

**ANALISIS KOMPARATIF *PARTICLE SWARM OPTIMIZATION* (PSO) DAN *GENETIKA ALGORITMA* (GA) UNTUK MENINGKATKAN ALGORITMA *NAÏVE BAYES* DALAM MEMREDIKSI PENYAKIT HEPATITIS**

Suherman<sup>1)</sup>, Meliana Firdhaus<sup>2)</sup>

Program Studi Teknik Informatika Sekolah Tinggi Teknologi Pelita Bangsa  
[suherman@pelitabangsa.ac.id](mailto:suherman@pelitabangsa.ac.id)

Disetujui, 28 Desember 2018

**Abstraksi**

Penelitian ini bertujuan untuk memprediksi jenis penyakit hepatitis, dengan cara memprediksi jenis tipe hepatitis yang diderita. Hasil penelitian menggunakan algoritma *naïve bayes* berbasis *Partile Swarm Optimization* (PSO) dan *naïve bayes* berbasis *Genetika Algoritma* (GA). Hasil akurasi *naïve bayes* tanpa optimasi sebesar 83,08% dan nilai AUC 0,826, menggunakan optimasi PSO sebesar 84,50% dan nilai AUC 0,883, sedangkan *naïve bayes* menggunakan optimasi GA sebesar 85,79% dan nilai AUC 0,901. Dari penelitian ini bahwa *naïve bayes* menggunakan optimasi GA mendapat nilai akurasi tertinggi dengan peningkatan 2,71% dalam memprediksi penyakit hepatitis.

**Kata kunci:** Hepatitis, Data Mining, Algoritma *naïve bayes*, *Particle Swarm Optimization*, Genetika Algoritma

**Abstract**

*The aims of this study is to predict the types of hepatitis, by predicting what types of hepatitis suffered. The results of the study used the Partile Swarm Optimization (PSO) based naïve bayes algorithm and the naïve bayes algorithm based on the Genetic Algorithm (GA). The results of the accuracy of naïve Bayes without optimization are 83.08% and the AUC value is 0.826, using PSO optimization of 84.50% and AUC value of 0.883, while naïve Bayes uses GA optimization of 85.79% and an AUC value of 0.901. From this study that naïve Bayes using GA optimization got the highest accuracy value with an increase of 2.71% in predicting hepatitis.*

**Keywords:** Hepatitis, Data Mining, Naïve Bayes Algorithm, Particle Swarm Optimization, Genetic Algorithm

**1. Pendahuluan**

Penyakit Hepatitis adalah termasuk jenis penyakit menular, penyebab utamanya adalah pola hidup yang tidak sehat, contohnya kebiasaan minum *alcohol*, merokok, serta zat racun atau obat-obatan tertentu.

Menurut penelitian Septiyani tahun 2017, penyakit ini dimulai dari yang tidak bergejala, bergejala dan sembuh sendiri, menjadi kronis dan yang paling berbahaya adalah berkembang menjadi gagal ginjal. mendorong peneliti untuk mengembangkan suatu teknologi yang mampu diterapkan pada *computer* dan dapat dipergunakan dengan mudah, karen belum adanya sistem mendeteksi hepatitis sehingga penderita tidak menyadari secara dini bahwa dirinya terkena penyakit hepatitis.

Seiring dengan berjalannya waktu teknologi informasi berkembang dan penyakit dapat diteteksi menggunakan teknologi yang lebih canggih, oleh karena itu dengan kehadiran teknologi baru seperti data *mining* menarik banyak perhatian dalam dunia sistem informasi dan dunia medis, seperti Naive Bayes, yaitu sebuah klasifikasi probabilistic sederhana yang menghitung sekumpulan data dengan menjumlahkan frekuensi dan kombinasi nilai dari data set yang diberikan, *naïve bayes* sendiri merupakan metode yang hanya membutuhkan jumlah data pelatihan atau yang sering kita sebut dengan *training data*. Belum adanya optimasi yang efektif untuk meningkatkan nilai akurasi penyakit hepatitis. Maka dari itu peneliti akan memprediksi penyakit Hepatitis menggunakan algoritma *naïve bayes* dengan dua optimasi yaitu *Particle Swarm Optimization* (PSO) dan *Genetika Algoritma* (GA) untuk mencari nilai akurasi yang tinggi sehingga dapat dapat digunakan secara efektif.

**2. Tinjauan Studi**

**2.1. Data Mining**

*Data mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengidentifikasi informasi dari berbagai database besar. Data mining menurut

Ridwan, Suyono, & Sarosa (2013) juga disebut sebagai *Knowledge Discovery in Database (KDD)*.

