

# PREDIKSI PENYAKIT DIA BETES MELLITUS DENGAN METODE *SUPPORT VECTOR MACHINE* BERBASIS *PARTICLE SWARM OPTIMIZATION*

Frisma Handayanna

**Abstract**—Diabetes at this time has increased the number of its patient. Diabetes is a disease that can cause complications even can causes death. In this research made model algorithm *Support Vector Machines* and model algorithm *Support Vector Machines* based on *Particle Swarm Optimization* to get the rule to predict the disease diabetes and give a more accurate value of the accuracy. Because there are still a lot of research using *Support Vector Machines* in predicting diabetes but the accuracy of the resulting value is still less accurate.. After the testing with two models that *Support Vector Machines* algorithms and *Support Vector Machines* based on *Particle Swarm Optimization* and so test the results are by using *Support Vector Machines* are get accuracy values 74.21% and AUC values was 0.758, while testing by *Support Vector Machines* based *particle swarm optimization* are get value accuracy 77.36 % and value AUC is 0.765 to level diagnose good classification. The two this method having the different levels of accuracy is as much as 3.15 % and the difference in value AUC of 0,017.

**Intisari**—Penyakit diabetes saat ini semakin lama semakin meningkat jumlah penderitanya. Penyakit diabetes adalah salah satu penyakit yang dapat menyebabkan komplikasi bahkan dapat menyebabkan kematian. Dalam penelitian ini dibuatkan model algoritma *Support Vector Machines* dan model algoritma *Support Vector Machines* berbasis *Particle Swarm Optimization* untuk mendapatkan rule dalam memprediksi penyakit diabetes dan memberikan nilai akurasi yang lebih akurat. Dikarenakan masih banyak penelitian yang menggunakan metode *Support Vector Machines* dalam memprediksi penyakit diabetes tetapi nilai akurasi yang dihasilkan masih kurang akurat. Setelah dilakukan pengujian dengan dua model yaitu Algoritma *Support Vector Machines* dan *Support Vector Machines* berbasis *Particle Swarm Optimization* maka hasil yang didapat adalah algoritma sehingga didapat pengujian dengan menggunakan *Support Vector Machines* dimana didapat nilai *accuracy* adalah 74.21% dan nilai AUC adalah 0.758, sedangkan pengujian dengan menggunakan *Support Vector Machines* berbasis *Particle Swarm Optimization* didapatkan nilai *accuracy* 77.36% dan nilai AUC adalah 0.765 dengan tingkat diagnosa *good classification*. Sehingga kedua metode tersebut memiliki perbedaan tingkat akurasi yaitu sebesar 3.15% dan perbedaan nilai AUC sebesar 0,017.

**Kata Kunci**— Diabetes, *Particle Swarm Optimization*, *Support*

---

Program Studi Teknik Informatika STMIK Nusamandiri Jakarta,  
Jl. Damai No. 8 Warung Jati Barat (Margasatwa) Jakarta  
Selatan. Telp. (021) 78839513 Fax. (021) 78839421, email:  
frisma.fha@nusamandiri.ac.id

*Vector Machine*.

## I. PENDAHULUAN

Diabetes adalah penyakit yang salah satunya disebabkan oleh pola makan yang tidak sehat, produksi insulin yang berlebihan, dan keturunan apabila tidak terdeteksi secara cepat akan menyebabkan masalah kesehatan bagi tubuh anda. Penyakit diabetes disebabkan oleh peningkatan kadar glukosa dalam darah, apabila kadar glukosa darah meningkat dalam jangka waktu yang lama maka akan menyebabkan komplikasi seperti gagal ginjal, kebutaan dan serangan jantung [12]. Kontrol glukosa darah merupakan hal terpenting dalam praktek medis penyakit diabetes dan penyakit kritis lainnya [10]. Kelainan darah diabetes dan gula lain disebabkan oleh apa yang kita makan dan bagaimana cara kita hidup [16]. Perkiraan terakhir populasi penderita penyakit diabetes menunjukkan 171 juta orang di dunia pada tahun 2000 dan diperkirakan akan meningkat menjadi 366 juta pada 2030 [18].

Penyakit diabetes perlu diprediksi dengan akurat karena penyakit diabetes merupakan penyakit sosial yang serius dan bisa terkena orang dalam jumlah besar, serta menyebabkan komplikasi dan melibatkan biaya yang tinggi serta dapat meningkatkan keadaan sakit melalui penyakit diabetes terutama pada anak-anak dan anak muda [11]. Sehingga untuk menghindari penyakit diabetes diupayakan kita memiliki gaya hidup yang sehat serta tidak makan berlebihan dari apa yang diperlukan oleh tubuh.

