Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 3, No. 4, August 2022, Page 405−409
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v3i4.2064

Analisis Penerapan Metode Ensembled Learning Decision Tree Pada Klasifikasi Virus Hepatitis C

Rifqi Alfinnur Charisma, Sofiyudin Pamungkas, Rifqi Akmal Saputra, Nur Ghaniaviyanto Ramadhan*, Faisal Dharma Adhinata

Fakultas Informatika, Program Studi Rekayasa Perangkat Lunak, Institut Teknologi Telkom Purwokerto, Indonesia
Email: [1]19104031@ittelkom-pwt.ac.id, [2]19104001@ittelkom-pwt.ac.id, [3]19104022@ittelkom-pwt.ac.id, [4*]ghani@ittelkom-pwt.ac.id, [5]faisal@ittelkom-pwt.ac.id
Submitted: **08/08/2022**; Accepted: **23/08/2022**; Published: **30/08/2022**

Abstrak−Virus hepatitis C merupakan salah satu virus mematikan yang mana menyerang organ dalam hati. Virus ini dapat menyebabkan infeksi yang kronis, bahkan 80% penderitanya telah mengalami infeksi. Untuk meminimalkan resiko terpaparnya infeksi yang diakibatkan virus hepatitis C dapat dilakukan konsultasi ke dokter atau dengan menggunakan sistem deteksi cerdas. Tentunya jika menggunakan sistem cerdas maka membutuhkan data yang telah berisikan parameter-parameter terkait hepatitis C. Pada penelitian ini menggunakan dataset publik yang dapat diakses umum. Sehingga tujuan dari penelitian ini yaitu untuk klasifikasi penderita virus hepatitis C dengan menggunakan algoritma berbasis tree. Hasil yang didapatkan dengan menerapkan algoritma usulan sebesar 93% untuk akurasi, presisi 92%, dan recall 91%. Pada penelitian ini juga melakukan perbandingan dengan metode lainnya yaitu *naïve bayes*. Hasil menunjukkan metode berbasis *tree* lebih unggul.

Kata Kunci: Decision Tree; Hepatitis C; Klasifikasi; Data Mining

Abstract−Hepatitis C virus is a deadly virus that attacks the liver. This virus can cause chronic infections, even 80% of sufferers have experienced an illness. To minimize the risk of exposure to disease caused by the hepatitis C virus, consultation with a doctor or using an intelligent detection system can be conducted. Of course, if used a smart strategy, our need data that already contains parameters related to hepatitis C. This study uses a public dataset that the public can access. So, the purpose of this study is to classify patients with hepatitis C virus using a tree-based algorithm. The results obtained by applying the proposed algorithm are 93% accuracy, 92% precision, and 91% recall. This study also performs comparisons with other methods, namely naive bayes. The results show that the tree-based way is superior.

Keywords: Decision Tree; Hepatitis C; Classification; Data Mining

1. PENDAHULUAN

Infeksi virus hepatitis adalah suatu infeksi dimana organ utama yang diserang adalah hati. Penamaan hepatitis A, B maupun C baru ditemukan pada tahun 1974 akan tetapi penyebab dari infeksi tersebut baru ditemukan pada tahun 1989. Hepatitis C atau biasa disingkat HCV merupakan persoalan yang serius dimana resiko tinggi infeksi disebabkan oleh transfusi darah yang berulang. Penyakit hepatitis C disebabkan oleh virus hepatitis C (VHC) termasuk dalam famili Flaviviridae genus Hepacivirus [1]. Virus hepatitis C dibedakan menjadi enam genotipe utama, yaitu genotipe 1 sampai 6, dan terdapat variasi dalam urutan nukleotida. Genotipe diklasifikasikan lebih lanjut ke dalam subtipe a, b, c, dan lain-lain. Genotipe yang paling banyak didistribusikan adalah 1 dan 2, dengan genotipe 1 yang paling umum [2].