Data mining menurut Alimuddin (2017) adalah proses menemukan teknik baru yang bermakna, pola, tren, dengan memilah-milah sejumlah data yang tersimpan dalam *repository*, menggunakan teknologi penerapan pola serta teknik statistik dan matematika. Sedangkan versi Jananto (2013) alasan utama mengapa data mining sangat di perlukan adalah karena adanya data yang dapat digunakan untuk menghasilkan informasi dan pengetahuan yang berguna.

Data merupakan teknologi baru menurut Wibiwo, Stenrianto, & Ninggolan(2015) yang sangat berguna untuk membantu perusahaan menemukan informasi yang sangat penting dari gudang data. Dapat disimpulkan bahwa data mining merupakan proses analisis dengan menggunakan basis data dalam menentukan suatu pola pengetahuan baru dan akan bermanfaat pada masa yang akan datang.

## 2.2. Naive Bayes

*Naive bayes* adalah salah satu metode klasifikasi probabilistik paling sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan. Definisi lain dari *naive bayes* merupakan pengklasifikasian dengan metode probabilitas dan statistik dikemukakan oleh ilmuwan Inggris, Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman dimasa sebelumnya.

*Naive bayes* bentuk sederhana dari bayesian menurut Juwita (2018) yang merupakan jaringan untuk klasifikasi, sebuah jaringan yang sangat diarahkan sebagai distribusi probabilitas gabungan lebih dari satu set dikrit.

*Naive bayes* didasarkan padaasumsi penyederhanaan bahwa nilai atribut secara kondisional saling bebas jika diberikan nilai output. Dengan kata lain probabilitas mengamati secara bersama adalah produk dari probabilitas individu.

Rumus dari teorema bayes sebagai berikut :

$$P(H|X) = \frac{p(X|H)P(H)}{P(X)}$$

Keterangan :

- X : Data dengan kelas yang belum diketahui  
H : Hipotesis data merupakan suatu kelas yang spesifik

## 2.3. Algoritma Genetika (AG)

*Genetika Algoritma (GA)* merupakan Salim (2017) adalah teknik yang didasarkan pada prinsip genetik dan seleksi alam. Dalam algoritma genetika populasi terbentuk dari banyak individu yang berkembang sesuai aturan seleksi spesifik dengan memaksimalkan fitness. Algoritma ini juga digunakan untuk mendapatkan nilai global optimum dengan cara melakukan perulangan pada konsep evolusi Darwin. Sedangkan menurut Widodo & Mahmudy (2010) adalah merupakan salah satu metode *heuristic* yang merupakan cabang dari *evolutionary algoritma* yaitu salah satu teknik untuk memecahkan masalah optimasi yang rumit. Algoritma ini sendiri berkembang seiring kemajuan teknologi maka dari itu algoritma ini juga banyak digunakan dalam bidang fisika, biologi, ekonomi dan lainnya.

Algoritma genetika menurut Juwita (2018) memiliki kelebihan yaitu mampu mengatasi berbagai jenis objektif dari berbagai kromosom. Algoritma ini juga mudah untuk dikombinasikan dengan metode lain, dapat menangani permasalahan yang kompleks.

Dalam algoritma genetika hasil penelitian Mutakhirroh, Saptono, Hasanah, & Wiryadinata (2007) diperlukan beberapa proses untuk menentukan jalur terpendek yaitu :

1. Proses Pengkodean ( Encoding)
2. Proses Seleksi
3. Proses Rekomendasi
4. Proses Mutasi

Rumus Algoritma Genetika untuk mencari nilai maksimal :

$$F(x) = e^{-2x} \cdot \sin(3x)$$

Adapun Tahapan Genetika Algoritma :

1. Membentuk populasi awal terdiri dari beberapa kromosom yang didalamnya memuat gen kromosom pada algoritma genetika yang digunakan untuk menunjukkan kandidat sekelompok gen yang dapat digunakan sebagai solusi permasalahan.

2. Masing-masing kromosom dalam populasi dievaluasi kemampuannya dengan menggunakan fungsi fitness, fungsi fitness berupa kesalahan klasifikasi.

#### 2.4. Particle Swarm Optimization (PSO)

*Particle swarm optimization* (PSO) menurut Ramanda (2015) adalah teknik optimasi berbasis populasi yang dikembangkan oleh Eberhart dan Kennedy pada tahun 1995, yang terinspirasi oleh perilaku social kawanan burung atau ikan. Dapat diasumsikan juga sebagai kelompok burung yang mencari makanan disuatu daerah.

