

# IMPLEMENTASI ALGORITMA NAÏVE BAYES BERBASIS PARTICLE SWARM OPTIMIZATION UNTUK MEMPREDIKSI PENYAKIT HEPATITIS

Hilda Farida Husniah<sup>a1</sup>, Toni Arifin<sup>a2</sup>

<sup>a</sup>Universitas Adhirajasa Reswara Sanjaya  
Jl Sekolah Internasional No 1-2, Antapani, Bandung, (022) 7100124  
<sup>1</sup>[hildahilda157@gmail.com](mailto:hildahilda157@gmail.com)

<sup>b</sup> Universitas Adhirajasa Reswara Sanjaya  
Jl Sekolah Internasional No 1-2, Antapani, Bandung, (022) 7100124  
<sup>2</sup>[toni.arifin@ars.ac.id](mailto:toni.arifin@ars.ac.id)

## Abstract

*Hepatitis disease is an inflammatory disease of the liver cells, caused by infections (viruses, bacteria, parasites), medicines (including traditional medicines), consuming alcohol, excessive fats and autoimmune diseases. The cause of hepatitis is often caused by Hepatitis B and C Virus. The Hepatitis prevalence in Indonesia in 2013 amounted to 1.2% increased twice compared to the year 2007 Riskesdas of 0.6%. East Nusa Tenggara is the province with the highest prevalence of Hepatitis in 2013 of 4.3%. Researchers are trying to make a breakthrough by making research for the prediction classification of Hepatitis patients with data mining technique. Naïve Bayes is a method used to predict the probability of the future based on past experience and proved to have a high level of accuracy and high speed of calculation. Particle Swarm Optimization is used to improve the accuracy of the method. The research aims to determine if the Naïve Bayes-based Particle Swarm Optimization method can improve the accuracy of the good. The results of using Naïve Bayes-based Particle Swarm Optimization has a confusion matrix accuracy of 91.90% and an AUC of 0.946 proved that has good results than Naïve Bayes has a confusion matrix accuracy of 88.52% and AUC 0.896.*

*Keywords: Data Mining, Hepatitis Disease Prediction, Naïve Bayes, Particle Swarm Optimization*

**Abstrak :** Penyakit Hepatitis merupakan penyakit peradangan pada sel-sel hati, yang disebabkan oleh infeksi (virus, bakteri, parasite), obat-obatan (termasuk obat tradisional), mengkonsumsi alkohol, lemak yang berlebihan dan penyakit autoimmune. Penyebab terjadinya Hepatitis adalah sering disebabkan oleh Virus Hepatitis B dan C. Prevalensi Hepatitis di Indonesia pada tahun 2013 sebesar 1,2% meningkat dua kali dibandingkan Riskesdas tahun 2007 yang sebesar 0,6%. Nusa Tenggara Timur merupakan provinsi dengan prevalensi Hepatitis tertinggi pada tahun 2013 yaitu sebesar 4,3%. Para peneliti berusaha membuat terobosan dengan membuat penelitian untuk klasifikasi prediksi pasien Hepatitis dengan teknik data mining. Naïve bayes merupakan metode yang digunakan untuk memprediksi probabilitas dimasa depan berdasarkan pengalaman dimasa lalu dan terbukti memiliki tingkat akurasi tinggi dan kecepatan yang tinggi dalam perhitungannya. *Particle Swarm Optimization* digunakan untuk meningkatkan akurasi dari metode. Penelitian ini bertujuan untuk mengetahui apakah metode Naïve Bayes berbasis *Particle Swarm Optimization* dapat meningkatkan akurasi yang baik. Hasil penelitian menggunakan Naïve Bayes berbasis *Particle Swarm Optimization* memiliki akurasi *confusion matrix* sebesar 91.90% dan AUC sebesar 0.946 terbukti bahwa memiliki hasil yang bagus dibanding Naïve Bayes memiliki akurasi *confusion matrix* sebesar 88.52% dan AUC 0.896.

**Kata Kunci:** Data Mining, Prediksi Penyakit Hepatitis, Naïve Bayes, Particle Swarm Optimization

## 1. Pendahuluan

Penyakit Hepatitis merupakan penyakit peradangan pada sel-sel hati, yang disebabkan oleh infeksi (virus, bakteri, parasite), obat-obatan (termasuk obat tradisional), mengkonsumsi alkohol, lemak yang berlebihan dan penyakit autoimmune (Kemenkes, 2014). Penyebab terjadinya Hepatitis adalah sering disebabkan oleh Virus Hepatitis B dan C. Persentase penderita Hepatitis di Indonesia pada tahun 2013 sebesar 1,2% meningkat dua kali dibandingkan Risesdas tahun 2007 yang sebesar 0,6%. Provinsi Nusa Tenggara Timur merupakan provinsi dengan persentase penderita Hepatitis tertinggi pada tahun 2013 yaitu sebesar 4,3%. Berdasarkan kuintil indeks kepemilikan (yang menggambarkan status ekonomi), kelompok kuintil indeks kepemilikan terbawah menempati prevalensi Hepatitis tertinggi dibandingkan dengan kelompok lainnya. Prevalensi semakin meningkat pada penduduk berusia di atas 15 tahun. Jenis Hepatitis yang banyak menginfeksi penduduk Indonesia adalah Hepatitis B (21,8%), Hepatitis A (19,3%) dan Hepatitis C (2,5%) (Kemenkes, 2017).

Seiring dengan banyaknya pasien penderita penyakit hepatitis sehingga banyak peneliti yang meneliti tentang penyakit hepatitis, Para peneliti berusaha membuat terobosan dengan membuat penelitian untuk klasifikasi prediksi pasien Hepatitis dengan teknik data mining. Data mining adalah proses mencari pola atau menganalisa data yang sudah ada dengan menggunakan teknik atau metode tertentu. Teknik, metode atau algoritma data mining sangat bervariasi (Muzakir & Wulandari, 2016). *Naïve Bayes Classifier* sering disebut juga dengan *Bayesian Classification* merupakan metode pengklasifikasian statistik yang didasarkan pada teorema bayes yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu kelas. *Bayesian Classification* terbukti memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam database yang besar (Muhamad et al., 2017).

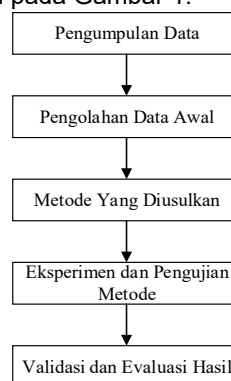
Seperti penelitian yang dilakukan oleh Novianti (2019), melakukan penelitian terhadap penyakit hepatitis menggunakan *Naïve Bayes* dengan akurasi sebesar 76.77%, dengan hasil *Class Precision* yang tertinggi sebesar 98.88% untuk prediksi "Life", dan *Class Recall* sebesar 96.88% untuk Prediksi "Die". Kemudian penelitian yang dilakukan oleh, Oktaviani, Ramadhani, Laksana, & Amalia (2018), melakukan perbandingan metode C.45 dengan *Support Vector Machine* (SVM), dari penelitian didapatkan hasil akurasi C4.5 sebesar 80,6452% dan SVM dengan akurasi 80,3279%.

