

**PENERAPAN *PARTICLE SWARM OPTIMIZATION* UNTUK  
SELEKSI ATRIBUT PADA METODE *SUPPORT VECTOR  
MACHINE* UNTUK PREDIKSI PENYAKIT LIVER**



**Tesis**

**Diajukan sebagai Salah satu syarat untuk memperoleh gelar**

**Magister Ilmu Komputer (M.Kom)**

**NU'MAN MUSYAFFA**

**14000959**

**PROGRAM PASCASARJANA MAGISTER ILMU KOMPUTER  
SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER**

**NUSAMANDIRI**

**2015**

## SURAT PERNYATAAN ORISINALITAS

Yang bertanda tangan di bawah ini :

Nama : Nu'man Musyaffa  
NIM : 14000959  
Program Studi : Magister Ilmu Komputer  
Jenjang : Strata Dua (S2)  
Konsentrasi : *MIS*

Dengan ini menyatakan bahwa tesis yang telah saya buat dengan judul: “Penerapan *Particle Swarm Optimization* Untuk Seleksi Atribut Pada Metode *Support Vector Machine* Untuk Prediksi Penyakit Liver” adalah hasil karya sendiri, dan semua sumber baik yang kutip maupun yang dirujuk telah saya nyatakan dengan benar dan tesis belum pernah diterbitkan atau dipublikasikan dimanapun dan dalam bentuk apapun.

Demikianlah surat pernyataan ini saya buat dengan sebenar-benarnya. Apabila dikemudian hari ternyata saya memberikan keterangan palsu dan atau ada pihak lain yang mengklaim bahwa tesis yang telah saya buat adalah hasil karya milik seseorang atau badan tertentu, saya bersedia diproses baik secara pidana maupun perdata dan kelulusan saya dari Program Pascasarjana Magister Ilmu Komputer Sekolah Tinggi Manajemen dan Komputer Nusa Mandiri dicabut/dibatalkan.

Jakarta, 04 Desember 2015  
Yang menyatakan,

Materai Rp. 6.000,-

Nu'man Musyaffa

HALAMAN PENGESAHAN

Tesis ini diajukan oleh :

Nama : Nu'man Musyaffa

NIM : 14000959

Program Studi : Magister Ilmu Komputer

Jenjang : Strata Dua (S2)

Konsentrasi : MIS

Judul Tesis : "Penerapan *Particle Swarm Optimization* Untuk Seleksi Atribut Pada Metode *Support Vector Machine* Untuk Prediksi Penyakit Liver"

Telah berhasil dipertahankan dihadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Magister Ilmu Komputer (M.Kom) pada Program Pascasarjana Magister Ilmu Komputer Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri (STMIK Nusa Mandiri).

Jakarta, 04 Desember 2015  
Pascasarjana Magister Ilmu Komputer  
STMIK Nusa Mandiri  
Direktur

Prof. Dr. Ir. Ricardus Eko Indrajit, M.Sc, MBA

**DEWAN PENGUJI**

Penguji I : Dr. Windu Gata, M.Kom



.....

Penguji II : Dr. Dwiza Riana, S,si, MM, M.Kom



.....

Penguji III /  
Pembimbing : Dr. H. Mochamad Wahyudi, MM, M.Kom, M.Pd .....



## DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PERNYATAAN ORISINALITAS.....	ii
HALAMAN PENGESAHAN.....	iii
KATA PENGANTAR .....	iv
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI KARYA... ILMIAH UNTUK KEPENTINGAN AKADEMIS .....	vii
ABSTRAK.....	viii
ABSTRACT.....	ix
DAFTAR ISI.....	x
BAB I. PENDAHULUAN .....	1
1.1. Latar Belakang Penulisan .....	1
1.2. Identifikasi Masalah.....	3
1.3. Tujuan Penelitian .....	4
1.4. Ruang Lingkup Penelitian.....	4
1.5. Hipotesis .....	4
1.6. Sistematika Penulisan .....	5
BAB II. LANDASAN TEORI/KERANGKA PENELITIAN .....	6
2.1. Tinjauan Pustaka.....	6
2.1.1. Liver.....	6
2.1.2. Data Mining .....	6
2.1.3. Support Vector Machine.....	10
2.1.4. Feature Selection.....	14
2.1.5. Particle Swarm Optimization.....	15
2.1.6. Pengujian Evaluasi dan Validasi.....	16
2.2. Tinjauan Studi.....	20
2.3. Obyek Penelitian.....	22
BAB III. METODE PENELITIAN.....	23
3.1. Desain Penelitian .....	23
3.2. Pengumpulan Data.....	24
3.3. Pengolahan Awal Data.....	25
3.4. Metode Yang Diusulkan .....	26
3.5. Esperimen dan Pengujian Metode .....	27
3.6. Evaluasi dan Validasi Hasil .....	28
BAB IV. HASIL ANALISA DAN PEMBAHASAN.....	29
4.1. Ekperimen dan Pengujian Model.....	29
4.1.1. SVM.....	29
4.1.2. SVM Berbasis PSO.....	29
4.2. Evaluasi dan Validasi Hasil .....	30
4.2.1. SVM.....	30

