JITE (Journal of Informatics and Telecommunication Engineering)

Available online <http://ojs.uma.ac.id/index.php/jite> DOI : 10.31289/jite.v4i1.3793

Received: 20 Mei 2020

Accepted: 30 Juni 2020

Published: 20 Juli 2020

Attribute Selection in Naive Bayes Algorithm Using Genetic Algorithms and Bagging for Prediction of Liver Disease

Dwi Yuni Utami^{1)*}, Elah Nurlalah²⁾, Noer Hikmah²⁾

1) Prodi Teknologi Komputer, Fakultas Teknik dan Informatika, Universitas Bina Sarana Informatika, Indonesia

2) Prodi Sistem Informasi, Fakultas Teknik dan Informatika, Universitas Bina Sarana Informatika, Indonesia

*Corresponding Email: dwi.dyu@bsi.ac.id

Abstrak

Penyakit Liver merupakan penyakit peradangan pada organ hati dan dapat mengakibatkan hati tidak mampu berfungsi seperti biasanya bahkan menyebabkan kematian. Menurut data WHO (*World Health Organization*) menunjukkan hampir 1,2 juta orang per tahun khususnya di Asia Tenggara dan Afrika mengalami kematian akibat terserang penyakit liver. Permasalahan yang biasanya terjadi adalah sulitnya mengenali penyakit liver sejak dini, bahkan ketika penyakit tersebut sudah menyebar. Penelitian ini bertujuan untuk melakukan komparasi dan evaluasi algoritma Naive Bayes sebagai algoritma terpilih dan Algoritma Naive Bayes berbasis *Genetic Algorithm* (GA) dan Bagging untuk mengetahui algoritma yang memiliki keakuratan lebih tinggi dalam memprediksi penyakit liver dengan mengolah dataset yang diambil dari database *Machine Learning Repository UCI (Universitas California Invene)*. Dari hasil pengujian dengan dilakukan evaluasi baik secara *confusion matrix* maupun *ROC curve* terbukti bahwa pengujian yang dilakukan Optimasi algoritma Naive Bayes menggunakan Algoritma Genetika dan Bagging memiliki nilai akurasi yang lebih tinggi dibanding hanya menggunakan algoritma Naive Bayes. Nilai akurasi untuk model algoritma Naive Bayes sebesar 66,66% dan nilai akurasi model Naive Bayes dengan seleksi atribut menggunakan Algoritma Genetika dan Bagging sebesar 72,02%. Berdasarkan nilai tersebut diperoleh selisih akurasi sebesar 5,36%.

Kata kunci: Penyakit Liver, Naïve Bayes, Algoritma Genetika, Bagging.

Abstract

Liver disease is an inflammatory disease of the liver and can cause the liver to be unable to function as usual and even cause death. According to WHO (*World Health Organization*) data, almost 1.2 million people per year, especially in Southeast Asia and Africa, have died from liver disease. The problem that usually occurs is the difficulty of recognizing liver disease early on, even when the disease has spread. This study aims to compare and evaluate Naive Bayes algorithm as a selected algorithm and Naive Bayes algorithm based on *Genetic Algorithm* (GA) and Bagging to find out which algorithm has a higher accuracy in predicting liver disease by processing a dataset taken from the *UCI Machine Learning Repository database (GA. University of California Invene)*. From the results of testing by evaluating both the *confusion matrix* and the *ROC curve*, it was proven that the testing carried out by the Naive Bayes Optimization algorithm using Algoritma Genetics and Bagging has a higher accuracy value than only using the Naive Bayes algorithm. The accuracy value for the Naive Bayes algorithm model is 66.66% and the accuracy value for the Naive Bayes model with attribute selection using Genetic Algorithms and Bagging is 72.02%. Based on this value, the difference in accuracy is 5.36%.

Keywords: Liver Disease, Naïve Bayes, Genetic Algorithms, Bagging.