Virus hepatitis C menyebabkan infeksi akut dan kronis. Infeksi HCV akut biasanya tanpa gejala dan sebagian besar tidak menyebabkan penyakit yang mengancam jiwa. Sekitar 30% orang yang terinfeksi secara spontan dapat sembuh dengan sendirinya tanpa pengobatan selama 6 bulan sedangkan 70% lainnya akan berubah menjadi kronis. Sekitar 80% pasien yang terpapar Hepatitis C mengalami infeksi kronis. Selain itu pasien yang telah mengidap Hepatitis selama 30 tahun akan mengalami sirosis. Sirosis umumnya terjadi pada pasien yang terpapar Hepatitis B atau HIV dan pecandu alkohol. Pasien yang mengidap sirosis memiliki resiko kemungkinan terkena penyakti kanker hati. Sekitar 27% kasus sirosis dan 25% kasus kanker hati disebabkan oleh Hepatitis C. Hepatitis C merupakan penyakit yang memiliki dampak global yang signifikan. Menurut Organisasi Kesehatan Dunia atau WHO terdapat 130 hingga 150.000.000 orang mengalami infeksi kronis virus hepatitis C (VHC) yang mena berjumlah 2-2,5% dari total populasi dunia.

Seiring dengan berkembangnya teknologi kehadiran kecerdasan buatan atau artificial intelligence telah menarik banyak perhatian dalam bidang kesehatan khususnya untuk memprediksi penyakit tertentu seperti Hepatitis C, kanker payudara [3], malaria [4], dan diabetes [5]. Cabang ilmu tersebut dibuat khusus guna membantu aktivitas penelitian ketika menentukan pilihan dengan menjelaskan input hingga output data dalam jangka waktu tertentu secara independen. Beberapa penelitian yang telah melakukan riset terkait deteksi virus hepatitis ini seperti penelitian membahas tentang penerapan teknik data mining dalam klasifikasi pasien yang terpapar virus hepatitis C dengan menerapkan metode KNN [6]. Penulis lain membahas tentang penerapan model machine learning SVM, XGBoost dan logistic untuk melakukan klasifikasi infeksi hepatitis C, fibrosis, dan cirrhosis [7]. Study terkait juga membahas tentang klasifikasi penyakit hepatitis C dengan menerapkan beberapa model machine learning seperti SVM, KNN, random forest, naïve bayes, dan logistic regression [8]. Penerapan model lain dengan menggunakan pendekatan deep learning yaitu artificial neural network (ANN) untuk mendeteksi hepatitis C mampu menghasilkan akurasi 97,78% [9]. Study lain membahas tentang evaluasi

Copyright © 2022 **Rifqi Alfinnur Charisma**, Page 405
This Journal is licensed under a Creative Commons Attribution 4.0 International License

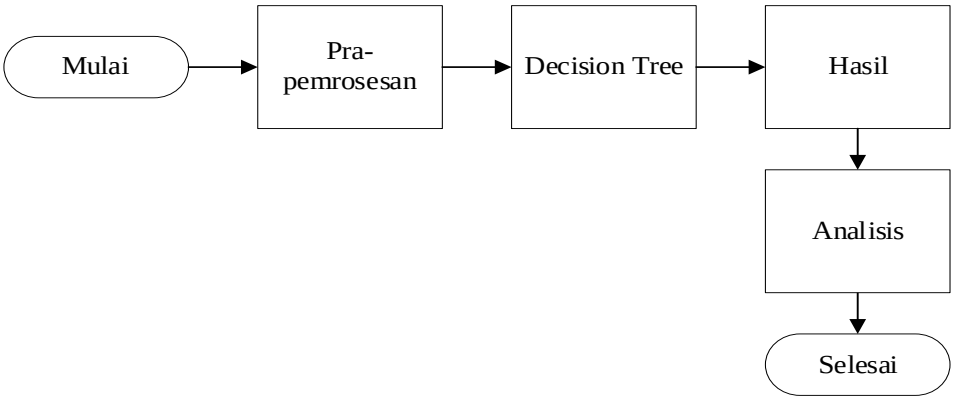
Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 3, No. 4, August 2022, Page 405−409
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v3i4.2064

penerapan berbagai algoritma klasifikasi machine learning untuk deteksi virus hepatitis C, algoritma tersebut adalah naïve bayes, random forest, dan KNN [10].

Berdasarkan pemaparan permasalahan tersebut maka penelitian ini akan melakukan deteksi klasifikasi virus hepatitis C dengan menggunakan metode decision tree. Penelitian ini diharapkan menghasilkan akurasi yang tinggi guna mengurangi resiko keasalahan dalam melakukan deteksi virus hepatitis C.

