoritma Decision Tree dan Naive Bayes menjadi dua metode yang relevan untuk dibandingkan karena masing-masing memiliki keunggulan tersendiri. Algoritma Decision Tree memiliki struktur hierarkis yang tersusun dari tiga jenis simpul: simpul akar (root) sebagai titik awal, simpul perantara yang digunakan untuk pengujian, dan simpul daun (leaf) yang merepresentasikan hasil keputusan [4]. Decision Tree memiliki kemampuan interpretasi yang baik dengan struktur berbentuk pohon yang mudah dipahami, sedangkan Naive Bayes bekerja dengan pendekatan probabilistik sederhana namun efisien dalam pengolahan data berukuran besar. Sementara itu, fungsi utama Algoritma Naïve Bayes adalah untuk menemukan pola atau karakteristik serupa yang terdapat dalam kelompok atau kelas data tertentu [5].

Tujuan utama dari riset ini bertujuan untuk melakukan analisis dan komparasi performa algoritma Decision Tree dan Naive Bayes dalam memprediksi kondisi hepatitis menggunakan data Hepatitis C. Studi ini diharapkan mampu mengidentifikasi model yang memberikan tingkat akurasi prediksi tertinggi, sekaligus menentukan atribut-atribut laboratorium mana yang paling berpengaruh terhadap hasil klasifikasi. Lebih lanjut, temuan ini diharapkan berkontribusi pada pengembangan ilmu komputer di sektor kesehatan, khususnya dalam utilisasi teknologi machine learning untuk memfasilitasi proses diagnosis penyakit yang lebih efisien dan akurat.

2. Tinjauan Pustaka

Penelitian mengenai penggunaan algoritma machine learning dalam klasifikasi penyakit hepatitis telah banyak dilakukan dan menunjukkan bahwa algoritma seperti Decision Tree dan Naive Bayes memiliki potensi besar dalam membantu diagnosis medis berbasis data laboratorium.

Hepatitis adalah salah satu penyakit yang menjadi pusat perhatian dalam bidang kesehatan masyarakat karena menyerang organ hati dan dapat menimbulkan dampak jangka panjang, termasuk sirosis dan kanker hati. Menurut Kusuma dan Wulandari [6], penyakit hepatitis umumnya disebabkan oleh infeksi virus, namun juga dapat dipicu oleh faktor gaya hidup, kebersihan, serta paparan bahan beracun. Penularannya dapat terjadi melalui makanan, darah, maupun kontak dengan cairan tubuh penderita. Oleh karena itu, pencegahan dini menjadi langkah penting untuk menekan angka penyebaran penyakit ini di masyarakat.

Penelitian oleh Damayanti dan Testiana [7] menggunakan algoritma Naive Bayes untuk memprediksi penyakit Hepatitis C dengan data yang diambil dari Kaggle berjumlah 586 sampel dengan 13 fitur biokimia. Hasil penelitian menunjukkan akurasi sebesar 85,71%, yang dikategorikan sebagai Good Classification. Penelitian ini menegaskan bahwa metode Naive



Lisensi

Lisensi Internasional Creative Commons Attribution-ShareAlike 4.0.

Bayes memiliki kinerja yang efisien dalam menangani dataset medis berskala besar serta mampu melakukan klasifikasi dengan tingkat kesalahan yang rendah.

Sementara itu, penelitian oleh Gunawan dan Pratama [8] membandingkan enam algoritma machine learning, yaitu Naive Bayes, Decision Tree, Logistic Regression, K-Nearest Neighbor, Support Vector Machine, dan Random Forest untuk klasifikasi penyakit hepatitis. Hasil penelitian menunjukkan bahwa algoritma Random Forest memberikan akurasi tertinggi sebesar 98,37%, diikuti oleh Decision Tree dan Naive Bayes dengan hasil yang relatif stabil. Faktor-faktor seperti nilai enzim AST dan ALP ditemukan memiliki kontribusi signifikan terhadap hasil prediksi, yang menegaskan relevansi parameter biokimia dalam diagnosis penyakit hati.

