

PENERAPAN METODE *SUPPORT VECTOR MACHINE* BERBASIS *PARTICLE SWARM OPTIMIZATION* UNTUK PREDIKSI PENYAKIT JANTUNG

Siti Nurajizah

AMIK BSI Jakarta
Manajemen Informatika
Jalan RS Fatmawati No 24 Pondok Labu, Jakarta Selatan
<http://www.bsi.ac.id>
siti.snz@bsi.ac.id

ABSTRACT

Heart disease is one of the world's deadliest diseases. Heart disease occurs due to narrowing or blockage of the coronary arteries caused by the buildup of fatty substances (cholesterol, triglycerides), more and more and accumulate beneath the inner lining of the arteries. Several studies have been conducted to diagnose patients is not yet known but the exact method to predict heart disease. This study uses support vector machine and support vector machine -based method particle swarm optimization to get the rules for the prediction of cardiovascular disease and provide a more accurate value of the result accuracy. After testing two models of Support Vector Machine and Support Vector Machine -based Particle Swarm Optimization and the results by using Support Vector Machine get accuracy values 81.85 % and AUC values 0.899, while testing with Support Vector Machine -based particle swarm optimization to get accuracy values 88.61 % and AUC values 0.919. Both of these methods have difference values of 6.76 % and the difference in AUC value of 0.02.

Keywords: *Heart disease, Data Mining, Support Vector Machine, Particle Swarm Optimization.*

I. PENDAHULUAN

Dunia kesehatan saat ini berkembang dengan begitu pesatnya. Industri kesehatan memiliki sejumlah besar data kesehatan, namun beberapa data kesehatan masih tersembunyi, padahal informasi tersebut dibutuhkan untuk membuat sebuah keputusan yang efektif. Dalam bidang medis, di dunia industri kesehatan memerlukan keputusan yang efektif dalam pengambilan keputusan dan keakuratan untuk prediksi suatu penyakit, seperti prediksi penyakit jantung. Penyakit jantung disebut juga dengan penyakit jantung coroner, yaitu penyakit yang terjadi bila darah ke otot jantung terhenti/tersumbat, sehingga mengakibatkan kerusakan berat pada jantung (Rajkumar dan Reena, 2010). Penyebab utama penyakit jantung adalah penggunaan tembakau, fisik tidak aktif, diet yang tidak sehat dan penggunaan alkohol, resiko penyakit jantung bertambah dengan meningkatnya usia, tekanan darah tinggi, mempunyai kolesterol tinggi, dan kelebihan berat badan.

Data mining (Subbalakshmi, et al, 2011) adalah proses identifikasi yang valid. Potensi yang berguna dan akhirnya dimengerti pada suatu pola dalam data yang luas dengan menggunakan *database* dan pertumbuhan eksplosif dalam ukuran mereka. Data mengacu pada pengetahuan dari jumlah data yang besar. Data mining adalah pencarian untuk hubungan dan pola global

yang ada di *database* yang besar tapi tersembunyi diantara sejumlah besar data. Sedangkan menurut (Santoso, 2007) data mining adalah kegiatan pengumpulan, pemakaian data *history* untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar.

II. LANDASAN TEORI

2.1 *Data Mining*

Data Mining adalah rangkaian proses untuk menggali nilai tambah berupa informasi yang belum terekplorasi dari sebuah basis data, melakukan eksplorasi dengan cara-cara tertentu untuk memanipulasi data menjadi informasi yang lebih berharga dengan cara mengekstraksi dan mengenali pola penting dari basis data (Han dan Kamber, 2006).

Konsep *data mining*, menemukan pola berharga dalam data, adalah respon yang jelas untuk pengumpulan dan penyimpanan volume data yang besar (Weiss, et al, 2010). Secara khusus, koleksi metode yang dikenal sebagai '*data mining*' menawarkan metodologi dan solusi teknis untuk mengatasi analisis data medis dan konstruksi dari prediksi (Bellazzi dan Zupanb, 2008). Untuk semua aplikasi *data mining*, akurasi prediksi tergantung pada kualitas prediksi atribut (Weiss, et al, 2010).

2.2 *Support Vector Machines*

Support Vector Machine (SVM) adalah sebuah metode seleksi yang membandingkan

parameter standar seperangkat niali diskrit yang disebut kandidat set, dan mengambil salah satu yang memiliki akurasi klasifikasi terbaik (Dong, Xia, Tu, & Xing, 2007). SVM pertama kali diperkenalkan oleh Vapnik, Boser dan Guyon pada tahun 1992. SVM adalah salah satu teknik baru dibandingkan dengan teknik lain, tetapi memiliki performansi yang lebih baik di berbagai bidang aplikasi seperti bioinformatika, pengenalan tulisan tangan, klasifikasi teks, klasifikasi diagnosis penyakit dan lain sebagainya (Feng-Chia, 2009). Support Vector Machines (SVM) adalah seperangkat metode yang terkait untuk suatu metode pembelajaran, untuk kedua masalah klasifikasi dan regresi (Maimon, 2010). Dengan berorientasi pada tugas, kuat, sifat komputasi mudah dikerjakan, SVM telah mencapai sukses besar dan dianggap sebagai state-of-the-art classifier saat ini (Huang, Yang, King, & Lyu, 2008).

