

PEMILIHAN ATRIBUT PADA ALGORITMA C4.5 MENGGUNAKAN PARTICLE SWARM OPTIMIZATION UNTUK MENINGKATKAN AKURASI PREDIKSI DIAGNOSIS PENYAKIT LIVER

Elah Nurlelah¹; M. Sukrisno Mardiyanto²

¹Program Studi Sistem Informasi
Universitas Bina Sarana Informatika
<http://bsi.ac.id/>
elah.enl@bsi.ac.id

²Departemen Teknik Informatika
Institut Teknologi Bandung (ITB)
<http://itb.ac.id/>
sukrisno@informatika.org

Abstract— *The liver is a vital human organ that has complex and diverse functions, one of which is to maintain the needs of organs in the body, especially the brain, because the brain is a complex liver function, therefore liver health needs to be considered early so that the body remains healthy. Liver or liver disease is one of the 10 biggest diseases that cause death in Indonesia, but the public's understanding of liver disease is still very low. As a result, many of them do not get early treatment appropriately. In this study the C4.5 algorithm was optimized by using Particle Swarm Optimization to improve the accuracy of predictions of liver disease diagnosis. After testing with two models, namely C4.5 Algorithm and C4.5 Optimization using Particle Swarm Optimization, the results obtained by testing using C4.5 where the accuracy value is 78,86% and the AUC value is 0,815%, while testing using Optimization of C4.5 with Particle Swarm Optimization obtained 82,08% accuracy value and AUC value was 0,829 with good classification diagnosis. So that the two methods have different levels of accuracy that is equal to 3,22% and the difference in the AUC value is 0,014%. It can be concluded that the application of the Particle Swarm Optimization optimization technique is able to select attributes on C4.5, resulting in a better level of diagnosis of liver disease than using the individual method C4.5 algorithm.*

Keywords: *Liver, Algorithm C4.5, Selection Attributes, Particle Swarm Optimization.*

Intisari— Hati adalah organ vital manusia yang memiliki fungsi kompleks dan beragam, salah satunya adalah dengan menjaga kebutuhan organ dalam tubuh, khususnya otak, karena otak adalah fungsi hati yang kompleks, oleh karena itu, kesehatan hatipun perlu diperhatikan sejak dini

agar tubuh tetap sehat. Penyakit hati atau liver merupakan salah satu dari 10 penyakit terbesar penyebab kematian di Indonesia, tetapi pemahaman masyarakat mengenai penyakit liver ini masih sangat rendah. Akibatnya banyak dari mereka yang tidak mendapatkan penanganan dini secara tepat. Karenanya, penting bagi kita untuk dapat mengenali gejala dini penyakit liver sebelum terlambat. Selain itu, deteksi dini terhadap penyakit liver memungkinkan penderita penyakit ini dapat disembuhkan lebih cepat. Sudah banyak penelitian untuk memprediksi penyakit liver, salah satunya yaitu dengan menerapkan metode C4.5. Pada penelitian ini dilakukan optimasi algoritma C4.5 dengan menggunakan Particle Swarm Optimization untuk meningkatkan akurasi prediksi diagnosis penyakit liver. Setelah dilakukan pengujian dengan dua model yaitu Algoritma C4.5 dan Optimasi C4.5 menggunakan Particle Swarm Optimization maka hasil yang didapat pengujian dengan menggunakan C4.5 dimana didapat nilai accuracy adalah 78,86% dan nilai AUC adalah 0,815%, sedangkan pengujian dengan menggunakan Optimasi C4.5 dengan Particle Swarm Optimization didapatkan nilai accuracy 82,08% dan nilai AUC adalah 0,829 dengan tingkat diagnosa *good classification*. Sehingga kedua metode tersebut memiliki perbedaan tingkat akurasi yaitu sebesar 3,22% dan perbedaan nilai AUC sebesar 0,014%. Dapat disimpulkan bahwa penerapan teknik optimasi Particle Swarm Optimization mampu menyeleksi atribut pada C4.5, sehingga menghasilkan tingkat akurasi diagnosis penyakit liver yang lebih baik dibanding dengan menggunakan metode individual algoritma C4.5.

