

ANALISA ALGORITMA *DECISION TREE* DAN *NAÏVE BAYES* PADA PASIEN PENYAKIT LIVER

Nadia Tiara Rahman¹⁾

¹Fakultas Ilmu Komputer, Universitas Darwan Ali Sampit
email: nadiatiara0505@gmail.com

Abstract

Liver disease is a disease caused by consuming alcohol, liver fat, and genetics from the elderly, diabetes, obesity in the body and chemical compounds derived from drugs. Liver or called liver is the largest and most important human organ. Liver disease can carry out physical examinations of parts of the body including the lungs, heart, skin, brain, nervous system, and stomach which can provide clues to the causes of liver disease. The liver functions as a detoxification in poisons in the body. The data studied were the results of examination of the data obtained as many as 583 data with 416 people detected "positive" liver disease and 167 people "negative" liver disease. Therefore we needed a data mining analysis system with the Decision Tree Algorithmic method which has an optimal accuracy value of 70.29% with AUC 0.757 while using the Naive Bayes algorithm has an optimal accuracy value of 67.05% with AUC 0.714. The results of the discussion can be concluded that the method that provides fractions for problems in identifying liver disease is the Decision Tree (C4.5).

Keywords: Data Mining, Decision Tree (C4.5), Naïve Bayes, Liver disease, RapidMiner

Abstrak

Penyakit liver merupakan penyakit yang disebabkan mengkonsumsi alkohol, pelemakan hati, genitika dari orang tua, diabetes, obesitas pada tubuh dan senyawa kimia yang berasal dari obat-obatan. Liver atau disebut hati merupakan organ manusia yang terbesar dan penting. Penyakit liver dapat melakukan pemeriksaan fisik bagian pada tubuh termasuk paru-paru, jantung, kulit, otak, system saraf, dan perut yang dapat memberikan petunjuk untuk penyebab penyakit hati. Hati berfungsi sebagai detoksifikasi pada racun dalam tubuh. Data yang diteliti merupakan hasil pemeriksaan terhadap data di peroleh sebanyak 583 data dengan 416 orang terdeteksi "positif" penyakit liver dan 167 orang "negative" penyakit liver. Oleh karena itu diperlukan sistem analisa data mining dengan metode Algoritma Decision Tree yang mempunyai nilai optimal akurasi sebesar 70.29% dengan AUC 0.757 sedangkan menggunakan Algoritma Naïve Bayes mempunyai nilai optimal akurasi sebesar 67.05% dengan AUC 0.714. Hasil pembahasan dapat disimpulkan bahwa metode yang memberikan pecahan untuk permasalahan dalam mengidentifikasi penyakit liver ialah Decision Tree (C4.5).

Keywords: Data Mining, Decision Tree (C4.5), Naïve Bayes, Penyakit Liver, RapidMiner

PENDAHULUAN

Penyakit liver merupakan penyakit peradangan pada hati manusia, yang mana kesehatan organ hati sangat penting bagi tubuh manusia. Hati merubah zat beracun menjadi nutrisi dan kemudian tubuh menggunakan untuk mengendalikan hormon dalam tubuh (Abdar, M., Zomoradi-Mogham, M., Das, R., & Ting, I.-H. 2016). Selain itu, menghasilkan

protein untuk membantu darah membeku, dan menguraikan dalam sel darah merah dan menguraikan zat-zat kimia. Faktor-faktor munculnya penyakit liver merupakan mengkonsumsi alkohol, pelemakan hati, genitika dari orangtua, diabetes dan obesitas pada tubuh. Dampak cedera pada liver ini mencakup radang, pembekuan darah, dan gagal hati. Kinerja hati adalah penetral racun dalam tubuh menjadikan racun-racun yang selama ini

masuk melalui makanan yang mampu menetralkan melalui hati (Global, Citra Yustitya.2018).

Penyakit liver, salah satu penyakit penyumbang kematian karena dianggap sebagai pembunuh diam-diam tanpa gejala. Menurut WHO (World Health Organization) tahun 2013, angka penderita penyakit liver di Indonesia mencapai 28 juta orang. Penyakit liver di indoneisa adalah salah satu dari 10 penyakit terbesar tingkat kemaitiannya (Pura May Candra Abrianto. 2018).

