

# Metode SVM Berbasis PSO untuk Meningkatkan Prediksi Ketepatan Waktu Kelulusan Mahasiswa

Endang Supriyadi

**Abstract**— This scientific article discusses how the swarm optimization particle method can improve accuracy in predicting student graduation accuracy. This is done in order to anticipate the decrease in the number of students in their lectures. The experimental results have shown that particle swarm optimization (PSO) can improve the accuracy of the support vector machine method from 80.14% accuracy to 82.05% so that there is an increase of accuracy of 1.91%.

**Intisari**— Artikel ilmiah ini membahas bagaimana metode partikel swarm optimization dapat meningkatkan akurasi dalam memprediksi ketepatan kelulusan mahasiswa. Hal tersebut dilakukan agar dapat mengantisipasi penurunan jumlah mahasiswa dalam kegiatan perkuliahannya. Hasil eksperimen yang telah dilakukan terbukti bahwa particle swarm optimization ( PSO ) dapat meningkatkan akurasi metode support vector machine dari akurasi sebesar 80.14% menjadi 82,05% sehingga terbukti ada peningkatan akurasi sebesar 1.91 %.

**Kata Kunci**— SVM, PSO, akurasi, prediksi, data mining.

## I. PENDAHULUAN

Semua lembaga pendidikan tinggi memiliki tujuan utama diantaranya adalah menyediakan pendidikan berkualitas bagi mahasiswanya [1] oleh karena itu keterlambatan mahasiswa lulus merupakan juga salah satu hal yang mengakibatkan penurunan kualitas pendidikan yang diberikan oleh lembaga pendidikan terhadap peserta didiknya.

Pengukuran kualitas perguruan tinggi, khususnya program studi di Indonesia diukur berdasarkan akreditasi yang dilaksanakan oleh BAN PT atau Badan Akreditasi Nasional Perguruan Tinggi. Menurut BAN PT [2] kualitas perguruan tinggi diukur berdasarkan 7 standar utama, salah satunya adalah mahasiswa dan Lulusan, khusus mengenai standard evaluasi mahasiswa dan lulusan, komponen yang dinilai adalah sistem rekrutmen mahasiswa baru dan lulusan dalam hal ini adalah IPK dan masa studi.

Menurut UU No.12 Thn. 2012, yang dimaksud dengan kelulusan mahasiswa adalah : “mahasiswa dapat menyelesaikan program pendidikan sesuai dengan kecepatan belajar masing-masing dan tidak melebihi ketentuan batas waktu yang ditetapkan oleh Perguruan Tinggi” [3].

---

*Program Studi Administrasi Bisnis, Institut Ilmu Sosial dan Manajemen STIAMI Jakarta, Jl. Pangkalan Asem Raya No. 55 Cempaka Putih Jakarta Pusat 10530 INDONESIA (Telp. 021-4213380; e-mail: endangs2013@gmail.com)*

Untuk mendeteksi ketepatan waktu kelulusan mahasiswa dapat digunakan dengan berbagai macam cara, salah satunya dengan menggunakan teknik data mining. Teknik data mining sebagai sebuah teknik analisis berbasis database dan statistik, sangatlah tepat untuk menggali berbagai potensi informasi yang terkandung dalam sebuah database akademik.

Beberapa peneliti yang telah melakukan riset terhadap database dengan berbagai bidang ilmu menggunakan data mining dengan teknik support vector machine dan particle swarm optimization untuk mendeteksi berbagai kepentingan, termuat dalam beberapa jurnal berikut ini diantaranya : Saraswati [3], jacobus [4], satsiou [5], Yenaeng [6]. Oleh karena itu metode data mining dengan menggunakan teknik support vector machine dan particle swarm optimization diharapkan dapat memprediksi mahasiswa telat waktu sebagai antisipasi penurunan peserta didik dalam suatu perguruan tinggi khususnya di PLJ LP3I kampus Cibinong.

## II. KAJIAN LITERATUR

### A. Kelulusan Mahasiswa

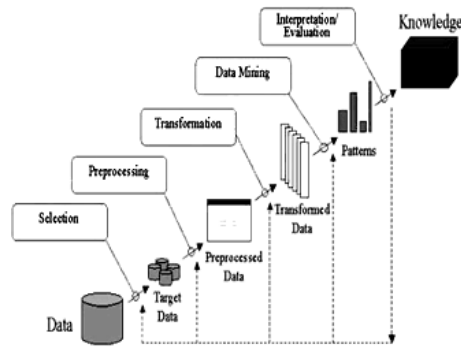
Kelulusan mahasiswa adalah “mahasiswa dapat menyelesaikan program Pendidikan sesuai dengan kecepatan belajar masing-masing dan tidak melebihi ketentuan batas waktu yang ditetapkan oleh Perguruan Tinggi” [7].