4.2.2. SVM Berbasis PSO.....	31
4.2.3. Analisa Hasil Komparasi .....	37
BAB V. PENUTUP.....	42
5.1. Kesimpulan .....	42
5.2. Saran .....	42
DAFTAR PUSTAKA .....	43

## **BAB I**

### **PENDAHULUAN**

#### **1.1 Latar Belakang Penulisan**

Liver atau hati adalah organ terbesar dalam tubuh padat dan juga dianggap sebagai kelenjar, karena di antara banyak fungsinya yang membuat dan mengeluarkan empedu. Liver ini terletak di bagian kanan atas perut yang dilindungi oleh tulang rusuk. Penyakit liver adalah suatu gangguan pada setiap fungsi liver yang menyebabkan penyakit. Liver bertanggung jawab untuk fungsi-fungsi kritis dalam tubuh, dimana hilangnya fungsi-fungsi tersebut dapat menyebabkan kerusakan yang signifikan pada tubuh. Liver adalah satu-satunya organ dalam tubuh yang dapat dengan mudah mengganti sel-sel yang rusak, tetapi jika sel-sel itu hilang, maka liver tidak mungkin dapat memenuhi kebutuhan tubuh (Salindeho, 2013).

Masalah yang ditimbulkan oleh penyakit liver adalah sulitnya mengenali penyakit liver sejak dini, bahkan ketika penyakit tersebut sudah menyebar. Diagnosa penyakit liver yang lebih awal dapat meningkatkan tingkat kelangsungan hidup pasien (Lin, 2009). Diagnosa awal merupakan permasalahan yang sangat penting dalam penyakit hati. Untuk mendiagnosa penyakit hati perlu dilakukan tes darah. Berdasarkan hasil tes darah tersebut dapat diketahui seorang pasien menderita penyakit hati atau tidak.

Artificial intelligence sudah lama digunakan pada bidang medis sejak awal tahun 1960an (Praetorius, 2006). Awalnya artificial intelligence in medicine (AIM) digunakan sebagai alat untuk membentuk kembali sistem informasi kesehatan. Selama bertahun-tahun, AIM mengalami perubahan perspektif dari tadinya hanya sebagai sistem informasi kesehatan berubah menjadi Clinical Decision Support System (CDSS). Sebagai CDSS, AIM digunakan pada sektor kesehatan yang berbeda-beda, seperti mendukung obat resep pada lingkungan laboratorium dan pendidikan serta untuk diagnosa penyakit. AIM mampu melakukan learning atau pembelajaran suatu pola penyakit (Praetorius, 2006). Salah satu AIM yang paling banyak digunakan adalah machine learning.

Penelitian yang dilakukan oleh Babu Ramana dan Venkateswarlu dengan judul *A Critical Study of Selected Classification Algorithms for Liver Disease Diagnosis*, algoritma klasifikasi populer dianggap dapat mengevaluasi kinerja dalam hal akurasi, presisi, sensitivitas dan spesififikasi dalam mengklasifikasi dataset pasien liver AP dibandingkan UCLA dataset liver dengan semua algoritma yang dipilih, dengan memakai atribut umum untuk dataset liver seperti Umur, jenis kelamin, SGOT, SGPT, ALP, Total Bilirubin, Direct Bilirubin, Total Protein dan Albumin sangat penting dalam memutuskan status penyakit pada kasus dataset liver. Dengan dataset yang dipilih, algoritma KNN, Back Propagation dan SVM memberikan hasil yang lebih baik dengan semua kombinasi set fitur. (Ramana, Babu & Venkateswarlu, 2011).

Selanjutnya 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 (Chunjiang & Yan, 2009).

Neural Network mempunyai kelebihan dalam hal kemampuan generalisasi tergantung pada seberapa baik Neural Network meminimalkan resiko empiris namun Neural Network mempunyai kelemahan dimana menggunakan data pelatihan cukup besar (Vapnik, 1999). Decision tree dan ID3 mempunyai

kelebihan untuk keputusan pengklasifikasi memiliki akurasi yang baik namun memiliki kelemahan karena perlu mengumpulkan lebih banyak data (Han, Rodriguze, & Beheshti, 2008). SVM adalah kasus khusus dari keluarga algoritma yang kita sebut sebagai regularized metode klasifikasi linier dan metode yang kuat untuk minimalisasi resiko (Weiss, Indurkha, & Zhang, 2010). Dan kelebihan SVM lainnya adalah dapat meminimalkan kesalahan melalui memaksimalkan margin dengan misahkan antara hyper-lane dan satu set data bahkan dengan jumlah sample yang kecil (Chunjiang & Yan, 2009).

Namun demikian masalah aplikasi tertentu, tidak semua fitur ini sama-sama penting dan kinerja yang lebih baik dapat dicapai dengan membuang beberapa fitur dengan begitu fitur dalam SVM memiliki pengaruh penting dalam akurasi klasifikasi (Zhao, Fu, Ji, Tang, & Zhou, 2011). Dataset yang tidak penting, fitur yang banyak atau sangat berhubungan secara signifikan akan mengurangi tingkat akurasi klasifikasi dengan menghapus fitur ini, dengan begitu tingkat akurasi efisiensi dan klasifikasi dapat diperoleh (Lin a, Shiue b, & Chen, 2009).