Diagnosis merupakan proses untuk menentukan sifat suatu penyakit atau kelainan yang akan membedakannya dari keadaan yang mungkin terjadi, mengidentifikasi penyakit liver dapat melakukan pemeriksaan fisik bagian pada tubuh termasuk paru-paru, jantung, kulit, otak, system saraf, dan perut yang dapat memberikan petunjuk untuk penyebab penyakit hati (Mandal, Anya. 2019). Dalam bidang kesehatan, kesalahan dalam mendiagnosa penyakit yang dialami pasien adalah tanggung jawab yang paling berat unruk diemban oleh ahli kesehatan. Kesalahan dalam mendiagnosa penyakit dapat menyebabkan hal yang membahayakan bagi kesehatan pasien bahkan dapat menyebabkan kematian (Neshat, Mehdi Mehdi Sargozei, Adel Nadjran Toosi, and Azra Masoimo. 2012).

Data mining adalah proses menemukan pengetahuan menarik dari data dalam jumlah besar (Han, J. and Kamber, M. 2000). Adalah suatu proses ekstraksi atau penggalian data yang belum diketahui sebelumnya namun dapat dipahami dan berguna dari database yang besar serta digunakan untuk membuat suatu keputusan bisnis yang sangat penting (Connolly, Thomas and Begg, Carolyn. 2010). Data mining adalah ekstraksi informasi atau pola yang penting atau menarik dari data yang ada di database yang besar. Dan dalam jurnal data mining juga dikenal dengan *Knowledge Discovery in Database (KDD)* (Amril

Mutoi Siregar, S.Kom., M.Kom. dan Adam Puspabhuana, S.Kom., M.Kom)

Pada penelitian digunakan data mining dengan merumuskan masalah dalam mengidentifikasi penyakit liver menggunakan metode *Decision Tree* dan *Naïve Bayes* untuk melakukan komparasi untuk menentukan metode mana yang lebih akurat. Data dalam penelitian ini bersumber dari <https://archive.ics.uci.edu/ml> data yang diteliti merupakan hasil pemeriksaan terhadap 583 orang dari wilayah Andhra Pradesh, India.

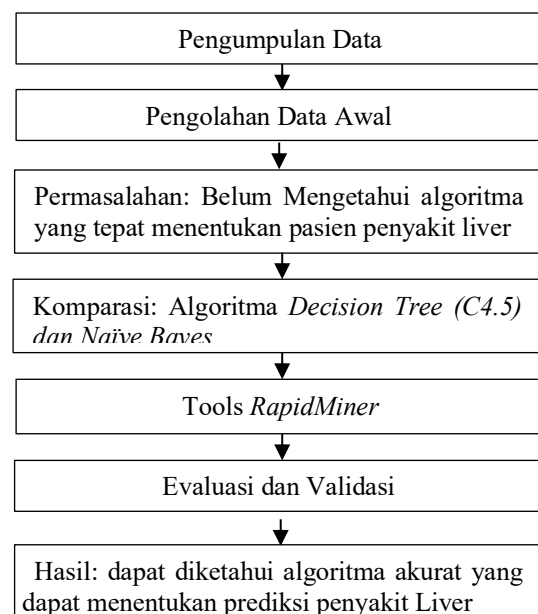
METODE PENELITIAN

Metode penelitian ini dilakukan sebagai pedoman peneliti untuk mnecapai hasil dan tujuan diharapkan.

1. Desain Penelitian

Jenis penelitian ini disebut dengan jenis eksperimen, bertujuan melakukan perbandingan pada algoritma Dessasion Tree dan Naïve Bayes. Penelitian eksperimen komparatif ini dilandasi kerangka pemikiran pemecahan masalah. Seperti pada gambar 1 dibawah ini:

Gambar 1. Langkah-langkah penelitian



2. Pengumpulan Data

Dalam pengumpulan data terdapat sumber data, sumber data yang terhimpun langsung oleh peneliti disebut dengan sumber primer, sedangkan apabila melalui tangan kedua disebut sumber sekunder (Amril Mutoi Siregar, S.Kom. M.Kom. dan Adam Puspabhuana, S.Kom. M.Kom). Data yang diperoleh adalah data sekunder karena diperoleh dari Data yang digunakan dalam penelitian ini bersumber dari alamat web: <http://archive.ics.uci.edu/ml/>. Data ini merupakan hasil pemeriksaan terhadap 583 orang dari wilayah India, Andhra Pradesh. Adapun atributnya *age*, *gender*, *Total Bilirubin*, *Direct Bilirubin*, *Alkaline Phosphatase*, *Alamine Aminotransferase*, *Aspartate Aminotransferase*, *Total Protiens*, *Albumin*, *Ratio Albumin and Globulin Ratio*, dan *class* atau atribut hasil prediksi.

