

Implementasi *Machine Learning* Untuk Memprediksi Hepatitis Mortality Menggunakan Algoritma Decision Tree

Rizky Aprilia¹

¹Fakultas Teknik, Teknik Informatika, Universitas Pamulang, Jl. Raya Puspiptek No. 46, Kel. Buaran, Kec. Serpong, Kota Tangerang Selatan, Banten 15310, Indonesia

Email: rizkyaprilia258@gmail.com

(* : coressponding author)

Abstrak–Hepatitis adalah penyakit yang menyerang hati. Hati adalah organ penting yang memproses nutrisi, menyaring darah dan melawan infeksi. Diagnosis medis memiliki implikasi penting untuk meningkatkan perawatan pasien, penelitian, dan kebijakan. Untuk diagnosis medis, para profesional kesehatan menggunakan berbagai jenis metode patologis untuk membuat keputusan tentang laporan medis dalam hal kondisi medis pasien. Penggunaan kecerdasan buatan dan machine learning yang dikombinasikan dengan temuan klinis telah semakin meningkatkan deteksi penyakit. Algoritma *machine learning* berdasarkan masalah tertentu dapat membantu seseorang untuk membuat keputusan. Machine learning (ML), algoritma berbasis data dapat digunakan untuk memvalidasi metode yang ada dan membantu peneliti untuk membuat keputusan baru yang potensial. Ada banyak teknik yang digunakan dalam machine learning, salah satunya adalah algoritma decision tree yang digunakan dalam penelitian ini. Decision tree adalah metode untuk mendekati fungsi target bernilai diskrit, di mana fungsi-fungsi yang dipelajari diwakili oleh pohon keputusan. Studi ini menyajikan visualisasi data yang dihasilkan dengan membangun program machine learning untuk mengumpulkan parameter yang diperlukan untuk pemodelan dan memvisualisasikan data didalam *machine learning web app*.

Kata Kunci: Hepatitis, *Machine Learning*, *Decision Tree*, *Machine Learning Web App*

Abstract–*Hepatitis is a disease that affects the liver. The liver is an important organ that processes nutrients, filters blood, and fights infections. Medical diagnosis has important implications for improving patient care, research, and policy. For medical diagnosis, healthcare professionals use different types of pathological methods to make decisions based on medical reports in terms of a patient's medical condition. The use of artificial intelligence and machine learning combined with clinical findings has further improved disease detection. Machine learning algorithms based on specific problems can help one make decisions. Machine learning (ML) and data-driven algorithms can be used to validate existing methods and help researchers make potential new decisions. There are many techniques used in machine learning, one of which is the decision tree algorithm used in this study. A decision tree is a method for approximating discrete-valued target functions, where the learned functions are represented by decision trees. This study presents the resulting data visualization by building a machine learning program to collect the parameters required for modeling and visualizing the data that demonstrates problem solving.*

Keywords: *Hepatitis, Machine Learning, Decision Tree, Machine Learning Web App*

1. PENDAHULUAN

Hepatitis adalah penyakit yang menyerang hati. Hati adalah organ penting yang memproses nutrisi, menyaring darah dan melawan infeksi. Virus hepatitis tipe B yang disebabkan oleh virus hepatitis B (HBV) adalah penyakit serius yang berpotensi mengancam jiwa yang dapat dicegah dengan vaksinasi. Kebanyakan orang yang terinfeksi HBV (baru) tetap tidak menunjukkan gejala dan tidak mengetahui status hepatitis mereka selama bertahun-tahun. Hanya beberapa orang yang baru terinfeksi HBV yang memiliki gejala hepatitis akut (Pattyn et al., 2021). Menurut data yang diterbitkan oleh WHO, Di kawasan Asia Tenggara, sekitar 100 juta orang menderita hepatitis B kronis dan 30 juta menderita hepatitis C kronis. Hepatitis B menyebabkan hampir 1,4 juta kasus baru dan 300.000 kematian di kawasan itu setiap tahun (Kemenkes RI, 2016). Oleh karena itu, perlu dilakukan prediksi mortality hepatitis untuk mengendalikan penyakitnya.