PSO sangat menarik untuk pemilihan fitur dimana kawanan partikel akan menemukan kombinasi fitur terbaik pada saat pencarian ruang masalah dan PSO dapat menemukan solusi yang optimal dengan cepat (Parimala & Nallaswamy, 2012). PSO banyak digunakan untuk memecahkan masalah optimasi, serta sebagai masalah seleksi fitur (Liu, Wang, Chen, Dong, Zhu, & Wang, 2011). Dalam teknik 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*.

## 1.2 Identifikasi Masalah

*Support Vector Machine* dapat menyelesaikan masalah *decision tree* khususnya sampel data yang kecil yang ada di ilpd liver, tetapi *Support Vector Machine* memiliki kelemahan pada sulitnya pemilihan fitur yang sesuai dan optimal pada bobot atribut yang digunakan sehingga menyebabkan tingkat akurasi prediksi menjadi rendah. Masalah yang diangkat pada penelitian ini adalah

seberapa besar akurasi metode *Support Vector Machine* yang ditingkatkan dengan *Particle Swarm Optimization* (PSO) dengan cara melakukan seleksi atribut untuk prediksi penyakit liver?

### **1.3 Tujuan Penelitian**

Tujuan dari penelitian ini adalah menerapkan *Particle Swarm Optimization* (PSO) untuk seleksi atribut yang optimal dan membobot atribut dari *dataset* pada metode *Support Vector Machine* (SVM) guna meningkatkan akurasi prediksi penyakit liver.

### **1.4 Ruang Lingkup Penelitian**

Ruang lingkup pembahasan dalam penelitian ini dibatasi pada perbandingan metode *Support Vector Machine* dengan *Support Vector Machine* berbasis *Particle Swarm Optimization* (PSO) dengan cara menganalisis atribut menjadi bobot atribut (*attribute weighting*) dalam prediksi penyakit liver.

### **1.5 Hipotesis**

Pada penelitian ini hipotesis yang diajukan adalah diduga *data mining* yang berbasis klasifikasi yang dikembangkan berdasarkan algoritma *Support Vector Machine* dengan *Support Vector Machine* berbasis *Particle swarm optimization* (PSO) dapat bermanfaat dalam memprediksi penyakit liver pada ILPD liver.

### **1.6 Sistematika Penulisan**

Dalam sistematika penulisan yang disajikan secara umum terdiri dari lima bab, dan masing-masing bab memiliki beberapa sub bab yang terdiri dari:

#### **Bab I      Pendahuluan**

Bab ini membahas tentang latar belakang penulisan, identifikasi masalah, tujuan penelitian, manfaat penelitian, ruang lingkup penelitian dan hipotesis.

#### **Bab II     Landasan Teori**

Bab ini membahas tentang landasan teori yang melandasi penelitian.

Bab III Metode Penelitian

Bab ini berisi tentang metode penelitian yang membahas tentang perancangan penelitian, tahap *computing approach* dan pengembangan sistem.

Bab IV Hasil Penelitian dan Pembahasan

Bab ini berisi tentang hasil dan pembahasan yang menguraikan tentang implementasi sistem, pengukuran serta implikasi penelitian.

Bab V Kesimpulan dan Saran

Bab ini membahas kesimpulan dari penelitian dan saran untuk penelitian selanjutnya

## BAB II

### LANDASAN TEORI

#### 2.1 Tinjauan Pustaka

Dalam penulisan tesis ini penulis melakukan tinjauan dengan menggunakan buku dan jurnal yang berhubungan dengan tema yang dipilih. Secara lebih detail tinjauan dalam penulisan tesis ini dapat dijelaskan sebagai berikut :