3. Pengumpulan Data

Pengolahan data ini menngolongkan berdasarkan Prediksi Pasien hasil atribut *dataset* dari atribut *class*. Dan data di peroleh sebanyak 583 data dengan 416 orang terdeteksi “positif” penyakit liver dan 167 orang “negative” penyakit liver.

4. Metode Yang Digunakan

Dalam penelitian ini metode yang digunakan adalah kalsifikasi algoritma *Decision Tree (C4.5)* dan *Naïve Baye*. Perhitungan terdapat yang manual dengan *Excel* dan menggunakan pengujian *Rapid Miner* yang nantinya akan dilakukan pengakurasian, apakah hasilnya akan sama atau beda dalam memprediksi penyakit liver.

5. Perhitungan Algoritma C4.5

Algoritma *Decision Tree* dapat digunakan untuk memprediksi atau mengkalasifikasi suatu kejadian dengan pembentukan pohon keputusan antara lain *Decision Tree* atau C4.5 yang merupakan suatu algoritma yang dikembangkan oleh J.Ross Quilan (Novita, Rina. 2003-2016). Pohon keputusan adalah hasil dari proses perhitungan entropy dan information gain,

setelah perhitungan berulang-ulang sampai atribut memiliki kelas dan tidak biasa lagi dilakukan proses perhitungan^[12]. Terdapat tahapan membuat pohon keputusan dengan algoritma *Decision Tree (C4.5)* yaitu:

1. Menyiapkan data training yang sudah dikelompokan dengan rumus entropy sebagai berikut:

$$Entropy(S) = - \sum_{i=1}^m p_i \log_2(p_i)$$

Sedngkan rumus untuk menghitung masing masing variabel sebagai berikut:

$$Entropy_A(S) = \sum_v \frac{|S_v|}{|S|} Entropy(S_v)$$

2. Selanjutnya menghitung gain
 $Gain(A) = Entropy(S) - Entropy_A(S)$
3. Hitung Nilai *Splt Info* sebagai pembagi untuk menghasilkan *Gain Raito*

$$SplitInfo(D) = - \sum_{j=1}^v \frac{|D_j|}{|D|} \log_2 \left(\frac{|D_j|}{|D|} \right)$$

4. Menghitung *Gain Raito* dan memilih atribut data training yang tertinggi.

$$GainRatio(A) = \frac{Gain(A)}{SplitInfo_A(D)}$$

6. Perhitungan Naïve Bayes

Naïve bayes merupakan klasifikasi teori *Bayesian* yang mana digunakan sebagai alat pengambil keputusan dari suatu informasi. Dengan mengkombinasikan Probabilitas *Bayes* untuk klasifikasi ketidakpastian data (Septiani, Dwi Wisati. 2017). Metode ini banyak diterapkan pada hal – hal yang berkenaan dengan diagnosa secara statistik yang berhubungan dengan probabilistik serta kemungkinan dari penyakit dan gejala gejala yang berkaitan.

Torema *Bayes* memiliki rumus sebagai berikut:

$$P(H|E) = \frac{P(E|H) \times P(H)}{P(E)}$$

Langkah-langkah Algoritma Naïve Bayes:

1. Menyiapkan Dataset

2. Lalu Hitung jumlah kelas yang terdapat di data Training,
3. Lalu Hitung jumlah kasus dengan kelas yang sama,
4. Kemudian dikalikan hasil sesuai pada uji yang di cari kelasnya, Lalu bandingkan perkelas, yang nilainya tinggi

7. Perhitungan Rapid Miner

Secara sederhana Rapidminer merupakan aplikasi yang dapat digunakan untuk pengolahan data dengan berbagai teknik metode dan salah satu perangkat lunak Rapid miner bersifat *openscource*. Rapidminer merupakan teknik menyediakan prosedur pada *data mining* dan *machine learning* yang termasuk preprosesing, modeling dan visualisasi (Novianti, Deny. 2019).

7. Analisa Data

Dalam keberhasilan transformasi untuk memudahkan pengelompokan nilai berdasarkan referensi hasil labotarium, berikut parameter atribut.