### B. Data Mining

Data mining dapat didefinisikan sebagai suatu proses untuk memecahkan masalah dengan cara menganalisa data yang berada pada database [8].

Sedangkan menurut Gartner Group, data mining adalah suatu proses menemukan hubungan yang berarti, pola, dan kecenderungan dengan memeriksa dalam sekumpulan besar data yang tersimpan dalam penyimpanan dengan menggunakan teknik pengenalan pola seperti teknik statistik dan matematika [9].

Berikut ini tahapan proses pencarian data pada sekumpulan database sehingga menghasilkan pengetahuan yang bermanfaat /KDD (Knowledge Discovery Data) [10].

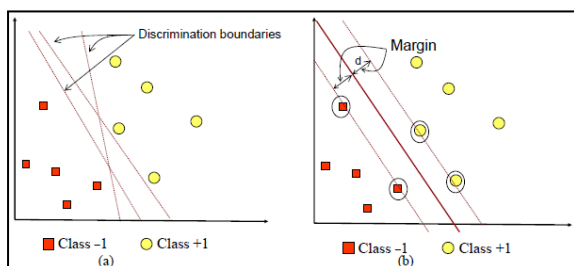


Gambar 1. Tahapan Proses KDD

- 1). *Pembersihan data (data cleaning)*, proses menghilangkan noise dan data yang tidak konsisten atau data yang tidak relevan.
- 2). *Melakukan integrasi data (data integration)*, merupakan penggabungan data dari berbagai database ke dalam satu database baru.
- 3). *Pemilihan data (data selection)*, pemilihan data yang akan di pakai, yang relevan yang di ambil dari database.
- 4). *Transformasi data (data transformation)*, data di ubah ke dalam format yang sesuai untuk diproses dalam data mining.
- 5). *Proses mining*, merupakan suatu proses utama, dimana metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data.
- 6). *Evaluasi pola (pattern evaluation)*, untuk mengidentifikasi pola-pola menarik untuk ke dalam knowledge based yang ditemukan.
- 7). *Presentasi pengetahuan (knowledge presentation)*, merupakan visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang di peroleh user. Tahap terakhir dari proses data mining adalah bagaimana memformulasikan keputusan atau aksi dari hasil analisis yang di dapat. Dalam presentasi, visualisasi juga dapat membantu untuk mengkomunikasikan hasil dari data mining.

C. *SUPPORT VECTOR MACHINE (SVM)*

Support Vector Machine adalah sebuah metode pembelajaran terpandu (supervised) yang digunakan untuk mengklasifikasikan dan analisa regresi pada suatu masalah tertentu [11]. SVM dapat diartikan sebagai suatu usaha untuk mencari hyperplane terbaik yang berfungsi sebagai pemisah dua buah kelas pada input space [12].



Gambar 2. SVM Mencari Hyperplane Terbaik [12]

Data yang tersedia dinotasikan sebagai  $x \in R^d$  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:

$$w \cdot x + b = 0 \tag{1}$$

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

$$w \cdot x + b = -1 \tag{2}$$

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

$$w \cdot x + b = +1 \tag{3}$$

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 4, dengan memperhatikan constraint persamaan 5.

$$\min \tau(w) = 1/2\|w\|^2 \tag{4}$$

$$y_i(x_i \cdot w + b) - 1 \geq 0, \quad i \tag{5}$$

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

$$L(w, b, \alpha) = \|w\|^2/2 - \sum_{i=1}^l \alpha_i (y_i (x_i \cdot w + b) - 1) \tag{6}$$

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

$$\text{Maximize: } \sum_{i=1}^l \alpha_i - \sum_{i,j=1}^l \alpha_i \alpha_j y_i \cdot y_j \cdot x_i \cdot x_j \tag{7}$$

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. *Pemilihan Parameter pada Support Vector Machine*

Untuk mendapatkan tingkat kinerja yang tinggi, beberapa parameter dari algoritma SVM harus diperbaiki, termasuk [13]:

- 1). *Pemilihan Fungsi Kernel.*
- 2). *Kinerja SVM tergantung pada pilihan fungsi kernel, besaran parameter kernel dan penentuan parameter C.* Fungsi kernel yang berbeda memperoleh tingkat

keberhasilan yang berbeda untuk berbagai jenis data aplikasi. Ketika nilai penentuan parameter C yang dipilih terlalu besar atau terlalu kecil, generalisasi SVM mungkin berkurang. Jika parameter kernel dan penentuan parameter yang tepat dipilih, kinerja SVM akan optimal.

3). *Parameter kernel(s)*.

4). *Parameter regularisasi* ( $C, \nu, \epsilon$ ) untuk tradeoff antara kompleksitas model dan akurasi mode.