##### 2.1.1 Liver

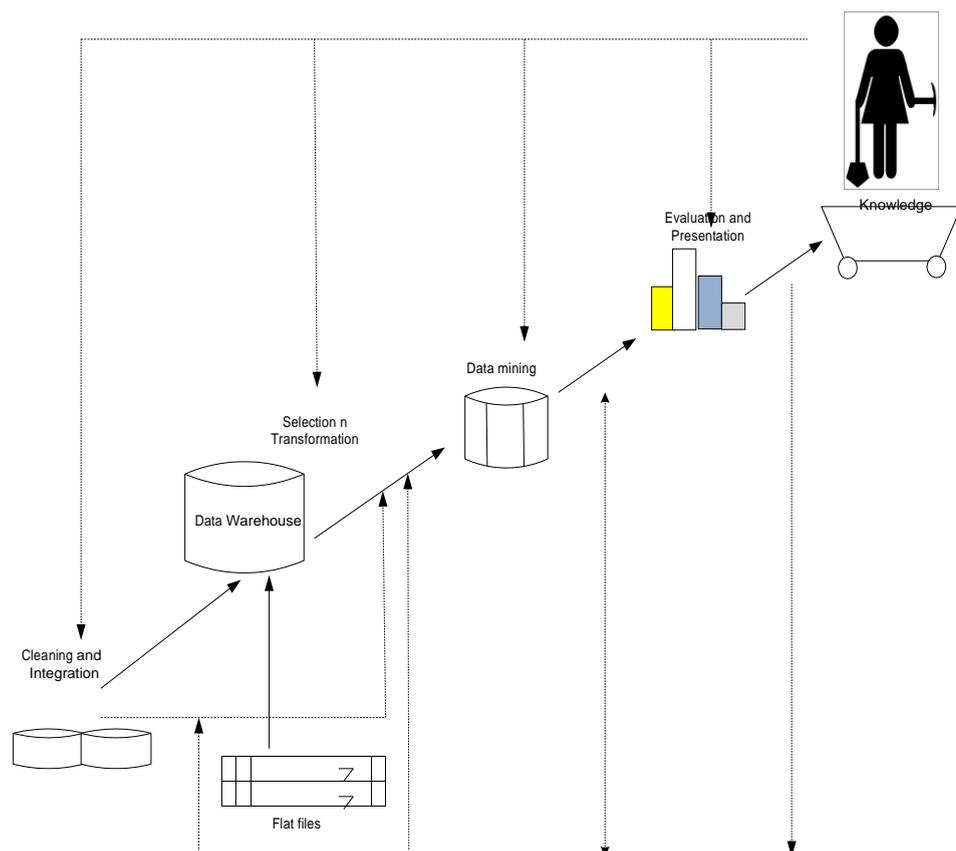
Liver merupakan organ yang paling besar dan penting bagi kita, berfungsi sebagai pembentukan dan sekresi empedu, tempat menyimpan glikogen, sintesa urea, metabolisme kolesterol, lemak dan detoksifikasi racun (Lumongga, 2009). Penyakit liver sulit ditemukan pada stadium awal, diagnosa awal merupakan permasalahan yang sangat penting dalam penyakit liver. (Lin, 2009).

##### 2.1.2 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 (Witten, Frank, & Hall, 2011). Kakas data mining meramalkan tren dan sifat-sifat perilaku bisnis yang sangat berguna untuk mendukung pengambilan keputusan penting. Analisis yang diotomatisasi yang dilakukan oleh data mining melebihi yang dilakukan oleh sistem pendukung keputusan tradisional yang sudah banyak digunakan (Moertini, 2002). Secara khusus, koleksi metode yang dikenal sebagai 'data mining' menawarkan metodologi dan solusi teknis untuk mengatasi analisis data medis dan konstruksi prediksi model (Bellazzi & Zupanb, 2008).

Secara umum, tugas data mining dapat diklasifikasikan menjadi dua kategori: deskriptif dan prediktif. Tugas pertambangan deskriptif mengkarakterisasi sifat umum data dalam database pertambangan prediktif tugas data pada saat ini untuk membuat prediksi (Han & Kamber, 2007).

Penerapan metode machine learning untuk database besar disebut data mining, dalam data mining volume besar data diproses untuk membangun sebuah model sederhana, misalnya memiliki akurasi prediksi yang tinggi. Area aplikasi data mining berlimpah, selain retail, bank menganalisis data masa lalu untuk membangun model aplikasi kredit, deteksi penipuan dan pasar saham. Dalam manufaktur, model pembelajaran digunakan untuk optimasi, kontrol, dan pemecahan masalah. Dalam pengobatan, program pembelajaran adalah digunakan untuk diagnosis medis. Dalam telekomunikasi, pola panggilan dianalisis untuk optimasi jaringan dan memaksimalkan kualitas layanan. Dalam ilmu pengetahuan, sejumlah besar data dalam fisika, astronomi, dan biologi dapat hanya dianalisis cukup cepat oleh komputer (Alpaydın, 2010). Model data mining memberikan contoh penerapannya pada berbagai algoritma dan pada data set yang besar (Larose, 2007).



Gambar 2.1 Tahapan data mining (Han & Kamber, 2007)

Tahapan data mining dalam proses penemuan pengetahuan (Han & Kamber,2007):

1. Pembersihan data (untuk menghilangkan noise dan data tidak konsisten)
2. Integrasi data (di mana beberapa sumber data dapat dikombinasikan)
3. Data seleksi (di mana data yang relevan dengan tugas analisis basis data yang akan diambil)
4. Data transformasi (dimana data diubah atau dikonsolidasikan ke dalam bentuk yang sesuai untuk pertambangan dengan melakukan operasi ringkasan atau agregasi)
5. Data mining (proses esensial dimana metode cerdas diaplikasikan untuk mengekstrak pola data)
6. Pola evaluasi (untuk mengidentifikasi pola yang benar-benar menarik yang mewakili pengetahuan didasarkan pada beberapa langkah-langkah interestingness)
7. Pengetahuan presentasi (dimana visualisasi dan teknik representasi pengetahuan digunakan untuk menyajikan pengetahuan ditambah kepada pengguna)

Terdapat empat pelompokan dalam data mining yaitu klasifikasi, asosiasi, clustering dan prediksi (Witten, Frank, & Hall, 2011):

#### 1. Klasifikasi

Proses klasifikasi didasarkan: Kelas-variabel dependen dari model-yang merupakan variabel kategori mewakili yang 'label' memakai objek setelah klasifikasin, contohnya loyalitas pelanggan, kelas bintang (galaksi), kelas gempa bumi (badai) (Gorunescu, 2011).