Ramdhani, (2016) melakukan penelitian menggunakan *Particle Swarm Optimization* (PSO) dengan menggunakan algoritma C.45 untuk seleksi atribut dalam meningkatkan akurasi prediksi penyakit hepatitis. Hasil penelitian untuk nilai akurasi algoritma C4.5 senilai 79,33%, akurasi Optimasi algoritma C4.5 menggunakan PSO sebesar 85,00% sehingga memiliki hasil akurasi lebih tinggi dari algoritma C.45. (Arifin & Ariesta, 2019)

## 2. Metode Penelitian

### 2.1. Perancangan Penelitian

Menurut Dawson dalam (Saifudin, 2018) mengemukakan bahwa "Metode penelitian yang digunakan adalah eksperimen, penelitian eksperimen memuat investigasi, hubungan sebab-akibat menggunakan pengujian yang dikontrol sendiri. Eksperimen biasanya dilakukan dalam pengembangan, evaluasi dan pemecahan masalah proyek. Maka pada penelitian ini dilakukan dengan mengikuti tahapan penelitian pada Gambar 1.



Gambar 1. Tahapan Penelitian

**2.2. Pengumpulan Data**

Pada penelitian ini dataset yang digunakan adalah *Hepatitis Data Set* yang diperoleh dari website *UCI Machine Learning Repository* pada alamat web <https://archive.ics.uci.edu/ml/datasets/Hepatitis> data tersebut bersifat publik jadi dapat diunduh oleh siapa saja.

**2.3. Pengolahan Data Awal**

Pada tahapan ini akan dilakukan pembersihan data, dimana dataset akan diperbaiki dan diperiksa hingga menjadi data yang siap di teliti. Selanjutnya akan dilakukan transformasi data yang nantinya akan diolah menggunakan model yang telah ditentukan pada *Microsoft Excel*.

**A. Data Cleaning**

Pada tahap ini dataset harus diolah dulu dengan melakukan proses pembersihan/*cleaning* untuk membersihkan data yang tidak sesuai, dan memperbaiki kesalahan pada data seperti kesalahan pada penulisan.

**B. Data Transformasi**

Penelitian ini menggunakan dataset hepatitis dimana terdapat 155 orang pasien terdiri dari 20 atribut dan 1 label. Namun dalam proses data mining menggunakan *naive bayes* semua atribut harus mengandung nilai kategorikal hingga diperlukannya inialisasi pada data tersebut. Serta perlunya proses transformasi data terhadap pengolahan *dataset* hepatitis untuk seleksi atribut yang digunakan. Pengubahan ini dilakukan pada *Micorosoft Excel*. Berikut merupakan table atribut yang telah ditransformasi terdapat pada Tabel 1.

**Tabel 1. Transformasi Data**

| No | Attribute        | Isi Dataset                        | Inialisasi    |
|----|------------------|------------------------------------|---------------|
| 1  | Class            | DIE, LIVE                          | DIE=1, LIVE=2 |
| 2  | Age              | 10, 20, 30, 40, 50, 60, 70, 80     |               |
| 3  | Steroid          | no, yes                            | no=1, yes=2   |
| 4  | Malaise          | no, yes                            | no=1, yes=2   |
| 5  | Anorexia         | no, yes                            | no=1, yes=2   |
| 6  | Liver Big        | no, yes                            | no=1, yes=2   |
| 7  | Liver Firm       | no, yes                            | no=1, yes=2   |
| 8  | Spleen Palpabale | no, yes                            | no=1, yes=2   |
| 9  | Ascites          | no, yes                            | no=1, yes=2   |
| 10 | Varices          | no, yes                            | no=1, yes=2   |
| 11 | Bilirubin        | 0.39, 0.80, 1.20, 2.00, 3.00, 4.00 |               |
| 12 | Alk Phosphate    | 33, 80, 120, 160, 200, 250         |               |
| 13 | Albumin          | 2.1, 3.0, 3.8, 4.5, 5.0, 6.0       |               |

**C. Populasi Dan Sampel Penelitian**

Pada *dataset* hepatitis terdapat populasi sebanyak 155 data dengan 20 atribut, data tersebut akan dibagi menjadi dua bagian yaitu data training dan data testing metode sampel yang digunakan adalah slovin. Untuk menghitung berapa banyak sampel yang akan di gunakan maka harus menghitung sampel dengan rumus, (Astuti & Wirama, 2016) berikut:

$$n = \frac{N}{1 + Ne^2}$$

Keterangan:

n = ukuran sampel

N= Ukuran populasi

E= persenan kelonggaran kesalahan pengambilan ampel yang dapat ditolerir atau yang di inginkan dalam penelitian ini (e=10%).

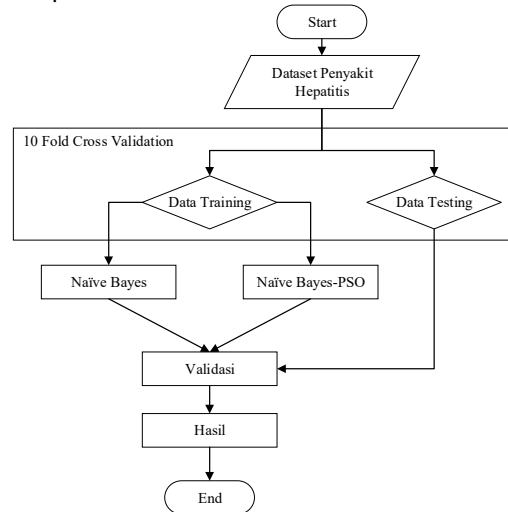
Perhitungan ukuran sampel:

$$n = \frac{155}{1+155(0,1)^2} = 60.78, \text{dibulatkan menjadi } 61$$

Berdasarkan dari hasil perhitungan sampel menggunakan rumus slovin, maka jumlah sampel yang akan diteliti adalah 61 sampel.

**2.4. Model Yang Diusulkan**

Pada tahapan ini model yang digunakan dalam penelitian ini adalah algoritma *Naïve Bayes* dan *Particle Swarm Optimization* untuk memprediksi penyakit hepatitis. Model yang diusulkan ditunjukkan dengan *flowchart* pada Gambar 2.