Penelitian oleh Sabriyanti et al. [9] melakukan perbandingan antara algoritma Naive Bayes, K-Nearest Neighbor (KNN), dan Support Vector Machine (SVM) dalam mendeteksi penyakit liver. Menggunakan metode SMOTE untuk mengatasi ketidakseimbangan data, hasil penelitian menunjukkan bahwa Naive Bayes memperoleh akurasi tertinggi yaitu 65,51%. Hal ini menunjukkan kemampuan Naive Bayes dalam menangani dataset yang tidak seimbang, serta efisiensinya dalam proses pelatihan model medis.

Selain itu, Gunawan et al. [10] menerapkan algoritma Decision Tree C4.5 pada dataset berjumlah 615 data pasien untuk mengklasifikasikan status kesehatan terkait hepatitis. Hasil yang diperoleh menunjukkan bahwa model Decision Tree menghasilkan akurasi sebesar 88,65%, dengan variabel AST, ALT, dan CHE sebagai atribut paling berpengaruh terhadap hasil prediksi. Model pohon keputusan dianggap unggul karena dapat memberikan visualisasi yang mudah dipahami serta interpretasi yang jelas terhadap hasil klasifikasi.

Dari kelima penelitian tersebut dapat disimpulkan bahwa baik Decision Tree maupun Naive Bayes merupakan algoritma yang efektif untuk mendeteksi penyakit hepatitis berbasis data biokimia pasien. Decision Tree unggul dalam hal interpretabilitas dan transparansi hasil, sedangkan Naive Bayes menonjol dalam efisiensi dan stabilitas model. Berdasarkan kajian tersebut, penelitian ini berfokus pada analisis perbandingan kedua algoritma untuk menentukan model dengan tingkat akurasi terbaik dalam memprediksi penyakit hepatitis.

3. Bahan & Metode

Penelitian ini menggunakan pendekatan kuantitatif berbasis machine learning untuk membandingkan performa algoritma Decision Tree dan Naive Bayes dalam memprediksi penyakit Hepatitis C. Alur umum penelitian meliputi: Pengambilan Data, Pra-pemrosesan, Split Data 70/30, Pelatihan Model, Evaluasi, Hasil dan Analisis.



Gambar 1. Tahapan Penelitian



Lisensi

Lisensi Internasional Creative Commons Attribution-ShareAlike 4.0.

3.1. Pengambilan Data

Dataset yang dipakai pada penelitian ini diambil dari platform Kaggle bertajuk “Hepatitis C Prediction” (<https://www.kaggle.com/code/mohamedzaghloula/hepatitis-c-prediction>). Dataset ini berisi 615 data pasien dengan 14 atribut biokimia darah, seperti ALT, AST, ALP, Bilirubin, Protein, dan Cholesterol. Setiap data pasien dikategorikan ke dalam empat kelas, yaitu Blood Donor, Suspect Blood Donor, Hepatitis, serta Fibrosis/Cirrhosis.