Gambar 2. Parameter atribut Dataset

Atribut	Nilai Rujukan
Age	< 20 Anak, <= 20 - <50 Muda, >= 50 Dewasa
Gender	Female (Perempuan) dan Male (Laki-Laki)
Total Bilirubin	<=1 Normal dan >1 Abnormal
Direct Bilirubin	<=0.2 Normal dan >0.2 Abnormal
Alkaline Phosphatase	<=30 <=120 Normal dan >120 Tinggi
Alamine Aminotransferase	<47 Normal dan >=47 Abnormal
Aspartate Aminotransferase	<37 Normal dan >=37 Abnormal
Total Proteins	<6 Rendah, <=6 - <=8 Normal, >8 Tinggi
Albumin	<3.4 Rendah, <=3.4 - <=4.8 Normal, >4.8 Tinggi
Ratio Albumin and Globulin Ratio	>1 Normal dan <=1 Abnormal
Dataset	1(Positif) dan 2(Negatif)

HASIL DAN PEMBAHASAN

1. Pengujian Metode *Decision Tree*

Data yang digunakan dalam penelitian ini adalah data pasien liver di indian sebanyak 583 dari 416 pasien positif penyakit liver dan 167 negatif penyakit

liver dengan menggunakan metode *Naïve Bayes* dan *Decision Tree* (C4.5).

Menghitung jumlah Atribut “Positif” Dan “Negatif. Dari jumlah kasus yang Positif sebanyak 416 data dan jumlah kasus Negatif sebanyak 167 data entropy secara keseluruhan.

$$\begin{aligned}
 Entropy(S) &= - \sum_{i=1}^m p_i \log_2(p_i) \\
 &= \left(- \frac{416}{583} \times \log_2 \left(\frac{416}{583} \right) \right) \\
 &\quad + \left(- \frac{167}{583} \times \log_2 \left(\frac{167}{583} \right) \right) \\
 &= 0.86409
 \end{aligned}$$

Kemudian hitung *entropy* dan *information gain*.

Gambar 3. Entropy Dan Information Gain

	JML KASUS	POSITIF	NEGATIF	ENTROPY	INFORMATION GAIN
TOTAL	583	416	167	0.8640904	
USIA					0.007734522
ANAK	38	22	16	0.9819408	
MUDA	315	219	96	0.8870342	
DEWASA	230	175	55	0.7935912	
GENDER					0.004770573
PEREMPUAN	142	92	50	0.9359407	
LAKI-LAKI	441	324	117	0.8346483	
TB					0.048523905
NORMAL	305	184	121	0.9690003	
ABNORMAL	278	232	46	0.6472309	
DB					0.041500535
NORMAL	257	152	105	0.9757383	
ABNORMAL	326	264	62	0.7018563	
ALKPHOE					2.28169E-05
NORMAL	20	14	6	0.8812909	
TINGGI	563	402	161	0.8634558	

Kemudian Menghitung Split info sebagai pembagi dari Gain untuk setiap atribut. Selanjutnya pilih *Gain Ratio* yang memiliki nilai paling tinggi untuk dijadikan *root node*.

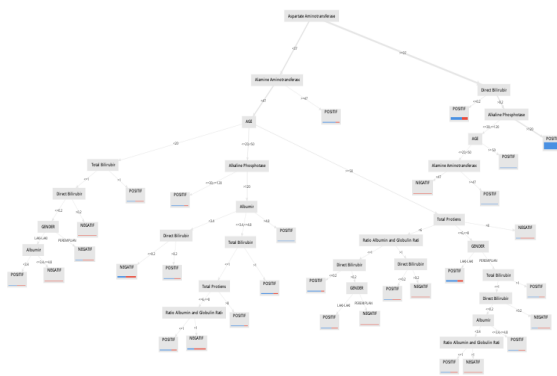
Gambar 4. Split info dan Gain

	JML KASUS	POSITIF	NEGATIF	ENTROPY	INFORMATION GAIN	SPLIT INFO	GAIN RATIO
TOTAL	583	416	167	0.8640904			
USIA					0.007734522	1.2660236	0.0061093
ANAK	38	22	16	0.9819408			
MUDA	315	219	96	0.8870342			
DEWASA	230	175	55	0.7935912			
GENDER					0.004770573	0.8009232	0.0059563
PEREMPUAN	142	92	50	0.9359407			
LAKI-LAKI	441	324	117	0.8346483			
TB					0.048523905	0.9984523	0.0485991
NORMAL	305	184	121	0.9690003			
ABNORMAL	278	232	46	0.6472309			
DB					0.041500535	0.989872	0.0419252
NORMAL	257	152	105	0.9757383			
ABNORMAL	326	264	62	0.7018563			
ALKPHOE					2.28169E-05	0.1182766	0.0001929
NORMAL	20	14	6	0.8812909			
TINGGI	563	402	161	0.8634558			