**Gambar 2. Model Yang diusulkan**

**2.5. Eksperimen dan Pengujian Model**

Eksperimen dan pengujian model dalam penelitian ini dilakukan dengan langkah-langkah sebagai berikut:

1. Menyiapkan dataset yang akan diuji, yaitu dataset hepatitis.
2. Melakukan proses data cleaning pada dataset.
3. Melakukan proses transformasi atribut.
4. Melakukan sampling untuk membagi dataset hepatitis menjadi 2 bagian yaitu, data *testing* dan data *training*
5. Kemudian data tersebut akan diuji menggunakan algoritma *Naïve Bayes* dan *Naïve Bayes* berbasis *Particle Swarm Optimization* sehingga mendapatkan hasil akurasi dan AUC untuk membandingkan tingkat akurasi dari setiap metode.

**2.6. Evaluasi dan Validasi Hasil**

Pada tahap ini dilakukan evaluasi terhadap *dataset* dengan menggunakan metode *Naïve Bayes* dan *Naïve Bayes* berbasis *Particle Swarm Optimization* yang akan menghasilkan *confusion matrix* dan performa AUC untuk nantinya dievaluasi mana nilai akurasi tertinggi dari setiap perhitungan metode tersebut. Evaluasi hasil klasifikasi dari prediksi tersebut akan dihitung menggunakan *tools RavidMiner versi 9.4* untuk melihat hasil akhir yang dihitung secara otomatis.

**3. Hasil dan Pembahasan**

**3.1. Algoritma Naïve Bayes**

*Naïve Bayes Classifier* atau disebut juga dengan *Bayesian Classification* merupakan metode pengklasifikasian statistik yang didasarkan pada teorema *bayes* yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu kelas. *Bayesian Classification* terbukti memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam *database* yang besar (Muhamad et al., 2017).

Persamaan dari teorema Bayes adalah :

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

Dimana :

- X = Data dengan *class* yang belum diketahui.
- H = Hipotesis Data X merupakan suatu *class* spesifik
- P(H|X) = probabilitas hipotesis H berdasarkan kondisi x (*posteriori prob.*)
- P(H) = Probabilitas hipotesis H (*prior prob.*)
- P(X|H) = probabilitas X berdasarkan kondisi tersebut
- P(X) = probabilitas dari X

Berikut ini merupakan langkah-langkan perhitungan manual algoritma *Naïve Bayes* adalah sebagai berikut:

- Menghitung *probabilitas* dari masing-masing kelas target yaitu dilakukan dengan membagi jumlah data perkelas dengan jumlah seluruh kelas. Pada kelas target yang bernilai *Die*(1) memiliki data sebanyak 14 data, sedangkan *Live*(2) memiliki data sebanyak 47 data dengan seluruh jumlah data 61 maka akan dihitung sebagai berikut:

$$P(X) = \frac{\text{jumlah kelas } i}{\text{jumlah seluruh data}}$$

$$P(C1) = \frac{14}{61} = 0.230$$

$$P(C2) = \frac{47}{61} = 0.770$$

**Tabel 1. Nilai Probabilitas Kelas**

| Attribute | Jumlah | Class |    | P(X Ci) |       |
|-----------|--------|-------|----|---------|-------|
|           |        | 1     | 2  | 1       | 2     |
| Total     | 61     | 14    | 47 | 0.230   | 0.770 |

- Setelah menghitung *probabilitas pior* kelas, selanjutnya menghitung jumlah *probabilitas prior* semua atribut perkelas.

Contoh perhitungan:

$$P(H|X) = \frac{\text{jumlah usia } i \text{ dengan keterangan } Ci}{\text{jumlah kelas } Ci}$$

$$P(\text{Age} = 70|C1)/C1) = 1/14 = 0.017$$

$$P(\text{Age} = 70|C2)/C2) = 3/47 = 0.064$$

Perhitungan tersebut dilakukam pada semua data sehingga dilihat hasil berikut:

**Tabel 2. Nilai Probabilitas Semua**

| Attribute               | Jumlah | Class |    | P(X Ci) |       |
|-------------------------|--------|-------|----|---------|-------|
|                         |        | 1     | 2  | 1       | 2     |
| Total                   | 61     | 14    | 47 | 0.230   | 0.770 |
| <b>Age</b>              |        |       |    |         |       |
| 20-30                   | 9      | 0     | 9  | 0.000   | 0.191 |
| 31-40                   | 22     | 3     | 19 | 0.214   | 0.404 |
| 41-50                   | 15     | 6     | 9  | 0.429   | 0.191 |
| 51-60                   | 10     | 4     | 6  | 0.286   | 0.128 |
| 61-70                   | 4      | 1     | 3  | 0.071   | 0.064 |
| 71-80                   | 1      | 0     | 1  | 0.000   | 0.021 |
| <b>Steroid</b>          |        |       |    |         |       |
| 1                       | 26     | 7     | 19 | 0.500   | 0.404 |
| 2                       | 35     | 7     | 28 | 0.500   | 0.596 |
| 0                       | 0      | 0     | 0  | 0.000   | 0.000 |
| <b>Malaise</b>          |        |       |    |         |       |
| 1                       | 21     | 7     | 14 | 0.500   | 0.298 |
| 2                       | 39     | 7     | 32 | 0.500   | 0.681 |
| 0                       | 1      | 0     | 1  | 0.000   | 0.021 |
| <b>Anorexia</b>         |        |       |    |         |       |
| 1                       | 9      | 2     | 7  | 0.143   | 0.149 |
| 2                       | 51     | 12    | 39 | 0.857   | 0.830 |
| 0                       | 1      | 0     | 1  | 0.000   | 0.021 |
| <b>Liver Big</b>        |        |       |    |         |       |
| 1                       | 12     | 2     | 10 | 0.143   | 0.213 |
| 2                       | 44     | 9     | 35 | 0.643   | 0.745 |
| 0                       | 5      | 3     | 2  | 0.214   | 0.043 |
| <b>Liver Firm</b>       |        |       |    |         |       |
| 1                       | 25     | 6     | 19 | 0.429   | 0.404 |
| 2                       | 30     | 5     | 25 | 0.357   | 0.532 |
| 0                       | 6      | 3     | 3  | 0.214   | 0.064 |
| <b>Spleen Palpabale</b> |        |       |    |         |       |
| 1                       | 10     | 7     | 3  | 0.500   | 0.064 |
| 2                       | 49     | 6     | 43 | 0.429   | 0.915 |
| 0                       | 2      | 1     | 1  | 0.071   | 0.021 |
| <b>Ascites</b>          |        |       |    |         |       |
| 1                       | 8      | 7     | 1  | 0.500   | 0.064 |
| 2                       | 51     | 6     | 45 | 0.429   | 0.915 |
| 0                       | 2      | 1     | 1  | 0.071   | 0.021 |
| <b>Varices</b>          |        |       |    |         |       |
| 1                       | 6      | 5     | 1  | 0.357   | 0.021 |
| 2                       | 53     | 8     | 45 | 0.571   | 0.957 |
| 0                       | 2      | 1     | 1  | 0.071   | 0.021 |
| <b>Bilirubin</b>        |        |       |    |         |       |
| 0-0.39                  | 2      | 0     | 2  | 0.000   | 0.043 |
| 0.40-0.80               | 18     | 1     | 17 | 0.071   | 0.362 |
| 0.81-1.20               | 21     | 3     | 18 | 0.214   | 0.383 |
| 1.21-2.00               | 13     | 4     | 9  | 0.286   | 0.191 |
| 2.01-3.00               | 2      | 2     | 0  | 0.143   | 0.000 |
| 3.01-4.60               | 5      | 4     | 1  | 0.286   | 0.021 |
| <b>Alk Phosphate</b>    |        |       |    |         |       |
| 0-33                    | 10     | 1     | 9  | 0.071   | 0.191 |
| 34-80                   | 19     | 1     | 18 | 0.071   | 0.383 |
| 81-120                  | 17     | 5     | 12 | 0.357   | 0.255 |
| 121-160                 | 8      | 4     | 4  | 0.286   | 0.085 |
| 161-200                 | 4      | 2     | 2  | 0.143   | 0.043 |
| 201-295                 | 3      | 1     | 2  | 0.071   | 0.043 |
| <b>Albumin</b>          |        |       |    |         |       |
| 0-2.1                   | 7      | 2     | 5  | 0.143   | 0.106 |
| 2.1-3.0                 | 10     | 5     | 5  | 0.357   | 0.106 |
| 3.1-3.8                 | 15     | 6     | 9  | 0.429   | 0.191 |
| 3.9-4.5                 | 25     | 1     | 24 | 0.071   | 0.511 |
| 4.6-5.0                 | 3      | 0     | 3  | 0.000   | 0.064 |
| 5.1-6.0                 | 1      | 0     | 1  | 0.000   | 0.021 |

