

BAB IV HASIL DAN PEMBAHASAN

Dalam peningkatan akurasi algoritma K-Nearest Neighbor dengan menerapkan Particle Swarm Optimization untuk penyakit jantung memiliki beberapa tahap penelitian. Adapun penjelasan yang lebih lengkap terkait tahapan penelitian tersebut akan diuraikan sebagai berikut

4.1 Persiapan Data

Untuk memprediksi penyakit jantung, dataset yang berisi 1025 kejadian, data yang digunakan berasal dari dataset publik *heart disease* yang di peroleh dari www.kaggle.com/johnsmith88/heart-disease-dataset. dataset bisa dilihat pada gambar 4.1 berikut

age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
52	1	0	125	212	0	1	168	0	1	2	2	3	0
53	1	0	140	203	1	0	155	1	3.1	0	0	3	0
70	1	0	145	174	0	1	125	1	2.6	0	0	3	0
61	1	0	148	203	0	1	161	0	0	2	1	3	0
62	0	0	138	294	1	1	106	0	1.9	1	3	2	0
58	0	0	100	248	0	0	122	0	1	1	0	2	1
58	1	0	114	318	0	2	140	0	4.4	0	3	1	0
55	1	0	160	289	0	0	145	1	0.8	1	1	3	0
46	1	0	120	249	0	0	144	0	0.8	2	0	3	0
54	1	0	122	286	0	0	116	1	3.2	1	2	2	0
71	0	0	112	149	0	1	125	0	1.6	1	0	2	1
43	0	0	132	341	1	0	136	1	3	1	0	3	0
34	0	1	118	210	0	1	192	0	0.7	2	0	2	1
51	1	0	140	298	0	1	122	1	4.2	1	3	3	0
52	1	0	128	204	1	1	156	1	1	1	0	0	0
34	0	1	118	210	0	1	192	0	0.7	2	0	2	1
51	0	2	140	308	0	0	142	0	1.5	2	1	2	1
54	1	0	124	266	0	0	109	1	2.2	1	1	3	0
50	0	1	120	244	0	1	162	0	1.1	2	0	2	1
58	1	2	140	211	1	0	165	0	0	2	0	2	1
60	1	2	140	185	0	0	155	0	3	1	0	2	0
67	0	0	106	223	0	1	142	0	0.3	2	2	2	1

Gambar 4.1 Dataset Penyakit Jantung

Dataset yang digunakan memiliki 13 atribut terdiri dari data nominal dan numerik. Kelas data sasaran adalah tidak adanya penyakit jantung dan adanya penyakit jantung. Untuk atribut yang digunakan dapat dilihat pada tabel 4.1 dibawah ini :

Tabel 4.1 Atribut Dataset

No.	Atribut	Skala Pengukuran
1.	<i>Age</i> (Usia)	[29,0;77,0]
2.	<i>Sex</i> (Jenis Kelamin)	[0,0;1,0] 0,0 = wanita 1,0 =pria
3.	<i>Chest Pain</i> <i>real</i> (Nyeri dada)	[0,1,2,3] 0 = Tipikal angina 1 = Angina atipikal 2 = Nyeri non angina 3 = Tanpa gejala
4.	<i>Trestbps</i> (<i>Resting blood pressure</i> /Tekanan Darah)	[94,0;200,0]
5.	<i>Chol</i> (serum kolestoral)	[126,0;564,0]
6.	<i>Fbs</i> (<i>fasting blood sugar</i>)/Gula Darah	[0,1] 0 = <i>false</i> 1 = <i>true</i>
7.	<i>Restecg</i> (hasil electrocardiografi)	[0,1,2] - Nilai 0 = normal. - Nilai 1 = memiliki kelainan gelombang ST-T (inversi gelombang T dan / atau elevasi atau depresi ST > 0,05 mV). - Nilai 2 = menunjukkan hipertrofi ventrikel kiri yang mungkin atau pasti berdasarkan kriteria Estes.
8.	<i>Thalach</i> (denyut jantung maksimum tercapai)	[71,0;202,0]