Kata Kunci: *Liver, Algoritma C4.5, Pemilihan Atribut, Particle Swarm Optimization.*

PENDAHULUAN

Hati merupakan organ vital manusia yang memiliki fungsi kompleks dan beragam seperti menawarkan dan menetralkan zat-zat racun yang tidak bisa diserap oleh usus dan menyaring darah yang datang dari usus melalui vena porta, kemudian bahan makanan dari vena porta tersebut disimpan dan diubah yang selanjutnya bahan makanan dikirim sesuai kebutuhan ke dalam darah. (Falatehan, Hidayat, & Brata, 2018). Di dalam hati, makanan yang mengandung racun akan dinetralkan terlebih dahulu sehingga jika telah melewati hati makanan tersebut tidak akan mengandung racun. Hati itu sendiri memiliki peran yang sangat vital di dalam tubuh manusia, salah satunya adalah dengan menjaga kebutuhan organ dalam tubuh, khususnya otak, karena otak adalah fungsi hati yang kompleks dan beragam.

Penyebab penyakit hati yang paling umum adalah konsumsi alkohol berlebihan, virus, kecanduan obat (khususnya dalam pembuluh darah), reaksi yang berlawanan dari berbagai macam obat (seperti analgesik, obat-obat anti peradangan, beberapa antibiotik, obat-obatan anti jamur dan penekan kekebalan (Latifatul Khairiah, Tursina, 2017). Oleh karena itu penyakit hati harus ditangani secara dini karena jika tidak maka akan berkembang menjadi kanker hati bahkan dapat berujung pada kematian.

Angka pasti prevalensi dan kejadian penyakit liver di Indonesia pada tahun 2007 belum diketahui tingkat kenaikannya, tetapi berdasarkan data WHO (World Health Organization) pada tahun 2007 menunjukkan bahwa Indonesia termasuk dalam peringkat endemik yang tinggi (Restiani, 2018).

Sudah banyak penelitian untuk memprediksi penyakit liver, seperti penelitian yang telah dilakukan (Bahramirad, Mustapha, & Eshraghi, 2013) tentang klasifikasi diagnosis penyakit liver dengan cara melakukan komparasi 11 algoritma, dan yang lebih akurat yaitu menggunakan metode Logistic, penelitian selanjutnya yang dilakukan oleh (Sontakke, Lohokare, & Dani, 2017) yaitu tentang diagnosis penyakit liver menggunakan *machine learning*, dengan menggunakan metode *Support Vector Machine* dan *Back Propagation*. Penelitian yang dilakukan oleh (Widodo, 2014) tentang *rule-based classifier* untuk mendeteksi penyakit liver, dan di dalamnya menggunakan beberapa algoritma yaitu zeroR, oneR, RIPPER dan juga algoritma C4.5 dengan cara membandingkan algoritma tersebut. Keuntungan pengklasifikasian menggunakan pohon keputusan memiliki kelebihan dalam memecahkan struktur kompleks menjadi struktur yang lebih sederhana sehingga lebih mudah untuk

diimplementasikan (Cho, Lee, & Jun, 2011), (Abellán & Masegosa, 2012). Akan tetapi pohon keputusan memiliki kelemahan dalam menangani data yang besar dan ketidakseimbangan data yang disebabkan oleh banyaknya atribut pada sebuah dataset (Sarkar, Sana, & Chaudhuri, 2011) dan muncul noise data ketika salah pelabelan (Abellán & Masegosa, 2012).

Tujuan dari penelitian ini adalah menerapkan *Particle Swarm Optimization* untuk mengeliminasi atribut input pada metode Algoritma C4.5, untuk meningkatkan akurasi prediksi penyakit liver. Hasil dari penelitian ini berupa hasil proses pengolahan kualitatif dan kuantitatif yang telah dikumpulkan dengan perhitungan berdasarkan model yang diusulkan. Penelitian akan dilakukan terhadap semua dataset yang tersedia. Eksperimen dan pengujian dalam penelitian dilakukan melakukan prediksi terhadap dataset dengan C4.5 tanpa PSO, dan metode seleksi atribut dengan menggunakan PSO. Eksperimen akan dilakukan terhadap dataset yang telah divalidasi.

BAHAN DAN METODE

A. Jenis Penelitian

Jenis penelitian yang dilakukan pada penelitian ini adalah jenis penelitian *experiment*, yaitu penelitian yang dilakukan dengan cara menguji keberanian sebuah hipotesis dengan statistik yang melibatkan penyelidikan beberapa variabel dengan menggunakan tes tertentu dan menghubungkannya dengan masalah penelitian. Data utama diperoleh dari data rekap medik rumah sakit pasien yang positif liver dan pasien yang negatif liver dengan jumlah 279 *record* yang terdiri dari 199 *record* (71,32%) pasien liver dan 80 *record* (28,68%) pasien negatif liver. Dengan jumlah 11 atribut.

