

Implementasi Algoritma Klasifikasi *Logistic Regression* dan *Naïve Bayes* untuk Diagnosa Penyakit Hepatitis

Amrin¹, Omar Pahlevi²

^{1,2} Universitas Bina Sarana Informatika

¹amrin.ain@bsi.ac.id

²omar.opi@bsi.ac.id

Diterima	Direvisi	Disetujui
09-02-2022	06-06-2022	13-06-2022

Abstrak - Hepatitis merupakan penyakit kronis dan berbahaya yang dapat menyebabkan kematian. Diagnosis dini penyakit ini penting dilakukan agar dapat segera ditangani dan diobati. Tahapan kegiatan diagnosis yang tepat dan prediksi penyakit yang akurat pada waktu dapat menyelamatkan banyak pasien. Penyakit ini biasanya disebabkan oleh virus. Virus yang menyebabkan Hepatitis A, B dan C, yaitu Hepatitis Virus tipe A (HVA), Hepatitis Virus tipe B (HVB) dan Hepatitis Virus tipe C (HVC). Besaran masalah penyakit ini di Indonesia dapat diketahui dari berbagai studi, kajian, maupun kegiatan pengamatan penyakit. Dalam penelitian ini, peneliti akan menerapkan dan membandingkan metode klasifikasi *data mining* yaitu metode *Logistic Regression* dan *Naïve Bayes* untuk mendiagnosis penyakit hepatitis. Berdasarkan penelitian diketahui bahwa metode *Logistic Regression* memiliki tingkat akurasi sebesar 84,62% dan nilai *under the curve* (AUC) sebesar 0,841, kemudian metode *Naïve Bayes* memiliki tingkat akurasi sebesar 83,71% dan nilai AUC sebesar 0,816. Dari hasil uji-t dapat diketahui bahwa tidak terdapat perbedaan yang signifikan antara metode *Logistic Regression* dan *Nave Bayes*, karena nilai = 0,821 > 0,05. Hal ini menunjukkan bahwa metode *Logistic Regression* memiliki performansi yang sama dibandingkan dengan metode *Naïve Bayes*.

Kata Kunci: Hepatitis, *Logistic Regression*, *Naïve Bayes*, *Confusion Matrix*, *ROC Curve*

Abstract - Hepatitis is a chronic and dangerous disease that can cause death. Early diagnosis of this disease is important so that it can be treated quickly and treated. Accurate stage of diagnostic activities and accurate disease prediction on time can save many patients. This disease is usually caused by a virus. Viruses that cause Hepatitis A, B and C, namely Hepatitis Virus type A (HVA), Hepatitis Virus type B (HVB) and Hepatitis Virus type C (HVC). The magnitude of this disease problem in Indonesia can be known from various studies and disease observation activities. In this research, the researchers will apply and compare data mining classification methods, namely *Logistic Regression* and *Naïve Bayes* methods to diagnose hepatitis. Based on the research, it is known that the *Logistic Regression* method has an accuracy rate of 84.62% and an *under the curve* (AUC) value of 0.841, then the *Naïve Bayes* method has an accuracy rate of 83.71% and an AUC value of 0.816. From the results of the t-test, it can be seen that there is no significant difference between *Logistic Regression* and *Nave Bayes*, because the value = 0.821 > 0.05. This shows that the *Logistic Regression* method has the same performance compared to the *Naïve Bayes* method.

Keywords: Hepatitis, *Logistic Regression*, *Naïve Bayes*, *Confusion Matrix*, *ROC Curve*

PENDAHULUAN

Hati merupakan salah satu organ yang sangat penting dalam tubuh manusia. Organ vital dalam tubuh ini disebut juga dengan kelenjar eksokrin, dikarenakan memiliki fungsi mengeluarkan empedu dalam usus. Namun dengan segala fungsi dan manfaatnya, hati pun tidak luput dari gangguan yang dapat menimbulkan penyakit. Sebagai contoh gangguan pada hati adalah hepatitis dan menjadi masalah utama pada bidang kesehatan di dunia, termasuk di negara Indonesia (Zuama, 2021).