- a. Prediksi variabel bebas dari model diwakili oleh karakteristik (atribut) dari data yang harus diklasifikasikan dan berdasarkan klasifikasi yang dibuat. Contoh prediktor tersebut adalah: merokok, konsumsi alkohol, tekanan darah, frekuensi pembelian, status perkawinan, karakteristik (satelit) gambar, dan kecepatan arah angin (Gorunescu, 2011).
- b. Pelatihan dataset yang merupakan sekumpulan data yang berisi nilai untuk dua sebelumnya komponen, dan digunakan untuk 'pelatihan' model untuk mengenali sesuai kelas, berdasarkan prediksi tersedia. Contoh set tersebut

adalah: kelompok pasien diuji pada serangan jantung, kelompok pelanggan dari supermarket (diselidiki oleh internal polling), database yang berisi gambar untuk pemantauan dan pelacakan objek teleskopik astronomi (Gorunescu, 2011).

- c. Pengujian dataset yang berisi data baru yang akan diklasifikasikan oleh (classifier) model dibangun di atas, dan akurasi klasifikasi (kinerja model) sehingga dapat dievaluasi (Gorunescu, 2011). Model klasifikasi yang populer adalah *Decision/classification trees, Bayesian classifiers/Naive Bayes classifiers, Neural networks, Algoritma Genetika, Support vector machines*.

## 2. Asosiasi

Setiap asosiasi antara fitur-fitur yang dicari, bukan hanya satu yang memprediksi nilai kelas tertentu (Witten, Frank, & Hall, 2011). Pada prinsipnya, penemuan aturan asosiasi/asosiasi mempelajari aturan bagaimana kita memahami proses mengidentifikasi aturan antara ketergantungan yang berbeda dari fenomena kelompok. Dengan demikian, mari kita perkirakan kumpulan set yang kita punya masing-masing berisi sejumlah objek/benda-benda. Jadi tujuan kita untuk mencari peraturan yang menghubungkan (asosiasi), obyek ini berdasarkan peraturan ini, untuk dapat memprediksi terjadinya objek/item, berdasarkan kejadian lain (Gorunescu, 2011).

## 3. Clustering

Cluster adalah menemukan kelompok (kelompok) objek, berdasarkan kemiripan (semacam kemiripan), sehingga dalam setiap kelompok ada kemiripan yang besar, sementara kelompok cukup berbeda dari satu sama lain (Gorunescu, 2011).

## 4. Prediksi

Prediksi/perkiraan model yang berkaitan dengan kemampuan untuk memprediksi tanggapan terbaik (output), yang paling dekat ke kenyataan, berdasarkan input data.

Dengan demikian, 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 (Gorunescu, 2011).

Konsep data mining, menemukan pola berharga dalam data, adalah respon yang jelas untuk pengumpulan dan penyimpanan volume data yang besar (Weiss, Indurkha, & Zhang, 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 & Zupanb, 2008). Untuk semua aplikasi data mining, akurasi prediksi tergantung pada kualitas prediksi atribut (Weiss, Indurkha, & Zhang, 2010).

### **2.1.3 Support Vector Machine**

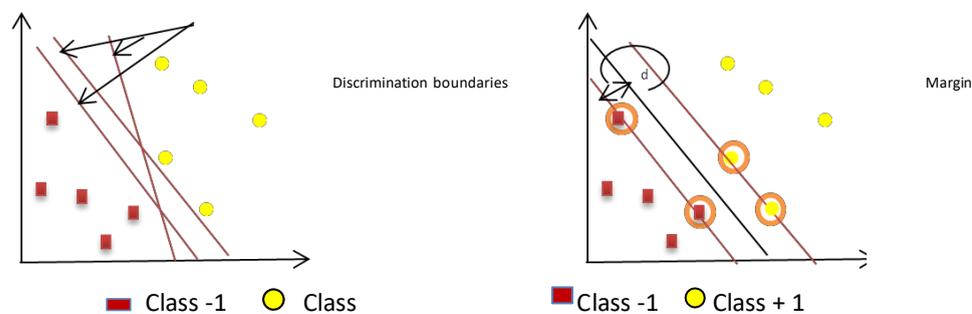
SVM adalah sebuah metode seleksi yang membandingkan parameter standar seperangkat nilai diskrit yang disebut kandidat set, dan mengambil salah satu yang memiliki akurasi klasifikasi terbaik (Dong, Xia, Tu, & Xing, 2007). SVM adalah salah satu alat yang paling berpengaruh dan kuat untuk memecahkan klasifikasi (Burgess, 1998). 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).