#### E. *Algoritma Particle Swarm Optimization (PSO)*

Particle Swarm Optimization (PSO) banyak digunakan untuk memecahkan masalah optimasi serta sebagai masalah seleksi fitur [14]. Optimasi adalah proses menyesuaikan kepada masukan atau karakteristik perangkat, proses matematis, atau percobaan untuk menemukan output minimum atau maksimum 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 [15]. Dalam teknik Particle Swarm Optimization (PSO) 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. Particle Swarm Optimization (PSO) adalah suatu teknik optimasi yang sangat sederhana untuk menerapkan dan memodifikasi beberapa parameter [16].

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:

Dimana:

t = menunjukkan counter iterasi

Vij = kecepatan partikel i pada dimensi ke-j (nilainya terbatas antara [-vmax, vmax],

p = posisi partikel i pada j dimensi (nilainya terbatas [-pmax, pmax]

pbestij = posisi pbest partikel i pada dimensi ke-j

gbestij = posisi gbest dari dimensi ke-j

w = berat inersia (menyeimbangkan eksplorasi global dan lokal eksploitasi) rand1 dan rand2 = fungsi acak di rentang [0, 1]

$\beta$  = faktor kendala untuk mengontrol kecepatan berat (nilainya ke 1)

c1 dan c2 adalah faktor pembelajaran pribadi dan sosial (nilainya ke 2).

#### F. *Pengujian Dan Evaluasi Serta Validasi Algoritma Data Mining*

Pengujian dan evaluasi algoritma data mining dilakukan dengan cara :

1). *Pengujian K-fold Cross-validation*

K-Fold Cross Validation adalah teknik validasi yang membagi data ke dalam k bagian dan kemudian masing-masing bagian akan dilakukan proses klasifikasi. Dengan menggunakan K-Fold Cross Validation akan dilakukan percobaan sebanyak k. Tiap percobaan akan menggunakan satu data testing dan k-1 bagian akan menjadi data training, kemudian data testing itu akan ditukar dengan satu buah data training sehingga untuk tiap percobaan akan didapatkan data testing yang berbeda-beda. Data training adalah data yang akan dipakai dalam melakukan pembelajaran sedangkan data testing adalah data yang belum pernah dipakai sebagai pembelajaran dan akan berfungsi sebagai data pengujian kebenaran atau keakurasian hasil pembelajaran [8]

2). *Evaluasi & Validasi Metode*

a. *Confusion Matrik*

Confusion Matrix adalah alat (tools) visualisasi yang biasa digunakan pada supervised learning. Tiap kolom pada matriks adalah contoh kelas prediksi, sedangkan tiap baris mewakili kejadian di kelas yang sebenarnya [17]. Confusion matrix berisi informasi aktual (actual) dan prediksi (predicted) pada sistem klasifikasi.

b. *ROC (Receiver Operating Characteristic)*

Kurva ROC menunjukkan akurasi dan membandingkan klasifikasi secara visual. ROC mengekspresikan confusion matrix. ROC adalah grafik dua dimensi dengan false positives sebagai garis horizontal dan true positives untuk mengukur perbedaan performansi metode yang digunakan. ROC Curve adalah cara lain untuk menguji kinerja pengklasifikasian [17].

Performance keakurasian AUC dapat diklasifikasikan menjadi lima kelompok yaitu [17]:

0.90 – 1.00 = Excellent Clasification

0.80 – 0.90 = Good Clasification

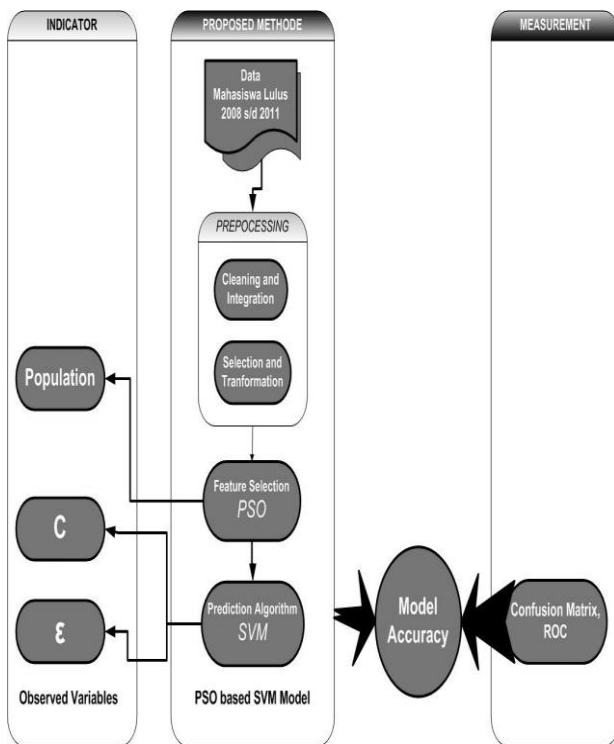
0.70 – 0.80 = Fair Clasification

0.60 – 0.70 = Poor Clasification

0.50 – 0.60 = Failure

#### G. *Kerangka Kerja Penelitian*

Model kerangka pemikiran yang digunakan adalah *method improvement* (perbaikan metode), yang sering digunakan pada penelitian di bidang sains dan teknik, termasuk bidang *computing* didalamnya. Komponen dari model kerangka pemikiran perbaikan metode (*method improvement*) adalah **Indicators, Proposed Method, Objectives, dan Measurements** [18]. Kerangka pemikiran pada penelitian ini dimulai dari prediksi hasil pemilihan umum. Maka dengan ini penulis ingin membuktikan bahwa Particle Swarm Optimization dapat meningkatkan tingkat akurasi Support Vector Machine.