Metode optimasi yang didasarkan pada swarm intelligence menurut Cholissodin, (2016) ini disebut juga *algoritma behaviorally inspired* sebagai alternative dari algoritma genetika, yang sering disebut evolution-based procedures. Dalam konteks optimasi ini mempunyai ukuran tertentu atau tetapan dengan setiap partikel posisi awalnya terletak disuatu lokasi yang acak dalam ruang multidimensi

Mempunyai kelebihan dengan konsep yang sederhana, menurut Habiburrahman, Hakim, Cholissodin, & Widodo, (2017) dinyatakan bahwa PSO mudah untuk diimplementasikan dan efisien dalam perhitungan berbeda dengan algoritma yang lain seperti algoritma matematika dan *optimasi heuristic*.

Algoritma ini memiliki tiga kompoen utama yaitu : partikel, komponen kongitif, dan komponen social serta kecepatan partikel. Dan setiap partikel mempresentasika solusi penyelesaian. Terdapat beberapa istilah dalam algoritma PSO berdasarkan penelitian Erviana & Handoko, (2012) diantaranya adalah sebagai berikut :

1. *Swarm*, merupakan jumlah partikel dalam populasi pada suatu algoritma.
2. Partikel, merupakan individu dalam suatu swarm yang merepresentikan solusi penyelesaian masalah
3. *Personal best* (pBest), merupakan posisi terbaik yang pernah di capai dengan membandingkan fitness pada posisi partikel sekarang dengan sebelumnya.
4. *Global best* (gBest), merupakan posisi terbaik partikel yang diperoleh dengan membandingkan nilai fitness terbaik dari keseluruhan partikel dalam swarm.
5. *Velocity*, merupakan kecepatan yang menggerakkan proses optimasi yang menentukan arah dimana suatu partikel diperlukan untuk memperbaiki posisi semula.
6. *Inertia weight*, merupakan parameter yang digunakan untuk mengontrol dampak dari adanya *velocity* yang diberikan oleh partikel

Rumusnya sebagai berikut :

$$V(t+1) = wV(t) + c_1r_1(P_{best}(t) - X(t)) + c_2r_2(g_{best}(t) - X(t)); k = 1, 2, \dots, P \quad X(t+1) = X(t) + V(t+1)$$

Di mana:

X	: posisi partikel
V	: kecepatan partikel
W	: bobot inersia
C <sub>1</sub> , C <sub>2</sub>	: koefisien akselerasi
P	: jumlah partikel dalam <i>swarm</i>
r <sub>1</sub> , r <sub>2</sub>	: nilai acak pada rentang

#### 2.5. Hepatitis

Hepatitis adalah salah satu penyakit yang mematikan, ditahun 2007 sendiri penyakit ini masuk kedalam deretan penyakit yang berbahaya. Hepatitis sendiri adalah penyakit yang menyerang organ hati yang disebabkan oleh infeksi virus karena kebiasaan mengkonsumsi *alcohol*, serta mengkonsumsi obat-obatan tertentu.

Penyakit ini menurut penelitian Trisulistyo & Noviyanto (2014) akan berkembang dan menjadi sirosis atau pengerasan *hepar*, sebagian besar penyakit ini tidak mudah dideteksi karena tidak menimbulkan gejala dan akan mulai terlihat gejala sekitar 10-30 tahun kemudian saat infeksi sudah parah.

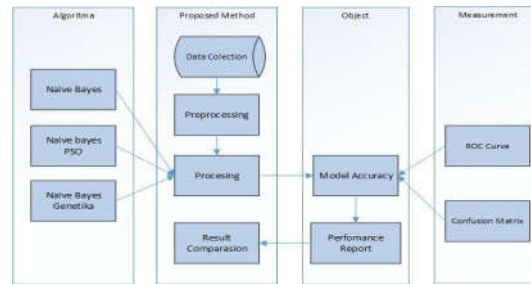
Hepatitis digolongkan menjadi beberapa macam yaitu sebagai berikut :

1. Hepatitis A (HAV) Menyebabkan infeksi hati menular, dapat menularkan melalui konsumsi air atau makanan dan berhubungan intim tanpa pengaman.
2. Hepatitis B (HBV) Virus yang ditularkan melalui kontak dengan darah, air mani dan cairan tubuh lainnya

3. Hepatitis C (HCV) Virus ini sering ditularkan melalui jarum suntik narkoba yang digunakan bersama-sama
4. Hepatitis D (HD)
5. Hepatitis E (HEV) Memasak daging setengah matang bisa menimbulkan virus hepatitis E.