Hepatitis didefinisikan sebagai peradangan pada hati dan paling sering disebabkan oleh virus infeksi menyebabkan 1,5 juta kematian di seluruh dunia setiap tahun (Trishna et al., 2019). Salah satu penyakit menular yang menjadi masalah kesehatan masyarakat yang berpengaruh terhadap angka kematian, status kesehatan masyarakat, angka harapan hidup, dan dampak sosial ekonomi lainnya. Besaran masalah penyakit ini di Indonesia dapat diketahui dari berbagai studi, kajian, maupun kegiatan pengamatan penyakit (Kemenkes RI,

2017). Penyakit ini biasanya disebabkan oleh virus. Virus yang menyebabkan Hepatitis A, B dan C, yaitu Hepatitis Virus tipe A (HVA), Hepatitis Virus tipe B (HVB) dan Hepatitis Virus tipe C (HVC) (Hendri et al., 2021).

Diagnosis yang tepat dan prediksi penyakit hepatitis yang akurat pada waktu dapat menyelamatkan banyak pasien. Algoritma *data mining* dapat digunakan secara efisien dalam prediksi dan klasifikasi. Tujuan dari analisis ini adalah untuk mengklasifikasikan dan mengukur akurasi hepatitis (Ramasamy et al., 2015). Pada tahapan diagnosis penyakit hepatitis ini berupa gejala-gejala yang muncul atau dialami oleh pasien serta faktor resiko yang dirasakan oleh pasien (Papuangan, 2018).

Kegiatan prediksi awal mengenai penentuan penyakit hepatitis perlu dilakukan karena banyak yang menyepelekan penyakit hepatitis. Hal ini dapat dilakukan dengan pengumpulan data pasien ataupun data hasil cek kesehatan (Sulastrri et al., 2020) Melihat banyaknya kasus kematian yang disebabkan oleh penyakit hepatitis, maka diperlukan satu langkah dini sebagai upaya pencegahan penyakit hepatitis. Seiring dengan perkembangan dalam dunia IT (*Information and Technology*). Kehadiran cabang ilmu di bidang *data mining* telah menarik banyak perhatian, seperti melakukan prediksi penyakit hepatitis (Wibowo & Indriyawati, 2020).

Data mining merupakan proses penggalian informasi baru berdasarkan data yang secara signifikan dapat meningkatkan kualitas keputusan dan memberikan peran penting untuk sistem kecerdasan di bidang medis. Proses ini banyak digunakan dalam tahapan engineering termasuk klasifikasi, pengelompokan, regresi, analisis asosiasi, dan sebagainya (Syafa'ah et al., 2021).

Algoritma klasifikasi pada *data mining* pertama-tama menggunakan *data training* untuk membangun model dan kemudian model yang dihasilkan diterapkan pada data uji untuk mendapatkan prediksi. Metode klasifikasi yang berbeda telah diterapkan untuk diagnosis penyakit dan temuannya sangat menjanjikan. Teknik ini dapat meminimalkan kesalahan diagnostik dan hasil dapat diperoleh dalam waktu singkat (Ibrahim & Abdulazeez, 2021). Pada penelitian ini, peneliti akan menerapkan dua buah metode klasifikasi *data mining*, yaitu metode *Logistic Regression* dan *Naïve Bayes* untuk mendiagnosa penyakit hepatitis. Alasan penggunaan algoritma *Logistic Regression* dikarenakan merupakan salah satu metode klasik yang biasanya digunakan dalam teknik klasifikasi. Pada algoritma *Logistic Regression* dapat memberikan hubungan antara variabel independen dan variabel dependen jika variabel dependen bersifat kategorikal (Bayrak et al., 2019). Serta menggunakan *Naïve Bayes*, dikarenakan memiliki kelebihan berupa algoritmanya yang sederhana namun mampu menghasilkan akurasi yang tinggi (Pusporani et al.,

2019). Meskipun bergantung pada kondisi *bold assumption*, *Naïve Bayes Classifier* biasanya mencapai hasil yang baik dalam banyak situasi penerapan aplikasi klasifikasi secara praktis (Chen et al., 2017). *Tools* yang digunakan pada penelitian ini adalah Rapid Miner Studio versi 9.8. Penelitian ini diharapkan dapat membantu para tenaga kesehatan untuk mendiagnosa secara dini penyakit hepatitis.