B. Instrumen Penelitian

Beberapa hal penting yang menjadi instrumen pada penelitian ini sebagai berikut:

1. Penelitian ini menggunakan data primer berupa data pasien yang menderita penyakit liver ataupun tidak yang akan digunakan sebagai instrumen guna memperoleh data dalam proses mendiagnosa penyakit liver.
2. Data disajikan dalam bentuk tabulasi model dan variabel masing-masing sebanyak 279 pasien yang terdiri dari pasien positif liver sebanyak 199 dan pasien tidak positif liver sebanyak 80 pasien yang diambil pada tahun 2016-2018.
3. Perangkat lunak yang digunakan untuk menganalisis adalah Rapid Miner dan

Graphical User Interface (GUI) untuk menguji rule algoritma terpilih adalah PHP.

C. Metode Analisa dan Pengujian Data

Dua pendekatan utama dalam penelitian yaitu pendekatan kualitatif dan pendekatan kuantitatif. Dalam penelitian ini metode yang digunakan yaitu metode penelitian kuantitatif. Jenis penelitian yang dilakukan pada penelitian ini adalah jenis penelitian *experiment*, dan dalam penelitian ini menggunakan metode *experiment* dengan model CRISP-DM (*Cross- Industry Standard Process for Data Mining*) (Nuraeni, 2017). Ada 6 fase yang saling berhubungan untuk menggambarkan proses data mining yaitu:

1. Tahap Pemahaman Bisnis (*Business Understanding*)

Dalam penelitian ini, berdasarkan data yang berasal dari data internal Rumah Sakit Betha Medika tahun 2016-2018 yang menderita penyakit liver itu sekitar 279 orang, untuk mengurangi jumlah resiko penyakit liver maka harus dilakukan pemeriksaan sejak dini terhadap pasien agar dapat dideteksi penyakitnya.

2. Tahap Pemahaman Data (*Data Understanding*)

Data yang digunakan pada penelitian ini menggunakan data primer. Data penelitian ini diambil dari data rekam medik penyakit liver Positif dan Liver Negatif, yang di dapat dari Rumah Sakit Betha Medika, dengan jumlah 279 record yang terdiri dari 199 record (71,32%) positif liver dan 80 record (28,68%) negatif liver.

3. Tahap Persiapan Data (*Data Preparation*)

Jumlah data yang diperoleh pada penelitian ini sebanyak 279 *record*, baik pasien *liver* dan pasien tidak *liver*, akan tetapi data tersebut masih mengandung anomali atau inkonsisten data.

4. Tahap Pemodelan (*Modelling*)

Tahap *modelling* dilakukan untuk menerapkan teknik yang tepat guna mendapatkan hasil yang optimal dalam memprediksi penyakit liver. Pada penelitian ini model yang digunakan yaitu algoritma terpilih pohon keputusan C4.5 dan algoritma pohon keputusan C4.5 berbasis *Particle Swarm Optimization* (PSO). Tahap *modelling* dilakukan dengan dua cara yaitu cara manual dan dengan menggunakan *software* rapid miner. Perhitungan manual untuk pembuatan model dengan menggunakan algoritma pohon keputusan C4.5 dilakukan dengan cara mencari nilai gain tertinggi dari setiap atribut, sedangkan perhitungan manual untuk model algoritma pohon keputusan C4.5 berbasis PSO diperoleh dari dataset dengan candidate split yang berbeda.

5. Tahap Evaluasi (*Evaluation*)

Pada tahap ini dilakukan pengujian terhadap model-model untuk mendapatkan informasi model yang akurat. Evaluasi dan validasi

menggunakan metode *Confision Matrix* dan kurva ROC.

6. Tahap *Deployment*

Pada tahap ini diterapkan model yang memiliki akurasi tinggi atau yang paling baik pada departemen kesehatan atau instansi yang relevan untuk mendiagnosa atau memprediksi penyakit Liver dengan menggunakan data baru.

HASIL DAN PEMBAHASAN

Dalam membuat pohon keputusan terlebih dahulu kita harus menghitung jumlah *class* yang terkena penyakit liver yang *Positif* dan yang *negative* dan juga *entropy* dari masing-masing *class* berdasarkan atribut yang telah ditentukan.

Hasil pemodelan dengan Algoritma C4.5

Pemodelan menggunakan algoritma C4.5 dilakukan pada dataset yang terdiri dari 11 atribut yang merupakan atribut dari diagnosis penyakit liver dan *class* yang merupakan hasil akhir prediksi. Data tersebut kemudian di validasi dengan tujuan agar proses pelatihan dapat berjalan dengan cepat dan mampu digunakan untuk melakukan pelatihan.