How to Cite: Utami, D.Y., Nurlalah, E. & Hikmah, N. (2020). Attribute Selection in Naive Bayes Algorithm Using Genetic Algorithms and Bagging for Prediction of Liver Disease. *JITE (Journal Of Informatics And Telecommunication Engineering)*. 4 (1): 76-85

I. PENDAHULUAN

Hati merupakan organ yang sangat penting dalam pengaturan homeostasis tubuh meliputi metabolisme, biotransformasi, sintesis, penyimpanan dan imunologi. (Rafsanjani et al., 2018)

Organ hati juga dapat terserang penyakit yang mengakibatkan hati tidak mampu berfungsi seperti biasanya bahkan menyebabkan kematian, penyakit liver merupakan penyakit hati yang sudah lama ada dan cukup umum di masyarakat. Menurut data WHO (*World Health Organization*) menunjukkan hampir 1,2 juta orang per tahun khususnya di Asia Tenggara dan Afrika mengalami kematian akibat terserang penyakit liver. (Ayudhitama & Pujianto, 2020)

Permasalahan yang biasanya terjadi adalah sulitnya mengenali penyakit liver sejak dini, bahkan ketika penyakit tersebut sudah menyebar. (Pusporani et al., 2019)

Pada dasarnya, sebuah Penyakit dapat diketahui atau diidentifikasi jenisnya dilihat dari gejala-gejala yang diderita pasien. (Zulfikar & Lukman, 2016) Kegiatan dalam melakukan prediksi terhadap berbagai penyakit telah banyak dilakukan dalam berbagai bidang keilmuan salah satunya bidang ilmu komputer science. (Noviandi, 2018) Banyak penelitian yang telah dilakukan

dibidang ilmu komputer science tentang penyakit liver.

Penelitian (Ghosh & Waheed, 2017) menyatakan bahwa dalam hal akurasi, presisi, sensitivitas dan spesifisitas algoritma K* ditemukan unggul karena memiliki tingkat kesalahan terendah dengan akurasi tertinggi dibandingkan dengan NBC, Bagging, logistik dan Rep Tree dengan dataset AP dan UCLA. Oleh karena itu algoritma K* yang paling cocok untuk diagnosis penyakit liver.

Penelitian (Dhamodharan, 2014) menggunakan Naïve Bayes dan akurasi algoritma FT Tree untuk prediksi penyakit liver sedangkan pengolahan data menggunakan sebuah perangkat lunak WEKA. Berdasarkan hasil ditemukan lebih baik menggunakan algoritma Naïve Bayes dibandingkan dengan algoritma lain dalam memprediksi hasil yang sangat akurat.

Penelitian (Nahar & Ara, 2018) membahas tentang penggunaan Teknik pohon keputusan (*Decision Tree*) seperti J48, LMT, Random Forest, Random Tree, REPTree, Decision Stump, dan Hoeffding untuk memprediksi penyakit liver dengan dataset dari UCI Machine Learning Repository. Dari hasil analisis, Decision Stump lebih baik daripada algoritma lainnya dengan akurasi mencapai 70,67%.

Penelitian (Pakhale & Xaxa, 2016) mensurvei teknik data mining yang telah digunakan untuk mengklasifikasi pasien liver dan memberikan ringkasan penggunaan Teknik data mining untuk klasifikasi pasien liver.

Penelitian (Priya et al., 2018) menggunakan Dataset ILPD (Indian Liver Patient Dataset) yang diambil dari UCI Machine Learning Repository (<http://archive.ics.uci.edu/ml/>) untuk menganalisis penyakit liver dengan menggunakan algoritma seperti J48, MLP, SVM, Random Forest, dan Bayesnet Classification berdasarkan model pemilihan fitur PSO. Dengan hasil algoritma J48 memberikan hasil yang lebih baik dengan akurasi 95,04% dibandingkan dengan algoritma lainnya.

Penelitian ini berbeda dengan penelitian lain dikarenakan penelitian memprediksi penyakit liver menggunakan algoritma Naïve Bayes berbasis *Genetic Algorithm* (GA) dan Bagging belum pernah ada. Untuk itu peneliti melakukan penelitian ini dengan tujuan untuk melakukan komparasi dan evaluasi algoritma Naive Bayes sebagai algoritma terpilih dan Algoritma Naive Bayes berbasis *Genetic Algorithm* (GA) dan Bagging untuk mengetahui algoritma yang memiliki keakuratan lebih tinggi dalam memprediksi penyakit liver.