Dua kelas data yang digambarkan sebagai lingkaran dan padat titik-titik yang disajikan di angka ini. Secara intuitif diamati, ada banyak keputusan hyperplanes yang dapat digunakan untuk memisahkan kedua kelompok data. Namun, yang digambarkan dengan angka ini dipilih sebagai yang menguntungkan memisahkan bidang, karena mengandung maksimal margin antara dua kelas. Karena itu, dalam tujuan fungsi svm, sebuah istilah regularization mewakili margin muncul. Apalagi seperti yang terlihat di angka ini, hanya mereka yang penuh poin disebut mendukung vektor terutama menentukan memisahkan bidang, sementara poin lain tidak memberi kontribusi untuk margin di semua. Dalam kata lain, hanya sejumlah titik penting untuk klasifikasi tujuan dalam

kerangka svm dan dengan demikian harus diambil (Huang, Yang, King, & Lyu, 2008).

Konsep SVM dapat dijelaskan secara sederhana sebagai usaha mencari hyperplane terbaik yang berfungsi sebagai pemisah dua buah class pada input space. Untuk n-dimensional space, input data  $x_i$  ( $i=1..k$ ), dimana milik kelas 1 atau kelas 2 dan label yang terkait menjadi -1 untuk kelas 1 dan +1 untuk kelas 2. Gambar 1a memperlihatkan beberapa pattern yang merupakan anggota dari dua buah class: positif (dinotasikan dengan +1) dan negatif (dinotasikan dengan -1). Pattern yang tergabung pada class negatif disimbolkan dengan kotak, sedangkan pattern pada class positif, disimbolkan dengan lingkaran. Jika data input dapat dipisahkan secara linear, pemisahan *hyper plane* dapat diberikan dalam:

Proses pembelajaran dalam problem klasifikasi diterjemahkan sebagai upaya menemukan garis (hyperplane) yang memisahkan antara kedua kelompok tersebut. Berbagai alternatif garis pemisah (discrimination boundaries) ditunjukkan pada Gambar 2.2 (Nugroho, 2008).



Gambar 2.2 SVM Berusaha Menemukan Hyperplane Terbaik Yang Memisahkan Kedua Class Negatif Dan Positif 2 (Nugroho, 2008).

Hyperplane pemisah terbaik antara kedua class dapat ditemukan dengan mengukur margin hyperplane tsb. dan mencari titik maksimalnya. Margin adalah jarak antara hyperplane tersebut dengan data terdekat dari masing-masing class. Subset data training set yang paling dekat ini disebut sebagai support vector. Garis solid pada Gambar 2.2 menunjukkan hyperplane yang terbaik, yaitu yang terletak tepat pada tengah-tengah kedua class, sedangkan titik kotak dan lingkaran yang berada dalam lingkaran hitam adalah support vector. Upaya mencari

lokasi hyperplane optimal ini merupakan inti dari proses pembelajaran pada SVM (Nugroho, 2008).

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

$$\vec{w} \cdot \vec{x} + b = 0 \quad (2.1)$$

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

$$\vec{w} \cdot \vec{x} + b = -1 \quad (2.2)$$

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

$$\vec{w} \cdot \vec{x} + b = +1 \quad (2.3)$$

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 2.4, dengan memperhatikan constraint persamaan 2.5.

$$\min_{\vec{w}} \quad \frac{1}{2} \|\vec{w}\|^2 \quad (2.4)$$

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

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

$$L(\vec{w}, b, \alpha) = \frac{1}{2} \|\vec{w}\|^2 + \sum_{i=1}^l \alpha_i (y_i (\vec{x}_i \cdot \vec{w} + b) - 1) \quad (i=1, 2, \dots, l) \quad (2.6)$$

$\alpha_i$  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 2.6 dapat dimodifikasi sebagai maksimalisasi problem yang hanya mengandung  $\alpha_i$  saja, sebagaimana persamaan

2.7.

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

Subject to:

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

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*.

### 2.1.3.1 Contoh Penerapan *Support Vector Machine*

Untuk ilustrasi bagaimana SVM bekerja, mari kita ikuti dua contoh berikut. Satu adalah contoh dimana data yang ada bisa dipisahkan secara linier. Untuk contoh ini kita gunakan problem AND. Contoh yang kedua adalah contoh untuk problem yang tidak bisa dipisahkan secara linier. Untuk contoh ini kita gunakan problem Exclusive OR (XOR). Problem AND adalah klasifikasi dua kelas dengan empat data (lihat Tabel 2.1). Karena ini problem linier, kernelisasi tidak diperlukan. Menggunakan data di Tabel 2.1

Tabel 2.1 AND Problem (Gorunescu, 2011)

X1	X2	y
1	1	1
-1	1	-1
1	-1	-1
-1	-1	-1

dapatkan formulasi masalah optimisasi sebagai berikut:

$$\min \frac{1}{2}(w_1^2 + w_2^2) + C(t_1 + t_2 + t_3 + t_4) \text{ Subject to}$$

$$w_1 + w_2 + b + t_1 \geq 1$$

$$w_1 - w_2 - b + t_2 \geq 1$$

$$-w_1 + w_2 - b + t_3 \geq 1$$

$$w_1 + w_2 - b + t_4 \geq 1$$

$$t_1, t_2, t_3, t_4 \geq 0$$

Karena fungsi AND adalah kasus klasifikasi linier, maka bisa dipastikan nilai variable slack  $t_i=0$ . Jadi Kita bisa masukkan nilai  $C=0$ . Setelah menyelesaikan problem optimasi di atas didapat solusi