Unnamed: 0	Category	Age	Sex	ALB	ALP	ALT	AST	BIL	CHE	CHOL	CREA	GGT	PROT
1	0=Blood Donor	32	m	38.5	52.5	7.7	22.1	7.5	6.93	3.23	106.0	12.1	69.0
2	0=Blood Donor	32	m	38.5	70.3	18.0	24.7	3.9	11.17	4.8	74.0	15.6	76.5
3	0=Blood Donor	32	m	46.9	74.7	36.2	52.6	6.1	8.84	5.2	86.0	33.2	79.3
4	0=Blood Donor	32	m	43.2	52.0	30.6	22.6	18.9	7.33	4.74	80.0	33.8	75.7
5	0=Blood Donor	32	m	39.2	74.1	32.6	24.8	9.6	9.15	4.32	76.0	29.9	68.7
6	0=Blood Donor	32	m	41.6	43.3	18.5	19.7	12.3	9.92	6.05	111.0	91.0	74.0
7	0=Blood Donor	32	m	46.3	41.3	17.5	17.8	8.5	7.01	4.79	70.0	16.9	74.5
8	0=Blood Donor	32	m	42.2	41.9	35.8	31.1	16.1	5.82	4.6	109.0	21.5	67.1
9	0=Blood Donor	32	m	50.9	65.5	23.2	21.2	6.9	8.69	4.1	83.0	13.7	71.3
10	0=Blood Donor	32	m	42.4	86.3	20.3	20.0	35.2	5.46	4.45	81.0	15.9	69.9
11	0=Blood Donor	32	m	44.3	52.3	21.7	22.4	17.2	4.15	3.57	78.0	24.1	75.4
12	0=Blood Donor	33	m	46.4	68.2	10.3	20.0	5.7	7.36	4.3	79.0	18.7	68.6
13	0=Blood Donor	33	m	36.3	78.6	23.6	22.0	7.0	8.56	5.38	78.0	19.4	68.7
14	0=Blood Donor	33	m	39.0	51.7	15.9	24.0	6.8	6.46	3.38	65.0	7.0	70.4
15	0=Blood Donor	33	m	38.7	39.8	22.5	23.0	4.1	4.63	4.97	63.0	15.2	71.9
16	0=Blood Donor	33	m	41.8	65.0	33.1	38.0	6.6	8.83	4.43	71.0	24.0	72.7
17	0=Blood Donor	33	m	40.9	73.0	17.2	22.9	10.0	6.98	5.22	90.0	14.7	72.4
18	0=Blood Donor	33	m	45.2	88.3	32.4	31.2	10.1	9.78	5.51	102.0	48.5	76.5
19	0=Blood Donor	33	m	36.6	57.1	38.9	40.3	24.9	9.62	5.5	112.0	27.6	69.3
20	0=Blood Donor	33	m	42.0	63.1	32.6	34.9	11.2	7.01	4.05	105.0	19.1	68.1

Gambar 2. Dataset Keseluruhan

```
file_path = "/content/HepatitisCdata.csv"
if not os.path.exists(file_path):
    print("⚠️ File tidak ditemukan di {file_path}. Silakan upload terlebih dahulu.")
else:
    df = pd.read_csv(file_path)
    print("✅ Dataset berhasil dibaca!\n")
    print("Jumlah baris dan kolom:", df.shape)
```

Gambar 3. Code memuat dataset pada platform Google Colab

3.2. Lingkungan dan Alat Penelitian

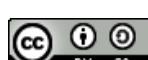
Seluruh eksperimen dilakukan menggunakan platform Google Colab, yang berfungsi sebagai cloud-based Jupyter Notebook. Alasan pemilihan platform ini adalah karena kemampuannya dalam menyediakan lingkungan pemrograman Python secara daring tanpa perlu instalasi lokal, serta dukungan GPU gratis untuk mempercepat proses komputasi.

Beberapa pustaka Python yang digunakan antara lain:

1. Pandas untuk pembersihan dan manipulasi data,
2. NumPy untuk perhitungan numerik,
3. Matplotlib dan Seaborn untuk visualisasi data,
4. Scikit-learn untuk implementasi algoritma Decision Tree, Naive Bayes, dan evaluasi performa model.

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier, plot_tree
from sklearn.naive_bayes import GaussianNB
from sklearn.preprocessing import LabelEncoder
from sklearn.metrics import accuracy_score, classification_report, confusion_matrix
import os
```

Gambar 4. Code Mengimpor Semua Library



Lisensi

Lisensi Internasional Creative Commons Attribution-ShareAlike 4.0.

3.3. Pra-pemrosesan Data

Tahapan ini bertujuan untuk memastikan data siap digunakan oleh model. Langkah-langkah yang dilakukan meliputi:

1. Pembersihan data, yaitu menghapus nilai kosong (missing values) dan duplikat.
2. Normalisasi fitur menggunakan metode Min-Max Scaling agar semua variabel memiliki skala yang seragam.
3. Encoding label untuk mengubah kategori penyakit menjadi nilai numerik agar dapat diproses oleh algoritma machine learning.