### 3. Kerangka Konsep

Model yang diberikan adalah dimulai dari pengumpulan dataset yang akan dijadikan bahan uji coba. Sebelum memasuki tahap proses data sebelumnya dilakukan tahapan preprocessing dataset mulai dari pengecekan *missing values* hingga *cleasing* hal ini dilakukan guna dataset yang akan diuji benar-benar siap dan tidak terdapat data *noise*. Setelah melakukan *preprocessing* maka penelitian pun siap dimulai, seperti pada gambar 2.1 dibawah ini :



Gambar 1. Kerangka Konsep

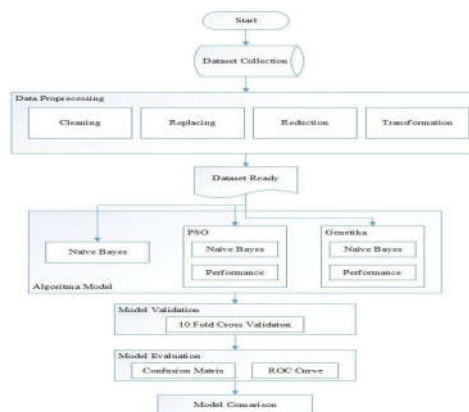
### 4. Desain Penelitian/ Metodologi

Penelitian ini dimulai dengan mengambil sebanyak 20 atribut data set, yaitu meliputi :

<i>Age</i>	11	<i>Spiders</i>
<i>Sex</i>	12	<i>Ascites</i>
<i>Streroid</i>	13	<i>Varices</i>
<i>Antivirals</i>	14	<i>Bilirubin</i>
<i>Fantigue</i>	15	<i>Alk Phosphate</i>
<i>Melaisa</i>	16	<i>Sgot</i>
<i>Anorexia</i>	17	<i>Albumin</i>
<i>Liver Big</i>	18	<i>Protime</i>
<i>Liver Firm</i>	19	<i>Histology</i>
<i>Spleen Palpable</i>	20	<i>Class</i>

Proses pengambilan data skunder berupa data publik yang berasal dari : datset Hepatitis dari UCI *Machine Learning repository* <https://archive.ics.uci.edu/ml/datasets/hepatitis> dengan jumlah data sebanyak 155 data. Setelah data terkumpul, dikelompokkan menjadi atribut class, yaitu *Life* dan *Died* yaitu Jumlah nilai *Died* ada 32 data sedangkan *Life* sebanyak 123 data. Data tersebut diolah sbb :

1. Data Cleaning, yaitu proses pembersihan data
2. Data Reduction, yaitu mengurangi jumlah data yg di hasilkan data asli
3. DataTransformation.
4. Pengolahan dengan k-folds cross validation
5. Proses evaluasi kinerja dengan teknik *confusion matrix* dan *Reciver Operating Characteristic (ROC Curve)*.



Gambar 2. Tahapan proses olah data

Adapun tahapan pengolahan data seperti digambar 2. dibawah ini :

## 5. Hasil Penelitian dan Pengujian

### 5.1. Hasil Penelitian

#### 5.1.1. Data Cleaning

Data cleaning merupakan penyaringan data atau pembersihan data yang tidak dibutuhkan karena adanya kesalahan data untuk menghasilkan data yang berkualitas. Pembersihan data dilakukan jika ada kesalahan berupa *Missing* data, yaitu data yang hilang penanganan data yang hilang dapat dilakukan dengan mengabaikan tuple atau baris data yang memiliki nilai hilang, mengisi nilai secara manual, menggunakan mean/median, mengganti dengan nilai terbanyak atau dengan menerapkan nilai prediksi. Berikut tabel 4. Data Missing dari 15 atribut yang diproses data cleaning dengan menggunakan nilai *mean*.