Machine learning telah menjadi sangat populer dalam beberapa tahun terakhir. Berbagai macam model tersedia untuk memahami dan memprediksi pola dan tren data di berbagai bidang. Interpretasi model *machine learning* saat ini menjadi topik hangat di komunitas visualisasi data, dan hasil menunjukkan bahwa model pembelajaran dapat memberikan prediksi yang lebih baik dan

meningkatkan keandalan hasil. Teknik *machine learning* dan *data mining* telah digunakan secara luas dan efektif oleh para peneliti di berbagai bidang dan bidang, namun sebagian besar informasi yang disajikan masih berupa data kasar, sehingga tidak ada gambaran yang jelas tentang konten atau skenario penelitian. pembaca akan menemukan dalam setiap studi.

Diagnosis medis memiliki implikasi penting untuk meningkatkan perawatan pasien, penelitian, dan kebijakan. Untuk diagnosis medis, para profesional kesehatan menggunakan berbagai jenis metode patologis untuk membuat keputusan tentang laporan medis dalam hal kondisi medis pasien. Baru-baru ini, para dokter telah secara aktif terlibat dalam meningkatkan diagnosis medis. Penggunaan kecerdasan buatan dan *machine learning* yang dikombinasikan dengan temuan klinis telah semakin meningkatkan deteksi penyakit. Di era modern, dengan keunggulan komputer dan teknologi, seseorang dapat mengumpulkan data dan memvisualisasikan banyak hasil tersembunyi seperti menangani data yang hilang dalam penelitian medis. Algoritma *machine learning* berdasarkan masalah tertentu dapat membantu seseorang untuk membuat keputusan. *Machine learning* (ML), algoritma berbasis data dapat digunakan untuk memvalidasi metode yang ada dan membantu peneliti untuk membuat keputusan baru yang potensial (Mostafa et al., 2021).

Ada banyak teknik yang digunakan dalam machine learning, salah satunya adalah algoritma decision tree yang digunakan dalam penelitian ini. *Decision tree* adalah metode untuk mendekati fungsi target bernilai diskrit, di mana fungsi-fungsi yang dipelajari diwakili oleh pohon keputusan. Pembelajaran *decision tree* adalah salah satu metode yang paling banyak yang paling banyak digunakan dan metode praktis untuk inferensi induktif (Krishnachandran, 2018).

Studi ini menyajikan visualisasi data yang dihasilkan dengan membangun program *machine learning* untuk mengumpulkan parameter yang diperlukan untuk pemodelan dan memvisualisasikan data yang menunjukkan pemecahan masalah. Oleh karena itu, penulis ingin mengimplementasikan model algoritma decision tree yang divisualisasikan sebagai aplikasi web machine learning.

2. METODOLOGI PENELITIAN

Metode penelitian adalah tahapan atau cara yang dilakukan dalam penelitian. Secara umum tahapan penelitian ini meliputi, *Data Preparation, Data Cleansing, Exploratory Data Analysis, Feature Engineering, Modeling, Evaluation, Deployment*. Data diolah menggunakan bahasa pemrograman python dengan jupyter serta framework streamlit untuk deployment aplikasi menjadi web app machine learning.

2.1 Dataset

Dataset yang digunakan merupakan dataset yang berasal dari repository UCI *machine learning* dengan menggunakan alamat website yang langsung mudah dapat di akses melalui <https://archive.ics.uci.edu/ml/datasets/Hepatitis>. Data disumbangkan oleh G.Gong (Carnegie Mellon University) melalui Bojan Cestnik di Yugoslavia pada tanggal 1 November 1988. Total data yang tersedia adalah 155 data dengan 123 pasien hepatitis hidup dan 32 pasien hepatitis sekarat. Atribut yang termasuk dalam dataset hepatitis tersebut seperti dalam tabel dibawah ini:

Tabel 1. Dataset

No	Atribute	Value
1	Age	0 - 80
2	Sex	Laki, Perempuan
3	Steroid	No, Yes
4	Antivirals	No, Yes
5	Fatigue	No, Yes
6	Malaise	No, Yes
7	Anorexia	No, Yes
8	Liver Big	No, Yes
9	Liver Firm	No, Yes
10	Spleen Palpable	No, Yes
11	Spiders	No, Yes

12	Ascites	No, Yes
13	Varices	No, Yes
14	Bilirubin	0 = 4.00
15	Alk Phosphate	30 - 250
16	Sgot	10 - 500
17	Albumin	2.0 - 6.0
18	Protime	10 - 100
19	Histology	No, Yes