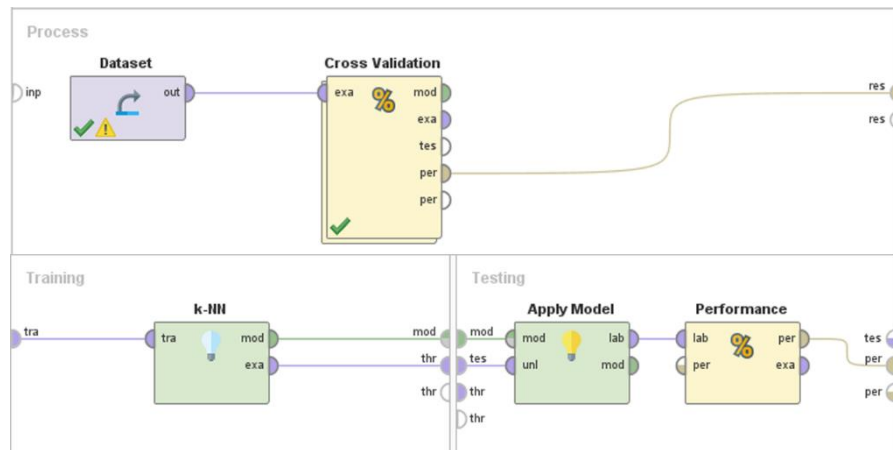
Tabel 4.1 Atribut Dataset

No.	Atribut	Skala Pengukuran
9.	<i>Exang</i> (latihan yang diinduksi angina)	[0;1] 0 = <i>No</i> 1 = <i>Yes</i>
10.	<i>Oldpeak</i> (Depresi ST disebabkan oleh olahraga relatif terhadap istirahat)	[0,0;6,2]
11.	<i>Slope</i> (kemiringan segmen latihan puncak ST)	[1,2,3] - Nilai 1 = Menanjak - Nilai 2 = Datar - Nilai 3 = Downsloping
12.	<i>Ca real</i> (number of major vessels (0-3) colored by flourosopy)	[0,1,2,3]
13.	<i>Thal real</i>	[0,1,2] 0 = <i>Normal</i> ; 1 = <i>Fixed defect</i> ; 2 = <i>Reversible defect</i>
14.	<i>Target</i> (diagnosis penyakit jantung)	{0,1} -Value 0: <i>No Disease</i> -Value 1: <i>Disease</i>

4.2 Pemodelan

4.2.1 Penelitian Menggunakan Algoritma *K-Nearest Neighbor*

Penerapan data pada Rapid Miner digunakan untuk Klasifikasi Penyakit Jantung menggunakan algoritma *K-Nearest Neighbor* ditunjukkan pada gambar 4.2 dibawah ini:



Gambar 4.2 Model RapidMiner tanpa menggunakan Seleksi Fitur

Gambar 4.2 menunjukkan proses penerapan klasifikasi menggunakan *tool rapidminer*. Data yang sudah disiapkan diimport ke rapidminer untuk melakukan pengujian menggunakan *Cross Validation* untuk membagi data menjadi data training dan data testing karena data yang digunakan adalah *supervised* dan algoritma yang digunakan adalah algoritma *K-Nearest Neighbors*. Hasil pengujian dapat kita lihat pada gambar 4.3 dibawah ini.

accuracy: 97.85% +/- 2.32% (micro average: 97.85%)

	true 0	true 1	class precision
pred. 0	496	19	96.31%
pred. 1	3	507	99.41%
class recall	99.40%	96.39%	

precision: 99.42% +/- 1.63% (micro average: 99.41%) (positive class: 1)

	true 0	true 1	class precision
pred. 0	496	19	96.31%
pred. 1	3	507	99.41%
class recall	99.40%	96.39%	

recall: 96.39% +/- 3.84% (micro average: 96.39%) (positive class: 1)

	true 0	true 1	class precision
pred. 0	496	19	96.31%
pred. 1	3	507	99.41%
class recall	99.40%	96.39%	

Gambar 4.3 Confusion Matrix tanpa menggunakan seleksi fitur

Gambar 4.3 menunjukkan hasil kinerja algoritma *K-Nearest Neighbors* dalam mengklasifikasi penyakit jantung, didalam pengujian tersebut menghasilkan nilai *accuracy*, *precision*, dan *recall*. *Accuracy* yang dihasilkan adalah sebesar 97,85%, *precision* yang dihasilkan adalah 99,42% dan *recall* yang dihasilkan adalah 96,39%.