Model dari algoritma C4.5 yaitu berupa pohon keputusan, untuk dapat membuat pohon keputusan, langkah pertama adalah menghitung jumlah *class* yang terkena penyakit liver yang *Positif* dan yang *negative* dari masing-masing *class* berdasarkan atribut yang telah ditentukan dengan menggunakan data *training*. Kemudian menghitung *Entropy* (Total):

$Entropy$ (Total)=

$$\left(\left(-\frac{199}{279}\right)\log_2\left(\frac{199}{279}\right)\right) + \left(\left(-\frac{80}{279}\right)\log_2\left(\frac{80}{279}\right)\right)$$

$Entropy$ (Total) = 0,864470605

Untuk mendapatkan *gain* tiap atribut, maka harus dihitung *entropy* atribut berdasarkan tiap-tiap kasus.

$Entropy$ (Age, <=14) =

$$\left(\left(-\frac{2}{279}\right)\log_2\left(\frac{2}{279}\right)\right) + \left(\left(-\frac{4}{279}\right)\log_2\left(\frac{4}{279}\right)\right)$$

$Entropy$ (Age, <=14) = 0,918295834

$Entropy$ (Age, 15-49) =

$$\left(\left(-\frac{112}{279}\right)\log_2\left(\frac{112}{279}\right)\right) + \left(\left(-\frac{44}{279}\right)\log_2\left(\frac{44}{279}\right)\right)$$

$Entropy$ (Age, 15-49) = 0,858230793

$Entropy$ (Age, >= 50) =

$$\left(\left(-\frac{85}{279}\right)\log_2\left(\frac{85}{279}\right)\right) + \left(\left(-\frac{32}{279}\right)\log_2\left(\frac{32}{279}\right)\right)$$

$Entropy$ (Age, >= 50) = 0,846448228

Setelah *entropy* dari atribut sudah di dapat maka langkah berikutnya menghitung *gain* sebagai berikut:

$$Gain(Total, Age) = 0,864470605 - \left(\left(\frac{6}{279} \right) * 0,918295834 \right) + \left(\left(\frac{156}{279} \right) * 0,858230793 \right) + \left(\left(\frac{117}{279} \right) * 0,846448228 \right)$$

$$Gain(Total, Age) = 0,009889167$$

Hasil dari perhitungan keseluruhan atribut dapat dilihat pada tabel 1 dibawah ini:

Tabel1. Perhitungan manual *entropy* dan *gain* total

Node	Atribut	Jml Kasus (S)	Positif Liver (Si)	Negatif Liver (Si)	Entropy	Gain
1	Total	279	199	80	0,864470605	
	AGE					
		<= 14	6	2	4	0,918295834
		15 - 49	156	112	44	0,858230793
		>= 50	117	85	32	0,846448228
						0,009889167
	GENDER					
		MALE	249	179	70	0,856978497
		FEMALE	30	20	10	0,918295834
						0,000898846
	TB					
		<= 1.0	121	65	56	0,996005523
		1.1 - 2.0	46	32	14	0,886540893
		2.1 - 3.0	24	20	4	0,650022422
		3.1 - 9.0	44	38	6	0,574635698
		>9.0	44	44	0	0
						0,13980373
	DB					
		<= 1.0	168	99	69	0,976874013
		1.1 - 2.0	41	36	5	0,534943699
		2.1 - 5.0	27	21	6	0,764204507
		>5.0	43	43	0	0
						0,123678321
	ALPKHOS					
		<= 140	13	10	3	0,779349837
		141 - 210	111	63	48	0,98678672
		211 - 280	70	60	10	0,591672779
		281 - 420	54	43	11	0,729273942
		>420	31	23	8	0,823811633
						0,054431459
	SGPT					
		<= 40	139	85	54	0,963817658
		41 - 80	72	54	18	0,811278124
		81 - 120	21	15	6	0,863120569
		121- 200	21	19	2	0,453716339
		>200	26	26	0	0
						0,075810195
	SGOT					
		<= 41	104	56	48	0,995727452
		42-82	71	48	23	0,908618598
		83 - 164	48	42	6	0,543564443
		> 164	56	53	3	0,301378644
						0,108069628
	TP					
		<5.0	16	11	5	0,896038233
		5.0 - 5.9	75	51	24	0,904381458
		>= 6.0	188	137	51	0,843290512
						0,001732837
	ALB					
		0.0 - 1.9	17	10	7	0,977417818
		2.0 - 2.9	103	81	22	0,748293286
		3.0 - 3.7	107	81	26	0,799978361
		3.8 - 6.0	52	27	25	0,998932655
						0,035680304
	A/G RATIO					
		0.0 - 0.97	156	115	41	0,830972808
		1.0 - 1.18	114	79	35	0,889700675
		>1.18	9	5	4	0,99107606
						0,004336844