3. Menghitung probabilitas keseluruhan dengan cara mengkalikan atribut yang telah dihitung dengan kelas yang sama.

**Tabel 3. Sampel Dataset Hepatitis Hitung *Posterior Probabilitas***

| Class | Age | Steroid | Malaise | Anorexia | Liver Big | Liver Firm | Spleen Palpable | Ascites | Varices | Bilirubin | Alk Phosphate | Albumin |
|-------|-----|---------|---------|----------|-----------|------------|-----------------|---------|---------|-----------|---------------|---------|
| ?     | 70  | 1       | 1       | 1        | 0         | 0          | 0               | 0       | 0       | 1.70      | 109           | 2.8     |
| ?     | 32  | 2       | 2       | 2        | 2         | 1          | 2               | 2       | 2       | 1.00      | 59            | 3.7     |
| ?     | 36  | 1       | 2       | 2        | 2         | 2          | 2               | 2       | 2       | 1.10      | 141           | 3.3     |
| ?     | 44  | 1       | 1       | 2        | 2         | 2          | 2               | 2       | 2       | 1.60      | 68            | 3.7     |
| ?     | 49  | 2       | 1       | 2        | 2         | 2          | 2               | 2       | 2       | 0.80      | 103           | 3.5     |

1. Data Ke-1

$$P(C1) = 0.017 * 0.500 * 0.500 * 0.143 * 0.214 * 0.214 * 0.071 * 0.071 * 0.071 * 0.286 * 0.357 * 0.357 * 0.230 = 0.000$$

$$P(C2) = 0.064 * 0.404 * 0.298 * 0.149 * 0.043 * 0.064 * 0.021 * 0.021 * 0.021 * 0.286 * 0.357 * 0.106 * 0.770 = 0.000$$

Dari perhitungan diatas, dapat kita lihat bahwa *Posterior Probabilitas* P(C1) and P(C2) mendapatkan hasil yang sama. Dilihat dari hasil perhitungan lebih condong dengan hasil klasifikasi bernilai **Die (1)**.

2. Data Ke-2

$$P(C1) = 0.214 * 0.500 * 0.500 * 0.857 * 0.643 * 0.429 * 0.429 * 0.429 * 0.429 * 0.571 * 0.214 * 0.071 * 0.429 * 0.230 = 1.999E - 06$$

$$P(C2) = 0.404 * 0.596 * 0.681 * 0.830 * 0.745 * 0.404 * 0.915 * 0.915 * 0.957 * 0.383 * 0.383 * 0.191 * 0.770 = 0.00011$$

Dari perhitungan diatas, dapat kita lihat bahwa *Posterior Probabilitas* P(C1) lebih kecil dari *Posterior Probabilitas* P(C2). Maka hasil klasifikasi data tersebut bernilai **Live (2)**.

3. Data Ke-3

$$P(C1) = 0.214 * 0.500 * 0.500 * 0.857 * 0.643 * 0.357 * 0.429 * 0.429 * 0.571 * 0.214 * 0.286 * 0.429 * 0.230 = 6.663E - 06$$

$$P(C2) = 0.404 * 0.404 * 0.681 * 0.830 * 0.745 * 0.532 * 0.915 * 0.915 * 0.957 * 0.383 * 0.085 * 0.191 * 0.770 = 0.00014$$

Dari perhitungan diatas, dapat kita lihat bahwa *Posterior Probabilitas* P(C1) lebih kecil dari *Posterior Probabilitas* P(C2). Maka hasil klasifikasi data tersebut bernilai **Live (2)**.

4. Data Ke-4

$$P(C1) = 0.429 * 0.500 * 0.500 * 0.857 * 0.643 * 0.357 * 0.429 * 0.429 * 0.571 * 0.286 * 0.071 * 0.429 * 0.230 = 6.663E - 06$$

$$P(C2) = 0.191 * 0.404 * 0.298 * 0.830 * 0.745 * 0.532 * 0.915 * 0.915 * 0.957 * 0.191 * 0.383 * 0.191 * 0.770 = 6.6E - 05$$

Dari perhitungan diatas, dapat kita lihat bahwa *Posterior Probabilitas* P(C1) lebih kecil dari *Posterior Probabilitas* P(C2). Maka hasil klasifikasi data tersebut bernilai **Live (2)**.

5. Data Ke-5

$$P(C1) = 0.429 * 0.500 * 0.500 * 0.857 * 0.643 * 0.357 * 0.429 * 0.429 * 0.571 * 0.071 * 0.357 * 0.429 * 0.230 = 5.553E - 06$$

$$P(C2) = 0.191 * 0.404 * 0.298 * 0.830 * 0.745 * 0.532 * 0.915 * 0.915 * 0.957 * 0.362 * 0.255 * 0.191 * 0.770 = 0.00012$$

Dari perhitungan diatas, dapat kita lihat bahwa *Posterior Probabilitas* P(C1) lebih kecil dari *Posterior Probabilitas* P(C2). Maka akan menghasilkan klasifikasi data tersebut bernilai **Live (2)**.