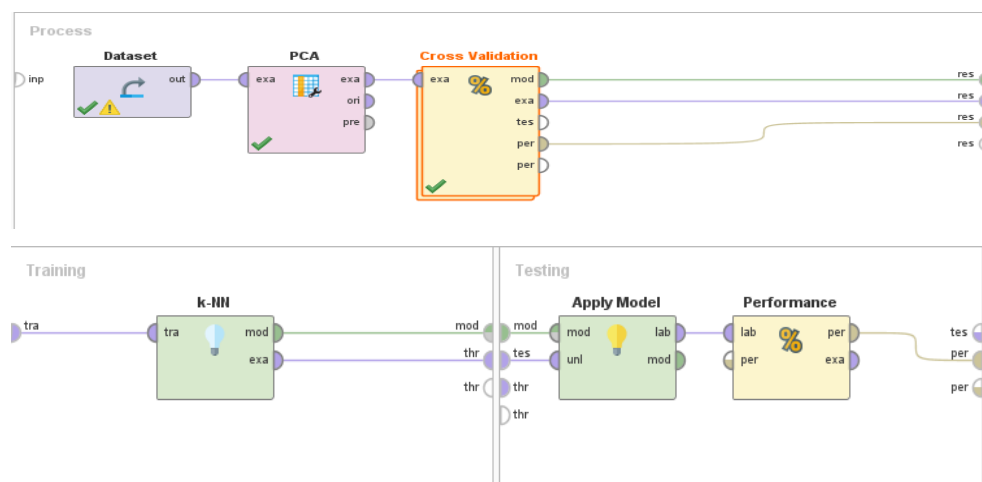
Perbandingan hasil akurasi dari penelitian sebelumnya dengan penelitian ini penggunaan algoritma *K-Nearest Neighbors* ditunjukkan pada tabel 4.2 dibawah ini:

Tabel 4.2 Perbandingan Hasil Akurasi

Penelitian	Akurasi
Penelitian sebelumnya	98,56%
Penelitian ini	97,76%

4.2.2 Penelitian Menggunakan *Principal Component Analysis (PCA)* dan Algoritma *K-Nearest Neighbors*

Metode yang dapat meningkatkan tingkat akurasi dalam klasifikasi adalah penggunaan teknik optimasi. pengujian lanjutan yang dilakukan pada penelitian ini adalah penggunaan teknik optimasi seleksi fitur menggunakan metode *Principal Component Analysis (PCA)* atau Analisis komponen utama. Penerapan pada rapidminer klasifikasi penyakit jantung menggunakan PCA dan Algoritma *K-Nearest Neighbors* dapat kita lihat pada gambar 4.4 dibawah ini:



Gambar 4.4 Model Rapidminer menggunakan PCA dan KNN

Pengujian pada gambar 4.4 menggunakan teknik validasi untuk membagi data training dan data testing menggunakan *cross validation*. Dari pengujian tersebut mendapatkan hasil yang dapat kita lihat pada gambar 4.5 dibawah ini:

accuracy: 98.54% +/- 1.79% (micro average: 98.54%)			
	true 0	true 1	class precision
pred. 0	499	15	97.08%
pred. 1	0	511	100.00%
class recall	100.00%	97.15%	

precision: 100.00% +/- 0.00% (micro average: 100.00%) (positive class: 1)			
	true 0	true 1	class precision
pred. 0	499	15	97.08%
pred. 1	0	511	100.00%
class recall	100.00%	97.15%	