Gambar 3. Model Prediksi Ketepatan Waktu Kelulusan

III. METODOLOGI PENELITIAN

A. Desain Penelitian

Pada umumnya ada empat metode penelitian yang dapat digunakan, yaitu tindakan penelitian, eksperimen, studi kasus dan survei. Dalam penelitian ini penulis menggunakan metode penelitian eksperimen dengan beberapa langkah yang dilakukan adalah sebagai berikut :

1) Pengumpulan Data

Teknik pengumpulan data ialah teknik atau cara-cara yang dapat digunakan untuk menggunakan data [19]. Dalam pengumpulan data terdapat sumber data, sumber data yang terhimpun langsung oleh peneliti disebut dengan sumber primer, sedangkan apabila melalui tangan kedua disebut sumber sekunder. Data yang dikumpulkan penulis merupakan data sekunder karena diperoleh dari database mahasiswa yang terkumpul pada file-file yang terpisah dalam format microsoft excel pada bagian pendidikan yang merupakan salah satu struktur dalam fungsi organisasi di PLJ LP3I kampus Cibinong.

Dataset yang dikumpulkan adalah dataset yang telah melalui proses preprosesing berupa data mahasiswa 3 periode tahun ajaran mulai dari tahun ajaran 2008-2009 s/d 2010-2011. sebanyak 518 mahasiswa yang terdiri dari atribut mahasiswa umur, prodi (program studi), kons (konsentrasi), ips (indeks prestasi semester), dan status kelulusan “tepat waktu” dan “telat”. Contoh sampel dataset seperti yang terlihat pada tabel berikut ini :

TABEL 1.  
DATASET AKADEMIK

No.	Atribut	Keterangan
1	Umur	Umur mahasiswa sesuai tahun ajaran
2	Prodi	[1]Manajemen Informatika
		[2]Administrasi Bisnis
		[3]Komputerisasi Akuntansi
		[4]Hubungan Masyarakat
3	Kons	[1]Informatika Komputer
		[2]Komputer Desain & Multimedia
		[3]Administrasi Perkantoran
		[4]Sekretaris
		[5]Komputer Akuntansi
		[6]Financial Syari'ah Banking
		[7]Marketing Bisnis Komunikasi
4	IPS	Indeks Prestasi Semester ( IPS_1 s/d IPS4
5	Status	0 = Telat Waktu1 = Lulus tepat waktu

2) Pengolahan data awal

Jumlah data awal yang diperoleh dari pengumpulan data yaitu sebanyak 518 data, namun tidak semua data dapat digunakan dan tidak semua atribut digunakan karena harus melalui beberapa tahap pengolahan awal data (preparation data). Untuk mendapatkan data yang berkualitas, beberapa teknik yang dilakukan sebagai berikut [20]:

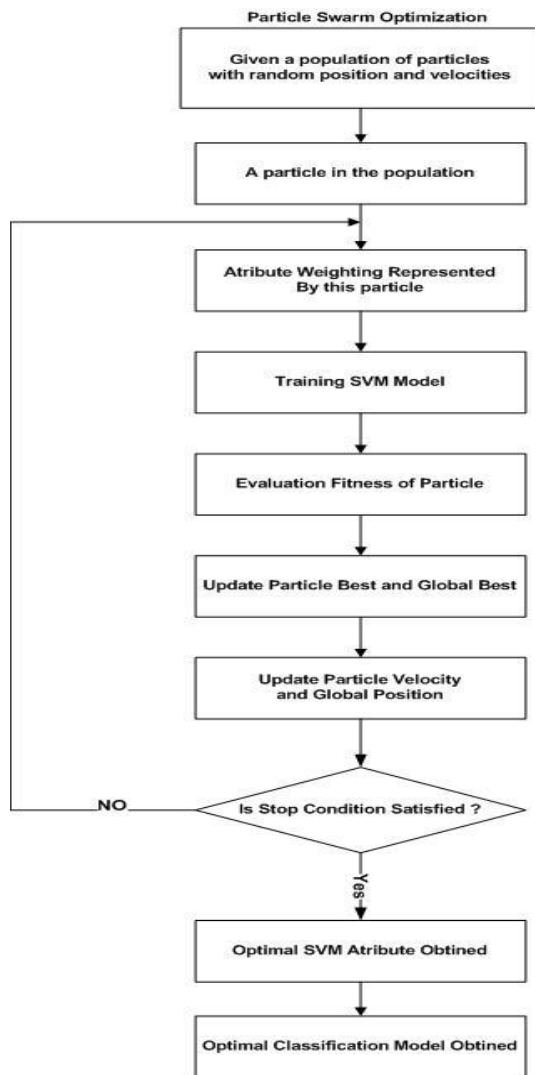
- a) *Data validation*, adalah digunakan untuk mengidentifikasi dan menghapus data yang ganjil (*outlier/noise*), data yang tidak konsisten, dan data yang tidak lengkap (*missing value*).
- b) *Data integration and transformation*, adalah untuk meningkatkan akurasi dan efisiensi algoritma. Data yang digunakan dalam penulisan ini bernilai kategorikal. Data ditransformasikan kedalam software Rapidminer.
- c) *Data size reduction and discretization*, adalah untuk memperoleh data set dengan jumlah atribut dan record yang lebih sedikit tetapi bersifat informatif.