II. STUDI PUSTAKA

A. *Data Mining*

Data mining merupakan suatu metode yang digunakan untuk pengolahan data guna menemukan pola yang tersembunyi dari data yang diolah. Data yang diolah dengan teknik data mining ini kemudian menghasilkan suatu pengetahuan baru yang bersumber dari data lama, hasil dari pengolahan data tersebut dapat digunakan dalam menentukan keputusan di masa depan.(Abrianto, 2018)

B. *Algoritma Genetika*

Algoritma genetika merupakan salah satu algoritma yang sangat tepat digunakan dalam menyelesaikan masalah optimasi kompleks, yang sulit dilakukan oleh metode konvensional. Sifat algoritma genetika adalah mencari kemungkinan-kemungkinan dari calon solusi untuk mendapatkan yang optimal bagi penyelesaian masalah.(Krisnandi & Agung, 2017)

C. *Naïve Bayes*

Naïve Bayes merupakan sebuah pengklasifikasian probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan (Saleh, 2015)

$$P(H \setminus X) = \frac{P(X \setminus H) P(H)}{P(X)} \quad (1)$$

Dimana:

X : Data dengan *class* yang belum diketahui

H : Hipotesis data merupakan suatu *class* yang spesifik

$P(H \setminus X)$: Probabilistik hipotesis H berdasar kondisi X (posteriori probabilistik)

$P(H)$: Probabilistik hipotesis H (prior probabilitas)

$P(X \setminus H)$: Probabilistik hipotesis X berdasar kondisi pada hipotesis H

$P(X)$: Probabilitas X

D. Bagging

Alfaro menjelaskan bahwa bagging bertujuan untuk meningkatkan akurasi pengklasifikasi dengan menggabungkan pengklasifikasi tunggal dan hasilnya lebih baik daripada random sampling. (Riyanto, 2018)

III. METODE PENELITIAN

Penelitian ini menggunakan data sekunder *Machine Learning Repository UCI* (*Universitas California Invene*) dengan alamat web: <http://archive.ics.uci.edu/ml/> dan jenis penelitian yang dilakukan pada penelitian ini adalah jenis penelitian *experiment*, yaitu penelitian yang dilakukan dengan cara menguji kebenaran sebuah hipotesis dengan statistik yang melibatkan penyelidikan beberapa variabel dengan menggunakan tes

tertentu dan menghubungkannya dengan masalah penelitian.

Langkah-langkah untuk melakukan optimasi dengan seleksi atribut *Genetic Algorithm* dan Bagging yaitu:

A. Pengumpulan Data

Data yang diteliti merupakan hasil pemeriksaan terhadap 583 orang dari wilayah Andhra Pradesh, India dengan 10 atribut dan satu atribut *class* yaitu *age*, *gender*, *Total Bilirubin*, *Direct Bilirubin*, *Alkaline Phosphotase*, *Sgpt Almine Aminotransferase*, *Sgot Aspartate Aminotransferase*, *Total Proteins*, *Albumin* dan *Albumin Globulin Ratio*.

B. Pengolahan Awal Data

Pengolahan awal data meliputi proses input data ke format yang dibutuhkan, pembersihan data, pengelompokan dan penentuan atribut data, serta pemecahan data (split) untuk digunakan dalam proses pembelajaran (training) dan pengujian (testing). Dari proses pengolahan data awal diperoleh sebanyak 583 data yang terdiri dari 416 positif liver dan 167 negatif liver, akan tetapi data tersebut masih mengandung anomali dan inkonsistensi data, maka dilakukanlah pembersihan data atau *cleaning data* karena masih adanya *missing values* dan juga karena data yang dipakai mengandung nilai-nilai yang salah

atau disebut noise atau outlier. Sehingga data yang diperoleh menjadi 579 data yang terdiri dari 414 record (71,50%) positif liver dan 165 record (28,50%) negatif liver.

C. Metode yang diusulkan

Dalam penelitian ini metode yang diusulkan adalah metode klasifikasi data mining yaitu Algoritma Naive Bayes dan Algoritma Naive Bayes dioptimasi dengan Algoritma Genetika dan Bagging.