Penelitian yang dilakukan oleh He Chunjian dan Zhang Cuilian Zhao Yan dengan judul *A New SVM Merged into Data Information*, dengan metode kernel fungsi dimana beberapa kernel dilatih dan kernel terbaik tampil di set validasi kemudian dipilih untuk pengujian dan kinerjanya dievaluasi pada set tes dan menunjukkan bahwa pendekatan secara efektif dapat meningkatkan klasifikasi akurasi [3].

*Support Vector Machines* adalah kasus khusus dari keluarga algoritma yang kita sebut sebagai *regularized* metode klasifikasi linier dan metode yang kuat untuk minimalisasi resiko [20]. Dan kelebihan *Support Vector Machines* lainnya adalah dapat meminimalkan kesalahan melalui memaksimalkan margin dengan misahkan antara hyper-lane dan satu set data bahkan dengan jumlah sample yang kecil [3].

*Particle Swarm Optimization* banyak digunakan untuk memecahkan masalah optimasi, serta sebagai masalah seleksi fitur [14]. Dalam teknik *Particle Swarm Optimization* terdapat beberapa cara untuk melakukan pengoptimasian diantaranya: meningkatkan bobot atribut (*attribute weight*) terhadap semua atribut atau variabel yang dipakai, menseleksi atribut (*attribute selection*), dan *feature selection*.

## II. KAJIAN LITERATUR

### a. Diabetes

Nama pendek untuk penyakit diabetes disebut mellitus, diabetes terjadi ketika tubuh tidak dapat menggunakan glukosa darah sebagai energi karena memiliki terlalu sedikit insulin atau tidak mampu menggunakan insulin [2]. Diabetes adalah salah satu penyebab utama kematian di banyak negara dan penyebab utama kebutaan, gagal ginjal, dan *nontraumatic* amputasi [19]. Diabetes dapat menyebabkan masalah kesehatan dari waktu ke waktu. Dapat melukai mata Anda, ginjal, dan saraf Anda. Hal ini dapat menyebabkan masalah dengan aliran darah dalam tubuh Anda. Bahkan gigi dan gusi dapat dirugikan. Diabetes pada kehamilan dapat menyebabkan masalah khusus [2]. Faktor penyebab diabetes adalah Gen diabetes dalam keluarga, insulin dan gula darah, kegemukan (Obesitas), asma, KB [17].

Ada tiga tipe utama diabetes:

1. Diabetes tipe 1  
Sebuah kondisi di mana pancreas membuat insulin begitu sedikit bahwa tubuh tidak dapat menggunakan darah glukosa sebagai energi. Orang dengan diabetes tipe 1 harus mengambil insulin setiap hari [2].
2. Diabetes tipe 2  
Kondisi di mana tubuh baik membuat terlalu sedikit insulin atau tidak dapat menggunakan insulin itu membuat menggunakan glukosa darah sebagai energy [2].
3. Diabetes gestasional  
Karena hanya mempengaruhi wanita hamil. Untuk beberapa alasan, wanita hamil lebih rentan terhadap diabetes daripada orang lain [16].

Faktor resiko diabetes gestasional dapat dilihat dari berapa kali keguguran, pernah melahirkan anak mati tanpa sebab, pernah melahirkan bayi 4000 gram, umur > 30 tahun, riwayat diabetes dalam keluarga, pernah terkena diabetes gestasional pada kehamilan sebelumnya, kegemukan, berat badan ibu waktu melahirkan > 5 kg, Infeksi saluran kemih berulang-ulang.

### b. Data Mining

Data Mining merupakan teknologi baru yang sangat berguna untuk membantu perusahaan-perusahaan menemukan informasi yang sangat penting dari gudang data

mereka. Beberapa aplikasi data mining fokus pada prediksi, mereka meramalkan apa yang akan terjadi dalam situasi baru dari data yang menggambarkan apa yang terjadi di masa lalu [21]. Secara khusus, koleksi metode yang dikenal sebagai 'data mining' menawarkan metodologi dan solusi teknis untuk mengatasi analisis data medis dan konstruksi prediksi model [1]. Model data mining memberikan contoh penerapannya pada berbagai algoritma dan pada data set yang besar [13].

Tahapan data mining dalam proses penemuan pengetahuan [7]:

1. Pembersihan data
2. Integrasi data
3. Data seleksi
4. Data transformasi
5. Data mining
6. Pola evaluasi
7. Pengetahuan presentasi

Terdapat empat pengelompokan dalam data mining yaitu klasifikasi, asosiasi, clustering dan prediksi [21]:

Konsep data mining, menemukan pola berharga dalam data, adalah respon yang jelas untuk pengumpulan dan penyimpanan volume data yang besar [20]. Semakin kecil perbedaan antara apa yang diharapkan terjadi (hasil yang diharapkan) dan apa yang sebenarnya terjadi (diamati hasil), semakin baik prediksi, contohnya prediksi ramalan cuaca (misalnya, untuk 24 atau 48 jam) atau diagnosis untuk penyakit tertentu yang diberikan kepada pasien tertentu, yang didasarkan pada data medis [6].