**Tabel 1.** Data Missing

No	Atribut	Jumlah <i>Missing Values</i>
1	Steroid	1
2	Fatigue	1
3	Melaisa	1
4	Anorexia	1
5	Liver big	10
6	Liver firm	11
7	Spleen palpable	5
8	Spiders	5
9	Ascites	5
10	Varices	5
11	Bilirubin	6
12	Alk Phosphate	29
13	Sgot	4
14	Albumin	16
15	Protine	68

#### 5.1.2. Data Reduction

Data reduction merupakan tahapan data preprocessing dengan mengurangi jumlah data yang dihasilkan dari data asli. Berkurangnya atribut tidak mempengaruhi hasil representasi akhir karena akan menunjukkan hasil yang sama. Tabel.. dibawah ini adalah merupakan tabel atribut yang digunakan dalam proses mining :

**Tabel 2.** Data Reduction

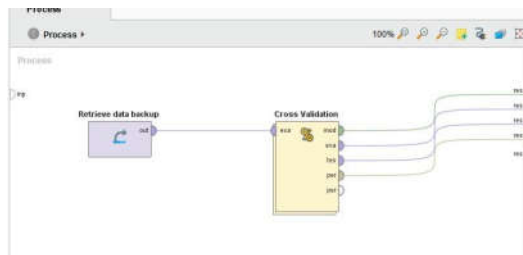
No	Atribut	Type	Keterangan
1	Age	<i>Numeric</i>	Umur
2	Sex	<i>Category</i>	Jenis Kelamin
3	Steroid	<i>Category</i>	Senyawa organik sterol
4	Antivirals	<i>Category</i>	Obat antivirus
5	Fatigue	<i>Category</i>	Suatu kondisi kelelahan yang dimiliki seseorang
6	Melaisa	<i>Category</i>	Seperti fatigue, penyakit ini juga disebut sebagai penyakit kelelahan
7	Anorexia	<i>Category</i>	Gangguan terhadap makanan
8	Liver big	<i>Category</i>	Penyakit yang merusak sistem hati besar
9	Liver Firm	<i>Category</i>	Tes Kekencangan Hati
10	Spleen palpable	<i>Category</i>	Pembesaran Limpa
11	Spiders	<i>Category</i>	Pembulu darah abnormal dekat permukaan kulit
12	Ascites	<i>Category</i>	Penumpukan cairan dirongga perut
13	Varices	<i>Category</i>	Pembengkakan dan pelebaran pembuluh darah vena
14	Bilirubin	<i>Category</i>	Pigmen berwarna kuning kecoklatan ditemukan didalam

			empedu
15	Alk phosphate	Numeric	Mengukur tingkat enzim fosfatase alkali dalam darah
16	Sgot	Numeric	Enzim yang biasa ditemukan pada hati,jantung,ginjal hingga otak
17	Albumin	Numeric	Protein utama yang terdapat dalam darah manusia
18	Prottime	Numeric	Waktu Protrombin
19	Histology	Category	struktur jaringan
20	Class	Category	Lebel LIFE atau DIE

## 5.2. Hasil Pengujian

### 5.2.1. Pengujian 1 dengan Naïve Bayes

Pengujian pertama ini akan menggunakan satu algoritma yaitu *naïve bayes* tanpa menggunakan optimasi dalam melakukan klasifikasi data. Berikut ini model pengujian pertama :



Gambar 3. Model pengujian pertama

### 5.2.2. Pengujian 2 dengan sub cross validation

Dalam sub proses *cross validation* terdapat dua bagian dimana ada training dan juga testing, dimana pada bagian traning terdapat algoritma *naïve bayes* hal tersebut agar dataset dibuat modelnya menggunakan algoritma *naïve bayes* dimana pada *cross validation* dengan nilai  $k = 10$  folds sehingga membagi data menjadi 10:90 yakni 10% dijadiandata testing dan 90% data training. Pada data testing terdapat dua fitur yakni *Apply model* digunakan untuk menerapkan model data yang sudah dilatih sebelumnya menggunakan dataset.

### 5.2.3. Evaluasi Kerja Naive bayes

mengevaluasi hasil kerja algoritma *naïve bayes* dengan parameter *confussion matrix* (*accuracy*, *recall*, *precision*) dan *Receiver Oprating Characteristic* (*ROC Curve*) :

#### 1. Hasil Accuracy

Hasil *accuracy* menggunakan algoritma klasifikasi *naïve bayes* dalam penguji pertama dengan *cross validation* yaitu sebesar 83.08%, lihat gambar 4

accuracy: 83.08% +/- 9.88% (micro average: 83.23%)			
	true LIVE	true DIE	class precision
pred. LIVE	108	11	90.76%
pred. DIE	15	21	58.33%
class recall	87.80%	65.62%	