D. Experimen dan pengujian model

Pada bagian ini pengujian model menggunakan *Cross Validation*. Dataset tersebut akan dibagi menjadi 10 bagian dan akan dilakukan pengulangan sebanyak 10 pengulangan.

E. Evaluasi dan Validasi Hasil

Pada bagian ini dilakukan pengujian terhadap model-model untuk mendapatkan informasi model yang akurat. Evaluasi dan validasi menggunakan metode *Confusion Matrix* dan *Curva ROC*.

Gambar 1. Kerangka Pemikiran

IV. HASIL DAN PEMBAHASAN

A. Eksperimen dan Pengujian Model Naive Bayes

Untuk dapat membentuk model diperlukan *data training*, dan tahap modelling menggunakan *software rapid miner* agar mampu membaca data dengan banyak format, yang paling umum digunakan diantaranya .csv, .xls dan .mdb. Dataset awal yang diambil dari UCI dalam bentuk format .txt kemudian ditransformasi menjadi format .xls.

1. Evaluasi Model Dengan Confusion Matrix

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 di bawah ini:

Tabel 1. Confusion Matrix Algoritma Klasifikasi Naive Bayes

	true YES	true NO	class precision
pred. YES	271	50	84.42%
pred. NO	143	115	44.57%
class recall	65.46%	69.70 %	

Berdasarkan tabel diatas dari data testing terdapat rincian jumlah True Positive (TP) 271, False Negative (FN) 115, False Positive (FP) adalah 143 dan True Negative (TN) 50. Dari data tersebut maka dapat dihitung nilai accuracy, sensitvity, specifity, PPV dan NPV.

Secara manual data tersebut dihitung menggunakan persamaan sebagai berikut:

$$Accuracy = \frac{271+115}{271+115+50+143} = 0,6666/66.66\% \quad (2)$$

$$Sensitivity = \frac{271}{271+115} = 0,7020 / 70,20\% \quad (3)$$

$$Specificity = \frac{50}{50+143} = 0,2590/25,90\% \quad (4)$$

$$PPV = \frac{271}{271+143} = 0,6546/65,46 \quad (5)$$

$$NPV = \frac{50}{50+115} = 0,3030/30,30\% \quad (6)$$

Maka dapat disimpulkan bahwa algoritma Naive Bayes dapat memprediksi penyakit liver Positif dan Negatif dengan akurasi yang baik sebesar 66,66%, untuk mengukur model yang paling baik dan model paling efisien maka digunakan *sensitivity* dan *specificity*. Sedangkan untuk mengukur proporsi kasus dengan hasil prediksi positif menggunakan *positive prediktive value* (PPV) dan untuk mengukur proporsi kasus dengan prediksi *negative* menggunakan *negative prediktive value* (NPV).

B. Eksperimen dan Pengujian model Naive Bayes, Algoritma Genetika dan Bagging

Pembuatan model Algoritma Naive Bayes, algoritma Genetika dan Bagging dilakukan pada dataset yang terdiri dari 10 atribut yang merupakan atribut dari prediksi penyakit liver. Data kemudian di validasi agar proses pelatihan dapat berjalan dengan cepat dan mampu digunakan untuk melakukan pelatihan. Tahap ini dibuatkan model pengolahan dengan menggunakan perangkat lunak aplikasi Rapidminer.. Evaluasi Model Dengan *Confusion Matrix*

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 di bawah ini:

Tabel 2. *Confusion Matrix* Algoritma Naive Bayes - Algoritma Genetika -Bagging

	true YES	true NO	class precision
pred. YES	365	113	76.36%
pred. NO	49	52	51.49%
class recall	88.16%	31.52%	

Berdasarkan tabel diatas terdapat rincian jumlah *True Positive (TP)* 365, *False Negative (FN)* 52, *False Positive (FP)* adalah 49 dan *True Negative (TN)* 113. Dari data tersebut maka dapat dihitung nilai *accuracy*, *sensitivity*, *specifity* dan *NPV*.