```
if 'Unnamed: 0' in df.columns:  
    df = df.drop(columns=['Unnamed: 0'])  
  
df.fillna(df.mean(numeric_only=True), inplace=True)  
  
le_sex = LabelEncoder()  
le_cat = LabelEncoder()  
df['Sex'] = le_sex.fit_transform(df['Sex'])  
df['Category'] = le_cat.fit_transform(df['Category'])
```

Gambar 5. Code Data Diolah Oleh Algoritma Machine Learning

3.4. Pembagian Data (Data Splitting)

Dataset dibagi menjadi dua bagian, yaitu:

1. 70% data latih (training set), digunakan untuk membangun model, dan
2. 30% data uji (testing set), digunakan untuk menguji kinerja model.

Proses pembagian dilakukan menggunakan fungsi `train_test_split()` dari pustaka scikit-learn dengan parameter `random_state` agar hasil pembagian data bersifat konsisten dan dapat direproduksi.

```
x = df.drop(columns=['Category'])  
y = df['Category']  
  
x_train, x_test, y_train, y_test = train_test_split(  
    x, y, test_size=0.3, random_state=42  
)
```

Gambar 6. Code Dataset Dibagi Menjadi 70% Training Set Dan 30% Testing Set

3.5. Pelatihan Model

Tahap pelatihan dilakukan dengan dua algoritma utama:

```
# Decision Tree Model  
tree_model = DecisionTreeClassifier(random_state=42)  
tree_model.fit(x_train, y_train)  
y_pred_tree = tree_model.predict(x_test)  
  
# Naive Bayes Model  
nb_model = GaussianNB()  
nb_model.fit(x_train, y_train)  
y_pred_nb = nb_model.predict(x_test)
```

Gambar 7. Code Membangun Model Berbasis Aturan Pemisahan dan Algoritma Probabilistik Sederhana



Lisensi

Lisensi Internasional Creative Commons Attribution-ShareAlike 4.0.

1. Decision Tree Classifier, algoritma berbasis pohon keputusan yang memecah data berdasarkan atribut paling informatif.
2. Naive Bayes Classifier (GaussianNB), algoritma probabilistik sederhana dengan asumsi independensi antar fitur.

Kedua model dilatih menggunakan data latih (70%) dan kemudian diuji menggunakan data uji (30%) untuk mengukur kemampuan prediksi.

3.6. Evaluasi Model

Kinerja masing-masing model diukur menggunakan beberapa metrik evaluasi, yaitu:

1. Akurasi (Accuracy), mengukur proporsi prediksi benar terhadap total data.
2. Presisi (Precision), menunjukkan ketepatan prediksi positif.
3. Recall (Sensitivity), mengukur kemampuan model mendeteksi kelas positif.
4. F1-Score, kombinasi presisi dan recall untuk hasil yang seimbang.
5. Confusion Matrix, digunakan untuk menganalisis kesalahan klasifikasi antar kelas.

```
from sklearn.metrics import classification_report, confusion_matrix, accuracy_score

acc_tree = accuracy_score(y_test, y_pred_tree)
acc_nb = accuracy_score(y_test, y_pred_nb)

print(f"牢固树立 Akurasi Decision Tree: {acc_tree:.4f}")
print(f"牢固树立 Akurasi Naive Bayes: {acc_nb:.4f}")
```

Gambar 8. Code Kinerja model diukur dengan beberapa metrik

3.7. Hasil dan Analisis

Tahap terakhir meliputi analisis hasil evaluasi dari kedua model. Nilai akurasi, presisi, recall, dan F1-score dikomparasikan untuk menentukan m terbaik dalam mendeteksi penyakit Hepatitis C. Analisis juga dilakukan terhadap fitur-fitur laboratorium yang paling berpengaruh terhadap hasil klasifikasi.

Untuk membantu interpretasi hasil, beberapa bentuk visualisasi digunakan, antara lain: grafik perbandingan akurasi antar model untuk melihat performa secara umum; confusion matrix dari masing-masing model untuk menganalisis distribusi prediksi benar dan salah; visualisasi struktur Decision Tree untuk memahami pola pengambilan keputusan model berbasis pohon; serta analisis Feature Importance guna mengidentifikasi atribut biokimia darah yang memiliki kontribusi paling signifikan terhadap hasil prediksi