TABEL 2.  
TABEL ATRIBUT DATASET AKADEMIK

No.	Atribut	Keterangan
1	Umur	Umur mahasiswa sesuai tahun ajaran
2	Prodi	[5]Manajemen Informatika
		[6]Administrasi Bisnis
		[7]Komputerisasi Akuntansi
		[8]Hubungan Masyarakat
3	Kons	[8]Informatika Komputer
		[9]Komputer Desain & Multimedia
		[10] Administrasi Perkantoran
		[11] Sekretaris
		[12] Komputer Akuntansi
		[13] Financial Syari'ah Banking
		[14] Marketing Bisnis Komunikasi
4	IPS	Indeks Prestasi Semester ( IPS_1 s/d IPS4
5	Status	0 = Telat Waktu1 = Lulus tepat waktu

3) *Metode yang diusulkan*

Metode yang diusulkan untuk menyelesaikan permasalahan pada penelitian ini seperti yang terlihat pada gambar dibawah ini :



Gambar 4. Model SVM - PSO

IV. HASIL DAN PEMBAHASAN

A. Hasil Eksperimen

1) *Support Vector Machine ( SVM )*

Penentuan nilai training cycles dalam penelitian ini dilakukan juga dengan cara uji coba memasukkan nilai C dan Epsilon, yaitu data dari periode tahun ajaran ( TA ) 2008/2009 s/d 2010/2011. Dengan hasil uji coba seperti yang terlihat pada tabel berikut ini :

TABEL 3.  
HASIL PENGUJIAN DATA MAHASISWA  
MENGUNAKAN METODE SVM

C	ε	TA. 2008/2009 s/d 2010/2011	
		Acc	AUC
0.0	0.0	79,36%	0,707
1.0	0.0	80.14%	0.715
0.0	1.0	75.67%	0.500
0.5	0.5	78.02%	0.721
0.6	0.6	70.25%	0.623
0.7	0.7	62.53%	0.602
0.8	0.8	60.99%	0.554
0.9	0.9	62.19%	0.625

Dari hasil pengujian / eksperimen yang telah dilakukan terhadap data mahasiswa periode tahun ajaran 2008/2009 hingga tahun ajaran 2010/2011 yang terkumpul pada tabel 3, dapat disimpulkan bahwa data dengan nilai parameter C = 1.0 dan epsilon = 0.0 memiliki akurasi yang paling tinggi yaitu sebesar 80.14 % dan AUC (Area Under Curve) sebesar 0.715.

2) *SVM berbasis Algoritma PSO (SVM-PSO)*

Penentuan nilai training cycles dalam penelitian ini dilakukan dengan cara uji coba memasukkan nilai C dan Epsilon pada data mahasiswa, yaitu data dari periode tahun ajaran 2008/2009 s/d 2010/2011. Dengan hasil uji coba seperti yang terlihat pada tabel berikut ini :

TABEL 4.  
HASIL PENGUJIAN DATA MAHASISWA  
MENGUNAKAN METODE SVM - PSO

C	(ε)	Pop size	TA. 2008/2009 s/d 2010/2011	
			ACC %	AUC
0.0	0.0	5	81,47	0,713
1.0	0.0	5	82.05	0.720
0.0	1.0	5	75.67	0.500
0.5	0.5	5	75.67	0.500
0.6	0.6	5	77.03	0.575
0.7	0.7	5	77.42	0.642
0.8	0.8	5	75.87	0.549
0.9	0.9	5	75.99	0.543

Dari hasil pengujian / eksperimen yang telah dilakukan terhadap data mahasiswa periode tahun ajaran 2008/2009 hingga tahun ajaran 2010/2011 yang terkumpul pada tabel 4, dapat disimpulkan bahwa data dengan nilai parameter C = 1.0 dan epsilon = 0.0 memiliki akurasi yang paling tinggi yaitu sebesar 82.05 % dan AUC (Area Under Curve) sebesar 0.720.