Karakteristik dari *Support Vector Machine* adalah sebagai berikut:

1. *Support Vector Machine* adalah *linier classifier*
2. *Pattern Recognition* dilakukan dengan mentransformasikan data pada *input space* ke ruang yang berdimensi lebih tinggi, dan optimasi dilakukan pada ruang *vector* yang baru.
3. Menerapkan strategi *Structural Risk Minimization* (SRM).
4. Prinsip kerja *Support Vector Machine* pada dasarnya hanya mampu menangani klasifikasi dua *class*.

Data yang tersedia dinotasikan sebagai $x \in R^d$ sedangkan label masing-masing dinotasikan $y_i \in \{-1,+1\}$ untuk $i = 1, 2, \dots, n$ yang mana n adalah banyaknya data. Diasumsikan kedua class -1 dan $+1$ dapat terpisah secara sempurna oleh *hyperplane* berdimensi d , yang didefinisikan:

$$\vec{w} \cdot x + b = 0 \tag{2.1}$$

Sebuah pattern x_i yang termasuk class -1 (sampel negatif) dapat dirumuskan sebagai pattern yang memenuhi pertidaksamaan:

$$\vec{w} \cdot x + b = -1 \tag{2.2}$$

sedangkan pattern yang termasuk class $+1$ (sampel positif):

$$\vec{w} \cdot x + b = +1 \tag{2.3}$$

Margin terbesar dapat ditemukan dengan memaksimalkan nilai jarak antara *hyperplane* dan titik terdekatnya, yaitu $1/\|w\|$. Hal ini dapat dirumuskan sebagai *Quadratic Programming* (QP) *problem*, yaitu mencari

titik minimal persamaan 2.4, dengan memperhatikan bentuk persamaan 2.5.

$$\min_w \tau(W) = \frac{1}{2} \left\| \vec{w} \right\|^2 \tag{2.4}$$

$$y_i (x_i \cdot w + b) - 1 \geq 0, \forall i \tag{2.5}$$

Problem ini dapat dipecahkan dengan berbagai teknik komputasi, diantaranya *Lagrange Multiplier* sebagaimana ditunjukkan pada persamaan 2.6:

$$L(\vec{w}, b, a) = \frac{1}{2} \left\| \vec{w} \right\|^2 + \sum_{i=1}^n \alpha_i (y_i (x_i \cdot w + b) - 1) \tag{2.6}$$

α_i adalah *Lagrange multipliers*, yang bernilai nol atau positif ($\alpha_i \geq 0$). Nilai optimal dari persamaan (6) dapat dihitung dengan meminimalkan L terhadap w dan b , dan memaksimalkan L terhadap α_i .

Dengan memperhatikan sifat bahwa pada titik optimal gradient $L=0$, persamaan langkah 2.6 dapat dimodifikasi sebagai maksimalisasi problem yang hanya mengandung α_i saja, sebagaimana persamaan 2.7

$$\text{maximize:} \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) \tag{2.7}$$

Subject to:

$$\alpha_i \geq 0 \quad (i=1,2,\dots,n) \quad \sum_{i=1}^n \alpha_i y_i = 0 \tag{2.8}$$

Dari hasil perhitungan ini diperoleh α_i yang kebanyakan bernilai positif. Data yang berkorelasi dengan α_i yang positif inilah yang disebut sebagai *Support Vector*.

2.3 Particle Swarm Optimization

Optimasi adalah proses menyesuaikan kepada masukan atau karakteristik perangkat, proses matematis, atau percobaan untuk menemukan output minimum atau maksimum atau hasil. Input terdiri dari variabel, proses atau fungsi dikenal sebagai fungsi biaya, fungsi tujuan, atau kemampuan fungsi, dan output adalah biaya atau tujuan, jika proses adalah sebuah percobaan, kemudian variabel adalah masukan fisik untuk percobaan (Haupt & Haupt, 2004).

Particle Swarm Optimization (PSO) adalah metode pencarian penduduk, yang berasal dari penelitian untuk pergerakan organisme dari kelompok burung atau ikan, seperti algoritma genetika, *Particle Swarm Optimization* (PSO) melakukan pencarian menggunakan populasi (*swarm*) dari individu (partikel) yang diperbaharui dari iterasi untuk iterasi (Fei, et al, 2009).