```
feature_importance = pd.Series(
    tree_model.feature_importances_, index=x.columns
).sort_values(ascending=True)

plt.figure(figsize=(10,6))
feature_importance.plot(kind='barh', color='teal')
plt.title("Feature Importance - Pengaruh Fitur terhadap Prediksi Hepatitis")
plt.xlabel("Nilai Importance")
plt.ylabel("Nama Fitur")
plt.show()
```

Gambar 9. Code Visualisasi Feature Importance

4. Hasil

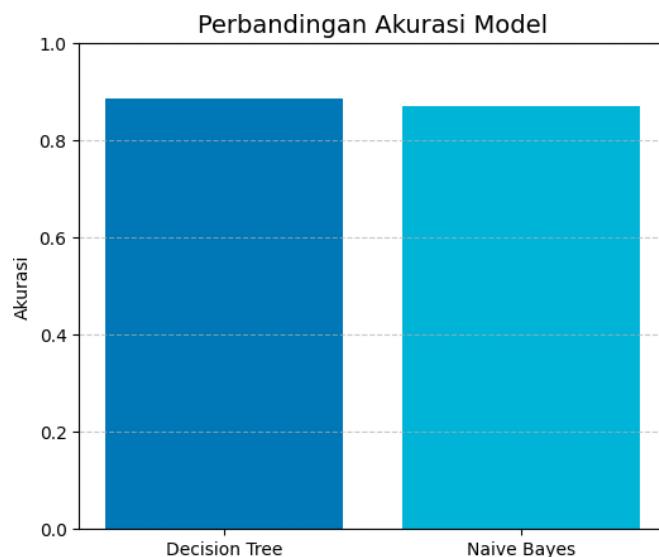


Lisensi

Lisensi Internasional Creative Commons Attribution-ShareAlike 4.0.

Penelitian ini menggunakan dua algoritma klasifikasi, yaitu Decision Tree dan Naive Bayes, untuk memprediksi status penyakit Hepatitis C berdasarkan data hasil laboratorium yang diambil dari platform Kaggle. Dataset terdiri dari 13 fitur yang meliputi parameter biokimia seperti kadar enzim hati (ALT, AST, ALP), protein, bilirubin, kolesterol, dan lainnya, serta 1 variabel target yaitu Category yang menunjukkan status kesehatan pasien. Data dibagi menjadi dua bagian, yaitu 70% data latih dan 30% data uji untuk memastikan model dapat diuji pada data yang tidak pernah dilihat sebelumnya. Proses pelatihan dan pengujian model dengan menggunakan Google Colab.

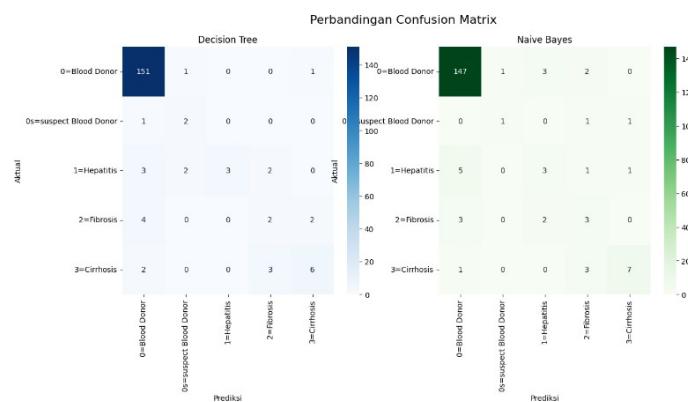
Hasil pengujian menunjukkan bahwa model Decision Tree memperoleh akurasi dengan nilai 88,65%, sedangkan model Naive Bayes memperoleh akurasi mencapai 87,03%. Hal ini menunjukkan bahwa kedua model memiliki performa yang cukup baik dalam mengklasifikasikan data pasien Hepatitis C, dengan Decision Tree sedikit lebih unggul dalam hal akurasi.



Gambar 20. Grafik Perbandingan Akurasi Model Decision Tree dan Naive Bayes

4.1. Analisis Confusion Matrix

Confusion Matrix digunakan untuk memahami distribusi klasifikasi model dalam membedakan setiap kategori pasien. Hasil menunjukkan bahwa Decision Tree memiliki tingkat klasifikasi yang lebih baik untuk kategori utama (sehat dan donor darah), sedangkan Naive Bayes masih menghasilkan kesalahan pada kategori minoritas seperti fibrosis dan sirosis.



Gambar 31. Grafik Perbandingan Akurasi Model Decision Tree dan Naive Baye

Dari analisis ini, dapat disimpulkan bahwa Decision Tree lebih stabil terhadap data dengan distribusi kelas yang tidak seimbang, sementara Naive Bayes lebih sensitif terhadap variasi data karena sifat probabilistiknya.