Gambar 4. model pengujian accuracy tahap pertama

#### 2. Hasil Recall

Hasil *recall* pada pengujian pertama menggunakan algoritma klasifikasi *naïve bayes* adalah 66.67%, dengan 87.80% nilai *class live* dan 65.62% untuk nilai *class die*

recall: 66.67% +/- 27.13% (micro average: 65.62%) (positive class: DIE)			
	true LIVE	true DIE	class precision
pred. LIVE	108	11	90.76%
pred. DIE	15	21	58.33%
class recall	87.80%	65.62%	

Gambar 5. Hasil uji Recall tahap pertama

#### 3. Hasil Precision

Hasil *precision* pada pengujian pertama menggunakan algoritma klasifikasi *naïve bayes* adalah 61.33% , dengan 90.76% *class live* sedangkan 58.33% untuk *class die*

Table View Plot View

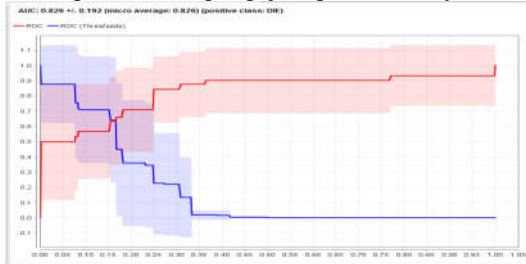
precision: 61.33% +/- 23.33% (micro average: 58.33%) (positive class: DIE)

	true LIVE	true DIE	class precision
pred. LIVE	108	11	90.76%
pred. DIE	15	21	58.33%
class recall	87.60%	65.62%	

**Gambar 6.** Hasil Uji Precision tahap pertama.

4. AUC (Area Under Curve)

AUC yang dihasilkan pada model pengujian pertama ini yaitu sebesar 0.826



**Gambar 7.** Hasil Uji AUC tahap pertama

**5.2.4. Pengujian 2 Naïve Bayes**

Hasil Pengujian 2 Naïve Bayes sebagai berikut :

1. Hasil Accuracy

Hasil *accuracy* menggunakan algoritma klasifikasi *naïve bayes* dalam pengujian kedua dengan *cross validation* yaitu sebesar 84,5 %

2. Hasil Recall

Hasil *recall* pada pengujian kedua menggunakan algoritma klasifikasi *naïve bayes* adalah PSO 67,5 %, dengan 89,43 % nilai *class live* dan 65.62% untuk nilai *class die*

3. Hasil Precision

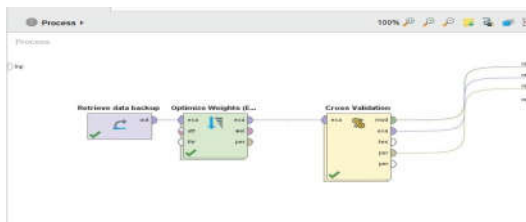
Hasil *precision* pada pengujian kedua menggunakan algoritma klasifikasi *naïve bayes* adalah PSO 66,5 % , dengan 90.91% *class live* sedangkan 61,76% untuk *class die*

4. AUC (Area Under Curve)

AUC yang dihasilkan pada model pengujian pertama ini yaitu sebesar 0.883

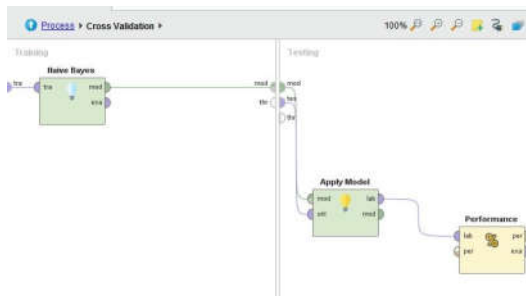
**5.2.5. Pengujian 3 Naïve Bayes – GA**

Pada pengujian ketiga model awal dari pengujian ini sama seperti pengujian pertama dan kedua yaitu menggunakan *naïve bayes* namun menggunakan penerapan genetika. Berikut adalah model pengujian ketiga pada gambar 8.