Dapat dilihat nilai tertinggi pada gain pada atribut *Aspartate Aminotransferase* yakni 0.054367 sehingga atribut *Aspartate*

Aminotransferase dapat disebut sebagai akar atau bisa disebut root dari pohon keputusan. Perhitungan menentukan dapat atribut *Aspartate Aminotransferase* <37 dan *Aspartate Aminotransferase* ≥ 37 . Untuk menentukan simpul selanjtnya dilakukan perhitungan nilai Entropy, Gain, SplitInfo, dan Gain Ratio dengan Cara yang Sama, sehingga diperoleh pohon keputusan seperti gambar dibawah ini:

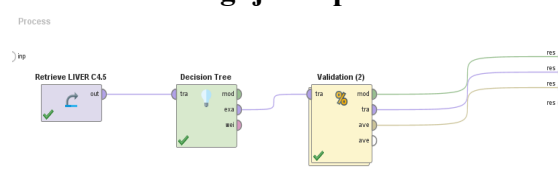
Gambar 5. Pohon Keputusan Hasil Algoritma Decision Tree



Sumber: Hasil Penelitian (2020)

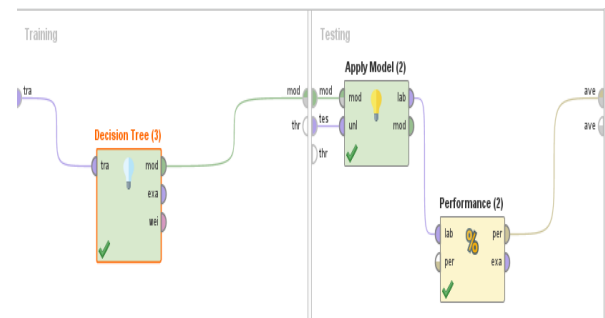
Dalam pengujian menggunakan Cross Validation untuk menggunakan Algoritma Decision Tree ini menggunakan RapidMiner sebagai berikut:

Gambar 6. Pengujian Split Validation



Sumber: Hasil Penelitian (2020)

Gambar 7. Validation Model Decision Tree



Sumber: Hasil Penelitian (2020)

2. Pengujian Metode Naïve Bayes

Naïve Bayes merupakan model kedua setelah *Decision Tree*. Langkah yang akan dilakukan adalah menghitung nilai probailitas nilai “Positif” dan “Negatif” dari seluruh data menggunakan persamaan 1 sebagai berikut:

$$P(\text{Positif}) = 416 : 583$$

$$= 0.714$$

$$P(\text{Negatif}) = 167 : 583$$

$$= 0.286$$

Hasil hitung menggunakan Naïve bayes dengan probalitas prior table dibawah ini.

Gambar 8. Perhitungan Probabalitas Prior Keseluruhan

ATRIBUT	NILAI RUJUKAN	JUMLAH KASUS	POSITIF	NEGATIF	P(H/E)	
			POSITIF	NEGATI F		
TOTAL		583	416	167	0.714	0.286
Age	< 20	38	22	16	0.053	0.096
	$\leq 20, < 50$	315	219	96	0.526	0.575
	≥ 50	230	175	55	0.421	0.329
Gender	Female	142	92	50	0.321	0.269
	Male	441	324	117	0.779	0.701
Total Bilirubin	≤ 1	305	184	121	0.442	0.725
	> 1	278	232	46	0.558	0.275
Direct Bilirubin	≤ 0.2	257	152	105	0.365	0.609
	> 0.2	326	264	62	0.635	0.371
Alkaline Phosphatase	$\leq 30, \leq 120$	20	14	6	0.034	0.036
	> 120	563	402	161	0.966	0.964
Albumin	≤ 47	369	231	138	0.555	0.626
Aspartate Aminotransferase	≥ 47	215	188	29	0.445	0.174
	< 37	250	187	109	0.333	0.653
	≥ 37	327	269	58	0.647	0.347
Total Proteins	≤ 6	171	125	46	0.300	0.275
	$\leq 5, \leq 8$	378	268	110	0.644	0.659
	> 8	34	23	11	0.055	0.066
Albumin	< 3.4	351	268	83	0.644	0.497
	$\leq 3.4, \leq 4.8$	225	143	82	0.344	0.491
	≥ 4.8	7	5	2	0.012	0.012
Ratio Albumin and Globulin	> 1	177	109	68	0.262	0.497
	≤ 1	406	307	99	0.738	0.593