Secara manual data tersebut dihitung menggunakan persamaan sebagai berikut:

$$Accuracy = \frac{365 + 52}{365 + 52 + 113 + 49} = 0,7202 / 72.02\% \quad (7)$$

$$Sensitivity = \frac{365}{365 + 52} = 0,8752 / 87,52\% \quad (8)$$

$$Specificity = \frac{113}{113 + 49} = 0,6975 / 69,75\% \quad (9)$$

$$PPV = \frac{365}{365 + 49} = 0,8816 / 88,16\% \quad (10)$$

$$NPV = \frac{113}{113 + 52} = 0,6848 / 68,48\% \quad (11)$$

Maka dapat disimpulkan bahwa algoritma Naive Bayes + Algoritma Genetika + Bagging dapat memprediksi penyakit liver Positif dan Negatif dengan akurasi yang baik sebesar 72.02%.

C. Tabel Probabilitas

1. Tabel Probabilitas Model Naive Bayes

Dalam membuat model Naive Bayes terlebih dahulu kita mencari probabilitas hipotesis untuk masing-masing Kelas $P(H)$. Hipotesis yang ada yaitu pasien yang menderita penyakit liver dan pasien yang tidak menderita penyakit liver, perhitungan probabilitas yaitu seperti dibawah ini:

$$P(\text{Yes}) = 414:579 = 0,7 \quad (12)$$

$$P(\text{No}) = 165:579 = 0,2 \quad (13)$$

Setelah probabilitas untuk tiap hipotesis diketahui, langkah selanjutnya adalah menghitung probabilitas kondisi tertentu (probabilitas X) berdasarkan probabilitas tiap hipotesis (probabilitas H) atau dinamakan probabilitas prior.

Tabel 2. Hasil perhitungan probabilitas prior dengan menggunakan Naïve Bayes.

Atribut	Jml Kasus (S)	Yes (Si)	No (Si)	P(X Ci)	
				Yes	No
Total	579	414	165		
Age	<=14	16	8	0,0193	0,0484
	15-49	330	228	0,5507	0,6181
	>=50	233	178	0,4299	0,3334
Gender	Female	140	91	0,2198	0,2670
	Male	439	323	0,7801	0,703
TB	<= 1.0	302	182	0,4396	0,7272
	1.0-2.0	105	73	0,1763	0,1939
	2.1-3.0	49	42	0,1014	0,0424
	3.1-9.0	72	66	0,1594	0,0363

	>9.0	51	51	0	0,1231	0
DB	<=1.0	414	259	155	0,6256	0,9393
	1.0-2.0	63	58	5	0,14	0,0303
	2.1-5.0	54	49	5	0,1183	0,0303
	>5.0	48	48	0	0,1159	0
Alpkhos	<=140	35	23	12	0,5555	0,0727
	141-210	261	151	110	0,3647	0,6666
	211-280	107	88	19	0,2125	0,1151
	281-420	98	82	16	0,198	0,0969
	>420	78	70	8	0,169	0,0484
Sgpt	<=40	332	204	128	0,4927	0,7757
	41-80	141	113	28	0,273	0,1696
	81-120	38	32	6	0,0772	0,0363
	121-200	34	31	3	0,0748	0,0181
	>200	34	34	0	0,0821	0
Sgot	<=41	288	172	116	0,4154	0,703
	42-82	137	104	33	0,2512	0,2
	83-164	82	69	13	0,1666	0,0787
	>164	72	69	3	0,1666	0,0181
TP	<5.0	52	40	12	0,0966	0,0727
	5.0-5.9	200	142	58	0,3429	0,3515
	>=6.0	327	232	95	0,5603	0,5757
ALB	0.0-1.9	38	31	8	0,0748	0,0484
	2.0-2.9	191	150	41	0,3623	0,2484
	3.0-3.7	202	144	58	0,3478	0,3515
	3.8-6.0	147	89	58	0,2149	0,3515
A/G	0.0-0.97	296	230	66	0,5555	0,4
	1.0-1.18	163	113	50	0,2739	0,303
	>1.18	120	71	49	0,1714	0,297

2. Tabel Probabilitas Model Naive Bayes+Algoritma Genetika+Bagging

Proses yang dilakukan sama dengan proses pembuatan model naive bayes namun perbedaannya adalah dari atribut yang digunakan, atribut yang digunakan adalah atribut yang terpilih dari hasil seleksi atribut terbaik menggunakan