$$w_1 = 1, w_2 = 1, b = -1$$

Persamaan fungsi pemisahannya adalah

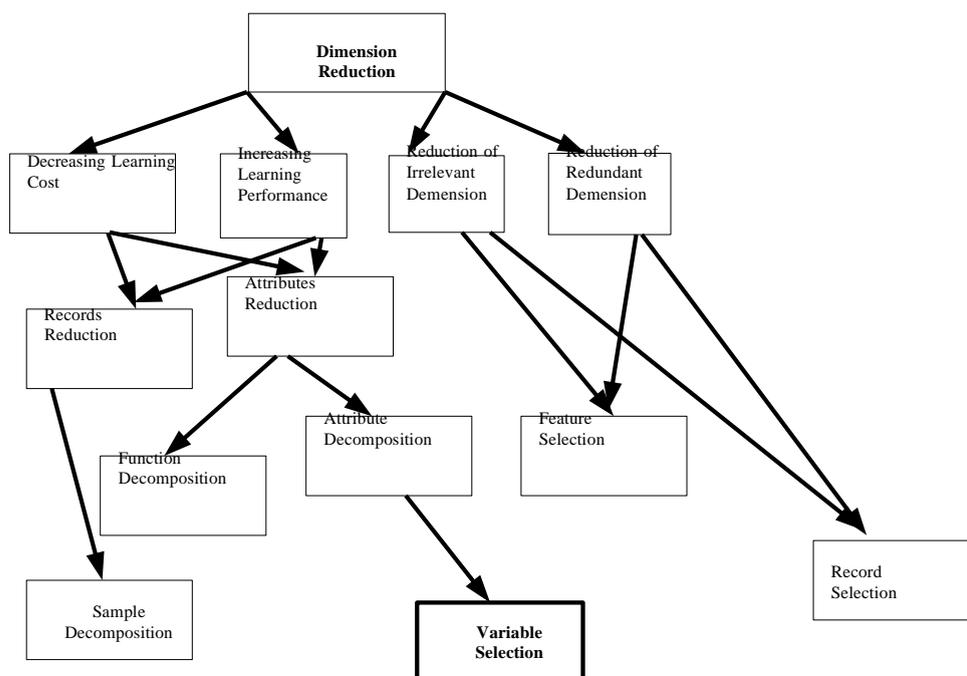
$$f(x) = x_1 + x_2 - 1.$$

Untuk menentukan output atau label dari setiap titik data/obyek kita gunakan fungsi  $g(x) = \text{sign}(x)$ . Dengan fungsi sign ini semua nilai  $f(x) < 0$  diberi label  $-1$  dan lainnya diberi label  $+1$ .

### 2.1.4 Feature Selection

Seleksi fitur adalah terkait erat dengan masalah pengurangan dimensi dimana tujuannya adalah untuk mengidentifikasi fitur dalam kumpulan data-sama pentingnya, dan membuang fitur lain seperti informasi yang tidak relevan dan berlebihan dan akurasi dari seleksinya pada masa depan dapat ditingkatkan (Maimon, 2010). Seleksi fitur adalah salah satu faktor yang paling penting yang dapat mempengaruhi tingkat akurasi klasifikasi karena jika dataset berisi sejumlah fitur, dimensi ruang akan menjadi besar dan non-bersih, merendahkan tingkat akurasi klasifikasi (Liu, Wang, Chen, Dong, Zhu, & Wang, 2011). Masalah dalam seleksi adalah pengurangan dimensi, dimana awalnya semua atribut diperlukan untuk memperoleh akurasi yang maksimal. Empat alasan utama untuk melakukan pengurangan dimensi (Maimon, 2010):

1. Decreasing the learning (model) cost (penurunan pembelajaran (model) biaya)



Gambar 2.3 Taksonomi masalah pengurangan dimensi (Maimon & Rokach, 2010)

2. Increasing the learning (model) performance (meningkatkan pembelajaran (model) kinerja)
3. Reducing irrelevant dimensions (mengurangi dimensi relevan)
4. Reducing redundant dimensions (mengurangi dimensi berlebihan)

Tujuan seleksi fitur adalah untuk pengurangan fitur, untuk menghilangkan dari dataset subset dari variabel yang tidak dianggap relevan untuk tujuan dari kegiatan data mining dan fitur metode seleksi 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.