Kegunaan probalitas prior adalah menentukan kelas pada kasus baru yang terdahulu dihitung. Probabilitas posterior disebut datanya bersumber dari data yang sebelumnya dan telah di sajikan. Pada Gambar 7 perhitungan probabalitas prior keseluruhan, sebagai contoh diambil kasus seperti Gambar 7. Probabilitas Prosterior

yang mana x tersebut adalah data yang diprediksi hasilnya.

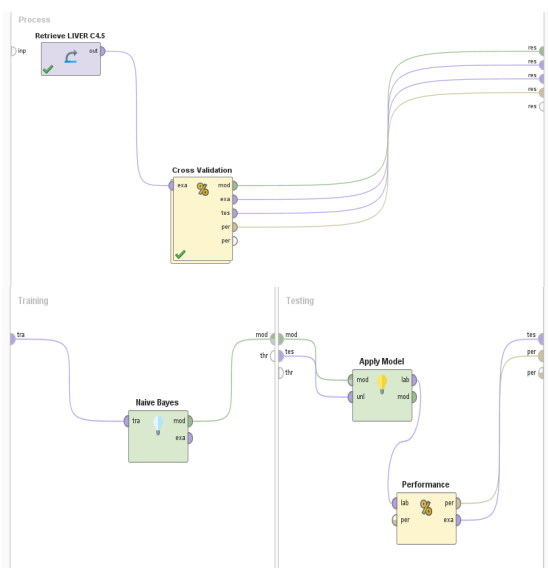
Berdasarkan dari table berikut diketahui probabilitas tiap atribut terhadap kelas. Maka selanjutnya menghitung keseluruhan probabilitas tiap rule yang diperoleh untuk atribut sebagai berikut:

1. Probabilitas diagnose Positif Liver
 $= 0.714 \times 0.421 \times 0.779 \times 0.558 \times 0.635 \times 0.034 \times 0.445 \times 0.647 \times 0.300 \times 0.644 \times 0.738 = 0.000114394$
2. Probabilitas diagnose Positif Liver
 $= 0.286 \times 0.329 \times 0.701 \times 0.275 \times 0.371 \times 0.036 \times 0.174 \times 0.347 \times 0.275 \times 0.497 \times 0.593 = 1.18858E-06$

Dikarenakan $0.000114394 > 1.18858E-06$ maka dapat disimpulkan bahwa data training dari hasil perhitungan nilai probabilitas Positif lebih besar dari probabilitas Negatif, dapat disimpulkan bahwa klasifikasi Positif Liver.

Pengujian metode cross validation ini menggunakan aplikasi RapidMiner seperti berikut:

Gambar 9. Pengujian Cross Validation naïve bayes



3. Evaluasi dan validasi hasil

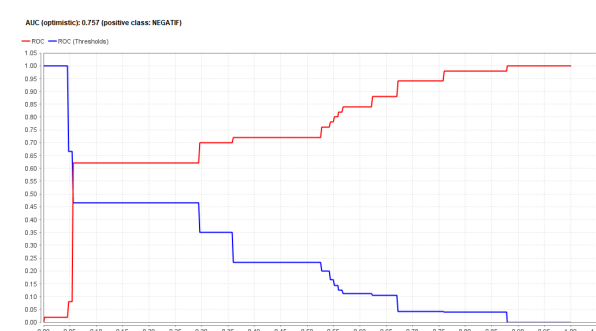
Setelah kedua data diolah maka dapat diuji tingkat akurasi untuk metode