### c. Support Vector Machine

*Support Vector Machine* (SVM) adalah seperangkat metode yang terkait untuk suatu metode pembelajaran, untuk kedua masalah klasifikasi dan regresi [15]. *Support Vector Machine* adalah sebuah metode seleksi yang membandingkan parameter standarseperangkatnilai diskrit yang disebut kandidat set, dan mengambilsalah satu yang memiliki akurasi klasifikasi terbaik [4]. Dengan berorientasi pada tugas, kuat, sifat komputasi mudah dikerjakan, SVM telah mencapai sukses besar dan dianggap sebagai *state-of-the-art classifier* saat ini [9].

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

$$\vec{w} \cdot \vec{x} + b = 0 \dots\dots\dots (1)$$

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

$$\vec{w} \cdot \vec{x} + b = -1 \quad \dots\dots\dots (2)$$

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

$$\vec{w} \cdot \vec{x} + b = +1 \quad \dots\dots\dots (3)$$

Margin terbesar dapat ditemukan dengan memaksimalkan nilai jarak antara hyperplane dan titik terdekatnya, yaitu  $1/\|\vec{w}\|$ . Hal ini dapat dirumuskan sebagai Quadratic Programming (QP) problem, yaitu mencari titik minimal persamaan (4), dengan memperhatikan constraint persamaan (5)

$$\min_w \tau(w) = \frac{1}{2} \|\vec{w}\|^2 \quad \dots\dots\dots (4)$$

$$y_i(\vec{x}_i \cdot \vec{w} + b) - 1 \geq 0, \quad \forall i \quad \dots\dots\dots (5)$$

Problem ini dapat dipecahkan dengan berbagai teknik komputasi, diantaranya *Lagrange Multiplier* sebagaimana ditunjukkan pada persamaan (6)

$$L(w, b, \alpha) = \frac{1}{2} \|\vec{w}\|^2 + \sum_{i=1}^l \alpha_i (y_i(\vec{x}_i \cdot \vec{w} + b) - 1) \quad \dots\dots\dots (6)$$

$\alpha_i$  adalah Lagrange multipliers, yang bernilai nol atau positif ( $\alpha_i \geq 0$ ). Nilai optimal dari persamaan (6) dapat dihitung dengan meminimalkan L terhadap  $\vec{w}$  dan b, dan memaksimalkan L terhadap  $\alpha_i$ . Dengan memperhatikan sifat bahwa pada titik optimal gradient L=0, persamaan langkah (6) dapat dimodifikasi sebagai maksimalisasi problem yang hanya mengandung  $\alpha_i$  saja, sebagaimana persamaan (7).

Maximize:

$$\sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j=1}^l \alpha_i \alpha_j y_i y_j x_i x_j \quad \dots\dots\dots (7)$$

Subject to:

$$\alpha_i \geq 0 (i = 1, 2, \dots, l \quad \sum_{i=1}^l \alpha_i y_i = 0) \quad \dots\dots\dots (8)$$

Dari hasil dari perhitungan ini diperoleh  $\alpha_i$  yang kebanyakan bernilai positif. Data yang berkorelasi dengan  $\alpha_i$  yang positif inilah yang disebut sebagai *support vector*.

d. *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 [8].

*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 [5]. 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:

$$v_{ij}(t+1) = w * v_{ij}(t) + c_1 * rand_1 * (pbest_{ij}(t) - p_{ij}(t)) + c_2 * rand_2 * (gbest_{ij}(t) - p_{ij}(t)) \quad \dots\dots\dots (9)$$

$$p_{ij}(t+1) = p_{ij}(t) + \beta * v_{ij}(t+1) \quad \dots\dots\dots (10)$$

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)  
 $rand_1$  dan  $rand_2$  = fungsi acak di rentang [0, 1]  
 $\beta$ = faktor kendala untuk mengontrol kecepatan berat (nilainya ke 1)  
 $c_1$  dan  $c_2$  adalah faktor pembelajaran pribadi dan sosial (nilainya ke 2)

III. METODE PENELITIAN

Dalam penelitian ini dilakukan beberapa langkah yang dilakukan dalam proses penelitian.