2. METODOLOGI PENELITIAN

Pada penelitian ini menggunakan sistem diagram seperti pada gambar 1.



Gambar 1. Sistem Diagram Penelitian

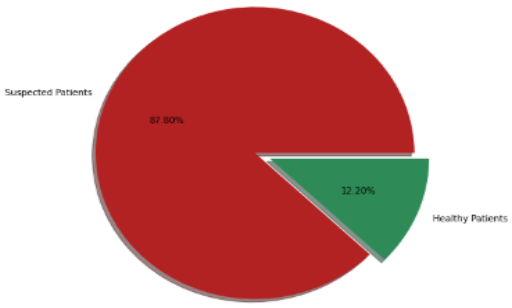
2.1 Dataset

Data yang kami gunakan dalam tugas ini yakni dataset yang berasal dari website Kaggle. Data berupa file csv yang berisi hasil tes fungsi hati pasien yang terinfeksi hepatitis dan pasien yang tidak terinfeksi. Pada gambar 2 merupakan contoh dataset yang digunakan.

Category	Age	Sex	ALB	ALP	ALT	AST	BIL	CHE	CHOL	CREA	GGT	PROT
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
0=Blood Donor	32	m	38.5	70.3	18.0	24.7	3.9	11.17	4.80	74.0	15.6	76.5
0=Blood Donor	32	m	46.9	74.7	36.2	52.6	6.1	8.84	5.20	86.0	33.2	79.3
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
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

Gambar 2. Dataset Hepatitis C

Pada dataset ini terdapat 4 kelas yakni blood donor, suspect blood donor, hepatitis, fibrosis, dan cirrhosis. Kami membagi 4 kategori tersebut menjadi 2 kategori saja yakni pasien yang sehat (blood donor dan suspect blood donor) dan pasien yang terinfeksi (hepatitis, fibrosis, dan cirrhosis). Setelah digabungkan kami menemukan total pasien sebanyak 615 dengan pembagian 540 pasien terinfeksi dan 75 pasien yang tidak terinfeksi. Data yang banyak akan membuat model klasifikasi semakin bagus.



Gambar 3. Persebaran Dataset

Copyright © 2022 **Rifqi Alfinnur Charisma**, Page 406
This Journal is licensed under a Creative Commons Attribution 4.0 International License

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 3, No. 4, August 2022, Page 405−409
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v3i4.2064

Data yang berjumlah 615 pasien tadi akan dibagi menjadi dua bagian yakni data training dan data testing dengan menggunakan perbandingan 80:20. Untuk data training digunakan sebanyak 492 data dan untuk data testing sebanyak 123 data. Sebelum proses pemisahan data, data sebaiknya diacak terlebih dahulu agar persebaran data merata. Kami menggunakan fungsi random state dan mengatur parameter menjadi 42

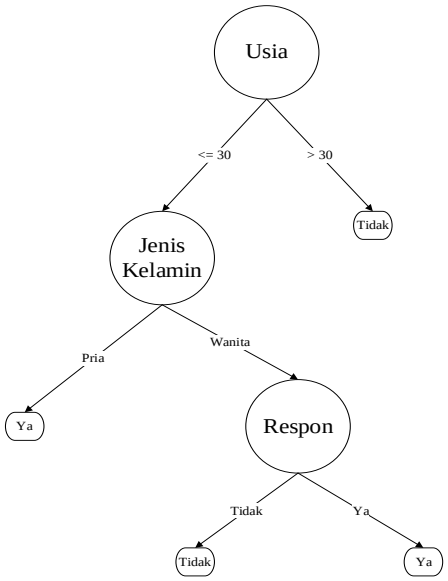
2.2. Pra-pemrosesan

Pada tahap ini dilakukan pemrosesan data untuk menghilangkan nilai kosong dan symbol-simbol special seperti (.,?!"'). Sehingga pada tahap ini didapatkan dataset yang bersih untuk dilakukan klasifikasi menggunakan decision tree.

2.3. *Decision Tree*

Decision tree atau pohon keputusan adalah pengklasifikasi yang dinyatakan sebagai partisi rekursif dari ruang. Pohon keputusan terdiri dari simpul-simpul yang membentuk pohon berakar, artinya itu adalah pohon terarah dengan simpul yang disebut "*root*" yang tidak memiliki saluran masuk tepi. Semua node lain memiliki tepat satu edge yang masuk. Sebuah node dengan keluar tepi disebut internal atau tes node. Semua simpul lainnya disebut daun (juga dikenal sebagai terminal atau simpul keputusan). Dalam pohon keputusan, setiap simpul internal membagi ruang instance menjadi dua atau lebih sub-ruang sesuai dengan tertentu fungsi diskrit dari nilai atribut input dalam cara yang paling sederhana dan paling bebas.