Decision Tree. Penelitian ini adalah untuk menentukan data yang paling akurat dengan sebagai acuan diagnosis pada penyakit liver. Pengujian tingkat akurasi dilakukan menggunakan *confussion matrix* dan kurva ROC/AUC (*Area Under Cover*). Hasil perhitungan data *training* menggunakan Algoritma *Decision Tree* (C4.5). Diketahui tingkat akurasi adalah 70.29%. Dari 583 data sebanyak 119 data diprediksikan yaitu 119 data “POSITIF” dan 46 data yang diprediksikan “POSITIF” ternyata “NEGATIF”. Dan sebanyak 6 data diprediksi “NEGATIF” ternyata termasuk klasifikasi “POSITIF” dan sebanyak 4 data diprediksikan sesuai yaitu “NEGATIF”. Tabel *Confussion Matrix* disajikan pada tabel 5 dan gambar 6 merupakan grafik ROC/AUC (*Area Under Cover*) dari model *Confussion Matrix* yaitu 0.757.

Gambar 10. Akurasi Algoritma Decision Tree

	true POSITIF	true NEGATIF	class precision
pred. POSITIF	119	46	72.12%
pred. NEGATIF	6	4	40.00%
class recall	95.20%	8.00%	

Gambar 11. Kurva ROC Algoritma Decision Tree



Hasil perhitungan akurasi menggunakan naïve bayes, tingkat akurasi adalah 67.07%. Data yang sebelumnya terdapat 583 data, dari 583 data sebanyak 298 data diprediksikan yaitu 298 data “POSITIF” dan 74 data yang diprediksikan “POSITIF” ternyata “NEGATIF”. Dan sebanyak 118 data diprediksi “NEGATIF” ternyata termasuk

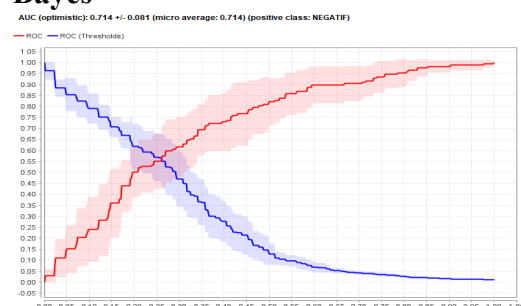
klasifikasi “POSITIF” dan sebanyak 93 data diprediksikan sesuai yaitu “NEGATIF”. Tabel *Confussion Matrix* disajikan pada tabel 6 dan gambar 7 merupakan grafik ROC/AUC (*Area Under Cover*). Berikut table data akurasi *Naïve Bayes*:

Gambar 12. Akurasi Naïve Bayes

accuracy: 67.05% +/- 4.04% (micro average: 67.07%)

	True POSITIF	True NEGATIF	class precision
pred. POSITIF	298	74	80.11%
pred. NEGATIF	118	93	44.02%
class recall	71.63%	55.69%	

Gambar 13. Kurva ROC dengan Naïve Bayes



Berdasarkan pengujian dari analisa masing-masing algoritma *Decision Tree* (C4.5) dan *Naïve Bayes* diatas dapat dirangkum sebagai berikut:

Gambar 14. Perbandingan Performance Metode

	Decision Tree (C4.5)	Naïve Bayes
ACCURAY	70.29%	67.05%
AUC	0.757	0.714

Performance untuk pengakurasian AUC dapat diklasifikasikan lima kelompok yaitu:

1. 0.90 – 1.00 = *Exellent Classification*
2. 0.80 – 0.90 = *Good Classification*
3. 0.70 – 0.80 = *Fair Classification*
4. 0.60 – 0.70 = *Poor Classification*
5. 0.70 – 0.80 = *Failure Classification*

Berdasarkan klasifikasi dapat disimpulkan bahwa algoritma *Decision Tree* (C4.5) dan *Naïve Bayes* termasuk dalam *Fair Classification* (0.70 – 0.80) untuk prediksi penyakit liver.

SIMPULAN DAN SARAN

Pada kesimpulan Dalam penelitian ini telah dilakukan pengujian pada data pasien penyakit liver menggunakan metode

Decision Tree (C4.5) dan *Naïve Bayes*. Untuk mengetahui komparasi algoritma yang paling baik dalam menentukan penyakit liver. Untuk mengukur kinerja kedua metode tersebut digunakan metode pengujian *Split Validation* dan *Cross Validation*. Dapat disimpulkan bahwa metode *Decision Tree* (C4.5) dalam klasifikasinya menghasilkan akurasi 70.29%. dan nilai AUC 0.757 yang termasuk dalam *Fair Classification*. *Naïve Bayes* menghasilkan akurasi 67.05% dan nilai AUC 0.714. Dengan demikian dapat disimpulkan metode yang memberikan pecahan untuk permasalahan dalam mengidentifikasi penyakit liver ialah *Decision Tree* (C4.5).