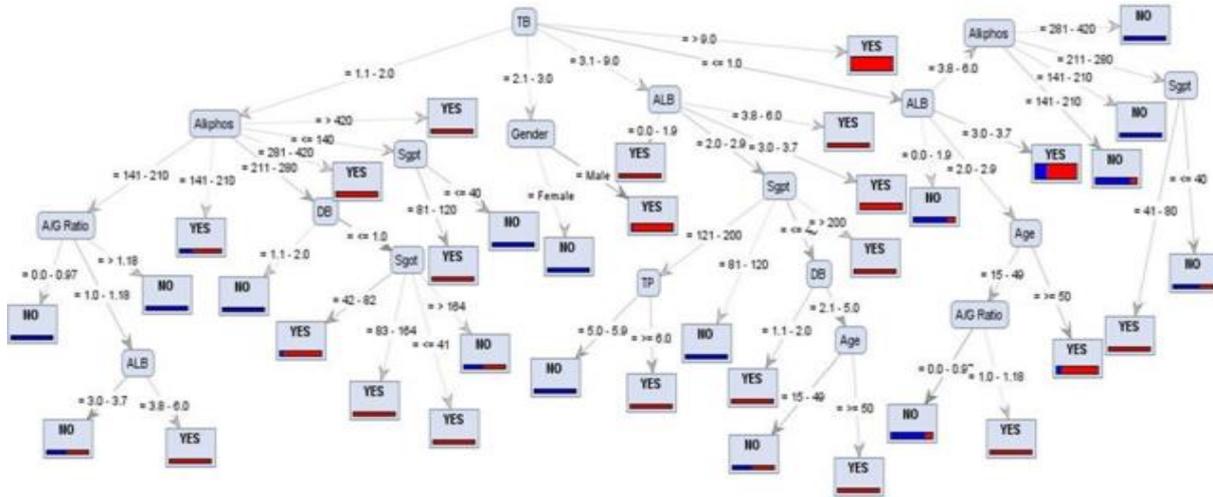
Sumber: (Nurlelah & Mardiyanto, 2019)

Dari hasil perhitungan *entropy* dan *gain* di atas, diketahui bahwa atribut TB memiliki nilai paling tinggi yaitu 0,13980373, oleh karena itu atribut TB menjadi akar atau node yang pertama dari pohon

keputusan yang terbentuk. Pada atribut ini terdapat empat cabang yaitu <=1.0, 1.1-2.0, 2.1-3.0, 3.1 - 9.0 dan >9.0. Cabang >9.0 sudah terklasifikasi sehingga tidak perlu lagi dilakukan

perhitungan *entropy* dan *gain* untuk cabang tersebut. Sedangkan untuk cabang ≤ 1.0 , $1.1-2.0$, $2.1-3.0$ dan $3.1-9.0$ masih belum terklasifikasi sehingga perlu dilakukan perhitungan *entropy* dan *gain* kembali. Setelah didapatkan hasil

perhitungan *entropy* dan *gain*, maka pohon keputusan yang terbentuk dapat dilihat seperti gambar dibawah ini:



Sumber:(Nurlelah & Mardiyanto, 2019)

Gambar 1. Pohon Keputusan Klasifikasi Penyakit Liver C4.5

Terdapat 35 rule yang dihasilkan dari pohon keputusan algoritma klasifikasi C4.5, dengan jumlah class NO sebanyak 16 rule dan 19 rule untuk class YES liver.

Hasil dari uji coba yang dilakukan yaitu untuk menghasilkan nilai *accuracy* dan nilai *AUC (Area Under Curve)*.

Model *confusion matrix* akan membentuk matrix yang terdiri dari *true positif* atau tupel positif dan *true negatif* atau tupel negatif, kemudian masukan data *testing* yang sudah disiapkan ke dalam *confusion matrix* sehingga didapatkan hasil pada tabel dibawah ini:

Tabel 2. *Confusion Matrix* Algoritma C4.5

	true YES	true NO	class precision
pred. YES	39	18	68.42%
pred. NO	41	181	81.53%
class recall	48.75%	90.95%	

Sumber: (Nurlelah & Mardiyanto, 2019)

Berdasarkan tabel diatas dari data *testing* terdapat rincian jumlah *True Positive (TP)* 39, *False Negative (FN)* 18, *False Positive (FP)* adalah 41 dan *True Negative (TN)* 181. Dari data tersebut maka dapat dihitung nilai *accuracy*, *sensitivity*, *specifity*, *PPV* dan *NPV*.