1. Pengumpulan data  
 Pada tahap ini dicari data yang tersedia, memperoleh data tambahan yang dibutuhkan, mengintegrasikan semua data kedalam data set, termasuk variabel yang diperlukan dalam proses.
2. Pengolahan data awal  
 Ditahap ini dilakukan penyeleksian data, data dibersihkan dan ditransformasikan ke bentuk yang diinginkan sehingga dapat dilakukan persiapan dalam pembuatan model.
3. Metode yang diusulkan  
 Pada tahap ini data dianalisis, dikelompokkan variabel mana yang berhubungan dengan satu sama lainnya. Setelah data dianalisis lalu diterapkan model-model yang sesuai dengan jenis data. Pembagian data kedalam data latihan (*training data*) dan data uji (*testing data*) juga diperlukan untuk pembuatan model.
4. Eksperimen dan pengujian metode  
 Pada tahap ini model yang diusulkan akan diuji untuk melihat hasil berupa *rule* yang akan dimanfaatkan dalam pengambilan keputusan.
5. Evaluasi dan validasi  
 Pada tahap ini dilakukan evaluasi terhadap model yang ditetapkan untuk mengetahui tingkat keakuratan model.

TABEL 1. ATRIBUT DAN DATA PENYAKIT DIABETES

Sumber: (UCI Repository)

No	Berapa Kali Hamil	Konsentrasi Glukosa	Tekanan darah	Lipatan kulit	Serum Insulin	Masssa Tubuh	Diabetes Silsilah Fungsi	Umur	Kelas
1	6	14	148	35	0	33,6	0,627	50	Ya
2	1	85	66	29	0	26,6	0,351	31	Tidak
3	8	18	64	0	0	23,3	0,672	32	Ya
4	1	89	66	23	94	28,1	0,167	21	Tidak
5	0	13	40	35	168	43,1	2,288	33	Ya
6	5	11	74	0	0	25,6	0,201	30	Tidak
7	3	78	50	32	88	31	0,248	26	Ya
8	10	115	115	0	0	35,3	0,134	29	Tidak
9	2	19	70	45	543	30,5	0,158	53	Ya
10	8	12	96	0	0	0	0,232	54	Ya
11	4	11	92	0	0	37,6	0,191	30	Tidak
12	10	168	74	0	0	38	0,537	34	Ya

**Teknik pengumpulan data**

Teknik pengumpulan data yang diperoleh adalah data sekunder karena diperoleh dari Pima Indian diabetes database dalam UCI (singkatan dari Pima Diabetes). Masalah yang harus dipecahkan di sini adalah prediksi terjadinya diabetes melitus dalam waktu 5 tahun dengan menggunakan Pima yang berisi 786 orang yang diperiksa dan sebanyak 500 pasien tidak terdeteksi terkena penyakit diabetes, sehingga 268 pasien terdeteksi penyakit diabetes. Data pasien penyakit diabetes bisa di lihat pada Tabel 1.

**Pengolahan data awal**

Jumlah data awal yang diperoleh dari pengumpulan data yaitu sebanyak 768 data, namun tidak semua data dapat digunakan dan tidak semua atribut digunakan karena harus melalui beberapa tahap pengolahan awal data (*preparation data*).

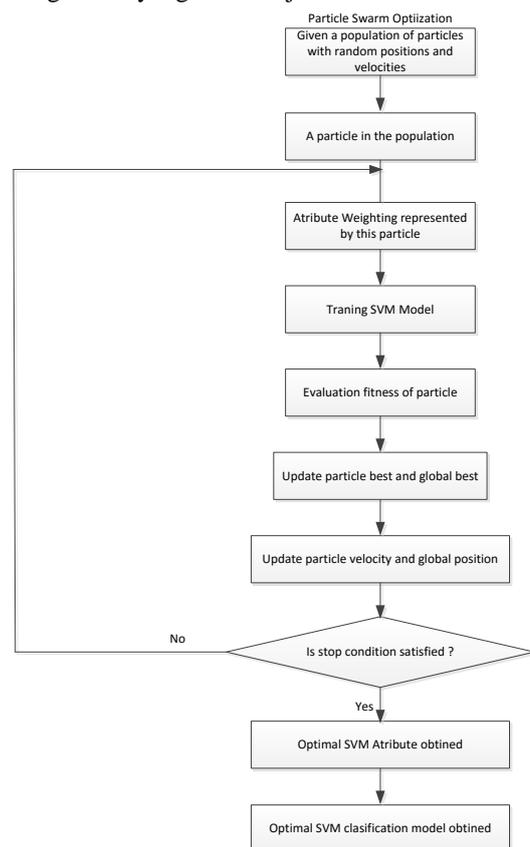
TABEL 2. TABEL ATRIBUT YANG DIGUNAKAN

Sumber: (UCI Repository)

No	Atribut	Nilai
1	Berapa Kali Hamil	Berapa kali wanita hamil
2	Konsentrasi Glukosa	Konsentrasi glukosa plasma 2 jam dalam tes toleransi glukosa oral
3	Tekanan Darah	Tekanan Darah diastolik (mmHg)
4	Lipatan Kulit	Triceps ketebalan lipatan kulit (mm)
5	Serum Insulin	2-Jam serum insulin (mu U / ml)
6	Masssa Tubuh	Indeks massa tubuh (berat dalam kg / (tinggi dalam m) ^ 2)
7	Diabetes Silsilah Fungsi	Diabetes silsilah fungsi
8	Umur	Umur (tahun)

**Metode yang diusulkan**

Dengan memasukan data penyakit diabetes kemudian dianalisa dan dikomparasi. Berikut ini bentuk gambaran metode algoritma yang akan diuji.