TERIMA KASIH

Dalam pelaksanaan dan pengerjaan laporan ini, penulis menyadari sepenuhnya bahwa selesainya jurnal yang tidak lepas dari dukungan, serta bimbingan dari berbagai pihak. Oleh karena itu, penulis ingin menyampaikan ucapan terimakasih kepada:

1. Bapak Nurrahman, S.Kom. M.M selaku Pembimbing I yang telah memberikan pengarahan, bimbingan, saran serta dorongan kepada penulis untuk menyusun laporan ini.
2. Bapak Abdul Aziz M.Pd selaku Pembimbing II yang senantiasa meluangkan waktunya telah memberikan bimbingan kepada penulis dalam menyusun laporan ini.
3. Rekan-rekan Program Studi Sistem Informasi Fakultas Ilmu Komputer Universitas Darwan Ali Sampit

DAFTAR PUSTAKA

Abdar, M., Zomoradi-Mogham, M., Das, R., & Ting, I.-H. (2016). *Perfromce Analysis Of Classopaction Algorithms On Early*

- Detection Of Liver Disease.Elsevier,1-27, 2016
- Global, Citra Yustitya. (2018).”Sistem Pakar Penyakit Liver Menggunakan K-Nearest Neighbors Algoritma Berbasis Website Volum 10.
- Pura May Candra Abrianto. (2018). Penerapan Metode K-Means Clustering Untuk Mengelompokkan Pasien Penyakit Liver Vol.2 No2.
- Mandal,Anyaa., MD. Liver Disease diagnosis, (2019) [https://www.news-medical.net/amp/health/Liver-disease-diagnosis.aspx]
- Neshat, Mehdi Mehdi Sargozei, Adel Nadjran Toosi, and Azra Masoimo. (2012). *Hepatitis Disease Diagnosis using hybrid case based reasoning and particle swarm optimization*. Internasional Scholary Research Network: ISRN Artificial Intelligence Volume.
- Han, J. and Kamber. (2020). M. Data Mining: Concepts and Techniques. Morgan Kaufmann, San Francisco, 2000
- Connolly, Thomas and Begg, Carolyn. (2010). Database Systems: A Practical Approach to Design, Implementation, and Management, Fifth Edition. Pearson Education, Boston.
- Amril Mutoi Siregar, S.Kom. M.Kom. dan Adam Puspabhuana, S.Kom. M.Kom.. DATA MINING: Pengolahan Data Menjadi Informasi dengan RapidMiner.
- Septiani, Dwi Wisati (2017). Komparasi Metode Klasifikasi Data Mining Algoritma C4.5 dan Naïve Bayes untuk Prediksi Penyakit Hepatitis. Volume 13 No.1,
- Novianti, Deny. (2019). Implementasi Algoritma Naïve Bayes Pada Dataset Hepatitis Menggunakan Rapid Miner. Vol XXI, No.1,
- Fathorazi Nur Fajri, “Perbandingan Sistem Klasifikasi Naïve Bayes dan Decision Tree Untuk diagnose Penyakit Diabetes”
- Eva Rahmawati. (2015) “Analisa Komparasi Algoritma Naive Bayes Dan C4.5 Untuk Prediksi Penyakit Liver”.VolXII,No2,
- Novita,Rina. (2003-2016) Teknik Data Mining Algoritma C4.5.IlmuKomputer.com,
- Ridwan.(2008) ”Metode dan Teknik Menyusun Tesis .Bandung:Alfabeta
- T.Sutojo, S.Si, M.Kom , Edy Mulyanto, S.Si, M.Kom dan Dr. Vincent Suhartono. (2011).
- Andi Cynithia,E.P and Ismanto,E (2018). ‘Metode Decision tree algoritma C4.5 dalam mengkalifikasi data penjualan ‘, Jurnal riset system informasi dan teknik informatika (JURASI), 3 juli, pp.1-13.
- Medha, Andro,Ied Dan Haspari,Puspita.,Dian. (2019) Implementing Fuzzy Decision Tree For Predicting Liver Disease Of ILPD (Indian Liver Patient Dataset). Vol5, No.2,