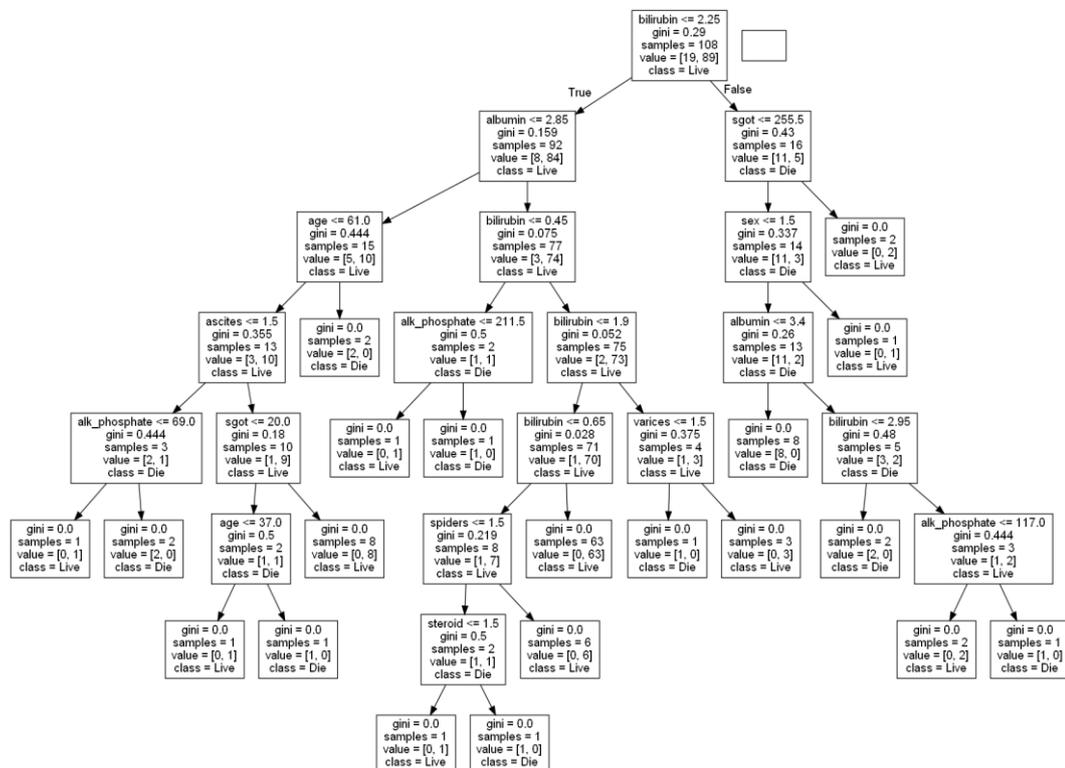
3. ANALISA DAN PEMBAHASAN

3.1 Graph

Dikarenakan gini cenderung lebih cepat dari entropy karena tidak perlu melakukan proses kalkulasi log, maka dalam membuat graph *decision tree* dengan menghitung nilai dari gini index tersebut. Gini Index mengukur nilai impurity dari sebuah node. Gini index memuat rentang nilai dari 0 hingga (1-1/n) dimana n adalah jumlah kategori dari suatu dependent variable.

$$Gini = 1 - \sum_{i=1}^c (P_i)^2$$

Dimana *c* adalah jumlah nilai dari masing-masing atribut dan *P_i* adalah jumlah masing-masing atribut berdasarkan kelas atau labelnya. Kemudian dilakukanlah proses penguadratan. Setelah itu dilakukanlah penjumlahan nilai *gini index* pada satu atribut. Maka graph yang terbentuk dari dataset hepatitis seperti dibawah ini :



Gambar 1. Graph yang Terbentuk

Dari graph diatas menunjukkan dari 155 data dibagi 70% data *training* dan 30% data *testing* sehingga terdapat tujuh *split* dengan hasil *variable bilirubin* menjadi *root node* karena memiliki nilai gini index paling yaitu sebesar 0,29.