3) Perbandingan Hasil Uji Coba Metode SVM dan SVM berbasis PSO

Dari hasil ujicoba data training dengan menggunakan metode yang berbeda menghasilkan perolehan data pada masing – masing metode baik SVM dan SVM berbasis PSO jika dilakukan perbandingan menghasilkan data training dengan akurasi sebagaimana yang terlihat pada tabel berikut ini :

TABEL 5.  
PERBANDINGAN METODE SVM DAN SVM-PSO

C	ε	SVM		Pop Size	SVM-PSO	
		ACC (%)	AUC		ACC (%)	AUC
0.0	0.0	79,36	0,707	5	81,47	0,713
1.0	0.0	80.14	0.715	5	82.05	0.720
0.0	1.0	75.67	0.500	5	75.67	0.500
0.5	0.5	78.02	0.721	5	75.67	0.500
0.6	0.6	70.25	0.623	5	77.03	0.575
0.7	0.7	62.53	0.602	5	77.42	0.642
0.8	0.8	60.99	0.554	5	75.87	0.549
0.9	0.9	62.19	0.625	5	75.99	0.543

Dari hasil pengujian / eksperimen dengan menggunakan metode SVM dan SVM-PSO yang telah dilakukan terhadap data mahasiswa maka jika dibandingkan dengan hasil uji coba tersebut, seperti yang terlihat pada tabel 5, dapat diketahui bahwa data dengan menggunakan metode SVM-PSO memiliki akurasi yang paling tinggi yaitu sebesar 82.05 % dan AUC (Area Under Curve) sebesar 0.720 dengan nilai parameter  $C = 1.0$  dan  $\epsilon = 0.0$ . Sedangkan data mahasiswa dengan menggunakan metode SVM memiliki akurasi dan AUC yang lebih rendah yaitu sebesar 80.14 % dan  $AUC = 0.715$ .

B. Pembahasan

Pengujian model untuk memprediksi ketepatan waktu kelulusan mahasiswa khususnya mahasiswa PLJ LP3I Jakarta Kampus Cibinong menggunakan metode SVM dan SVM-PSO dilakukan menggunakan rapidminer 5.3. Validasi model menggunakan Cross Validation dengan nilai parameter 10. Hasil evaluasi dan validasi ditampilkan dalam bentuk confusion matrix dan receiver operating curve (ROC) sebagaimana terlihat pada tabel dan gambar sebagai berikut :

- 1) Evaluasi dan Validasi Metode SVM
- a) Confusion Matrix

TABEL 6.  
MODEL CONFUSION MATRIX DENGAN METODE SVM

Accuracy : 80,14% +/-1,46% (mikro:75,63%)			
	True 1.0	True 0.0	Class Precision
Pred 1.0	382	93	80.42%
Pred 0.0	10	33	76.74%
Class recall	97.45%	26.19%	

Dari hasil pengujian model dengan menggunakan metode SVM pada data tabel 3, yang memiliki akurasi tertinggi

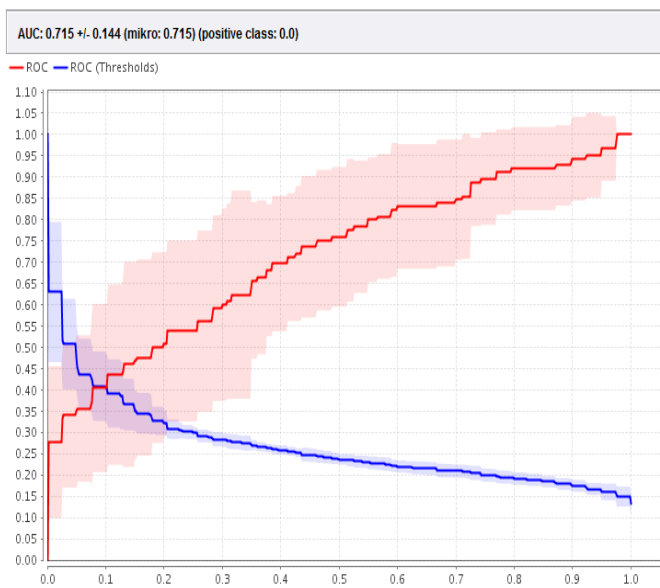
menghasilkan model confusion matrix tabel 6, yaitu data sebanyak 518 data terdiri dari 382 diklasifikasikan *lulus tepat waktu* ( 1.0 ) hasilnya sesuai prediksi yang dilakukan dengan metode SVM, 93 data diprediksi *lulus tepat waktu* (1.0) tetapi hasil prediksi *telat* sedangkan 10 data diprediksi *telat* (0.0) ternyata *lulus tepat waktu* dan 33 data diprediksi *telat* (0.0).