Gbr 1. Metode yang diusulkan

1. Eksperimen dan Pengujian Metode

Tahap modeling untuk menyelesaikan prediksi penyakit diabetes dengan menggunakan dua metode yaitu algoritma *Support Vector Machine*. *Support Vector Machine* yaitu suatu metode sebuah metode seleksi fitur, dan mengambil salah satu yang memiliki akurasi klasifikasi terbaik.

2. Evaluasi dan Validasi Hasil

Model yang diusulkan pada penelitian tentang prediksi penyakit diabetes adalah dengan menerapkan *Support Vector Machine* dan *Support Vector Machine* berbasis *Particle Swarm Optimization*. Penerapan algoritma *Support Vector Machine* dengan menentukan nilai *weight* terlebih dahulu. Setelah didapatkan nilai akurasi dan AUC terbesar, nilai *weight* tersebut akan dijadikan nilai yang akan digunakan untuk mencari nilai akurasi dan AUC tertinggi.

IV. HASIL DAN PEMBAHASAN

Hasil Eksperimen dan Metode

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 3. EKSPERIMENT PENENTUAN NILAI TRAINING CYCLE SVM

Sumber: Hasil Penelitian (2012)

C	epsilon	SVM	
		accuracy	AUC
0.0	0.0	74,21%	0,753
1.0	1.0	65%	0,500
1.0	1.0	65%	0,500
1.0	0.0	74,21%	0,758
1.0	1.0	65%	0,500
1.0	1.0	65%	0,500
1.0	1.0	65%	0,500
1.0	0.0	74,21%	0,758
1.0	0.0	74,21%	0,758

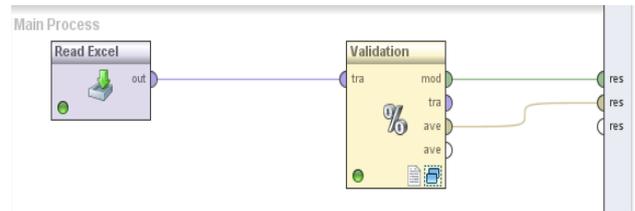
Hasil terbaik pada *eksperimen SVM* diatas adalah dengan C=0.0 dan Epsilon=0.0 dihasilkan *accuracy* 74,21% dan AUCnya 0.753 untuk SVM dengan C=1.0 dan Epsilon=0.0 dihasilkan *accuracy* 74,21% dan AUCnya 0.758.

1. Evaluasi dan Validasi Hasil

a. Hasil Pengujian Model Support Vector Machine

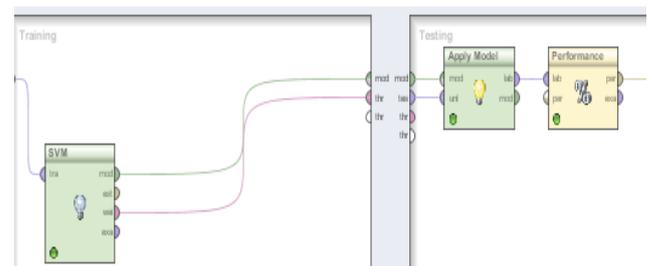
Hasil dari pengujian model yang dilakukan adalah memprediksi penyakit diabetes dengan *Support Vector Machine* untuk menentukan nilai *accuracy* dan AUC. Dalam menentukan nilai tingkat keakurasian dalam model *Support*

*Vector Machine*. Metode pengujiannya menggunakan *cross validation* dengan desain modelnya sebagai berikut.



Gambar 2. Desain model Validasi  
Sumber: Hasil Penelitian (2012)

Pada penelitian penentuan hasil penyakit diabetes menggunakan algoritma *Support Vector Machine* berbasis pada framework Rapid Miner sebagai berikut



Gbr 3. Model pengujian validasi Support Vector Machine  
Sumber: Hasil Penelitian (2012)

Nilai *accuracy*, *precision*, dan *recall* dari *data training* dapat dihitung dengan menggunakan Rapid Miner. Hasil pengujian dengan menggunakan model *Support Vector Machine* didapatkan hasil pada Tabel 4.

Tabel 4 diketahui dari 768 data, 118 diklasifikasikan ya sesuai dengan prediksi yang dilakukan dengan metode SVM, lalu 48 data diprediksi ya tetapi ternyata hasilnya prediksi tidak, 452 data *class tidak* diprediksi sesuai, dan 150 data diprediksi tidak ternyata hasil prediksinya ya.