3.2 Nilai Performance Matrix

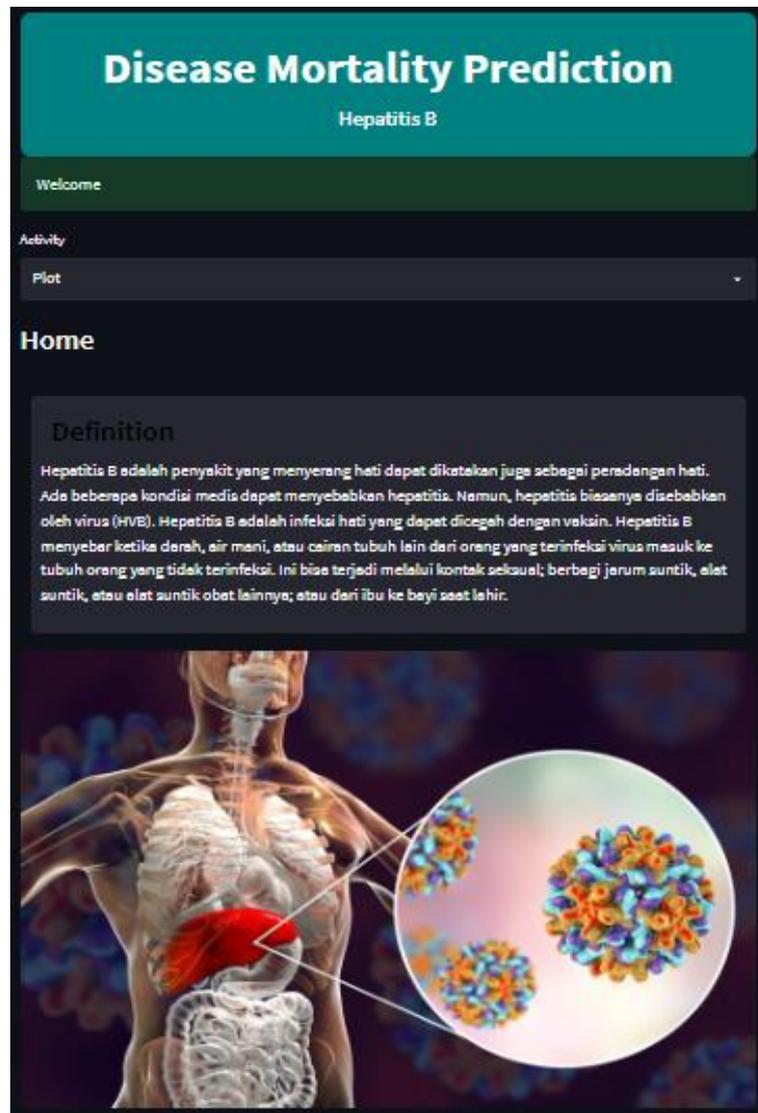
Salah satu teknik yang dapat digunakan untuk mengukur keefektifan suatu model, khususnya untuk klasifikasi (*supervised learning*) dalam *machine learning*, adalah dengan confusion matrix

Tabel 2. Nilai Performace Matrix

Nilai Confusion Matrix				
Accuracy	Precision		Recall	
	Live	Die	Live	Die
74%	79%	56%	88%	38%

4. IMPLEMENTASI

Dari model yang terbentuk dari tahapan *data science* adalah proses *deployment*. Dalam penelitian ini menggunakan *framework streamlit*. Berikut tampilan dari *web app machine learning* prediksi *hepatitis mortality*:



Gambar 2. Web App Machine Learning



5. KESIMPULAN

Berdasarkan hasil penelitian akurasi model menggunakan algoritma Decision tree didapatkan akurasi sebesar 74%, Precision pada kelas “Live” sebesar 79% dan kelas “Die” sebesar 56%. Recall pada kelas “Live” sebesar 88% dan kelas “Die” sebesar 38%. Pengujian selanjutnya Pengujian lebih lanjut sebaiknya menggunakan data sampel yang lebih banyak, dan berupa data primer serta dapat mencoba pemodifikasian metode lain agar akurasi model yang dibuat dapat meningkat.

REFERENCES

- Kemendes RI. (2016). Sebagian Besar Kematian Akibat Hepatitis Virus Berhubungan Dengan Hepatitis B Dan C Kronis. *Selasa, 26 April, April*.
- Krishnachandran, V. N. (2018). *Lecture Notes in MACHINE LEARNING*.
- Mostafa, F., Hasan, E., Williamson, M., & Khan, H. (2021). Statistical Machine Learning Approaches to Liver Disease Prediction. *Livers, 1*(4), 294–312. <https://doi.org/10.3390/livers1040023>
- Pattyn, J., Hendrickx, G., Vorsters, A., & van Damme, P. (2021). Hepatitis B Vaccines. *Journal of Infectious Diseases, 224*. <https://doi.org/10.1093/infdis/jiaa668>.