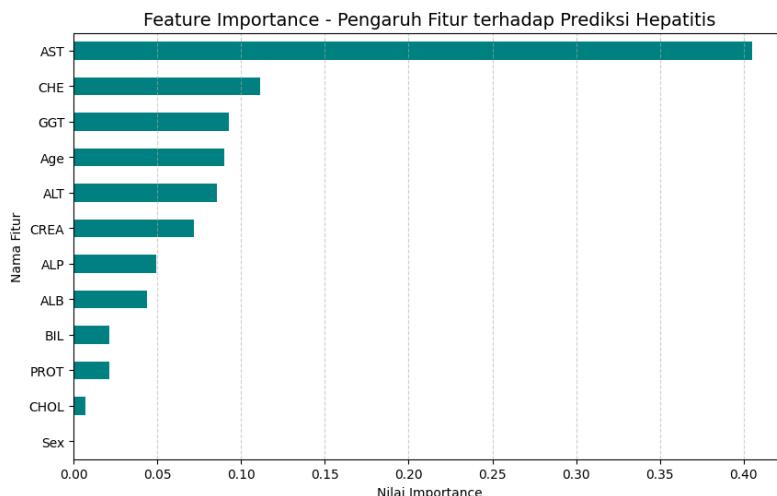


Lisensi

Lisensi Internasional Creative Commons Attribution-ShareAlike 4.0.

4.2. Analisis Feature Importance

Pada kedua model, fitur-fitur seperti AST (Aspartate Aminotransferase), ALT (Alanine Aminotransferase), dan Bilirubin menempati posisi tertinggi dalam tingkat kepentingan terhadap hasil klasifikasi. Fitur-fitur ini menunjukkan kontribusi yang dominan dibandingkan atribut lainnya dalam proses prediksi kategori pasien hepatitis.



Gambar 42. Feature Importance Model Decision Tree dan Naive Bayes

Meskipun kedua model menggunakan pendekatan yang berbeda Decision Tree berbasis pada pemisahan data menggunakan kriteria informasi (information gain) dan Naive Bayes berbasis pada probabilitas bersyarat hasil visualisasi menunjukkan bahwa fitur-fitur utama yang berpengaruh signifikan relatif serupa pada kedua model.

5. Kesimpulan

Studi ini berfokus pada implementasi algoritma machine learning untuk memprediksi kondisi hepatitis menggunakan data laboratorium pasien yang bersumber dari platform Kaggle. Kedua metode yang diuji, yakni Decision Tree dan Naive Bayes, terbukti efektif untuk mengklasifikasikan rekaman medis serta memberikan tingkat akurasi yang memadai.

Secara keseluruhan, Decision Tree menunjukkan kinerja yang lebih unggul dibandingkan Naive Bayes. Keunggulan ini disebabkan oleh kemampuan model Decision Tree untuk mengidentifikasi hubungan non-linear antara variabel-variabel biokimia secara lebih efektif. Meskipun demikian, Naive Bayes tetap menjadi pilihan metode yang efisien dalam hal kecepatan komputasi, membuatnya ideal untuk implementasi pada sistem dengan keterbatasan sumber daya.

Kesimpulan dari studi ini menunjukkan bahwa algoritma klasifikasi berbasis machine learning berperan signifikan dalam mendukung proses diagnosis dini penyakit hepatitis secara akurat dan efisien. Penelitian ini semoga bisa jadi pengembangan sistem pembantu keputusan klinis yang didorong oleh data

REFERENSI



Lisensi

Lisensi Internasional Creative Commons Attribution-ShareAlike 4.0.

- [1] J. Hasil, K. Sosialisasi, and M. Volume, “PERILAKU HIDUP BERSIH DAN SEHAT UNTUK PENCEGAHAN PENYAKIT HEPATITIS DI PANTI ASUHAN,” *J. Kreat. Pengabdi. Kpd. Masy.*, vol. 2, no. 1, pp. 1856–1866, 2024.