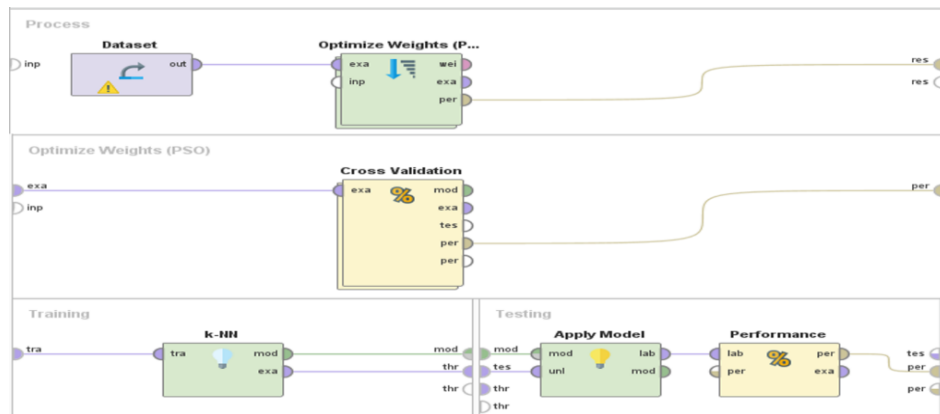
recall: 97.14% +/- 3.53% (micro average: 97.15%) (positive class: 1)			
	true 0	true 1	class precision
pred. 0	499	15	97.08%
pred. 1	0	511	100.00%
class recall	100.00%	97.15%	

Gambar 4.5 Confusion Matrix PCA dan KNN

Gambar 4.5 menunjukkan hasil kinerja algoritma *Principal Component Analysis* dengan *K-Nearest Neighbors* dalam mengklasifikasi penyakit jantung, pengujian tersebut menghasilkan nilai *accuracy*, *precision*, dan *recall*. *Accuracy* yang dihasilkan adalah 98,54%, *precision* yang dihasilkan adalah 100% dan *recall* yang dihasilkan adalah 97,14%

4.2.3 Penelitian Menggunakan *Particle Swarm Optimization (PSO)* dan Algoritma *K-Nearest Neighbors*

Untuk meningkatkan akurasi dalam klasifikasi penyakit jantung peneliti menambahkan teknik optimasi seleksi fitur menggunakan metode *Particle Swarm Optimization (PSO)*. Untuk skema desain dari tools rapid minernya dapat kita lihat pada gambar 4.6 dibawah ini:



Gambar 4.6 Model RapidMiner PSO dan KNN

Data yang sudah disiapkan diterapkan pada aplikasi rapidminer kalsifikasi dengan melakukan eksperimen menggunakan *cross validation* dan penggunaan teknik optimasi seleksi fitur *Particle Swarm Optimization* (PSO) dan algoritma yang digunakan tetap sama adalah *K-Nearest Neighbors*. Hasil pengujian dapat kita lihat pada gambar 4.7 dibawah ini.

accuracy: 99.22% +/- 1.18% (micro average: 99.22%)

	true 0	true 1	class precision
pred. 0	499	8	98.42%
pred. 1	0	518	100.00%
class recall	100.00%	98.48%	

precision: 100.00% +/- 0.00% (micro average: 100.00%) (positive class: 1)

	true 0	true 1	class precision
pred. 0	499	8	98.42%
pred. 1	0	518	100.00%
class recall	100.00%	98.48%	

recall: 98.47% +/- 2.32% (micro average: 98.48%) (positive class: 1)

	true 0	true 1	class precision
pred. 0	499	8	98.42%
pred. 1	0	518	100.00%
class recall	100.00%	98.48%	

Gambar 4.7 Confusion Matrix PSO dan KNN

Gambar 4.7 menunjukkan hasil kinerja algoritma *Particle Swarm Optimization* dengan *K-Nearest Neighbors* dalam mengklasifikasi penyakit jantung, pengujian tersebut menghasilkan nilai *accuracy*, *precision*, dan

recall. *Accuracy* yang dihasilkan adalah 99,22%, *precision* yang dihasilkan adalah 100% dan *recall* yang dihasilkan adalah 98,47%.