**Gambar 8.** Model Cross Validation Genetika Algoritma

Pada gambar 9 menggambarkan rangkaian model pengujian ketiga dimana pada *retrieve data backup* nantinya disesuaikan dengan dataset yang tersedia. Untuk lebih jelasnya bisa di lihat pada gambar 9 berikut dibawah ini:



**Gambar 9.** Retrieve Data Back up.

Hasil Pengujian 2 Naïve Bayes sebagai berikut :

1. Hasil Accuracy  
Hasil *accuracy* menggunakan algoritma klasifikasi *naïve bayes dengan GA* dalam pengujian ketiga yaitu sebesar 85,79 %
2. Hasil Recall  
Hasil *recall* pada pengujian pertama menggunakan algoritma dengan GA adalah 70, 83% dengan 89,43 % nilai *class live* dan 71,8% untuk nilai *class die*
3. Hasil Precision  
Hasil *precision* pada pengujian pertama menggunakan algoritma *naïve bayes dengan GA* 67 % , dengan 92,44% *class live* sedangkan 63,89% untuk *class die*
4. AUC (Area Under Curve)  
AUC yang dihasilkan pada model pengujian pertama ini yaitu sebesar 0.901

#### 5.2.4. Pembahasan

Penelitian ini dilakukan sebanyak tiga kali, yaitu dalam pengujian pertama hanya menggunakan algoritma *naïve bayes*, pengujian kedua menggunakan algoritma *naïve bayes* dengan metode optimasi PSO dan pengujian ketiga menggunakan algoritma *naïve bayes* dengan metode optimasi GA yang menghasilkan nilai *accuracy*, *recall*, *precision*, dan UAC. Berikut tabel hasil dari ketiga penelitian yang telah dilakukan :

**Tabel 3.** Hasil Penelitian

	Pengujian Pertama	Pengujian Kedua	Pengujian Ketiga
	Naïve Bayes	Naïve Bayes + PSO	Naïve Bayes + GA
Accuracy	83.08%	84.50%	85.79%
Recall	66.67%	67.50%	70.83%
Precesion	61.33%	66.50%	70.00%
AUC	0.826	0.883	0.901

Berdasarkan hasil yang diperoleh tabel 3. membuktikan bahwa adanya peningkatan nilai antara *accuracy*, *recall*, *precision*, dan AUC. Pada algoritma *naïve bayes* menggunakan optimasi lebih meningkat dibandingkan dengan *naïve bayes* tanpa optimasi berikut adalah gambar 10. peningkatan dibawah ini :



**Gambar 10.** Grafik Peningkatan

Berdasarkan gambar 10 menunjukkan hasil peningkatan pada grafik *accuracy* yaitu pengujian kedua meningkat sebanyak 1.42% sedangkan pengujian ke tiga meningkat sebanyak 2.71% . Pada grafik *recall* yaitu pengujian kedua meningkat sebanyak 0.83% sedangkan pengujian ketiga sebanyak 4.16%. Namun nilai pada *precision* meningkat sangat tinggi yaitu pada pengujian ke dua 5.17% sedangkan pada pengujian ke tiga sebanyak 8.67%. Untuk grafik yang terakhir

Penelitian yang telah dilakukan penulis bahwa nilai akurasi *naïve bayes* tanpa optimasi menghasilkan 83.08%, namun nilai akan lebih besar dengan menggunakan dua optimasi dibandingkan dengan tidak menggunakan optimasi, nilai akurasi algoritma *naïve bayes* dengan PSO 84.50% peningkatan akurasi yaitu sebesar 1.42%, sedangkan *naïve bayes* menggunakan optimasi GA nilai akurasi sebesar 85.79% dengan nilai peningkatan sebesar 2.71%.

## 5. Kesimpulan

Berdasarkan pada hasil penelitian yang telah dilakukan maka dapat disimpulkan sebagai berikut :

1. Pengaruh terhadap akurasi PSO dan GA pada meningkatnya algoritma *naïve bayes* ialah pada tahap *preprocessing* dimana pada tahapan tersebut hanya berfokus pada meningkatkan nilai akurasi.
2. Berdasarkan penelitian *naïve bayes* dengan nilai akurasi 83.08%, *naïve bayes* menggunakan optimasi PSO nilai akurasi sebesar 84.50% dan *naïve bayes* dengan optimasi GA sebesar 85.79%, maka nilai perbandingan 3 pengujian tersebut menunjukkan nilai GA lebih tinggi nilai



akurasinya di bandingkan dengan pengujian yang lain.

3. Penerapan optimasi genetika algoritma dapat meningkatkan akurasi pada algoritma *naive bayes* dalam mendeteksi penyakit hepatitis. Hasil akurasi diperoleh yaitu 85.79% dan meningkatkan sebesar 2.71% dari hasil tunggal *Naive Bayes* sebesar 83.08%.