Particle Swarm Optimization (PSO) merupakan algoritma pencarian berbasis

populasi dan diinisialisasi dengan populasi solusi acak dan digunakan untuk memecahkan masalah optimasi (Abraham et al, 2006). PSO adalah teknik yang terinspirasi oleh proses alami burung yang berkelompok, dan juga dikenal sebagai segerombolan intelijen dengan mempelajari perilaku sosial atau kelompok hewan. (Shukla, et al, 2010).

Untuk menemukan solusi yang optimal, masing-masing partikel bergerak ke arah posisi sebelumnya terbaik (pbest) dan terbaik posisi global (gbest). Kecepatan dan posisi partikel dapat diperbarui sebagai berikut persamaan:

$$v_{ij}(t+1) = w * v_{ij}(t) + c_1 * rand_1 * (pbest_{ij}(t) - p_{ij}(t)) + c_2 * rand_2 * (gbest_{ij}(t) - p_{ij}(t)) \quad (2.9)$$

$$p_{ij}(t+1) = p_{ij}(t) + \beta * v_{ij}(t+1) \quad (2.10)$$

Dimana:

t = menunjukkan *counter* iterasi

V_{ij} = kecepatan partikel i pada dimensi ke-j (nilainya terbatas antara [- v_{max}, v_{max}])

p_{ij} = posisi partikel i pada j dimensi (nilainya terbatas [-p_{max}, p_{max}])

$pbest_{ij}$ = posisi pbest partikel i pada dimensi ke-j

$gbest_{ij}$ = posisi gbest dari dimensi ke-j

w = berat inersia (menyeimbangkan eksplorasi global dan lokal eksploitasi)

rand1 dan rand2 = fungsi acak di rentang [0, 1]

β = faktor kendala untuk mengontrol kecepatan berat (nilainya ke 1)

c_1 dan c_2 adalah faktor pembelajaran pribadi dan sosial (nilainya ke 2)

2.4 Seleksi Atribut

Proses *data mining* membutuhkan biaya komputasi yang tinggi ketika berhadapan dengan kumpulan data dalam jumlah besar. Mengurangi dimensi yaitu jumlah atribut set data atau kelompok atribut, secara efektif dapat memotong biaya tersebut. Pengurangan dimensi tersebut dilakukan dengan menekan seminimal mungkin kerugian yang dapat terjadi akibat kehilangan sebagian informasi. Tujuan pengurangan dimensi dalam domain *data mining* adalah untuk mengidentifikasi biaya terkecil di mana algoritma *data mining* dapat menjaga tingkat kesalahan di bawah perbatasan garis efisiensi. Yang dimaksud dengan biaya adalah fungsi dari kompleksitas teoritis dari algoritma *data mining* yang berasal dari model, dan berkorelasi dengan waktu yang dibutuhkan algoritma tersebut dalam menjalankan model, serta ukuran dari kumpulan data (Maimon dan Rokach, 2010).

Seleksi atribut adalah masalah terkait erat dengan pengurangan dimensi. Tujuan

seleksi atribut adalah untuk mengidentifikasi tingkat kepentingan atribut dalam kumpulan data, dan membuang semua atribut lain seperti informasi yang tidak relevan dan berlebihan. Karena seleksi atribut mengurangi dimensi dari data, maka hal ini akan memungkinkan operasi algoritma data mining dapat berjalan lebih efektif dan lebih cepat. Dalam beberapa kasus dengan dilakukannya seleksi atribut dihasilkan peningkatan tingkat akurasi klasifikasi (Maimon dan Rokach, 2010).

Di sisi lain, seleksi atribut adalah proses yang mahal, dan juga bertentangan dengan asumsi awal yaitu bahwa semua informasi atau atribut diperlukan dalam rangka mencapai akurasi maksimal.

Ada empat alasan utama untuk melakukan pengurangan dimensi yaitu (Maimon dan Rokach, 2010):

1. Penurunan biaya model pembelajaran
2. Meningkatkan kinerja model pembelajaran
3. Mengurangi dimensi yang tidak relevan
4. Mengurangi dimensi yang berlebihan

Tujuan seleksi atribut adalah untuk pengurangan atribut dari *dataset* untuk menghilangkan variabel yang dianggap tidak relevan. Metode seleksi atribut dapat diklasifikasikan ke dalam tiga kategori utama (Vercellis, 2009):

1. Metode *filter*
Metode *Filter* adalah memilih atribut yang relevan sebelum pindah ke tahap pembelajaran berikutnya, atribut yang dianggap paling penting yang dipilih untuk pembelajar sedangkan sisanya dikecualikan
2. Metode *wrapper*
Metode *wrapper* menilai sekelompok variabel dengan menggunakan klasifikasi yang sama atau algoritma regresi digunakan untuk memprediksi nilai dari variabel target.
3. Metode *embedded*
Untuk metode *embedded*, proses seleksi atribut terletak di dalam algoritma pembelajaran, sehingga pemilihan set optimal atribut secara langsung dibuat selama fase generasi model.