Algoritma Genetika, kemudian untuk meningkatkan stabilitas dalam klasifikasi maka diterapkan bagging selain berguna untuk meningkatkan stabilitas klasifikasi, bagging juga berguna untuk mengurangi variansi data dan untuk menghindari overfitting data, berikut hasil probabilitas prior untuk model naive bayes + algoritma genetik+bagging:

Tabel 13. Probabilitas Prior setelah di seleksi menggunakan Algoritma Genetik dan Bagging

Atribut		Jml Kasus (S)	Yes (Si)	No (Si)	P(X Ci)	
					Yes	No
Total		579	414	165		
Age	<=14	16	8	8	0,0193	0,0484
	15-49	330	228	102	0,5507	0,6181
	>=50	233	178	55	0,4299	0,3334
Gender	Female	140	91	49	0,2198	0,2670
	Male	439	323	116	0,7801	0,703
TB	<= 1.0	302	182	120	0,4396	0,7272
	1.0-2.0	105	73	32	0,1763	0,1939
	2.1-3.0	49	42	7	0,1014	0,0424
	3.1-9.0	72	66	6	0,1594	0,0363
	>9.0	51	51	0	0,1231	0
DB	<=1.0	414	259	155	0,6256	0,9393
	1.0-2.0	63	58	5	0,14	0,0303
	2.1-5.0	54	49	5	0,1183	0,0303
	>5.0	48	48	0	0,1159	0
Sgpt	<=40	332	204	128	0,4927	0,7757
	41-80	141	113	28	0,273	0,1696
	81-120	38	32	6	0,0772	0,0363
	121-200	34	31	3	0,0748	0,0181
	>200	34	34	0	0,0821	0
Sgot	<=41	288	172	116	0,4154	0,703
	42-82	137	104	33	0,2512	0,2
	83-164	82	69	13	0,1666	0,0787
	>164	72	69	3	0,1666	0,0181
TP	<5.0	52	40	12	0,0966	0,0727

	5.0-5.9	200	142	58	0,3429	0,3515
	>=6.0	327	232	95	0,5603	0,5757

Probabilitas prior yang digunakan yaitu probabilitas hasil seleksi atribut menggunakan algoritma genetika, probabilitas priors ini digunakan untuk menentukan kelas pada kasus baru yang terlebih dahulu dihitung *probabilitas posterior* nya.

D. *Komparasi Model Algoritma Naive Bayes dan Algoritma Naive Bayes + Algoritma Genetika+Bagging*

Hasil pengujian menyimpulkan bahwa Atribut yang mempengaruhi pengujian adalah 10 atribut. Hasil pengujian Naive Bayes tanpa seleksi atribut dibandingkan dengan model Naive Bayes dengan seleksi atribut menggunakan Algoritma Genetika dan Bagging dapat dilihat pada tabel berikut:

Tabel 4. Pengujian Algoritma klasifikasi Naive Bayes dan Naive Bayes+Algoritma Genetika + Bagging

	Accuracy
Naive Bayes	66,66%
Naive Bayes + Algoritma Genetika + Bagging	72,02%

Berikut perbedaan akurasi model Algoritma Naive Bayes dan model Algoritma Naive Bayes dengan Algoritma Genetika + Bagging.

Gambar 2 Perbedaan Akurasi Algoritma Naive Bayes dan Algoritma Naive Bayes dengan Algoritma Genetika + Bagging

Dari hasil pengujian diatas, dengan dilakukan evaluasi baik secara confusion matrix maupun ROC curve terbukti bahwa pengujian yang dilakukan Optimasi algoritma Naive Bayes menggunakan Algoritma Genetika dan Bagging memiliki nilai akurasi yang lebih tinggi dibanding hanya menggunakan algoritma Naive Bayes. Nilai akurasi untuk model algoritma Naive Bayes sebesar 66,66% dan nilai akurasi model Naive Bayes dengan seleksi atribut menggunakan Algoritma Genetika dan Bagging sebesar 72,02%. Berdasarkan nilai tersebut diperoleh selisih akurasi sebesar 5,36%.