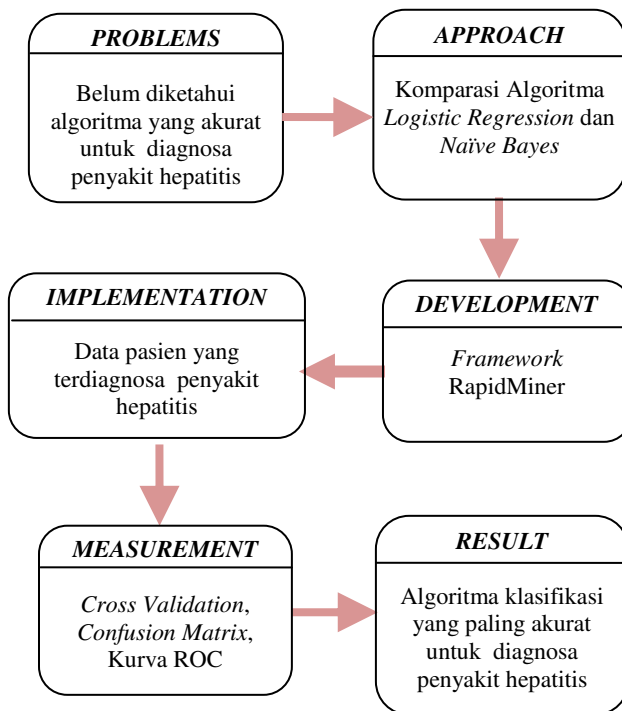
Sebagai bahan acuan dan perbandingan, peneliti melakukan studi literatur dengan beberapa penelitian terdahulu yang terkait dengan tema metode-metode yang pernah digunakan untuk menyelesaikan prediksi penyakit hepatitis, diantaranya penelitian yang dilakukan oleh (Bhargav et al., 2018) yang melakukan penelitian dengan judul *Application of Machine Learning Classification Algorithms on Hepatitis Dataset*. Pada penelitian ini membahas mengenai penerapan teknik klasifikasi seperti *Logistic Regression*, *Decision Tree*, *Linear Support Vector* dan *Naïve Bayes* dalam mendiagnosa penyakit hepatitis C. Dari hasil penelitian ini diperoleh bahwa algoritma *Logistic Regression* menghasilkan akurasi optimum sebesar 87,17%, algoritma *Decision Tree* dengan akurasi optimal sebesar 82,05%. Kemudian algoritma *Linear Support Vector Machine* dengan akurasi optimal sebesar 76,92%, dan yang terakhir adalah algoritma *Naïve Bayes* yang memiliki akurasi optimal 76,92%. Dimana pada akhirnya empat macam algoritma ini diharapkan dapat membantu dalam mengklasifikasikan apakah seorang pasien pengidap penyakit hepatitis C bertahan hidup atau meninggal. Penelitian yang terkait selanjutnya merupakan penelitian yang dilakukan oleh (Nivaan & Emanuel, 2020) dengan judul penelitian *Analytic Predictive of Hepatitis using The Regression Logic Algorithm*. Dalam penelitian ini, teknik regresi dalam *machine learning* yang digunakan adalah *Logistic Regression* untuk prediksi hepatitis dengan menggunakan *dataset* yang diperoleh dari UCI *Machine Learning Repository*. Dimana hasil akhir penelitian ini adalah penerapan algoritma *Logistic Regression* sebagai metode prediksi cukup baik dengan memperoleh tingkat akurasi sebesar 83,33%.