III. METODE PENELITIAN

3.1 Desain Penelitian

Menurut (Sugiyono, 2009) Metode penelitian dapat diklasifikasikan berdasarkan tujuan dan tingkat kealamian (*natural setting*) obyek yang diteliti. Berdasarkan tujuan, metode penelitian dapat diklasifikasikan

menjadi penelitian dasar (*basic research*), penelitian terapan (*applied research*) dan penelitian pengembangan (*research and development*). Selanjutnya berdasarkan tingkat kealamian, metode penelitian dapat dikelompokkan menjadi metode penelitian eksperimen, *survey* dan naturalistik.

Menurut (Kothari, 2004) Jenis penelitian eksperimen adalah metode yang menguji kebenaran sebuah hipotesis dengan statistik dan menghubungkan dengan masalah penelitian. Jenis penelitian eksperimen dibagi dua, yaitu eksperimen absolut dan eksperimen komparatif. Eksperimen absolut mengarah kepada dampak yang dihasilkan dari eksperimen, misalnya pengaruh honor dosen terhadap kinerja. Sedangkan eksperimen komparatif yaitu membandingkan dua objek yang berbeda, misalnya membandingkan dua algoritma yang berbeda dengan melihat hasil statistik masing-masing yang mana lebih baik (Kothari, 2004). Dalam penelitian ini digunakan jenis penelitian eksperimen untuk menghasilkan nilai akurasi untuk prediksi penyakit jantung.

Dalam penelitian ini metode pengumpulan data untuk mendapatkan sumber data yang digunakan adalah metode pengumpulan data sekunder. Data utama diperoleh dari *University of California Irvine machine learning data repository* sedangkan data pendukung didapatkan dari buku, jurnal dan publikasi lainnya.

3.2 Tahapan penelitian

Terdapat beberapa tahap dalam pengolahan data eksperimen, pada penelitian ini menggunakan model *Cross-Standard Industry for Data Mining (CRISP-DM)*, yaitu seperti terlihat pada diagram di bawah ini: (Sumathi dan Sivanandam,2006)

1. Tahap Business Understanding

Tahap *Business Understanding* merupakan pemahaman tentang substansi dari kegiatan data mining yang akan dilakukan serta kebutuhan data mining dari segi perspektif bisnis. Kegiatan yang dilakukan antara lain: menentukan sasaran atau tujuan bisnis, memahami situasi bisnis, menentukan tujuan data mining, serta membuat perencanaan strategis beserta jadwal penelitian. Dalam penelitian ini dilakukan prediksi untuk mendeteksi penyakit jantung menggunakan *Support Vector Machine dan Support Vector Machine* berbasis Algoritma *Particle Swarm Optimization*.

2. Tahap Data Understanding

Data Understanding adalah fase mengumpulkan data awal, mempelajari data untuk bisa mengenal data yang akan dipakai, mengidentifikasi masalah yang berkaitan dengan kualitas data, mendeteksi subset yang menarik dari data untuk membuat hipotesa awal. Pada fase ini akan ditentukan atribut-atribut yang digunakan untuk membangun model. Pada penelitian ini data yang diolah merupakan data pasien penyakit jantung pada *University of California Irvine* yang terdiri dari *field* sebagai berikut:

Tabel Atribut Penyakit Jantung

No	Atribut	Nilai	Kategori
1	Age	<40	Inventus
		40-50	Verilitas
		55-64	Prasenium
		>=65	Senium
2	Sex	1	Laki-laki
		2	Perempuan
3	Chest Pain Type	1	Typical Angina
		2	Atypica Angina
		3	Non Angina Pain
		4	Asymptomatic
4	Testing Blood Pressure	<120	Normal
		120-139	Prehipertensi
		140-159	Hipertensi Level I
		>159	Hipertensi Level II
5	Serum Cholesterol	<200	Normal
		200-239	Batas Normal Tinggi
		>239	Tinggi
6	Fasting Blood Sugar	1	Ya
		0	Tidak
7	Resting Electrocardiographic Result	0	Normal
		1	Abnormality
		2	Left Ventricular Hypertropy
8	Maximum Heart Rate Archeived	=220-usia	Normal
		<>220-usia	Tidak normal
9	ExerciseInduced Angina	0	Tidak
		1	Ya
10	Old peak	<1	0
		>=1 dan <2	1
		>=2 dan <3	2
		>=3 dan <4	3
		>=4	4
11	The Slope of the peak exercise ST segmen	1	Unsloping
		2	Flat
		3	Downsloping
12	Number of Major Vessel	1	1
		2	2
		3	3
13	Thal	3	Normal
		6	Fixed defect
		7	Reversible defect