V. SIMPULAN

Dari hasil penelitian yang telah dilakukan untuk akurasi algoritma klasifikasi naive bayes sebesar 66,66%, sedangkan untuk akurasi algoritma klasifikasi dengan Algoritma Genetik dan Bagging sebesar 72,02% dengan selisih akurasi sebesar 5,36%. Sehingga dapat

disimpulkan bahwa penerapan teknik optimasi Algoritma Genetik dan Bagging mampu menyeleksi atribut pada Algoritma Naive Bayes, dan dapat menghasilkan tingkat akurasi diagnosis penyakit liver yang lebih baik dibandingkan hanya menggunakan metode individual yaitu algoritma naive bayes.

DAFTAR PUSTAKA

- Abrianto, P. M. C. (2018). PENERAPAN METODE K-MEANS CLUSTERING UNTUK PENGELOMPOKAN PASIEN PENYAKIT LIVER. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 2(2), 247–255.
- Ayudhitama, A. P., & Pujianto, U. (2020). ANALISA 4 ALGORITMA DALAM KLASIFIKASI PENYAKIT LIVER MENGGUNAKAN RAPIDMINER. *JIP (Jurnal Informatika Polinema)*, 6(2), 1–9.
- Dhamodharan, S. (2014). Liver Disease Prediction Using Bayesian Classification. *4th National Conference on Advanced Computing, Applications & Technologies, May*, 1–3. <https://ijact.in/index.php/ijact/article/view/443/378>
- Ghosh, S. R., & Waheed, S. (2017). Analysis of classification algorithms for liver disease diagnosis. *Journal of Science Technology and Environment Informatics*, 5(1), 360–370. <https://doi.org/10.18801/jstei.050117.38>
- Krisnandi, K., & Agung, H. (2017). Implementasi Algoritma Genetika Untuk Memprediksi Waktu Dan Biaya Pengerjaan Proyek Konstruksi. *Jurnal Ilmiah FIFO*, 9(2), 90. <https://doi.org/10.22441/fifo.2017.v9i2.001>
- Nahar, N., & Ara, F. (2018). LIVER DISEASE PREDICTION BY USING DIFFERENT DECISION TREE TECHNIQUES. *International Journal of Data Mining & Knowledge Management Process*, 8(2), 01–09. <https://doi.org/10.5121/ijdkp.2018.8201>
- Noviandi. (2018). Implementasi Algoritma Decision Tree C4.5 Untuk Prediksi Penyakit Diabetes. *Inohim*, 6(1), 1–5.
- Pakhale, H., & Xaxa, D. K. (2016). A Survey on Diagnosis of Liver Disease Classification. *International Journal of Engineering and Techniques*, 2(3), 132–138. <http://www.ijetjournal.org>
- Priya, M. B., Juliet, P. L., & Tamilselvi, P. R. (2018). Performance Analysis of Liver Disease Prediction Using Machine Learning Algorithms. *International Research Journal of Engineering and Technology(IRJET)*, 5(1), 206–211. www.irjet.net
- Pusporani, E., Qomariyah, S., & Irhamah. (2019). Klasifikasi Pasien Penderita Penyakit Liver dengan Pendekatan Machine Learning. *Inferensi*, 2(March), 25–32.
- Rafsanjani, R. G., Hidayat, N., & Dewi, R. K. (2018). Diagnosis Penyakit Hati Menggunakan Metode Naive Bayes Dan Certainty Factor. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer (J-PTIIK)*, 2(11), 4478–4482.
- Riyanto, U. (2018). ANALISIS PERBANDINGAN ALGORITMA NAIVE BAYES DAN SUPPORT VECTOR MACHINE DALAM MENGLASIFIKASIKAN JUMLAH PEMBACA ARTIKEL ONLINE. *Jurnal Teknik Informatika (JIKA) Universitas Muhammadiyah Tangerang*, 62–72.
- Saleh, A. (2015). Implementasi Metode Klasifikasi Naive Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga. *Creative Information Technology Journal*, 2(3), 207–217.
- Zulfikar, W. B., & Lukman, N. (2016). Perbandingan Naive Bayes Classifier Dengan Nearest Neighbor Untuk Identifikasi Penyakit Mata. *Jurnal Online Informatika*, 1(2), 82–86. <https://doi.org/10.15575/join.v1i2.33>