Penelitian yang dilakukan oleh (Amrin & Pahlevi, 2020) tentang *Data Mining Model For Designing Diagnostic Applications Inflammatory Liver Disease*. Pada penelitian ini membandingkan beberapa metode klasifikasi *data mining*, antara lain algoritma C4.5, *Naïve Bayes*, dan *k-Nearest Neighbor* untuk mendiagnosis penyakit radang hati, kemudian membandingkan mana dari ketiga metode tersebut yang paling akurat. Berdasarkan hasil pengukuran performansi ketiga model menggunakan metode *Cross Validation*, *Confusion Matrix* dan *ROC Curve*, diketahui bahwa metode C4.5 merupakan metode terbaik dengan akurasi 70,99% dan *under the curva* (AUC) nilai 0,950, kemudian metode *k-Nearest Neighbor* dengan akurasi 67,19% dan nilai *under the curve* (AUC) 0,873, kemudian

metode *Naïve Bayes* dengan tingkat akurasi 66,14% dan nilai *under the curve* (AUC) sebesar 0,742. Selanjutnya penelitian yang dilakukan oleh (Prayoga, 2018) dan kawan-kawan yang meneliti tentang Diagnosis penyakit hepatitis dengan metode *Naïve Bayes*. Dari pengujian tingkat akurasi yang dilakukan, diketahui bahwa akurasi metode *Naïve Bayes* adalah sebesar 87,50%.

METODOLOGI PENELITIAN

Penelitian ini terdiri dari beberapa tahap seperti terlihat pada kerangka pemikiran gambar 1. Permasalahan pada penelitian ini adalah belum diketahui algoritma yang akurat untuk diagnosa penyakit hepatitis. Untuk itu dibuat *approach* (model) yaitu Algoritma *Logistic Regression* dan *Naïve Bayes* untuk memecahkan permasalahan kemudian dilakukan pengujian terhadap kinerja dari metode tersebut. Pengujian menggunakan metode *Cross Validation*, *Confusion Matrix* dan kurva *Receiver Operating Characteristic*. Pada tahapan pengembangan aplikasi (*development*) berdasarkan model yang dibuat menggunakan Rapid Miner.



Sumber : Hasil Penelitian
Gambar 1. Kerangka Pemikiran Penelitian

HASIL DAN PEMBAHASAN

1. Analisa Data

Pada penelitian ini, *dataset* yang digunakan adalah data pasien hepatitis dari *website UCI Machine Learning Repository*, diakses pada laman <http://archive.ics.uci.edu/ml/machine-learning-databases/hepatitis/>. Jumlah *dataset* sebanyak 155 data dengan 123 pasien penyakit hepatitis yang hidup dan 32 pasien penyakit hepatitis yang mati. Atribut input atau atribut yang menunjukkan gejala sebanyak 19 atribut dan 1 atribut sebagai *output* atau

kelas. Atribut *output* atau keputusan berisi nilai 1 untuk “*die*” dan 2 untuk “*live*”. Adapun deskripsi atribut seperti ditunjukkan pada tabel 1 berikut:

Tabel 1. Deskripsi Atribut

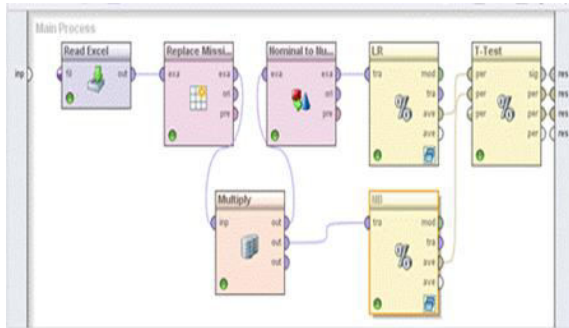
Atribut	Nilai Atribut	Keterangan
Age	Angka numerik	Umur pasien
Sex	Male, Female	Jenis kelamin pasien
Steroid	No, Yes	Apakah mendapatkan terapi steroid?
Antivirals	No, Yes	Apakah mendapatkan terapi antiviral?
Fatigue	No, Yes	Apakah mengalami symptoms atau gejala kelelahan akut?
Malaise	No, Yes	Apakah mengalami symptoms atau gejala malaise (rasa tidak nyaman)?
Anorexia	No, Yes	Apakah mengalami symptoms atau gejala anorexia (muntah setiap maka)?
Liver_Big	No, Yes	Apakah kondisi hati/liver membesar?
Liver_Firm	No, Yes	Apakah kondisi hati/liver mengeras?
Spleen_Palpa ble	No, Yes	Apakah ada gejala spleen palpable/ limfa lebih jelas/besar dari normal?
Spiders	No, Yes	Apakah ada gejala Spider/ pembuluh darah upnormal pada kulit (pembuluh darah mengumpul dan menonjol pada permukaan kulit)?
Ascites	No, Yes	Terjadi penumpukan cairan pada rongga perut?
Varices	No, Yes	Terjadi pembekakan vena esophagus (varises)?
Bilirubin	Angka numerik	Nilai kadar bilirubin dalam darah
Alk_Phosphate	Angka numerik	Kadar Alkalin Phosphate dalam liver
Sgot	Angka numerik	Nilai sgot
Albumin	Angka numerik	Kadar Albumin
Protime	Angka numerik	Uji Masa protrombhine
Histology	No, Yes	Apakah dilakukan pemeriksaan dengan histology (biopsy hati)?
Class	Die, Live	Class apakah pasien positif liver atau tidak

Sumber : Hasil Penelitian

2. Pengujian Model

Penelitian ini dilakukan dengan eksperimen pengujian pada model yang diusulkan. Kemudian dilakukan evaluasi dan validasi model untuk

menghasilkan nilai *accuracy* dan AUC. Pengujian menggunakan Rapid Miner dengan operator *10-fold cross-validation* untuk mendapatkan hasil *accuracy* dan AUC pada setiap algoritma yang diuji. Adapun model yang dibangun ditunjukkan oleh gambar 2 berikut ini:



Sumber : Hasil Penelitian

Gambar 2. Model yang Dibangun Evaluasi yang dilakukan adalah dengan *Confusion Matrix* dan *Receiver Operating Characteristic Curve* (ROC) atau *Area Under Curve* (AUC).

a. *Confusion Matrix*

1) Metode *Logistic Regression*

Tabel 2 adalah *confusion matrix* untuk metode *logistic regression*. Diketahui 113 data diklasifikasi “live (2)” diprediksi sesuai dengan data sebenarnya, lalu 10 data diprediksi “die (1)” tetapi ternyata “live”. Kemudian 18 data diklasifikasi “die (1)” diprediksi sesuai, dan 14 data diprediksi “live (2)” ternyata “die (1)”.

Tabel 2. Model *Confusion Matrix* untuk *Logistic Regression*

<i>accuracy</i> : 84.62% +/- 7.57% (mikro: 84.52%)		
	<i>true</i> 2.0	<i>true</i> 1.0
<i>pred</i> 2.0	113	14
<i>pred</i> 1.0	10	18
<i>class recall</i>	91.87%	56.25%

Sumber : Hasil Penelitian

2) Metode *Naïve Bayes*

Tabel 3 adalah *confusion matrix* untuk metode *naïve bayes*. Diketahui 109 data diklasifikasi “live (2)” tepat sesuai dengan data sebenarnya, lalu 14 data diprediksi “die (1)” tetapi ternyata “live (2)”. Kemudian 21 data diklasifikasi “die (1)” diprediksi tepat sesuai dengan kenyataan, dan 11 data diprediksi “live (2)” ternyata “die (1)”.