4.2.4 Perbandingan Penelitian Tidak Menggunakan PSO dan Menggunakan PSO

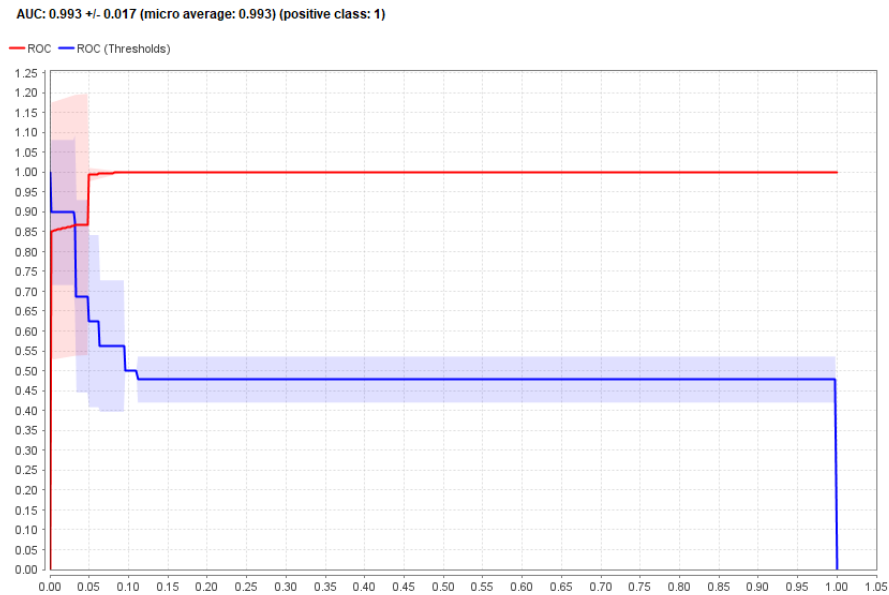
Perbandingan hasil eksperimen yang tidak menggunakan metode PSO dan eksperimen yang menggunakan PSO ditunjukkan pada tabel 4.3 dibawah ini

Tabel 4.3 Perbandingan Hasil Akurasi

Metode	K nilai				
	K=1	K=2	K=3	K=4	K=5
KNN tanpa PSO	99.71%	99.71%	97.85%	96.49%	82.43%
Setelah KNN+PSO	100%	100%	99.22%	98.44%	92.30%

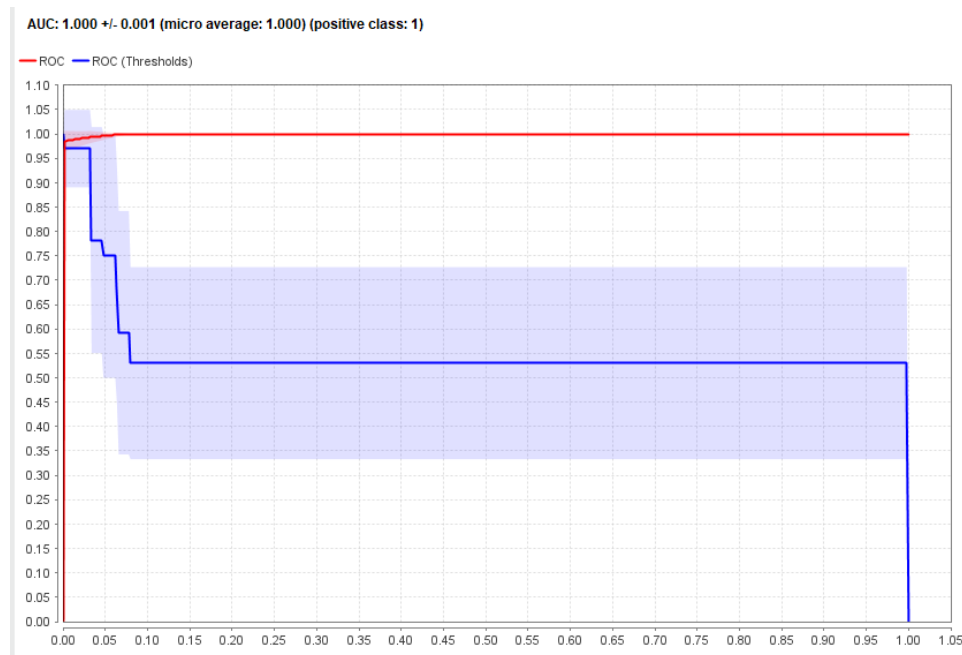
Sebelum pemilihan melakukan teknik optimasi seleksi fitur akurasi yang diperoleh adalah 97,76%. Penambahan fitur seleksi PSO ini dapat menyaring jumlah fitur dan memilih fitur yang berkontribusi lebih pada klasifikasi. Dengan menerapkan *K-Nearest Neighbors* dengan seleksi fitur PSO Ada peningkatan yang signifikan dalam akurasi dari penelitian sebelumnya yaitu 99,22%.