$$Accuracy = \frac{39+181}{39+181+41+18} = 0,7886 / 78.86\%$$

$$Sensitivity = \frac{39}{39+18} = 0,6842 / 68,42\%$$

$$Specificity = \frac{181}{181+41} = 0,8153 / 81,53\%$$

$$PPV = \frac{39}{39+41} = 0,4875 / 48,75\%$$

$$NPV = \frac{181}{181+18} = 0,9095 / 90,95\%$$

Hasil pengujian terhadap data yang digunakan untuk algoritma C4.5 terhadap nilai ROC diketahui pada gambar di bawah ini:



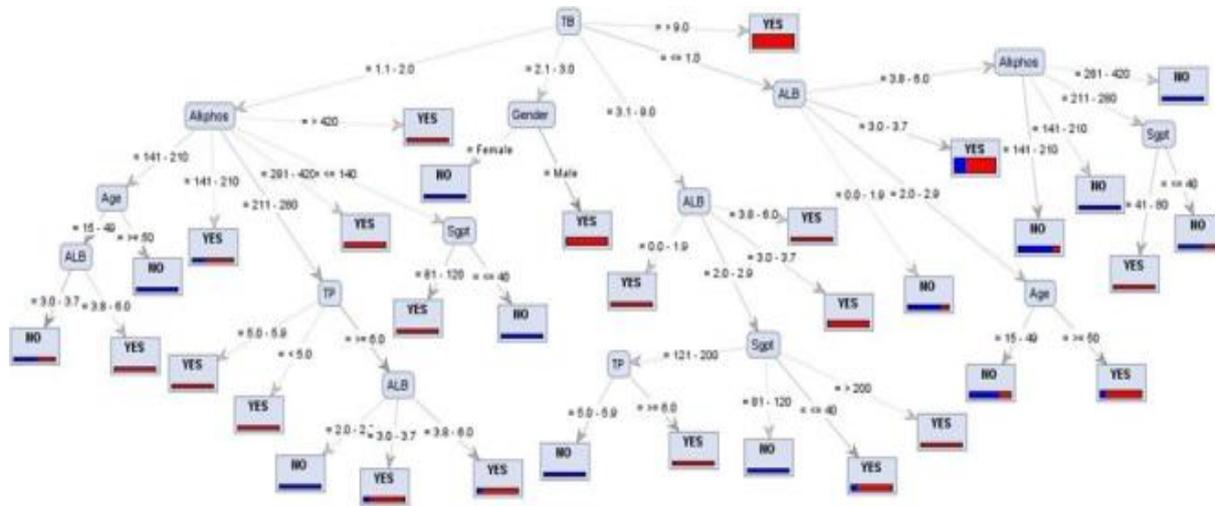
Sumber: (Nurlelah & Mardiyanto, 2019)

Gambar 2. Nilai AUC Algoritma C4.5 dalam Grafik ROC

Hasil pemodelan dengan Algoritma C4.5 – PSO

Pada PSO, Perhitungan pada keseluruhan data dilakukan dengan Rapid miner, model yang terbentuk dengan algoritma C4.5 sendiri akan dibandingkan dengan model yang terbentuk setelah proses pembobotan attribut. Nilai bobot atribut akan diberikan kepada setiap partikel PSO. Model yang terbentuk akan diuji sehingga dapat diketahui akurasi model yang terbentuk, setelah itu partikel PSO akan bergerak ke arah partikel terbaik

agar akurasi model yang dihasilkan dapat lebih tinggi dari sebelumnya.



Sumber: (Nurlelah & Mardiyanto, 2019)

Gambar 3. Pohon Keputusan Klasifikasi Penyakit Liver C4.5-PSO

Terdapat 31 rule yang dihasilkan dari pohon keputusan algoritma klasifikasi C4.5, dengan jumlah class NO sebanyak 11 rule dan 20 rule untuk class YES LIVER. Sedangkan atribut yang terseleksi dari 10 atribut menjadi 7 atribut, yang terdiri dari AGE, GENDER, TB, ALPKHOS, SGPT, TP dan ALB.