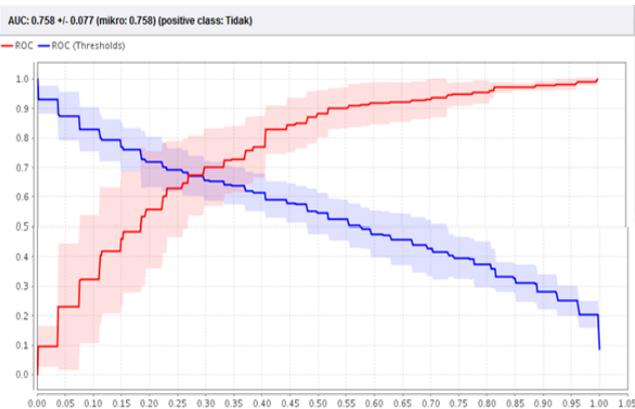
TABEL 4. MODEL CONFUSION MATRIX UNTUK METODE SUPPORT VECTOR MACHINE

accuracy:74.21% +/-5.79% (mikro: 74.22%)			
	True Ya	True Tidak	Class precission
pred. Ya	118	48	71.08%
pred. Tidak	150	452	75.08%
class recall	44.03	90.40%	

Sumber: Hasil Penelitian (2012)

Hasil perhitungan divisualisasikan dengan kurva ROC. Perbandingan kedua metode komparasi bisa dilihat pada Gambar 4 yang merupakan kurva ROC untuk algoritma *Support Vector Machine*es Kurva ROC pada gambar 4 mengekspresikan *confusion matrix* dari Tabel 4. Garis

horizontal adalah *false positives* dan garis vertikal *true positives*.



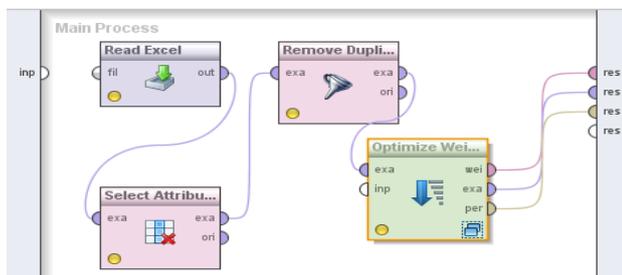
Gbr 4. Kurva ROC dengan Metode *Support Vector Machine*  
Sumber: Hasil Penelitian (2012)

Dari Gambar 4 terdapat grafik ROC dengan nilai AUC (*Area Under Curve*) sebesar 0.758 dimana diagnosa hasilnya *Fair classification*.

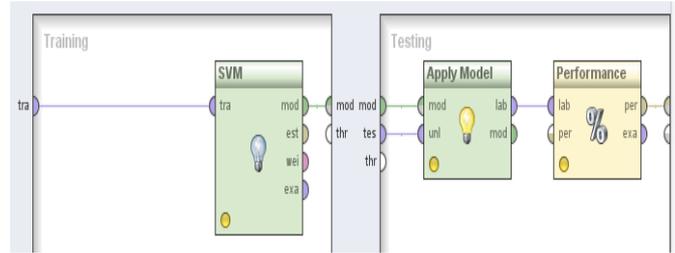
Berdasarkan hasil *eksperiment* yang dilakukan untuk memecahkan masalah prediksi hasil prediksi penyakit diabetes, dapat disimpulkan bahwa hasil *eksperiment* menggunakan metode *Support Vector Machine* mempunyai tingkat akurasi sebesar 74.21 % dan mempunyai nilai AUC sebesar 0.753. Setelah dilakukan penyesuaian pada parameter C dan epsilon didapat nilai akurasi terbaik untuk algoritma *Support Vector Machine* yaitu mempunyai akurasi sebesar 74.21 % dan nilai AUCnya sebesar 0.758.

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

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



Gbr 5. Desain model Validasi *Support Vector Machine* berbasis *particle* berbasis *Particle Swarm Optimization* (PSO)  
Sumber: Hasil Penelitian (2012)



Sumber: Hasil Penelitian (2012)

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

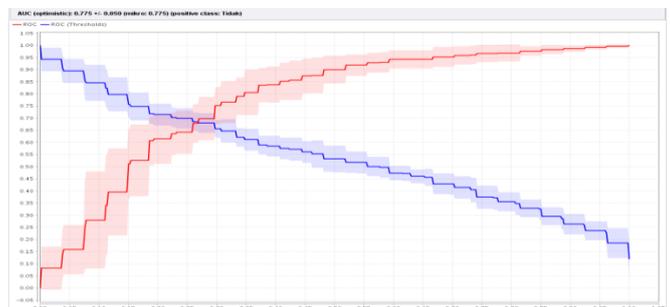
Tabel 5 diketahui dari 768 data, 129 diklasifikasikan ya sesuai dengan prediksi yang dilakukan dengan metode SVM, lalu 35 data diprediksi ya tetapi ternyata hasilnya prediksi tidak, 465 data *class tidak* diprediksi sesuai, dan 139 data diprediksi tidak ternyata hasil prediksinya ya.