**3.2. Algoritma Naïve Bayes berbasis Particle Swarm Optimization**

Particle Swarm Optimization (PSO) adalah algoritma optimasi yang di perkenalkan oleh Dr. James Kennedy dan Dr. Russell C Eberhart pada 1995. Algoritma ini terinspirasi oleh pola hidup populasi burung dan kan. Keuntungan metode ini adalah memiliki konsep yang sederhana, mudah diimplementasikan, dan efisien dalam perhitungan dari pada algoritma matematika dan teknik heuristic lainnya (Herliana, Arifin, Susanti, & Hikmah, 2018).

Berikut ini merupakan deskripsi dari rumus kecepatan dan posisi partikel PSO oleh Kennedy dan Eberhart:

$$v_i(t) = v_i(t - 1) + \varphi C_1(p_i - x_i(t - 1)) + \varphi C_2(g - x_i(t - 1))$$

$$x_i(t) = x_i(t - 1) + v_i(t)$$

Keterangan :

$V_i(t-1)$  = kecepatan partikel ke-i pada iterasi ke-i

$P_i$  = posisi terbaik sebelum partikel ke-i

$g$  = Informasi dari lingkungan partikel terbaik dalam satu populasi

$X_i(t-1)$  = posisi saat partikel ke-i pada iterasi ke-i

$\varphi$  = Vektor acak dengan rentang Skor 0,1

$C_1$  = Koefisien dari pembelajaran kognitif

$C_2$  = Koefisien pembelajaran sosial

Berikut ini adalah pengolahan data Hepatitis secara manual menggunakan Particle Swarm Optimization:

**Table 5. Sampel Dataset Hepatitis Hitung PSO**

| Age | Steroid | Malaise | Anorexia | Liver Bigger | Liver Firm | Spleen Palpable | Ascites | Variation | Bilirubin | Alkaline Phosphate | Albumin |
|-----|---------|---------|----------|--------------|------------|-----------------|---------|-----------|-----------|--------------------|---------|
| 70  | 1       | 1       | 1        | 0            | 0          | 0               | 0       | 0         | 1.70      | 109                | 2.8     |
| 32  | 2       | 2       | 2        | 2            | 1          | 2               | 2       | 2         | 1.00      | 59                 | 3.7     |
| 36  | 1       | 2       | 2        | 2            | 2          | 2               | 2       | 2         | 1.10      | 141                | 3.3     |
| 44  | 1       | 1       | 2        | 2            | 2          | 2               | 2       | 2         | 1.60      | 68                 | 3.7     |
| 49  | 2       | 1       | 2        | 2            | 2          | 2               | 2       | 2         | 0.80      | 103                | 3.5     |

1. Diketahui Atribut Age Memiliki:

Pbest<sub>1</sub>: 70 Pbest<sub>4</sub>: 44 C<sub>1</sub>=C<sub>2</sub>=1

Pbest<sub>2</sub>: 31 Pbest<sub>5</sub>: 49  $\varphi_1=0,4$

Pbest<sub>3</sub>: 36 Gbest: 70  $\varphi_2=0.5$

Penyelesaian:

$$V_1 = 0 + 0.4(70 - 70) + 0.5(70 - 70) = 1$$

$$V_2 = 0 + 0.4(31 - 31) + 0.5(70 - 30) = 20$$

$$V_3 = 0 + 0.4(36 - 36) + 0.5(70 - 36) = 18$$

$$V_4 = 0 + 0.4(44 - 44) + 0.5(70 - 44) = 14$$

$$V_5 = 0 + 0.4(49 - 49) + 0.5(70 - 49) = 11.5$$

Dari perhitungan diatas, dapat dilihat bahwa kecepatan partikel  $V_2$  lebih cepat dibandingkan partikel yang lain.

2. Diketahui Atribut Steroid Memiliki:

P<sub>1</sub>: 1 P<sub>4</sub>: 1 C<sub>1</sub>=C<sub>2</sub>=1

P<sub>2</sub>: 2 P<sub>5</sub>: 2  $\varphi_1=0,4$

P<sub>3</sub>: 1 g: 2  $\varphi_2=0.5$

Penyelesaian:

$$V_1 = 0 + 0.4(1 - 1) + 0.5(2 - 1) = 0.5$$

$$V_2 = 0 + 0.4(2 - 2) + 0.5(2 - 2) = 0$$

$$V_3 = 0 + 0.4(1 - 1) + 0.5(2 - 1) = 0.5$$

$$V_4 = 0 + 0.4(1 - 1) + 0.5(2 - 1) = 0.5$$

$$V_5 = 0 + 0.4(2 - 2) + 0.5(2 - 1) = 0$$

Dari perhitungan diatas, dapat dilihat bahwa kecepatan partikel  $V_1, V_3, V_4$  lebih cepat dibandingkan partikel  $V_2$ , dan  $V_5$ .

3. Diketahui Atribut Malaise Memiliki:

P<sub>1</sub>: 1 P<sub>4</sub>: 1 C<sub>1</sub>=C<sub>2</sub>=1

P<sub>2</sub>: 2 P<sub>5</sub>: 1  $\varphi_1=0,4$

P<sub>3</sub>: 2 g : 2  $\varphi_2=0.5$

Penyelesaian:

$$V_1 = 0 + 0.4(1 - 1) + 0.5(2 - 1) = 0.5$$

$$V_2 = 0 + 0.4(2 - 2) + 0.5(2 - 2) = 0$$

$$V_3 = 0 + 0.4(2 - 2) + 0.5(2 - 2) = 0$$

$$V_4 = 0 + 0.4(1 - 1) + 0.5(2 - 1) = 0.5$$

$$V_5 = 0 + 0.4(1 - 1) + 0.5(2 - 1) = 0.5$$

Dari perhitungan diatas, dapat dilihat bahwa kecepatan partikel  $V_1$ ,  $V_4$ ,  $V_5$  lebih cepat dibandingkan partikel  $V_2$ , dan  $V_3$ .

4. Diketahui Atribut Anorexia Memiliki:

Penyelesaian:

$$P_1: 1 \quad P_4: 2 \quad C_1=C_2=1$$

$$P_2: 2 \quad P_5: 2 \quad \varphi_1=0,4$$

$$P_3: 2 \quad g : 2 \quad \varphi_2=0.5$$

$$V_1 = 0 + 0.4(1 - 1) + 0.5(2 - 1) = 0.5$$

$$V_2 = 0 + 0.4(2 - 2) + 0.5(2 - 2) = 0$$

$$V_3 = 0 + 0.4(2 - 2) + 0.5(2 - 2) = 0$$

$$V_4 = 0 + 0.4(2 - 2) + 0.5(2 - 2) = 0$$

$$V_5 = 0 + 0.4(2 - 2) + 0.5(2 - 1) = 0$$

Dari perhitungan diatas, dapat dilihat bahwa kecepatan partikel  $V_1$  lebih cepat dibandingkan partikel yang lain.