#### Daftar Pustaka

- Alimuddin, Sadali, M., & Wasali, M. (2017). P. Jurnal Informatika Hamzanwadi, 2(1), 1–19.
- Cholissodin, I. (2016). *Swarm Intelligence (Teori & Case Study)*. Cholissodin, & Riyandani. (2016). swarm intelligence (Teori & case study).
- Erviana, M., & Handoko, S. (2012). *Rugi Daya Menggunakan Particle Swarm*, 1– 8.
- Habiburrahman, S., Hakim, F., Cholissodin, I., & Widodo, A. W. (2017). Seleksi Fitur Dengan Particle Swarm Optimization Untuk Pengenalan Pola Wajah Menggunakan Naive Bayes ( Studi Kasus Pada Mahasiswa Universitas Brawijaya Fakultas Ilmu Komputer Gedung A ). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 1(10), 1045–1057.
- Jananto, A. (2013). Algoritma Naive Bayes untuk Mencari Perkiraan Waktu Studi Mahasiswa. *Jurnal Teknologi Informasi DINAMIK*, 18(1), 9–16.
- Josi, A. (2017). Implementasi Algoritma Genetika Pada Aplikasi Penjadwalan Perkuliahan Berbasis Web Dengan Mengadopsi Model Waterfall. *Jurnal Pengembangan Ti*, 2(2), 77–83.
- Juwita, R. (2018). Klasifikasi Kelas Risiko Pasien Pneumonia Menggunakan Regresi Logika Ordinal, Hybrid Regresi Logistik Ordinal - Algoritma Genetika Dan Naive Bayes Classification.
- Mutakhirah, I., Saptono, F., Hasanah, N., & Wiryadinata, R. (2007). Pemanfaatan Metode Heuristik Dalam Pencarian Jalur Terpendek Dengan Algoritma Semut dan Algoritma Genetika. *Seminar Nasional Aplikasi Teknolohi Informasi, 2007(Snati)*.
- Pengestu, A., Arie, B., Jovi, E., Dwi, rina s, & Prasetyo, H. (2017). Optimizing Neural Network Classifier For Diabetes Data Using Metaheuristik Algorithms. *Jurnal Imiah Teknologi Dan Informasi*, 6(2), 85–91.
- Ramanda, K. (2015). *Penerapan Particle Swarm Optimization Sebagai Seleksi Fitur Prediksi Kelahiran*, I(2), 178–183.
- Ridwan, M., Suyono, H., & Sarosa, M. (2013). *Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier*, 7(1), 59–64.
- Saleh, A. (2015). Implementasi Metode Klasifikasi Naive Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga. *Citec Journal*, 2(3), 207.
- Salim, A. (2017). *Pengoptimalan Naive Bayes dan Regresi*.
- Septiani, wisti dwi. (2017). Koparasi Metode Klasifikasi Data Mining Algoritma C4.5 dan Naive Bayes Untuk Prediksi Penyakit Hepatitis. *Jurnal Pilar Nusa Mandiri*, 13(1), 76–77.
- Trisulistyo, F., & Noviyanto, F. (2014). Sistem Pakar Untuk Mendiagnosa Penyakit Hepatitis Dengan Metode Fuzzy Tsukamoto. *Jurnal Sarjana Teknik Informatika*, 2, 85–91.
- Wahono, R. S. (2015). Penerapan Algoritma Genetika untuk Optimasi Parameter pada Support Vector Machine untuk Meningkatkan Prediksi Pemasaran Langsung. *Journal of Intelligent Systems*, 1(2), 115–119.
- Wibiwo, K., Stenrianto, & Ninggolan, K. (2015). Klasifikasi Gangguan Motorik Kasar Anak Menggunakan Naive Bayes Serta Optimasi dengan PSO dan Adaboost. *Jurnal Ilmu Pengetahuan Dan Teknologi Komputer*, 1(1), 1–10.
- Widiastuti, N. A., Santosa, S., & Supriyanto, C. (2014). Algoritma Klasifikasi Data Mining Naive Bayes Berbasis Particle Swarm. *Jurnal Pseudocode*, 1(1), 11– 14.
- Widodo, A. W., & Mahmudy, wayan firdaus. (2010). penerapan algoritma genetika pada sistem rekomendasi wisata kuliner. *Jurnal Ilmiah Cursor*, 5(4), 205–211.
- Widiastuti, N. A., Santosa, S., & Supriyanto, C. (2014). Algoritma Klasifikasi Data Mining Naive Bayes Berbasis Particle Swarm. *Jurnal Pseudocode*, 1(1), 11– 14.
- Widodo, A. W., & Mahmudy, wayan firdaus. (2010). penerapan algoritma genetika pada sistem rekomendasi wisata kuliner. *Jurnal Ilmiah Cursor*, 5(4), 205–211.