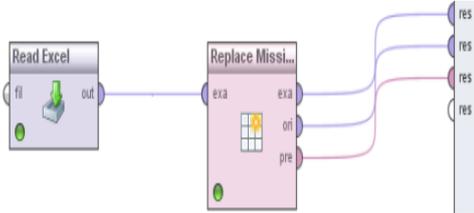
3. Tahap Data Preparation

Data preparation sering disebut sebagai fase yang padat karya. Aktifitas yang

dilakukan antara lain memilih tabel atau *field* sebagai bahan untuk data mining. Terdapat beberapa tehnik *data preprocessing*, diantaranya *data cleaning*, *data integration*, *data reduction* dan *data transformations* (Vecellis, 2009):

1. *Data validation*, untuk mengidentifikasi dan menghapus data yang ganjil (*outlier/noise*), data yang tidak konsisten, dan data yang tidak lengkap (*missing value*).
2. *Data integration and transformation*, untuk meningkatkan akurasi dan efisiensi algoritma.
3. *Data size reduction and discretization*, Teknik ini digunakan untuk mereduksi ukuran data jika dataset yang digunakan cukup besar yang berguna untuk membuat model pembelajaran algoritma menjadi lebih efisien tanpa mengurangi kualitas dari hasil yang diperoleh.

Jika data yang ada masih terdapat duplikasi atau anomaly maka terlebih dahulu dilakukan proses *replace missing* seperti gambar berikut:



Gambar 1 Model Desain *Replace Missing*

Hasil dari proses *replace missing* dapat terlihat pada tabel berikut:

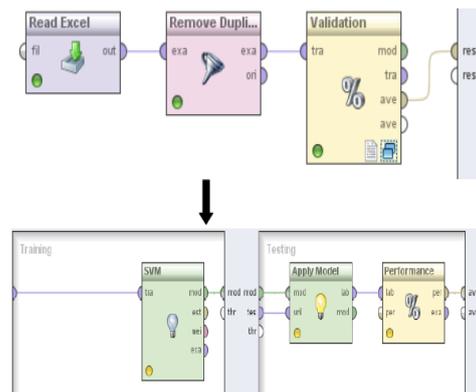
Role	Name	Type	Statistics	Range	Missings
label	CLASS	binomial	mode = y (150), least = 1 (120), y (150)		0
regular	AGE	integer	avg = 54.433 +/- 9.109	[29.000 ; 77.000]	0
regular	SEX	integer	avg = 0.678 +/- 0.468	[0.000 ; 1.000]	0
regular	CP	integer	avg = 3.174 +/- 0.950	[1.000 ; 4.000]	0
regular	RBP	integer	avg = 131.344 +/- 17.682	[94.000 ; 200.000]	0
regular	SC	integer	avg = 249.650 +/- 51.686	[129.000 ; 364.000]	0
regular	FBS	integer	avg = 0.148 +/- 0.356	[0.000 ; 1.000]	0
regular	RE	integer	avg = 1.022 +/- 0.988	[0.000 ; 2.000]	0
regular	IH	integer	avg = 149.678 +/- 23.166	[71.000 ; 202.000]	0
regular	EIVANG	integer	avg = 0.330 +/- 0.471	[0.000 ; 1.000]	0
regular	OLDPEAK	numeric	avg = 1.950 +/- 1.145	[0.000 ; 6.200]	0
regular	SLOPE	integer	avg = 1.585 +/- 0.614	[1.000 ; 3.000]	0

Gambar 2 Hasil *Replace Missing*

4. Tahap *Modelling*

Tahap *modelling* dilakukan untuk memilih dan menerapkan teknik pemodelan yang tepat, menentukan *tools data mining* yang digunakan, serta menentukan parameter dengan nilai yang optimal. Pada tahap ini juga dapat disebut sebagai tahap *learning* karena pada tahap ini data *training* dilatih oleh model yang dipilih.

Pada penelitian ini model yang digunakan adalah *Support Vector Machine* dengan algoritma *Particle Swarm Optimization*. Pengujian prediksi penyakit jantung dengan metode *Support Vector Machine* menggunakan *K-Fold Cross Validation* dengan desain sebagai berikut:



Gambar 3. Pengujian SVM dengan *K-Fold Cross Validation*

5. Tahap *Evaluation*

Evaluation adalah fase interpretasi terhadap hasil data mining yang ditunjukkan pada proses pemodelan fase sebelumnya. Evaluasi dilakukan secara mendalam dengan tujuan menyesuaikan model yang didapat agar sesuai dengan sasaran yang ingin dicapai dalam fase pertama. Pada tahap ini dilakukan pengujian terhadap model yang dipakai yaitu *Support Vector Machine* dan *Support Vector Machine* berbasis *Particle Swarm Optimization*.