Berikut model confusion matrix Algoritma C4.5 - PSO pada tabel dibawah ini:

Tabel 3. Confusion Matrix Algoritma C4.5 - PSO

	true YES	true NO	class precision
pred. YES	44	14	75.86%
pred. NO	36	185	83.71%
class recall	55.00%	92.96%	

Sumber: (Nurlelah & Mardiyanto, 2019)

Berdasarkan tabel diatas terdapat rincian jumlah True Positive (TP) 44, False Negative (FN) 14, False Positive (FP) adalah 36 dan True Negative (TN) 185. Dari data tersebut maka dapat dihitung nilai accuracy, sensitivity, specificity dan NPV.

$$Accuracy = \frac{44+185}{43+185+36+14} = 0,8208 / 82,08\%$$

$$Sensitivity = \frac{44}{44+14} = 0,7586 / 75,86\%$$

$$Specificity = \frac{185}{185+36} = 0,8371 / 83,71\%$$

$$PPV = \frac{44}{44+36} = 0,55 / 55,00\%$$

$$NPV = \frac{185}{185+14} = 0,9296 / 92,96\%$$

Hasil pengujian terhadap data testing untuk algoritma C4.5 - PSO terhadap nilai ROC diketahui pada gambar di bawah ini:



Sumber: (Nurlelah & Mardiyanto, 2019)

Gambar 4. Nilai AUC Algoritma C4.5 - PSO dalam grafik ROC

Uji Beda t-Test untuk Model Algoritma C4.5 dengan Model Algoritma C4.5 dan Particle Swarm Optimization

Untuk menjamin hasil hipotesis dari penelitian ini, dilakukan evaluasi terhadap hipotesis menggunakan metode yang paling umum dalam metode statistik tradisional, yaitu t-Test. Metode tersebut biasanya menggunakan nilai signifikan $\alpha=0.05$. Untuk mengetahui ada atau tidaknya perbedaan antara penggunaan metode Algoritma C4.5 dengan Algoritma C4.5 dan Particle Swarm Optimization membutuhkan pengujian, salah satunya dengan uji t-Test. Untuk mengukur seberapa signifikan hipotesis tersebut teruji. Jika nilai $\alpha<0,05$ maka menunjukkan hipotesis nol ditolak atau disebut hipotesis alternatif (Sumanto, 2014). Hipotesis nol menyatakan tidak dapat meningkatkan akurasi, sedangkan hipotesis alternatif menyatakan dapat meningkatkan akurasi.

Tabel 4. Uji t-Test Model C4.5 dengan Model C4.5 dan Particle Swarm Optimization

	C4.5		C4.5-PSO	
Accuracy	0.763	+/-	0.821	+/-
	0.056		0.060	
C4.5	0.763	+/-	0.045	
	0.056			
C4.5-PSO	0.821	+/-		
	0.060			

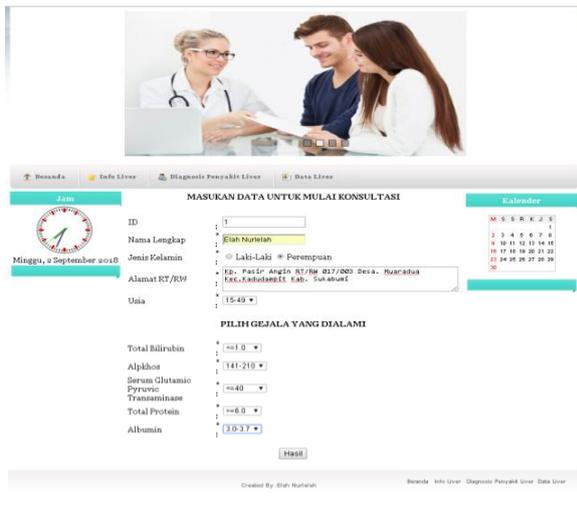
Sumber: (Nurlelah & Mardiyanto, 2019)

Pada Tabel 6 menampilkan t-Test untuk hasil Accuracy pada dataset liver menunjukkan liver nol ditolak (hipotesis alternatif) yaitu dengan nilai $\alpha < 0,05$ yaitu 0,045.

Hasil t-Test dengan hipotesis nol ditolak tersebut menunjukkan bahwa antara penggunaan metode algoritma C4.5 dengan algoritma C4.5 dan PSO menunjukkan adanya pengaruh atau perbedaan yang signifikan. Metode algoritma C4.5 dan PSO menghasilkan tingkat akurasi yang lebih baik dibanding dengan menggunakan metode algoritma saja.