TABEL 5.  
MODEL *CONFUSION MATRIX* UNTUK METODE *SUPPORT VECTOR MACHINE* BERBASIS *PARTICLE SWARM OPTIMIZATION*

accuracy:77.36% +/-4.06% (mikro: 77.34%)			
	True Ya	True Tidak	Class precission
pred. Ya	129	35	78.66%
pred. Tidak	139	465	76.99%
class recall	48.13%	93.00%	

Sumber: Hasil Penelitian (2012)

Berdasarkan Tabel 5.tersebut menunjukkan bahwa, tingkat akurasi dengan menggunakan algoritma SVM berbasis *Particle Swarm Optimization* (PSO) adalah sebesar 77,36%. Hasil perhitungan divisualisasikan dengan kurva ROC. Perbandingan kedua metode komparasi bisa dilihat pada Gambar 6 yang merupakan kurva ROC untuk algoritma *Support Vector Machine* berbasis *Particle Swarm Optimization* (PSO).



Gbr 7. Kurva ROC dengan Metode *Support Vector Machine* berbasis *Particle Swarm Optimization* (PSO)

Sumber: Hasil Penelitian (2012)

Dari hasil pengujian diatas, baik evaluasi menggunakan *counfusion matrix* maupun ROC curve terbukti bahwa hasil

pengujian algoritma SVM berbasis PSO memiliki nilai akurasi yang lebih tinggi dibandingkan dengan algoritma SVM Nilai akurasi untuk model algoritma SVM sebesar 74.21% dan nilai akurasi untuk model algoritma SVM berbasis PSO sebesar 77.36 % dengan selisih akurasi 3.15%, dapat dilihat pada Tabel 6 dibawah ini:

TABEL 6.  
PENGUJIAN ALGORITMA SVM DAN SVM BERBASIS PSO

	Accuracy	AUC
SVM	74.21%	0.758
SVM berbasis PSO	77.36 %	0.775

Sumber: Hasil Penelitian (2012)

Untuk evaluasi menggunakan ROC curve sehingga menghasilkan nilai AUC (*Area Under Curve*) untuk model algoritma SVM mengasilkan nilai 0.758 dengan nilai diagnosa *Fair Classification*, sedangkan untuk algoritma *Support Vector Machine* (SVM) berbasis PSO (*Particle Swarm Optimization*) menghasilkan nilai 0.775 dengan nilai *diagnose Fair Classification*, dan selisih nilai keduanya sebesar 0.017. Dengan demikian algoritma *Support Vector Machine* (SVM) berbasis *Particle Swarm Optimization* (PSO) dapat memberikan solusi untuk permasalahan dalam prediksi hasil prediksi penyakit diabetes.

#### V. KESIMPULAN

Berikut ini kesimpulan yang penulis ambil setelah melakukan penelitian

1. Pengujian model dengan menggunakan *Support Vector Machine* dan *Support Vector Machine* berbasis *Particle Swarm Optimization* dengan menggunakan data penyakit diabetes yang terkena penyakit atau tidak. Model yang dihasilkan diuji untukmendapatkan nilai *accuracy*, *precision* dan AUC dari setiap algoritmasehingga didapat pengujian dengan menggunakan *Support Vector Machine* didapat nilai *accuracy* adalah 74.21 % dengan nilai *precision* 74.75 % dan nilai AUC adalah 0.753
2. Pengujian dengan menggunakan support vector machines berbasis *Particle Swarm Optimization* (PSO) didapatkan nilai *accuracy* 77.36% dengan nilai dan nilai AUC adalah 0.775. Maka dapat disimpulkan pengujian pengujian data diabetes UCI data set menggunakan *Support Vector Support Vector Machines* berbasis *Particle Swarm Optimization* (PSO) lebih baik dari pada *Support Vector Machines* sendiri.

Pada bagian ini, penulis memberikan saran-saran berdasarkan permasalahan serta kesimpulan yang penulis dapat selama penelitian, yaitu :

1. Penelitian ini diharapkan dapat digunakan sebagai bahan pertimbangan memprediksi penyakit diabetes oleh pihak medis, sehingga dapat meningkatkan akurasi dalam prediksi prediksi penyakit diabetes.

2. Penelitian ini dapat dikembangkan dengan metode optimasi lainnya seperti *Ant Colony Optimization* (ACO), dan lainnya.
3. Penelitian ini dapat dikembangkan dengan metode klasifikasi data mining lainnya seperti *Neural Network*, *Naive Bayes*, *KNN* dan lainnya untuk melakukan perbandingan.