- [2] A. Damayanti and G. Testiana, “Penerapan Data Mining Untuk Prediksi Penyakit Hepatitis C Menggunakan Algoritma Naïve Bayes,” *J. Manaj. Inform. Jayakarta*, vol. 3, no. 2, pp. 177–186, 2023, [Online]. Available: <https://doi.org/10.52362/jmijayakarta.v3i2.1098>
- [3] R. G. Gunawan and M. Ilham Pratama, “Analisa Kinerja Algoritma Machine Learning Untuk Prediksi Virus Hepatitis C,” *J. CoSciTech (Computer Sci. Inf. Technol.)*, vol. 4, no. 3, pp. 772–777, 2024, doi: 10.37859/coscitech.v4i3.6513.
- [4] A. Sabriyanti and A. K. Batubara, “Jurnal Indonesia: Manajemen Informatika Dan Komunikasi,” *J. Indones. Manaj. Inform. dan Komun.*, vol. 4, no. 3, pp. 989–998, 2023, [Online]. Available: <https://journal.stmiki.ac.id/index.php/jimik/article/view/301>
- [5] R. G. Gunawan, M. Ilham Pratama, F. D. Prasetya, H. W. Nugroho, J. Triloka, and C. ROMADHON, “Analisa Data Mining Untuk Prediksi Penyakit Hepatitis C Menggunakan Algoritma Decision Tree C.45 Dengan Particle Swarm Optimization,” *Semin. Nas. Has. Penelit. dan Pengabdi. Masy.*, vol. 4, no. 3, pp. 772–777, 2024, [Online]. Available: <http://archive.ic>
- [6] R. E. Nugroho, W. Y. Pamungkas, and J. H. Jaman, “Pendeteksi Penyakit Hepatitis Menggunakan Cart Decision Tree,” *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 3S1, pp. 3690–3696, 2024, doi: 10.23960/jitet.v12i3s1.5184.
- [7] N. Sitohang, “Jurnal Sains Informatika Terapan (JSIT),” Penerapan Data Min. Untuk Peringatan Dini Banjir Menggunakan Metod. Klastering K-Means, vol. 2, no. 1, pp. 16–20, 2023.
- [8] Alfarizi M. Riziq Sirfatullah, Al-farish Muhamad Zidan, Taufiqurrahman Muhamad, Ardiansah Ginan, and Elgar Muhamad, “Penggunaan Python Sebagai Bahasa Pemrograman untuk Machine Learning dan Deep Learning,” *Karya Ilm. Mhs. Bertauhid (KARIMAH TAUHID)*, vol. 2, no. 1, pp. 1–6, 2023.
- [9] R. Puspita and A. Widodo, “Perbandingan Metode KNN, Decision Tree, dan Naïve Bayes Terhadap Analisis Sentimen Pengguna Layanan BPJS,” *J. Inform. Univ. Pamulang*, vol. 5, no. 4, p. 646, 2021, doi: 10.32493/informatika.v5i4.7622.
- [10] A. Damuri, U. Riyanto, H. Rusdianto, and M. Aminudin, “Implementasi Data Mining dengan Algoritma Naïve Bayes Untuk Klasifikasi Kelayakan Penerima Bantuan Sembako,” *JURIKOM (Jurnal Ris. Komputer)*, vol. 8, no. 6, p. 219, 2021, doi: 10.30865/jurikom.v8i6.3655.



Lisensi

Lisensi Internasional Creative Commons Attribution-ShareAlike 4.0.