Berdasarkan hasil tersebut dapat diketahui pula perhitungan nilai accuracy, sensitivity, specificity, ppv, npv yang dikumpulkan dalam tabel 7 sebagai berikut :

TABEL 7.  
HASIL PERHITUNGAN ACCURACY, SENSITIVITY, SPECIFICITY, PPV DAN NPV DENGAN METODE SVM.

	Nilai ( % )
Accuracy	80.14
Sensitivity	80,42
Specificity	76,74
PPV	97,45
NPV	26,19

b) Receiver Operating Curve (ROC)



Gambar 5. Visualisasi ROC dengan Metode SVM

Dari gambar 5. dapat diketahui bahwa pengujian data training terpilih dengan akurasi tertinggi menghasilkan kurva ROC dengan nilai AUC sebesar 0.715. Hal ini berarti bahwa klasifikasi data training tersebut memiliki hasil diagnosa *fair classification*.

- 2) Evaluasi dan Validasi Metode SVM-PSO

TABEL 8.

MODEL CONFUSION MATRIX DENGAN METODE SVM-PSO

Accuracy : 82,05% +/-3,23% (mikro:82,05%)			
	True 1.0	True 0.0	Class Precision
Pred 1.0	385	86	81,74%
Pred 0.0	7	40	85,11%
Class recall	98,21%	31,75%	

Dari hasil pengujian model dengan menggunakan metode SVM dengan optimasi PSO dapat diketahui bahwa dari jumlah data yaitu sebanyak 518 data terdiri dari 385 diprediksi lulus tepat waktu ( 1.0 ) hasilnya sesuai prediksi yang dilakukan dengan metode SVM-PSO, 86 diprediksi lulus tepat waktu (1.0) tetapi ternyata telat sedangkan 7 data diprediksi telat (0.0) ternyata lulus tepat waktu dan 40 data diprediksi telat (0.0) hasilnya sesuai.

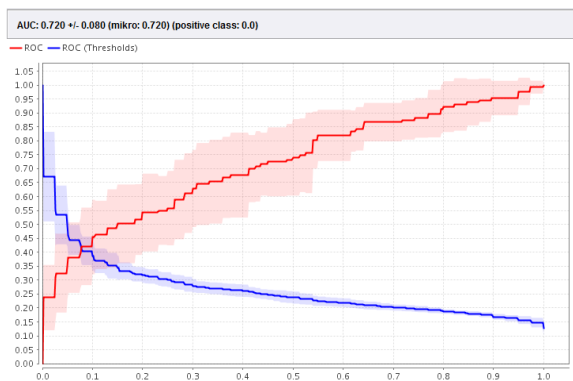
Berdasarkan hasil tersebut dapat diketahui pula perhitungan nilai accuracy, sensitivity, specificity, ppv, npv yang dikumpulkan dalam tabel 9 sebagai berikut :

TABEL 9.

HASIL PERHITUNGAN ACCURACY, SENSITIVITY, SPECIFICITY, PPV DAN NPV DENGAN METODE SVM.

	Nilai ( % )
Accuracy	82,05
Sensitivity	81,74
Specificity	85,11
PPV	98,21
NPV	31,75

a) Receiver Operating Curve (ROC)



Gambar 6. Visualisasi ROC dengan Metode SVM-PSO

Dari gambar 6, dapat diketahui bahwa pengujian data dengan akurasi tertinggi pada tabel 5, menghasilkan kurva ROC dengan nilai AUC sebesar 0.720. Hal ini berarti bahwa klasifikasi data training tersebut memiliki hasil diagnosa *fair classification*.

V. KESIMPULAN

Dari hasil eksperimen atau uji coba data training dengan menggunakan metode SVM atau SVM-PSO memiliki tingkat akurasi dan AUC yang berbeda. Metode SVM-PSO memiliki tingkat akurasi lebih tinggi sebesar 82,05% dan AUC = 0,720 dibandingkan dengan metode SVM yaitu Akurasi sebesar 80,14% dan AUC = 0,715 dengan selisih peningkatan akurasi sebesar 1.91 % dan AUC = 0,005. Sehingga metode SVM-PSO dapat digunakan untuk memprediksi ketepatan waktu kelulusan mahasiswa.

Metode SVM – PSO jika dapat diwujudkan dalam prototipe aplikasi ketepatan waktu kelulusan mahasiswa akan sangat membantu semua pihak yang berkepentingan dalam usaha meningkatkan jumlah kelulusan ditahun yang akan datang. Sehingga dapat meningkatkan kualitas lembaga pendidikan tersebut ke jenjang akreditasi yang lebih baik dari jenjang akreditasi sebelumnya.