Selain Confusion Matrix untuk mengetahui kinerja dari pengujian ini kami mengandalkan kurva *ROC/AUC*. (*Area Under Curve*) yang dihasilkan. Perbandingan hasil Kurva ROC-AUC pada penelitian yang tidak menggunakan PSO dan yang menggunakan PSO dapat kita lihat pada gambar 4.8 dibawah ini.



Gambar 4.8 Kurva ROC-AUC Hasil Pengujian Tanpa Menggunakan Seleksi Fitur

Gambar 4.8 menunjukkan visualisasi antara *true positive rate (TPR)* dan *false positive rate (FPR)*. Classifier yang memberikan kurva semakin mendekati ke sudut kiri atas (*perfect classifier*) menunjukkan kinerja yang semakin baik. Selanjutnya, sebagai dasar evaluasi kinerja model, dibuat *classifier* acak (*random classifier*) yang memberikan titik-titik yang terletak di sepanjang diagonal ($FPR = TPR$). Semakin dekat kurva ke diagonal 45 derajat dari ruang ROC, semakin tidak akurat *classifier* tersebut. Dari ROC curve diatas, kita bisa menyimpulkan bahwa model Algoritma K-Nearest Neighbors memberikan performa yang cukup baik dengan nilai AUC (Area Under the Curve) yaitu 0,993 (mendekati 1). Model dikatakan sempurna ketepatan prediksinya jika nilai AUC adalah 1 yang artinya 100% area dibawah kurva seperti pada gambar 4.9 dibawah ini.



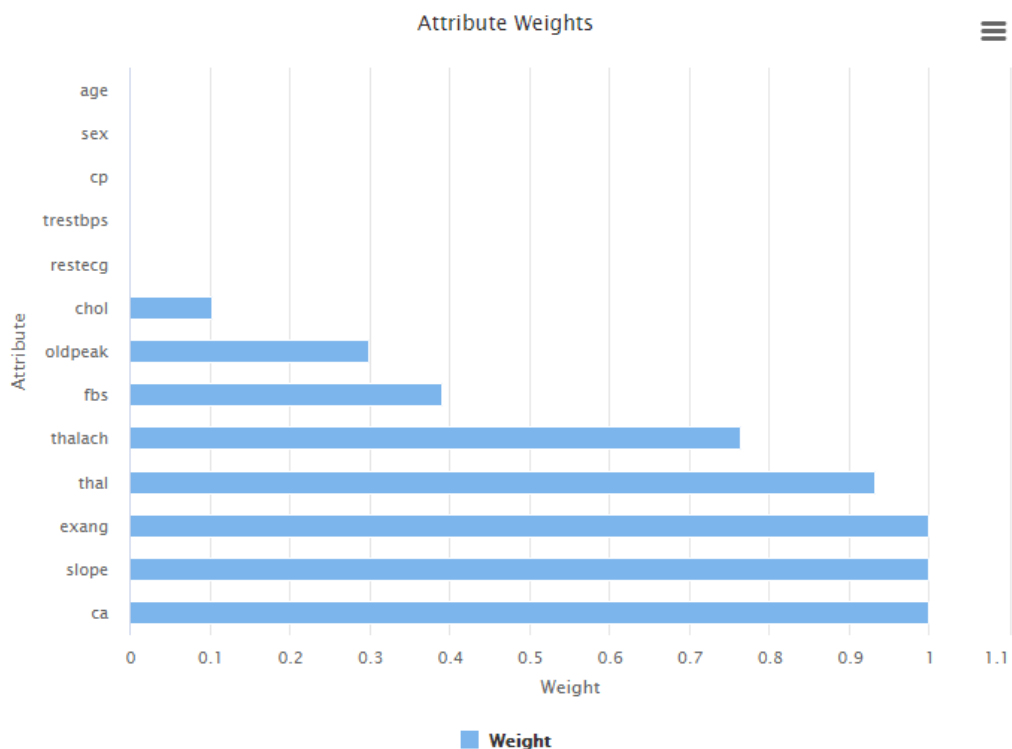
Gambar 4.9 Kurva ROC-AUC Hasil Pengujian Menggunakan PSO