Setiap daun ditugaskan ke satu kelas yang mewakili target yang paling tepat nilai. Sebagai alternatif, daun mungkin memiliki vektor probabilitas yang menunjukkan kemampuan atribut target yang memiliki nilai tertentu. Instance diklasifikasikan berdasarkan menavigasi mereka dari akar pohon ke daun, menurut hasil tes di sepanjang jalan. Node internal direpresentasikan sebagai lingkaran, sedangkan daun dilambangkan sebagai tri-sudut. Perhatikan bahwa pohon keputusan ini menggabungkan atribut nominal dan numerik upeti. Dengan pengklasifikasi ini, analis dapat memprediksi respons dari suatu potensi pelanggan (dengan menyortirnya ke bawah pohon), dan memahami karakter perilaku karakteristik dari seluruh populasi pelanggan potensial mengenai pengiriman langsung. Setiap node diberi label dengan atribut yang diujinya, dan cabangnya diberi label dengan nilai-nilainya yang sesuai. Gambar 4 merupakan contoh cara kerja dari model *decision tree*.



Gambar 4. Decision Tree

Formula pada *decision tree* yang digunakan ada pada formula (1).

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

Dimana:
C = jumlah kelas yang ada pada dataset (hepatitis/tidak hepatitis).
i = indeks data ke-i dalam kelas.
Pi = data yang akan dihitung pada indeks ke i.

3. HASIL DAN PEMBAHASAN

Pada tahap ini dilakukan pengukuran menggunakan nilai *confusion matrix* terhadap implementasi yang telah dilakukan menggunakan algoritma *decision tree*. Formula dapat dilihat pada (2), (3) dan (4) [11].

Copyright © 2022 **Rifqi Alfinnur Charisma**, Page 407
This Journal is licensed under a Creative Commons Attribution 4.0 International License

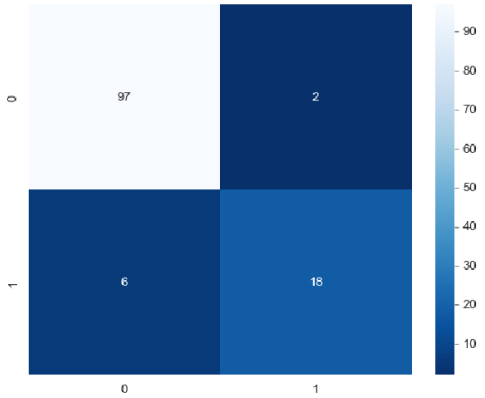
Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 3, No. 4, August 2022, Page 405−409
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v3i4.2064

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

$$Presisi = \frac{TP}{TP+FP} \quad (3)$$

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

Pada gambar 5 merupakan hasil implementasi confusion matrix yang dilakukan, selanjutnya akan dilakukan perhitungan menggunakan formula (2), (3), dan (4).



Gambar 5. Confussion Matrix

Pada tabel 1 merupakan hasil akurasi, presisi, dan recall yang dihasilkan.

Tabel 1. Hasil Penelitian

Metode	Akurasi	Presisi	Recall
Decision Tree	93%	92%	86%
Naive Bayes	90%	91%	84%

Dapat dilihat pada gambar 5 yang mana nilai 0 memiliki arti tidak hepatitis, sedangkan nilai 1 memiliki arti hepatitis. Nilai-nilai yang ada pada gambar 5 merupakan nilai dari *true positif* (TP), *false positif* (FP), *false negative* (FN), dan *true negative* (TN) yang dimana untuk nilai TP sebesar 97, FP sebesar 2, FN sebesar 6, dan TN sebesar 18. Besaran angka pada confusion matrix dapat mempengaruhi hasil pada tabel 1. Pada tabel 1 terlihat bahwa usulan metode penelitian ini berhasil menghasilkan akurasi yang lebih tinggi dibandingan metode lainnya. Selisih akurasi yang dihasilkan bisa mencapai 3%, untuk selisih presisi 1%, dan recall 2%. Hasil selisih tersebut tentunya jika tidak dilakukan dengan seksama maka akan membahayakan pengguna dalam menentukan apakah pasien terkena virus hepatitis C atau tidak. Metode *decision tree* pada penelitian ini unggul dibandingkan dengan metode *naïve bayes* dikarenakan pada metode *tree* menerapkan proses kesimpulan ya atau tidak yang mana sesuai dengan kelas data yang digunakan. Sehingga pada penelitian ini penerapan metode decision tree mampu menghasilkan akurasi tertinggi dibandingkan metode naïve bayes dalam klasifikasi virus hepatitis C.