REFERENSI

- [1] Ahmed, "Data Mining : A Prediction for Student's Performance Using Classification Method," *World Journal of Computer Application and Technology* 2, pp. (2) 43-47, 2014.
- [2] BAN-PT, *Akreditasi Institusi Perguruan Tinggi - Buku III Pedoman Penyusunan Borang.*, 2011.
- [3] N. W. S Saraswati, "Text mining dengan metode naïve bayes classifier dan support vector machines untuk sentiment analysis," 2014.
- [4] Agustinus Jacobus, "Penerapan Metode Support Vector Machine pada Sistem Deteksi Intrusi secara Real-time," *IJCCS*, pp. Vol.8, No.1, January 2014, pp. 13~24, 2014.
- [5] ana Satsiou, "Genetic Algorithms for the Optimization of Support Vector Machines in Credit Risk Rating," 2002.
- [6] S. Yenaeng, "Automatic Medical Case Study Essay Scoring by Support Vector Machine and Genetic Algorithms," *IJIET*, pp. Vol. 4, No. 2., 2014.
- [7] (2012, january) [www.academia.edu](http://www.academia.edu). [Online]. [http://www.academia.edu/2160186/Undang-Undang\\_R.I\\_nomor\\_12\\_tahun\\_2012\\_tentang\\_Pendidikan\\_Tinggi](http://www.academia.edu/2160186/Undang-Undang_R.I_nomor_12_tahun_2012_tentang_Pendidikan_Tinggi)
- [8] I.,H.Frank,E., dan Hall, M.,A Witten, *Data Mining: Practical Machine Learning and Tools*. United States: Morgan Kaufmann Publisher, 2011.
- [9] Daniel T Larose, *Discovering Knowledge in Data: An Introduction to Data Mining.*: John Willey & Sons. Inc, 2005.
- [10] J., & Kamber, M. Han, *Data Mining Concepts and Techniques*. San Fransisco: Mofgan Kaufan Publisher., 2007.
- [11] Sampson Adu-Poku, *Comparing Classification Algorithms in Data Mining*. Connecticut, new britain, 2012.
- [12] Anto Satrio Nugroho, "Support Vector Machine: Paradigma Baru Dalam Softcomputing," *Konferensi Nasional Sistem Dan Informatika*, pp. 92-99, 2008.
- [13] O. Maimon, *Data Mining And Knowledge Discovery Handbook*. New York Dordrecht Heidelberg London: Springer, 2010.
- [14] Y., Wang, G., Chen, H., Dong, H., Zhu, X., & Wang, S. Liu, "An Improved Particle Swarm Optimization for Feature Selection," *Journal of Bionic Engineering*, pp. 8(2), 191–200. doi:10.1016/S1672-6529(11)60020-6, 2011.
- [15] R. L., & Haupt, S. E. Haupt, *Practical Genetic Algorithms*. United State of America: A John Wiley & Sons Inc Publication, 2004.

- [16] A. S. H., Hussin, B., Ananta, I. G. P., & Zeniarja, J. Basari, "Opinion Mining of Movie Review using Hybrid Method of Support Vector Machine and Particle Swarm Optimization.," *Procedia Engineering*, pp. 53, 453–462. doi:10.1016 /j.proeng.2013.02.059, 2013.
- [17] F. Gorunescu, *Data Mining: Concepts and Techniques*. Verlag berlin Heidelberg: Springer, 2011.
- [18] G. Polancic, "Empirical Research Method Poster," 2007.
- [19] Riduwan, *Metode dan Teknik Menyusun Tesis*. Bandung: Alfabeta, 2008.
- [20] C Vercellis, *Business Intelligence Data Mining And Optimization For Decision Making*. United Kingdom: A John Wiley And Sons, Ltd., Publication., 2009.



**Endang Supriyadi**, lahir di Jakarta, 14 Nopember 1974. Memperoleh gelar Diploma III (D3), Jurusan Manajemen Informatika Universitas Gunadarma Depok, lulus tahun 1998 dan memperoleh gelar Sarjana Al-Quran (SQ) dan gelar Sarjana Agama (S.Ag) Jurusan Tarbiyah di Perguruan Tinggi Ilmu Al-qur'an Jakarta (PTIQ) lulus tahun 1999. Memperoleh gelar Magister Komputer (M.Kom) Program Pasca Sarjana Magister Ilmu

Komputer Konsentrasi e-Business STMIK Nusa Mandiri Jakarta, lulus tahun 2015. Pernah mengajar di STMIK Nusa Mandiri Jakarta dan saat ini aktif mengajar di beberapa perguruan tinggi swasta lainnya : STIAMI, STIMA IMMI, STMIK Triguna Utama, Politeknik LP3I Jakarta serta menjadi anggota tim asesor LSP Politeknik LP3I Jakarta untuk kompetensi bidang ilmu komputer. Aktif pula menulis buku panduan komputer aplikasi perkantoran dan pendiri lembaga kursus komputer LPK-PROTECH Depok.