Tabel 3. Model *Confusion Matrix* untuk Metode *Naïve Bayes*

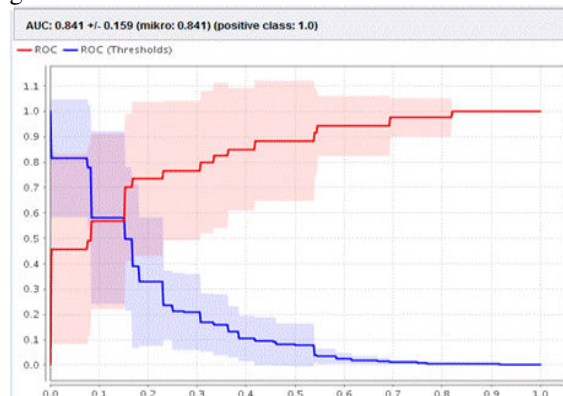
<i>accuracy</i> : 83.71% +/- 9.95% (mikro: 83.87%)		
	<i>true</i> 2.0	<i>true</i> 1.0
<i>pred</i> 2.0	109	11
<i>pred</i> 1.0	14	21
<i>class recall</i>	88.62%	65.62%

Sumber : Hasil Penelitian

b. *Kurva Receiver Operating Characteristic*

1) Metode *Logistic Regression*

Kurva Receiver Operating Characteristic untuk metode *Logistic Regression* seperti ditunjukkan oleh gambar 3 di bawah ini.



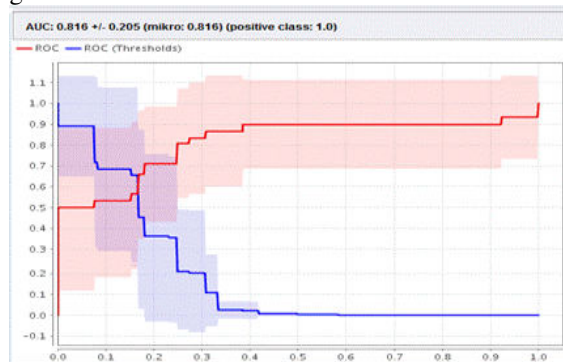
Sumber : Hasil Penelitian

Gambar 3. *Kurva Receiver Operating Characteristic* Metode *Logistic Regression*

Kurva Receiver Operating Characteristic pada gambar 3 mengekspresikan *confusion matrix*. Garis horizontal adalah *false positives* dan garis vertikal *true positives*.

2) Metode *Naïve Bayes*

Kurva Receiver Operating Characteristic untuk metode *Naïve Bayes* seperti ditunjukkan oleh gambar 4 di bawah ini.



Sumber : Hasil Penelitian

Gambar 4. *Kurva Receiver Operating Characteristic* Metode *Naïve Bayes*

3. Analisis Hasil Komparasi

Perbandingan nilai *accuracy* dan AUC untuk algoritma *Logistic Regression* dan *Naïve Bayes* ditunjukkan oleh tabel 4 di bawah ini.

Tabel 4. Komparasi Nilai *Accuracy* dan AUC

Metode	<i>Accuracy</i>	AUC
<i>Logistic Regression</i>	84.62%	0.841
<i>Naïve Bayes</i>	83.71%	0.816

Sumber : Hasil Penelitian

Tabel 4 membandingkan *accuracy* dan AUC dari tiap metode. Secara sekilas terlihat bahwa nilai *accuracy* metode *Logistic Regression* sedikit lebih tinggi, begitu pula dengan nilai AUC-nya memiliki akurasi yang sedikit lebih tinggi dibandingkan metode *Naïve Bayes*. Untuk menjamin klaim atau dugaan tersebut, maka dilakukan evaluasi menggunakan metode yang paling umum dalam metode statistik tradisional, yaitu dengan *t-Test*. Dapat dilihat pada Gambar 5 hasil *performance t-*

test. Dari hasil *t-test* dapat dilihat bahwa tidak ada perbedaan signifikan antara *Logistic Regression* dengan *Naïve Bayes*, karena nilai $\alpha = 0,821 > 0,05$. Hal tersebut menunjukkan bahwa metode *Logistic Regression* memiliki *performance* atau kinerja yang sama dibandingkan dengan metode *Naïve Bayes*.