4. KESIMPULAN

Pada penelitian ini berhasil menerapkan metode berbasis tree yaitu decision tree dalam permasalahan klasifikasi virus hepatitis C. Hasil pada penelitian ini menggunakan pengukuran akurasi, presisi, dan recall. Akurasi yang dihasilkan sebesar 93%, presisi 92%, dan recall 91%. Pada penelitian ini juga melakukan perbandingan dengan menerapkan metode *naïve bayes* yang dimana hasilnya lebih unggul metode *decision tree*. Untuk penelitian selanjutnya dapat dilakukan pendalaman pada pra-pemrosesan.

REFERENCES

[1] K. K. R. Indonesia, "Situasi dan Analisis HEPATITIS," *InfoDATIN Pus. Data dan Inf. Kementrian Kesehat. RI*, 2014.
[2] J. T. Dipiro, R. L. Talbert, G. C. Yee, G. R. Matzke, B. G. Wells, and L. M. Posey, "Pharmacotherapy: a pathophysiologic approach, ed," *Connect. Applet. Lange*, vol. 4, pp. 141–142, 2014.
[3] N. G. Ramadhan, F. D. Adhinata, and others, "Teknik SMOTE dan Gini Score dalam Klasifikasi Kanker Payudara," *RADIAL J. Perad. Sains, Rekayasa dan Teknol.*, vol. 9, no. 2, pp. 125–134, 2021.
[4] N. G. Ramadhan and A. Khoirunnisa, "Klasifikasi Data Malaria Menggunakan Metode Support Vector Machine," *J. MEDIA Inform. BUDIDARMA*, vol. 5, no. 4, pp. 1580–1584, 2021.
[5] N. G. Ramadhan, "Comparative Analysis of ADASYN-SVM and SMOTE-SVM Methods on the Detection of Type 2 Diabetes Mellitus," *Sci. J. Informatics*, vol. 8, no. 2, pp. 276–282, 2021.
[6] R. Safdari, A. Deghatipour, M. Gholamzadeh, and K. Maghooli, "Applying data mining techniques to classify patients with suspected hepatitis C virus infection," *Intell. Med.*, 2022.

Copyright © 2022 **Rifqi Alfinnur Charisma**, Page 408
This Journal is licensed under a Creative Commons Attribution 4.0 International License

ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 3, No. 4, August 2022, Page 405−409
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v3i4.2064

[7] L. Akter and others, "Detection of Hepatitis C Virus Progressed Patient's Liver Condition Using Machine Learning," in *International Conference on Innovative Computing and Communications*, 2022, pp. 71–80.
[8] F. J. Kaunang, "A Comparative Study on Hepatitis C Predictions Using Machine Learning Algorithms," *8ISC Proc. Technol.*, pp. 33–42, 2022.
[9] A. Drobo *et al.*, "Application of artificial neural networks in diagnosis of Hepatitis C," in *2022 XXVIII International Conference on Information, Communication and Automation Technologies (ICAT)*, 2022, pp. 1–5.
[10] L. Syafa'ah, Z. Zulfatman, I. Pakaya, and M. Lestandy, "Comparison of machine learning classification methods in hepatitis C virus," *J. Online Inform.*, vol. 6, no. 1, pp. 73–78, 2021.
[11] N. G. Ramadhan, A. G. Putrada, and M. Abdurohman, "Improving smart lighting with activity recognition using hierarchical hidden Markov model," *Indones. J. Comput.*, vol. 4, no. 2, pp. 43–54, 2019.

Copyright © 2022 **Rifqi Alfinnur Charisma**, Page 409
This Journal is licensed under a Creative Commons Attribution 4.0 International License