5. Diketahui Atribut Liver Big Memiliki:

$$P_1: 0 \quad P_4: 2 \quad C_1=C_2=1$$

$$P_2: 2 \quad P_5: 2 \quad \varphi_1=0,4$$

$$P_3: 2 \quad g : 2 \quad \varphi_2=0.5$$

Penyelesaian:

$$V_1 = 0 + 0.4(0 - 0) + 0.5(2 - 0) = 1$$

$$V_2 = 0 + 0.4(2 - 2) + 0.5(2 - 2) = 0$$

$$V_3 = 0 + 0.4(2 - 2) + 0.5(2 - 2) = 0$$

$$V_4 = 0 + 0.4(2 - 2) + 0.5(2 - 2) = 0$$

$$V_5 = 0 + 0.4(2 - 2) + 0.5(2 - 1) = 0$$

Dari perhitungan diatas, dapat dilihat bahwa kecepatan partikel  $V_1$  lebih cepat dibandingkan partikel yang lain.

6. Diketahui Atribut Liver Firm Memiliki:

$$P_1: 0 \quad P_4: 2 \quad C_1=C_2=1$$

$$P_2: 1 \quad P_5: 2 \quad \varphi_1=0,4$$

$$P_3: 2 \quad g : 2 \quad \varphi_2=0.5$$

Penyelesaian:

$$V_1 = 0 + 0.4(0 - 0) + 0.5(2 - 0) = 1$$

$$V_2 = 0 + 0.4(1 - 1) + 0.5(2 - 1) = 0.5$$

$$V_3 = 0 + 0.4(2 - 2) + 0.5(2 - 2) = 0$$

$$V_4 = 0 + 0.4(2 - 2) + 0.5(2 - 1) = 0$$

$$V_5 = 0 + 0.4(2 - 2) + 0.5(2 - 2) = 0$$

Dari perhitungan diatas, dapat dilihat bahwa kecepatan partikel  $V_1$  lebih cepat dibandingkan partikel yang lain.

7. Diketahui Atribut Spleen Palpable Memiliki:

$$P_1: 0 \quad P_4: 2 \quad C_1=C_2=1$$

$$P_2: 2 \quad P_5: 2 \quad \varphi_1=0,4$$

$$P_3: 2 \quad g : 2 \quad \varphi_2=0.5$$

Penyelesaian:

$$V_1 = 0 + 0.4(0 - 0) + 0.5(2 - 0) = 1$$

$$V_2 = 0 + 0.4(2 - 2) + 0.5(2 - 2) = 0$$

$$V_3 = 0 + 0.4(2 - 2) + 0.5(2 - 2) = 0$$

$$V_4 = 0 + 0.4(2 - 2) + 0.5(2 - 2) = 0$$

$$V_5 = 0 + 0.4(2 - 2) + 0.5(2 - 1) = 0$$

Dari perhitungan diatas, dapat dilihat bahwa kecepatan partikel  $V_1$  lebih cepat dibandingkan partikel yang lain.



8. Diketahui Atribut Arcites Memiliki:

$$\begin{array}{lll} P_1: 0 & P_4: 2 & C_1=C_2=1 \\ P_2: 2 & P_5: 2 & \varphi_1=0,4 \\ P_3: 2 & g : 2 & \varphi_2=0.5 \end{array}$$

Penyelesaian:

$$\begin{aligned} V_1 &= 0 + 0.4(0 - 0) + 0.5(2 - 0) = 1 \\ V_2 &= 0 + 0.4(2 - 2) + 0.5(2 - 2) = 0 \\ V_3 &= 0 + 0.4(2 - 2) + 0.5(2 - 2) = 0 \\ V_4 &= 0 + 0.4(2 - 2) + 0.5(2 - 2) = 0 \\ V_5 &= 0 + 0.4(2 - 2) + 0.5(2 - 1) = 0 \end{aligned}$$

Dari perhitungan diatas, dapat dilihat bahwa kecepatan partikel  $V_1$  lebih cepat dibandingkan partikel yang lain.

9. Diketahui Atribut Varices Memiliki:

$$\begin{array}{lll} P_{t1}: 0 & P_4: 2 & C_1=C_2=1 \\ P_2: 2 & P_5: 2 & \varphi_1=0,4 \\ P_3: 2 & g : 2 & \varphi_2=0.5 \end{array}$$

Penyelesaian:

$$\begin{aligned} V_1 &= 0 + 0.4(0 - 0) + 0.5(2 - 0) = 1 \\ V_2 &= 0 + 0.4(2 - 2) + 0.5(2 - 2) = 0 \\ V_3 &= 0 + 0.4(2 - 2) + 0.5(2 - 2) = 0 \\ V_4 &= 0 + 0.4(2 - 2) + 0.5(2 - 2) = 0 \\ V_5 &= 0 + 0.4(2 - 2) + 0.5(2 - 1) = 0 \end{aligned}$$

Dari perhitungan diatas, dapat dilihat bahwa kecepatan partikel  $V_1$  lebih cepat dibandingkan partikel yang lain.

10. Diketahui Atribut Bilirubin Memiliki:

$$\begin{array}{lll} P_1: 1.70 & P_4: 1.60 & C_1=C_2=1 \\ P_2: 1.00 & P_5: 1.70 & \varphi_1=0,4 \\ P_3: 1.10 & g : 1.70 & \varphi_2=0.5 \end{array}$$

Penyelesaian:

$$\begin{aligned} V_1 &= 0 + 0.4(1.70 - 1.70) + 0.5(1.70 - 1.70) = 0 \\ V_2 &= 0 + 0.4(1.00 - 1.00) + 0.5(1.70 - 1.00) = 0 \\ V_3 &= 0 + 0.4(1.10 - 1.10) + 0.5(1.70 - 1.10) = 0.35 \\ V_4 &= 0 + 0.4(1.60 - 1.60) + 0.5(1.70 - 1.60) = 0.3 \\ V_5 &= 0 + 0.4(1.10 - 1.10) + 0.5(1.70 - 1.10) = 0.35 \end{aligned}$$

Dari perhitungan diatas, dapat dilihat bahwa kecepatan partikel  $V_3$ , dan  $V_5$  lebih cepat dibandingkan partikel yang lain.