	0.846 +/- 0.076	0.837 +/- 0.099
0.846 +/- 0.076		0.821
0.837 +/- 0.099		

Probabilities for random values with the same result.
Bold values are smaller than alpha=0.050 which indicates a probably significant difference between the actual mean values!

Sumber : Hasil Penelitian

Gambar 5. *Performance t-test*

Menurut (Gorunescu, 2011) dalam (Amrin et al., 2021) nilai AUC pada klasifikasi *data mining*, dapat dibagi menjadi beberapa kelompok, antara lain :

- 0.90-1.00 = klasifikasi sangat baik
- 0.80-0.90 = klasifikasi baik
- 0.70-0.80 = klasifikasi cukup
- 0.60-0.70 = klasifikasi buruk
- 0.50-0.60 = klasifikasi salah

Berdasarkan pengelompokan di atas maka dapat disimpulkan bahwa metode *Logistic Regression* dan *Naïve Bayes* termasuk kategori klasifikasi baik.

KESIMPULAN

Kesimpulan yang dapat diambil berdasarkan penelitian ini adalah bahwa performa metode *Logistic Regression* untuk prediksi penyakit hepatitis memberikan tingkat akurasi kebenaran sebesar 84,62% dengan nilai *area under the curva* (AUC) sebesar 0,841. Sedangkan performa metode *Naïve Bayes* memberikan tingkat akurasi kebenaran sebesar 83,71% dengan nilai *area under the curve* (AUC) sebesar 0,816.

Kemudian dari hasil *t-test* dapat dilihat bahwa tidak ada perbedaan signifikan antara *Logistic Regression* dengan *Naïve Bayes*, karena nilai $\alpha = 0,821 > 0,05$. Hal tersebut menunjukkan bahwa metode *Logistic Regression* memiliki *performance* atau kinerja yang sama dibandingkan dengan metode *Naïve Bayes*.

Untuk pengembangan penelitian ini lebih lanjut, dapat dilakukan dengan menambahkan model optimasi pada metode model *Logistic Regression* dan model *Naïve Bayes*. Serta dapat menambahkan informasi lebih jelas dan detail tentang penyakit hepatitis.

REFERENSI

Amrin, & Pahlevi, O. (2020). Data Mining Model For Designing Diagnostic Applications Inflammatory Liver Disease. *Sinkron*, 5(1), 51.