Gambar 4.9 menunjukkan visualisasi antara *true positive rate (TPR)* dan *false positive rate (FPR)*. Classifier yang memberikan kurva semakin mendekati sudut kiri atas (*perfect classifier*) menunjukkan kinerja yang semakin baik. Selanjutnya, sebagai dasar evaluasi kinerja model, dibuat *classifier* acak (*random classifier*) yang memberikan titik-titik yang terletak di sepanjang diagonal ($FPR = TPR$). Semakin dekat kurva ke diagonal 45 derajat dari ruang ROC, semakin tidak akurat *classifier* tersebut. Dari ROC curve diatas, kita bisa menyimpulkan bahwa model Algoritma K-Nearest Neighbors dengan menambahkan Particle Swarm Optimization memberikan performa yang lebih baik dengan nilai AUC (Area Under the Curve) yaitu 1.000. Model dikatakan sempurna ketepatan prediksinya jika nilai AUC adalah 1 yang artinya 100% area dibawah kurva. Nilai AUC 1.000 termasuk dalam predikat Exellent Classification (0.90 – 1.00).

Akurasi yang didapat oleh model kami setelah melakukan teknik optimasi seleksi fitur dengan berbagai nilai k yang digunakan untuk menemukan tingkat akurasi terbaik yaitu 99,22% dan nilai AUC 1.000 menggunakan *number of fold* 8 dan nilai K=3.

4.3 Evaluasi

Pemodelan yang telah dilakukan berhasil menghasilkan akurasi yang tinggi. Penggunaan seleksi fitur *Particle Swarm Optimization* (PSO) dengan menggunakan 13 atribut dapat diperoleh data yang terseleksi oleh pembobotan atribut yang bisa dilihat pada gambar 4.10 dibawah ini:



Gambar 4.10 Hasil Seleksi Atribut

Penerapan algoritma PSO pada pembobotan atribut menghasilkan nilai bobot dari 13 atribut yang digunakan pada gambar 4.10. Atribut *Chol* menghasilkan bobot nilai 0,103, *Fbs* 0,390, *Thalach* 0,763, *Exan* 1, *Oldpeak* 0,298, *Slope* 1, *Ca real* 1 dan *Thal real* 0,932. Berdasarkan hasil pembobotan diputuskan bahwa delapan atribut memberikan kontribusi pada proses klasifikasi melihat gap tiap bobot atribut apabila salah satu atribut hilang maka akan mempengaruhi hasil dari akurasi, *precision* dan *recall*. Kemudian terdapat 5 atribut yang memiliki nilai 0 yang artinya atribut tersebut otomatis terseleksi dan tidak perlu digunakan karena tidak memiliki pengaruh dalam proses. Penerapan algoritma PSO pada pembobotan atribut di

Algoritma *K-Nearest Neighbors* berhasil meningkatkan akurasi dalam klasifikasi penyakit jantung dengan tingkat *accuracy* yang dihasilkan sebesar 99.22%, *precision* 100%, *recall* 98,47% dan nilai AUC 1.000. Selain itu kinerja algoritma pada penelitian ini juga menunjukkan performa yang sangat baik sekali dan tingkat akurasi yang sangat tinggi.