#### REFERENSI

- [1] Bellazzi, R., & Zupanb, B. (2008). Predictive Data Mining In Clinical Medicine: Current Issues And And Guidelines. *International Journal Of Medical Informatics* 77 , 81–97.
- [2] Centers for Disease Control and Prevention. *Take Charge of Your Diabetes*. 4th edition. Atlanta:U.S. Department of Health and Human Services, 2007.
- [3] Chunjiang, He. Cuilian, Zhang. Yan, Zhao. A New Svm Merged Into Data Information. *IEEE Asia-Pacific Conference*, 14-17. 2009.
- [4] Dong, Y., Xia, Z., Tu, M., & Xing, G. (2007). An Optimization Method For Selecting Parameters In Support Vector Machine. *Sixth International Conference On Machine Learning And Applications* , 1
- [5] Fei, S. W., Miao, Y. B., & 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.
- [6] Gorunescu, F. (2011). *Data Mining Concepts, Models And Techniques*. Verlag Berlin Heidelberg: Springer.
- [7] Han, J., dan Kamber, M. (2007). *Data Mining Concepts And Techniques*. San Francisco: Morgan Kaufmann Publisher.
- [8] Haupt, R. L., dan Haupt, S. E. (2004). *Practical Particle Swarm Optimizations*. Untied States Of America: A John Wiley & Sons Inc Publication.
- [9] Huang, K., Yang, H., King, I., & Lyu, M. (2008). *Machine Learning Modeling Data Locally And Globally*. Berlin Heidelberg: Zhejiang University Press, Hangzhou And Springer-Verlag GmbH.
- [10] Iancu, E., Iancu, I., & Sfredel, V. (2010). Predictive Control Of Blood Glucose In Diabetes Mellitus Patients. *International Conference On Automation, Quality And Testing, Robotics* , 1-6.
- [11] Iancu, I., Mota, M., & Iancu, E. (2008). Method For The Analysing Of Blood Glucose Dynamics In Diabetes Mellitus Patients. *International Conference On Automation, Quality And Testing, Robotics* , 60-65.
- [12] Jayalshmi dan Santhakumaran. Impact of Preprocessing for Diagnosis of Diabetes Mellitus Using Artificial Neural Networks. *IEEE International Conference* , 109-112. 2012
- [13] Larose, D. T. (2007). *Data Mining Methods And Models*. New Jersey: A John Wiley & Sons.
- [14] Liu, Y., Wang, G., Chen, H., Dong, H., Zhu, X., & Wang, S. (2011). An Improved Particle Swarm Optimization for Feature Selection. *Journal of Bionic Engineering Vol 8* , 1-10.
- [15] Maimon, O. (2012). *Data Mining And Knowledge Discovery Handbook*. New York Dordrecht Heidelberg London: Springer.
- [16] Mason, R. (2005). *The Natural Diabetes Cure*. Usa: 4th Printing Spring 2012.
- [17] Nurrahmani, U. (2012). *Stop! Diabetes Mellitus*. Yogyakarta: Familia. Moertini, V. S. (2002). *Data Mining Sebagai Solusi Bisnis*. Integral, Vol. 7 No. 1, April 2002 ,
- [18] Report Who. *Definition And Diagnosis Of Diabetes Mellitus And Intermediate Hyperglycemia*. Switzerland: Who Document Production Services. 2006.
- [19] Robert, F. G., Zgonis, T., & Driver, V. R. (2006). *Diabetic Foot Disorders: A Clinical Practice Guideline (2006 Revision)*. The Journal Of Foot & Ankle Surgery , 3.
- [20] Weiss, S. M., Indurkha, N., & Zhang, T. (2012). *Fundamentals Of Predictive Text Mining*. London: Springer.

- [21] Witten, I. H., Frank, E., & Hall, M. A. (2011). Data Mining Practical Machine Learning Tools And Techniques. Burlington, Usa: Morgan Kaufmann Publishers.



Friska Handayanna, M.Kom. lulus Tahun 2006 Diploma Tiga (DIII) Jurusan Komputer Akutansi AMIK BSI Jakarta. Tahun 2012 lulus dari Program Strata Satu (S1) Program Studi Sistem Informasi STMIK Nusa Mandiri Jakarta lulus dari Program Strata Dua (S2) Prodi Ilmu Komputer STMIK Nusa Mandiri Jakarta Tahun 2012. Aktif mengajar di STMIK Nusa Mandiri Jakarta. Telah melakukan penulisan paper di Jurnal STMIK Antarbangsa Jurnal Sistem Informasi No ISSN 2089-8711 Vol. IV No.1 Februari 2015, Jurnal STMIK Antarbangsa Jurnal Teknik Informatika Vol. I No. 2 Agustus 2015 ISSN. 2442-2444.