11. Diketahui Atribut Alk Phosphat Memiliki:

$$\begin{array}{lll} P_1: 109 & P_4: 68 & C_1=C_2=1 \\ P_2: 59 & P_5: 103 & \varphi_1=0,4 \\ P_3: 141 & g : 141 & \varphi_2= 0.5 \end{array}$$

Penyelesaian:

$$\begin{aligned} V_1 &= 0 + 0.4(109 - 109) + 0.5(141 - 109) = 16 \\ V_2 &= 0 + 0.4(59 - 59) + 0.5(141 - 59) = 41 \\ V_3 &= 0 + 0.4(141 - 141) + 0.5(141 - 141) = 0 \\ V_4 &= 0 + 0.4(68 - 68) + 0.5(141 - 68) = 36.5 \\ V_5 &= 0 + 0.4(103 - 103) + 0.5(141 - 103) = 19 \end{aligned}$$

Dari perhitungan diatas, dapat dilihat bahwa kecepatan partikel  $V_2$  lebih cepat dibandingkan partikel yang lain

12. Diketahui Atribut Albumin Memiliki:

$$\begin{array}{lll} P_1: 2.8 & P_4: 3.7 & C_1=C_2=1 \\ P_2: 3.7 & P_5: 3.5 & \varphi_1=0,4 \\ P_3: 3.3 & g : 3.7 & \varphi_2= 0.5 \end{array}$$

Penyelesaian:

$$\begin{aligned} V_1 &= 0 + 0.4(2.8 - 2.8) + 0.5(3.7 - 2.8) = 0.45 \\ V_2 &= 0 + 0.4(3.7 - 3.7) + 0.5(3.7 - 3.7) = 0 \\ V_3 &= 0 + 0.4(3.3 - 3.3) + 0.5(3.7 - 3.3) = 0.2 \\ V_4 &= 0 + 0.4(3.7 - 3.7) + 0.5(3.7 - 3.7) = 0 \end{aligned}$$

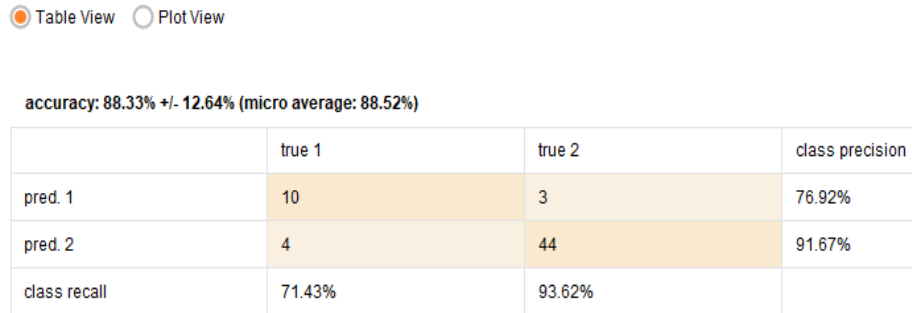
$$V_5 = 0 + 0,4(3,5 - 3,5) + 0,5(3,7 - 3,5) = 0,1$$

Dari perhitungan diatas, dapat dilihat bahwa kecepatan partikel  $V_1$  lebih cepat dibandingkan partikel yang lain.

### 3.3. Evaluasi dan Validasi Hasil

#### A. Hasil menggunakan Naïve Bayes

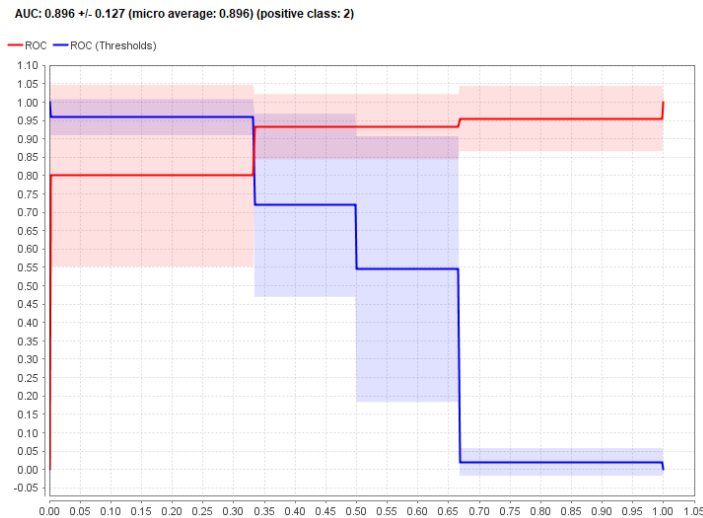
Pada tahap penelitian ini dilakukan eksperimen menggunakan *10-fold Cross Validation* sebagai teknik untuk evaluasi pada algoritma *Naïve Bayes*. Dalam eksperimen ini dilakukan pengujian algoritma *Naïve Bayes* dengan teknik *10-Fold Cross Validation* yang akan diimplementasikan pada *dataset* Hepatitis. Setelah dilakukan pengujian maka hasil yang diperoleh dapat dilihat pada gambar IV.14.



**Gambar 3. Hasil Confusion Matrix Akurasi Naïve Bayes**

Dari *confusion matrix* pada gambar 3. dapat diukur tingkat akurasi dari klasifikasi sebagai berikut :

$$\begin{aligned}
 Accuracy &= \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \\
 &= \frac{10 + 44}{10 + 44 + 3 + 4} \times 100\% \\
 &= 88.33\%
 \end{aligned}$$



**Gambar 4. Grafik AUC Algoritma Naïve Bayes**

Selain menghasilkan *confusion matrix*, pengujian tersebut menghasilkan nilai AUC sebesar 0.896 dan termasuk ke dalam *good classification*.

#### B. Hasil menggunakan Naïve Bayes Berbasis Particle Swarm Optimization

Pada tahap penelitian ini dilakukan eksperimen menggunakan *Particle Swarm Optimization* sebagai seleksi fitur hingga terjadi pembobotan dalam memilih bobot atribut yang bagus dan relevan. Sehingga mendapatkan hasil seperti pada tabel 6.

**Tabel 6. Hasil Seleksi Atribut**

| attribute   | weight |
|-------------|--------|
| Age         | 1      |
| Steroid     | 0      |
| Malaise     | 0      |
| Anorexia    | 0.174  |
| Liver Blg   | 1      |
| Liver Firm  | 0      |
| Spleen P... | 0.211  |
| Ascites     | 0.183  |
| Varices     | 0.343  |
| Bilirubin   | 0.705  |
| Alk Phos... | 1      |
| Albumin     | 1      |

Hasil seleksi atribut tersebut memperoleh beberapa atribut terendah yang memiliki nilai *weight* 0 yaitu Steroid, Malaise, dan Liver Firm yang berarti atribut tersebut tidak memiliki pengaruh dalam nilai akurasi. Sehingga atribut ini dapat dihilangkan agar dapat meningkatkan nilai akurasi.