- <https://doi.org/10.33395/sinkron.v5i1.10589>
- Amrin, Pahlevi, O., & Satriadi, I. (2021). Optimasi Algoritma C4 . 5 dan Naïve Bayes Berbasis Particle Swarm Optimization Untuk Diagnosa Penyakit Peradangan Hati. *Insantek*, 2(1), 10–14.
<http://jurnal.bsi.ac.id/index.php/insantek/article/view/399>
- Bayrak, E. A., Kirci, P., & Ensari, T. (2019). Performance Analysis of Machine Learning Algorithms and Feature Selection Methods on Hepatitis Disease. *International Journal of Multidisciplinary Studies and Innovative Technologies*, 3(2), 135–138.
- Bhargav, K. S., Kumari, T. D., Thota, D. S. S. B., & B, V. (2018). Application of Machine Learning Classification Algorithms on Hepatitis Dataset. *International Journal of Applied Engineering Research*, 13(16), 12732–12737.
- Chen, Y., Luo, Y., Huang, W., Hu, D., Zheng, R. qin, Cong, S. zhen, Meng, F. kun, Yang, H., Lin, H. jun, Sun, Y., Wang, X. yan, Wu, T., Ren, J., Pei, S. F., Zheng, Y., He, Y., Hu, Y., Yang, N., & Yan, H. (2017). Machine-learning-based classification of real-time tissue elastography for hepatic fibrosis in patients with chronic hepatitis B. *Computers in Biology and Medicine*, 89(January), 18–23.
<https://doi.org/10.1016/j.compbio.2017.07.012>
- Hendri, A., Jones, S., & Makmun, M. S. (2021). Implementasi Metode CART untuk Klasifikasi Diagnosis Penyakit Hepatitis Pada. *Journal of Informatics, Information System, Software Engineering and Applications*, 3(2), 61–70.
- Ibrahim, I., & Abdulazeez, A. (2021). The Role of Machine Learning Algorithms for Diagnosing Diseases. *Journal of Applied Science and Technology Trends*, 2(01), 10–19.
<https://doi.org/10.38094/jast20179>
- Kemenkes RI. (2017). *Situasi Penyakit Hepatitis B Di Indonesia*. InfoDatin Kemenkes RI.
- Nivaan, G. V., & Emanuel, A. W. R. (2020). Analytic Predictive of Hepatitis using the Regression Logic Algorithm. *2020 3rd International Seminar on Research of Information Technology and Intelligent Systems, ISRITI 2020*, 106–110.
<https://doi.org/10.1109/ISRITI51436.2020.9315365>
- Papuangan, M. (2018). Penerapan Case Based Reasoning Untuk Sistem Diagnosis Penyakit Hepatitis. *JIKO (Jurnal Informatika Dan Komputer)*, 1(1), 7–12.
<https://doi.org/10.33387/jiko.v1i1.1165>
- Prayoga, N. D. (2018). Sistem Diagnosis Penyakit

- Hati Menggunakan Metode Naïve Bayes. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2(8), 2666–2671.
- Pusporani, E., Qomariyah, S., & Irhamah. (2019). Klasifikasi Pasien Penderita Penyakit Liver dengan Pendekatan Machine Learning. *Inferensi*, 2(1), 25–32. <https://doi.org/10.12962/j27213862.v2i1.6810>
- Ramasamy, M., Selvaraj, S., & Mayilvaganan, M. (2015). An empirical analysis of decision tree algorithms: Modeling hepatitis data. *ICETECH 2015 - 2015 IEEE International Conference on Engineering and Technology, March*, 18–21. <https://doi.org/10.1109/ICETECH.2015.7275013>
- Sulastri, S., Hadiono, K., & Anwar, M. T. (2020). Analisis Perbandingan Klasifikasi Prediksi Penyakit Hepatitis Dengan Menggunakan Algoritma K-Nearest Neighbor, Naïve Bayes Dan Neural Network. *Dinamik*, 24(2), 82–91. <https://doi.org/10.35315/dinamik.v24i2.7867>
- Syafa'ah, L., Zulfatman, Z., Pakaya, I., & Lestandy, M. (2021). Comparison of Machine Learning Classification Methods in Hepatitis C Virus. *Jurnal Online Informatika*, 6(1), 73. <https://doi.org/10.15575/join.v6i1.719>
- Trishna, T. I., Emon, S. U., Ema, R. R., Sajal, G. I. H., Kundu, S., & Islam, T. (2019). Detection of Hepatitis (A, B, C and E) Viruses Based on Random Forest, K-nearest and Naïve Bayes Classifier. *2019 10th International Conference on Computing, Communication and Networking Technologies, ICCCNT 2019*, 1–7. <https://doi.org/10.1109/ICCCNT45670.2019.8944455>
- Wibowo, R., & Indriyawati, H. (2020). Top-k Feature Selection Untuk Deteksi Penyakit Hepatitis Menggunakan Algoritme Naïve Bayes. *Jurnal Buana Informatika*, 11(1), 1. <https://doi.org/10.24002/jbi.v11i1.2456>
- Zuama, R. A. (2021). Pembelajaran Mesin untuk diagnosis tingkat kerusakan hati akibat hepatitis C. *Journal Speed – Sentra Penelitian Engineering Dan Edukasi*, 13(3), 24–28.