IV. HASIL PENELITIAN

Pengukuran dari hasil penelitian ini adalah sebagai berikut:

4.1 *Support Vector Machine*

Nilai *training cycles* dalam penelitian ini ditentukan dengan cara melakukan uji coba memasukkan C, epsilon. Berikut ini adalah hasil dari percobaan yang telah dilakukan untuk penentuan nilai *training cycles*:

Tabel 2. Eksperimen penentuan nilai *training cycle SVM*

C	Epsilon	SVM	
		Accuracy	AUC

		(%)	
0.0	0.0	79.55	0.850
1.0	1.0	75	0.756
1.0	1.0	75	0.756
1.0	0.0	81.85	0.899
0.0	0.0	79.55	0.750
1.0	0.0	81.85	0.899

Hasil terbaik pada *eksperiment* SVM diatas adalah dengan C=0.0 dan Epsilon=0.0 dihasilkan accuracy 79,55 dan AUCnya 0,750 untuk SVM dengan C=1.0 dan Epsilon=0.0 dihasilkan accuracy 81,85 dan AUCnya 0,899.

4.2 Support Vector Machine Berbasis Particle Swarm Optimization

Nilai *training cycles* dalam penelitian ini ditentukan dengan cara melakukan uji coba memasukkan C, epsilon dan population size. Berikut ini adalah hasil dari percobaan yang telah dilakukan untuk penentuan nilai *training cycles*.

Tabel 3. Eksperiment penentuan nilai *training cycle* SVM berbasis PSO

C	Epsilon	SVM		Population Size	SVM-PSO	
		Accuracy	AUC		Accuracy	AUC
0.0	0.0	79.55	0.850	5	80.03	0.872
1.0	1.0	75	0.756	5	75.15	0.783
1.0	1.0	75	0.756	10	78.28	0.783
1.0	0.0	81.85	0.899	20	88.61	0.919
0.0	0.0	79.55	0.750	30	82.50	0.825
1.0	0.0	81.85	0.899	40	88.61	0.923

Hasil terbaik pada *eksperiment* SVM berbasis PSO diatas adalah dengan C=1.0 dan Epsilon=0 serta population size = 5 yang dihasilkan accuracy 88.61 % dan AUCnya 0.923 dan dengan C=1.0 dan Epsilon=0 serta population size = 20 untuk SVM berbasis PSO dihasilkan accuracy 88.61 dan AUCnya 0.919.

Tahap selanjutnya adalah menyeleksi atribut yang digunakan yaitu *age, sex, chest paint type, testing blood pressure, serum cholesterol, fasting blood sugar, resting electrocariographic result, maximum heart rate archived, exercise induced angina, old peak, the slope of the peak exercise ST segmen, number of major vessel, thal* dan 1 atribut sebagai label yaitu *class* . Dari hasil *eksperiment* dengan menggunakan algoritma *support vector machine* berbasis *particle swarm optimization* diperoleh hasil seperti dalam tabel dibawah ini:

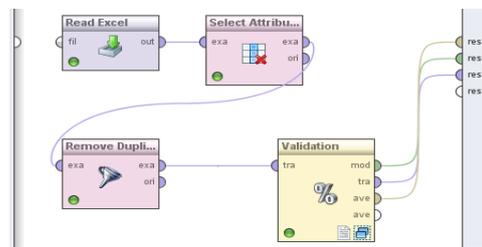
Tabel 4 Hasil Seleksi Atribut

Atribut	Weight
Age	0
Sex	0
Chest pain type	1
Testing Blood Pressure	0,223
Serum Cholesterol	0,029
Fasting Blood Sugar	1
Resting Electrocardiographic Result	0,16
Maximum heart rate Archeived	0,345
Exercise Induced Angina	0,333
Old Peak	0
The Slope of the peak exercise ST segmen	0,9
Number of Major Vessel	0
Thal	0,544

Dari 13 variabel yang telah dilakukan seleksi atribut sehingga menghasilkan 9 atribut terpilih yang digunakan, yaitu: *chest pain type, testing blood pressure, serum cholesterol, fasting blood pressure, resting electrocardiographic, maximum heart rate archeived, exercise induced angina, slope of the peak exercise ST segemn, thal*. Sedangkan *age, sex, old peak, dan number of major vessel* tidak berpengaruh terhadap bobot atribut.