Table View Plot View

accuracy: 91.90% +/- 8.56% (micro average: 91.80%)

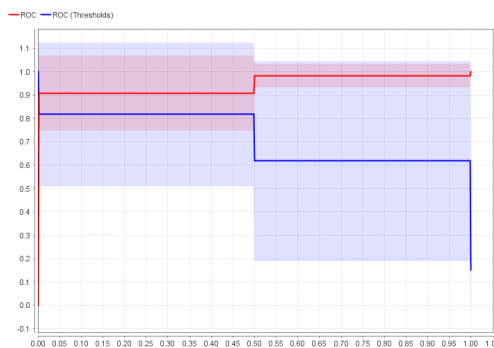
|              | true 1 | true 2 | class precision |
|--------------|--------|--------|-----------------|
| pred. 1      | 12     | 3      | 80.00%          |
| pred. 2      | 2      | 44     | 95.65%          |
| class recall | 85.71% | 93.62% |                 |

**Gambar 5. Hasil Confusion Matrix Akurasi Naïve Bayes-PSO**

Dari *confusion matrix* pada gambar IV.13. dapat diukur tingkat akurasi dari klasifikasi sebagai berikut :

$$\begin{aligned}
 Accuracy &= \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \\
 &= \frac{12 + 44}{12 + 44 + 3 + 2} \times 100\% \\
 &= 91.90\%
 \end{aligned}$$

AUC: 0.946 +/- 0.092 (micro average: 0.946) (positive class: 2)



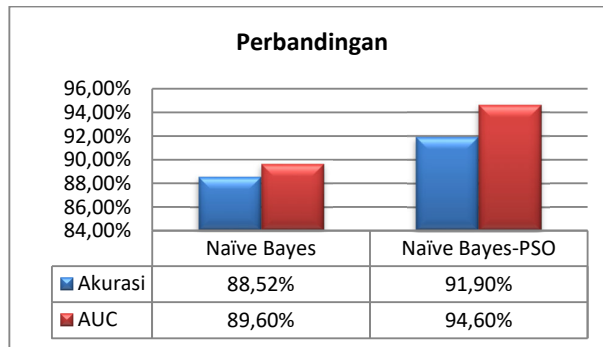
**Gambar 6. Grafik AUC Algoritma Naïve Bayes-PSO**

Selain menghasilkan *confusion matrix*, pengujian tersebut menghasilkan nilai AUC sebesar 0.946 dan termasuk ke dalam *excellent classification*.

**3.4. Pembahasan**

Tahap ini adalah tahapan untuk mengevaluasi hasil dari eksperimen yang dilakukan menenggunakan *Naïve Bayes* dan *Naïve Bayes-PSO* yang diimplementasikan menggunakan *Software RapidMiner* untuk memprediksi penyakit Hepatitis dan mengetahui hasil akurasi yang baik dari penelitian sebelumnya. Apakah hasil akurasi dari penelitian ini mendapatkan hasil yang

maksimal atau tidak. Mekanisme untuk mengetahuinya maka harus di evaluasi terlebih dahulu hasil dari eksperimen yang telah dilakukan, dapat dilihat seperti yang dibawah ini:



**Gambar 7. Grafik Perbandingan Hasil Pengujian**

Dilihat dari perbandingan hasil eksperimen diatas menunjukkan bahwa Algoritma Naïve Bayes-PSO lebih unggul dikarenakan nilai dari akurasi prediksinya 91.90% dengan nilai AUC 94.60%/0.960 termasuk kedalam klasifikasi *Excellent Classification* menunjukan bahwa hasil yang didapat sangat baik dinandingkan dengan algoritma Naïve Bayes yang memiliki akurasi 88.52% dengan nilai AUC sebesar 89.60%/0.896.

Dengan adanya hasil penelitian ini diharapkan dapat memberikan informasi mengenai penyakit hepatitis, sehingga kedepanya dapat digunakan sebagai bahan pembuatan aplikasi untuk memprediksi penyakit hepatitis dengan menggunakan algoritma *Naïve Bayes* berbasis *Particle Swarm Optimization*.

#### 4. Kesimpulan

Pada penelitian ini dilakukan eksperimen untuk memprediksi penyakit Hepatitis dengan menggunakan algoritma *Naïve Bayes* dan *Naïve Bayes* berbasis *Particle Swarm Optimization*. Berikut ini kesimpulan yang didapat dari hasil penelitian ini adalah:

1. Hasil penelitian menggunakan *Naïve Bayes* menghasilkan nilai akurasi sebesar 88.33% dengan nilai AUC sebesar 0.896, setelah ditambahkan algoritma *Particle Swarm Optimization* hasil dari akurasi meningkat menjadi 91.90% dengan nilai AUC sebesar 0.946. Sehingga dapat disimpulkan bahwa *Particle Swarm Optimization* dapat meningkatkan nilai Akurasi dan AUC dari algoritma *Naïve Bayes* dapat meningkatkan hasil akurasi dalam memprediksi penyakit Hepatitis.
2. Penelitian menggunakan metode *Naïve Bayes* berbasis *Particle Swarm Optimization* menghasilkan nilai akurasi 91.90% diharapkan dapat digunakan oleh para ahli dalam membuat aplikasi yang dapat memprediksi penyakit Hepatitis.
3. Dengan adanya penelitian ini dapat digunakan sebagai acuan dalam pembuatan penelitian oleh peneliti lain sehingga bermanfaat sebagai literatur dalam penelitiannya.

#### Daftar Pustaka

- Arifin, T., & Ariesta, D. (2019). *Prediksi Penyakit Ginjal KRONIS Menggunakan Algoritma Naive Bayes Classifier Berbasis Particle Swarm Optimization*. 13(1), 26–30.
- Astuti, N. M. A., & Wirama, D. G. (2016). *pengaruh Ukuran Perusahaan, Tipe Industri dan Intensitas Research and Development Pada Pengungkapan Modal Intelektual*. 15, 522–548.
- Herliana, A., Arifin, T., Susanti, S., & Hikmah, A. B. (2018). *Feature Selection of Diabetic Retinopathy Disease Using Particle Swarm Optimization and Neural Network*. (Citsm), 2016–2019.
- Muhamad, H., Prasajo, C. A., Sugianto, N. A., Surtiningsih, L., Cholissodin, I., Ilmu, F., ... Optimization, P. S. (2017). *Optimasi Naive Bayes Classifier Dengan Menggunakan Particle Swarm Optimization Pada Data Iris*. 4(3), 180–184.
- Novianti, D. (2019). *Implementasi Algoritma Naïve Bayes Pada Data Set Hepatitis Menggunakan Rapid Miner*. XXI(1). <https://doi.org/10.31294/p.v20i2>
- Oktaviani, P. S., Ramadhani, R. D., Laksana, T. G., & Amalia, A. E. (2018). *Komparasi Tingkat Akurasi Support Vector Machine (SVM) dan C4.5 dalam Mengklasifikasikan*

- Keberlangsungan Hidup Pasien Hepatitis*. 163–167.
- Ramdhani, L. S. (2016). *Penerapan Particle Swarm Optimization ( PSO ) Untuk Seleksi Atribut Dalam Meningkatkan Akurasi Prediksi Diagnosis Penyakit Hepatitis Dengan Metode Algoritma C4* . 5. IV(1), 1–15.
- Saifudin, A. (2018). *Metode Data Mining Untuk Seleksi Calon Mahasiswa Pada Penerimaan Mahasiswa Baru di Universitas Pamulang*. 10(1), 25–36.