Pengembangan Aplikasi Model dari Algoritma C4.5 dan Particle Swarm Optimization

Berdasarkan evaluasi pada optimasi algoritma klasifikasi C4.5 berbasis PSO diketahui bahwa hasilnya lebih tinggi dibandingkan dengan algoritma klasifikasi C4.5, sehingga rule yang dihasilkan dari algoritma C4.5 berbasis PSO dijadikan sebagai rule untuk pembuatan Graphical User Interface guna membantu dokter, tim medis atau masyarakat awam dalam mendiagnosis penyakit liver. Interface yang digunakan dalam penelitian ini dibuat dengan menggunakan aplikasi berbasis web seperti pada gambar dibawah ini:



Sumber: (Nurlelah & Mardiyanto, 2019)

Gambar 5. Pengujian Aplikasi Model Algoritma C4.5 dan Particle Swarm Optimization

KESIMPULAN

Dari hasil penelitian untuk akurasi algoritma klasifikasi C4.5 sebesar 78,86%, sedangkan untuk akurasi algoritma klasifikasi dengan PSO sebesar 82,08%, sehingga didapat selisih peningkatan akurasi sebesar 3,22%. Sedangkan evaluasi menggunakan ROC curve diperoleh hasil untuk algoritma C4.5 bernilai 0,815 dengan tingkat diagnosa Good Classification dan Optimasi algoritma C4.5 menggunakan PSO bernilai 0,829 dengan tingkat diagnosa Good Classification, didapatkan selisih nilai AUC sebesar 0,014%. Dapat disimpulkan bahwa penerapan teknik optimasi particle swarm optimization mampu menyeleksi atribut pada C4.5, sehingga menghasilkan tingkat akurasi diagnosis penyakit liver yang lebih baik dibanding dengan menggunakan metode individual algoritma C4.5.

REFERENSI

Abellán, J., & Masegosa, A. R. (2012). Bagging schemes on the presence of class noise in classification. *Expert Systems with Applications*, 39(8), 6827–6837. <https://doi.org/10.1016/j.eswa.2012.01.013>

Bahramirad, S., Mustapha, A., & Eshraghi, M. (2013). Classification of liver disease diagnosis: A comparative study. *2013 2nd International Conference on Informatics and Applications, ICIA 2013*, 42–46. <https://doi.org/10.1109/ICoIA.2013.6650227>

Cho, Y.-J., Lee, H.-S., & Jun, C.-H. (2011). Optimization of Decision Tree for Classification Using a Particle Swarm. *Industrial Engineering and Management Systems*, 10(4), 272–278. <https://doi.org/10.7232/iems.2011.10.4.272>

Falatehan, A. I., Hidayat, N., & Brata, K. C. (2018). Sistem Pakar Diagnosis Penyakit Hati Menggunakan Metode Fuzzy Tsukamoto Berbasis Android. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer (J-PTIIK) Universitas Brawijaya*, 2(8), 2373–2381. Retrieved from <http://j-ptiik.ub.ac.id>

Latifatul Khairiah, Tursina, T. R. (2017). Sistem Diagnosis Penyakit Hati Dengan Metode Dempstershafer Berbasis Android. *Coding Jurnal Komputer Dan Aplikasi*, 5(2), 57–66. Retrieved from <http://jurnal.untan.ac.id/index.php/jcskonnipa/article/view/19889>

- Nuraeni, N. (2017). Penentuan Kelayakan Kredit Dengan Algoritma Naïve Bayes Classifier : Studi Kasus Bank Mayapada Mitra Usaha Cabang PGC. *Jurnal Teknik Komputer AMIK BSI (JTK)*, 3(1), 9–15.
- Nurlelah, E., & Mardiyanto, M. S. (2019). *Laporan Akhir Penelitian - Pemilihan Atribut Pada Algoritma C4.5 Menggunakan Particle Swarm Optimization Untuk Meningkatkan Akurasi Prediksi Diagnosis Penyakit Liver*.
- Restiani, D. (2018). Kombinasi Algoritma C-Ripper untuk Mendiagnosis Penyakit Liver. *JTI (Jurnal Teknik Informatika) UIN Syarif Hidayatullah*, 11(1), 31–36. Retrieved from <http://journal.uinjkt.ac.id/index.php/ti/article/download/6660/pdf>
- Sarkar, B. K., Sana, S. S., & Chaudhuri, K. (2011). Selecting informative rules with parallel genetic algorithm in classification problem. *Applied Mathematics and Computation*, 218(7), 3247–3264. <https://doi.org/10.1016/j.amc.2011.08.065>
- Sontakke, S., Lohokare, J., & Dani, R. (2017). Diagnosis of liver diseases using machine learning. *2017 International Conference on Emerging Trends and Innovation in ICT, ICEI 2017*, 129–133. <https://doi.org/10.1109/ETIICT.2017.7977023>
- Widodo, P. (2014). Rule-Based Classifier untuk Mendeteksi Penyakit Liver. *Bianglala Informatika*, II(1), 71–80.