4.3 Evaluasi dan Validasi Hasil

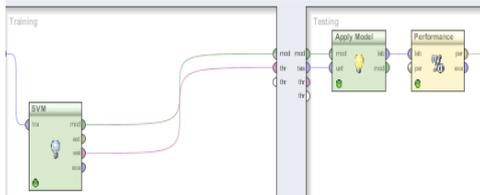
Hasil dari pengujian model yang dilakukan adalah memprediksi penyakit jantung dengan *support vector machine* dan *support vector machine* berbasis *particle swarm optimization* untuk menentukan nilai *accuracy* dan *AUC*. Dalam menentukan nilai tingkat keakurasian dalam model *support vector machine* dan *support vector machine* berbasis *particle swarm optimization*. Metode pengujiannya menggunakan *cross validation* dengan desain modelnya sebagai berikut:



Gambar 4. Desain Model Validasi

1. Hasil Pengujian Model Support Vector Machine

Pada penelitian penentuan hasil penyakit jantung menggunakan algoritma *Support Vector Machine* berbasis pada framework RapidMiner sebagai berikut:



Gambar 5. Model Pengujian validasi *Support Vector Machine*

Nilai *accuracy*, *precision*, dan *recall* dari *data training* dapat dihitung dengan menggunakan RapidMiner. Hasil pengujian dengan menggunakan model *Support Vector Machine* didapatkan hasil pada tabel berikut:

Confusion Matrix

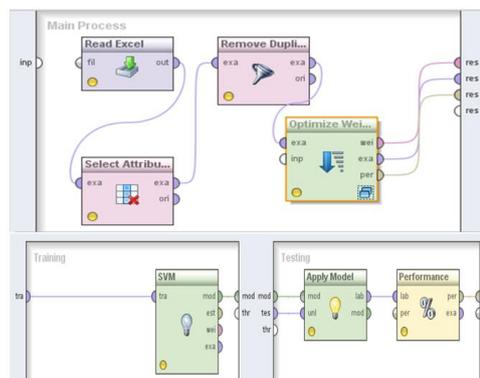
Pada tabel 5 diketahui dari 270 data, sebanyak 79 data diklasifikasikan ya sesuai dengan prediksi yang dilakukan metode SVM, kemudian 8 data diprediksi ya tetapi hasilnya ternyata tidak, 142 data class tidak diprediksi sesuai, dan 41 data diprediksi tidak ternyata hasil prediksinya ya.

Tabel 5 Model *Confusion Matrix* untuk Metode *Support Vector Machine*

accuracy: 81.85% +/- 7.49% (mikro: 81.85%)			
	True Ya	True Tidak	Class Precision (%)
Prediksi Ya	79	8	90.80
Prediksi Tidak	41	142	77.60
Class recall (%)	65.83	94.67	

2. Hasil Pengujian Model *Support Vector Machine* berbasis *Particle Swarm Optimization*

Pada penelitian penentuan hasil penyakit jantung menggunakan algoritma *Support Vector Machine* berbasis *Particle Swarm Optimization* (PSO) pada framework RapidMiner sebagai berikut:



Gambar 6. Model pengujian validasi *Support Vector Machine* berbasis *particle* berbasis *Particle Swarm Optimization* (PSO)

Gambar 6. Model pengujian validasi *Support Vector Machine* berbasis *particle* berbasis *Particle Swarm Optimization* (PSO)

Nilai *accuracy*, *precision*, dan *recall* dari *data training* dapat dihitung dengan menggunakan RapidMiner. Hasil pengujian dengan menggunakan model *Support Vector Machine* berbasis *Particle Swarm Optimization* didapatkan hasil sebagai berikut:

Confusion Matrix

Pada tabel 6 diketahui dari 270 data, sebanyak 85 data diklasifikasikan ya sesuai dengan prediksi yang dilakukan metode SVM dengan *Particle Swarm Optimization*, kemudian 5 data diprediksi ya tetapi hasilnya ternyata tidak, 154 data class tidak diprediksi sesuai, dan 26 data diprediksi tidak ternyata hasil prediksinya.

Tabel 6 Model *Confusion Matrix* untuk Metode *Support Vector Machine* berbasis *Particle Swarm Optimization*

accuracy: 88.61% +/- 4.06% (mikro: 88.60%)			
	True Ya	True Tidak	Class Precision (%)
Prediksi Ya	85	5	94.44
Prediksi Tidak	26	154	85.55
Class recall (%)	76.57	96.85	

Dari hasil pengujian diatas, baik evaluasi menggunakan *confusion matrix* terbukti bahwa hasil pengujian metode SVM berbasis PSO memiliki nilai akurasi yang lebih tinggi dibandingkan dengan metode SVM tunggal. Nilai akurasi untuk metode SVM sebesar 81.85% dan nilai akurasi untuk metode SVM berbasis PSO sebesar 88,61% dengan selisih akurasi sebesar 6.76% dapat dilihat pada tabel berikut:

Tabel 7 Pengujian algoritma SVM dan SVM berbasis PSO

	<i>Support Vector Machine</i>	<i>Support Vector Machine</i> dengan PSO	Peningkatan
Accuracy	81.85%	88.61%	6.76%
AUC	0.899	0.919	0.02

V. KESIMPULAN

Berdasarkan hasil *eksperiment* yang dilakukan untuk memecahkan masalah hasil prediksi penyakit jantung, dapat disimpulkan bahwa hasil *eksperiment* menggunakan metode *support vector machine* mempunyai tingkat akurasi sebesar 79.55 % dan mempunyai nilai AUC sebesar 0.850. Setelah dilakukan penyesuaian pada parameter C dan

epsilon didapat nilai akurasi terbaik untuk *support vector machine* yaitu mempunyai akurasi sebesar 81.85 % dan nilai AUCnya sebesar 0.899. Sedangkan *eksperiment* kedua yang dilakukan dengan menggunakan metode *support vector machine* berbasis *particle swarm optimization* mempunyai nilai akurasi sebesar 82.50 % dan nilai AUC sebesar 0.825. Setelah dilakukan penyesuaian pada parameter C dan epsilon dan *population* didapat nilai akurasi terbaik untuk metode *support vector machine* berbasis *particle swarm optimization* yaitu mempunyai akurasi sebesar 88.61 % dan nilai AUC sebesar 0.919.

DAFTAR PUSTAKA

- Abraham, Grosan. 2006. *Swarm Intelligence In Data Mining*. Verlag Berlin Heidelberg: Springer.
- Bellazzi, R., dan Zupanb, B. 2008. Predictive Data Mining In Clinical Medicine: Current Issues And And Guidelines. *International Journal Of Medical Informatics* 77, 81-97.
- Fei, S. W, Miao, Y. B, dan Liu, C. L. 2009. Chinese Grain Production Forecasting Method Based On Particle Swarm Optimization-Based Support Vector Machine. *Recent Patents On Engineering* 2009, 3, 8-12.
- Feng-Chia, L. 2009. Comparison of the Primitive Classifiers without Features Selection in Credit Scoring. *Management and Service Science*.
- Han, J., dan Kamber, M. 2006. *Data Mining Concept and Tehniques*. San Fransisco: Morgan Kauffman.
- Haupt, R. L., dan Haupt. 2004. *Practical Genetic Algorithms*. Untied States Of America: A John Wiley & Sons Inc Publication.
- Huang, K., Yang, H., King, I., dan Lyu, M. 2008. *Machine Learning Modeling Data Locally And Globally*. Berlin Heidelberg: Zhejiang University Press, Hangzhou And Springer-Verlag GmbH.
- Kothari, C. R. 2004. *Research Methology Methods and Techniques*. India: New Age International Limited.
- Liu, Y., Wang, G., Chen, H., dan Dong, H., 2011. An Improved Particle Swarm Optimization for Feature Selection. *Journal of Bionic Engineering Vol 8*, 1-10.
- Maimon, Oded dan Rokach, Lior. 2005. *Data Mining and Knowledge Discovey Handbook*. New York: Springer.
- Rajkumar, A., dan Reena, G. S. 2010. Diagnosis Of Heart Disease Using Data Mining Algorithm. *Global Journal of Computer Science and Technology, Vol. 10 Issue 10*, 38-43.
- Santoso, B. 2007. *Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu.
- Subbalakshmi, G., Ramesh, K., dan Chinna Rao, M. 2011. Decision Support in Heart Disease Prediction System using Naive Bayes. *Indian Journal of Computer Science and Engineering (IJCSSE)*, 170-176.
- Sugiyono. 2009. *Metode Penelitian Bisnis*. Bandung: Alfabeta
- Vercellis, C. 2009. *Business Intelligent: Data Mining and Optimization for Decision Making*. Southern Gate, Chichester, West Sussex: John Willey dan Sons, Ltd.
- Weiss, S. M., Indurkha, N., dan Zhang. 2010. *Fundamentals Of Predictive Text Mining*. London: Springer.

Biodata Penulis

Siti Nurajizah, dilahirkan di Jakarta pada tanggal 21 April 1985. Menamatkan sarjana (S1) jurusan Sistem Informasi pada tahun 2010 kemudian menyelesaikan pendidikan Pascasarjana pada tahun 2013 di STMIK Nusa Mandiri Jakarta. Saat ini beliau bekerja sebagai Staf Akademik di Bina Sarana informasi dan mengajar mata kuliah bidang aplikasi komputer. Untuk menghubungi beliau silahkan berkirin email ke : siti.snz@bsi.ac.id