### **2.1.5 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, Miao, & Liu, 2009).

*Particle Swarm Optimization* (PSO) merupakan algoritma pencarian berbasis populasi dan diinisialisasi dengan populasi solusi acak dan digunakan untuk memecahkan masalah optimasi (Abraham, Grosan, & Ramos, 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, Tiwari, & Kala, 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:

$$v_{ij}(t+1) = w * v_{i,j}(t) + c1 * rand1 * (pbest_{ij}(t) - p_{ij}(t)) + c2 * rand2 * (gbest_{ij}(t) - p_{ij}(t)) \quad (2.9)$$

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

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]

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

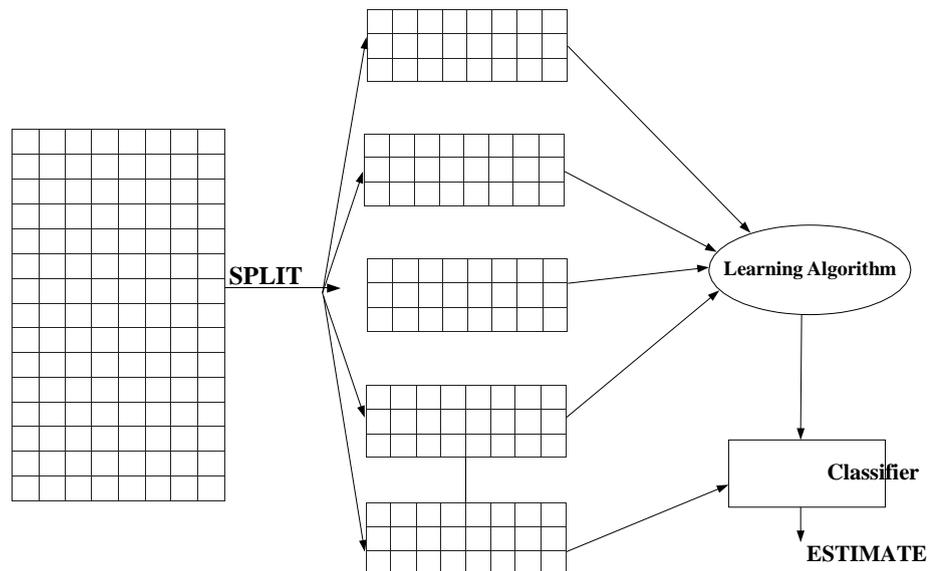
c1 dan c2 adalah faktor pembelajaran pribadi dan sosial (nilainya ke 2)

## 2.1.6 Pengujian Evaluasi dan Validasi Metode Klasifikasi Data Mining

### 2.1.6.1 Pengujian *K-fold Cross validation*

Pendekatan alternatif untuk 'training dan testing' yang sering diadopsi ketika sejumlah contoh kecil (dan yang banyak yang memilih menggunakan terlepas dari

ukuran) dikenal sebagai k kali lipat cross-validasi. Jika dataset terdiri kasus N, ini dibagi menjadi bagian-bagian k sama, k biasanya menjadi sejumlah kecil seperti 5 atau 10. (Jika N tidak tepat habis dibagi oleh k, bagian akhir akan memiliki kasus lebih sedikit dari k lain - 1 bagian) Serangkaian berjalan k kini dilakukan.. Setiap bagian k pada gilirannya digunakan sebagai ujian menetapkan dan k lainnya - 1 bagian digunakan sebagai training set (Bramer, 2007).



Gambar 2.4 *K-fold Cross-validation* (Bramer, 2007)

### 2.1.6.2 *Confusion matrix*

*Confusion matrix* merupakan dataset hanya memiliki dua kelas, kelas yang satu sebagai positif dan kelas yang lain sebagai negatif (Bramer, 2007). Metode ini menggunakan tabel matrix seperti pada tabel 2.2.

Tabel 2.2 Model *Confusion Matrix* (Vercellis, 2009)

		Predictions		Total
		-1(negative)	+(positif)	
Examples	-1 (negatif)	p	q	p + q
	+1 (positif)	u	v	u + v
	Total	p + u	q + v	m

Keterangan:

- a.  $p$  adalah jumlah prediksi yang tepat bahwa *instance* bersifat negatif. b.  $q$  adalah jumlah prediksi yang salah bahwa *instance* bersifat positif. c.  $u$  adalah jumlah prediksi yang salah bahwa *instance* bersifat negatif. d.  $v$  adalah jumlah prediksi yang tepat bahwa *instance* bersifat positif.

Berikut adalah persamaan model *confusion matrix*:

- a. Nilai akurasi (*acc*) adalah proporsi jumlah prediksi yang benar. Dapat dihitung dengan menggunakan persamaan:

$$\text{acc} = \frac{p + v}{(p + q + u + v)} = \frac{p + v}{m} \quad (2.1)$$

- b. Tingkat negatif benar (*tn*) didefinisikan sebagai proporsi kasus negatif yang diklasifikasikan dengan benar, yang dihitung dengan menggunakan persamaan:

$$\text{tn} = \frac{p}{P + q} \quad (2.6)$$

- c. Tingkat negatif palsu (*fn*) adalah proporsi kasus positif yang salah diklasifikasikan sebagai negatif, yang dihitung dengan menggunakan persamaan:

$$\text{fn} = \frac{u}{u + v} \quad (2.7)$$

- d. Tingkat negatif palsu (*fp*) adalah proporsi kasus negatif yang salah diklasifikasikan sebagai positif, yang dihitung dengan menggunakan persamaan:

$$\text{fp} = \frac{q}{p + q} \quad (2.8)$$

- e. Penarikan kembali (*recall*) atau tingkat positif benar (*tp*) adalah proporsi kasus positif yang diklasifikasikan dengan benar, yang dihitung dengan menggunakan persamaan:

$$\text{tp} = \frac{u}{u + v} \quad (2.9)$$

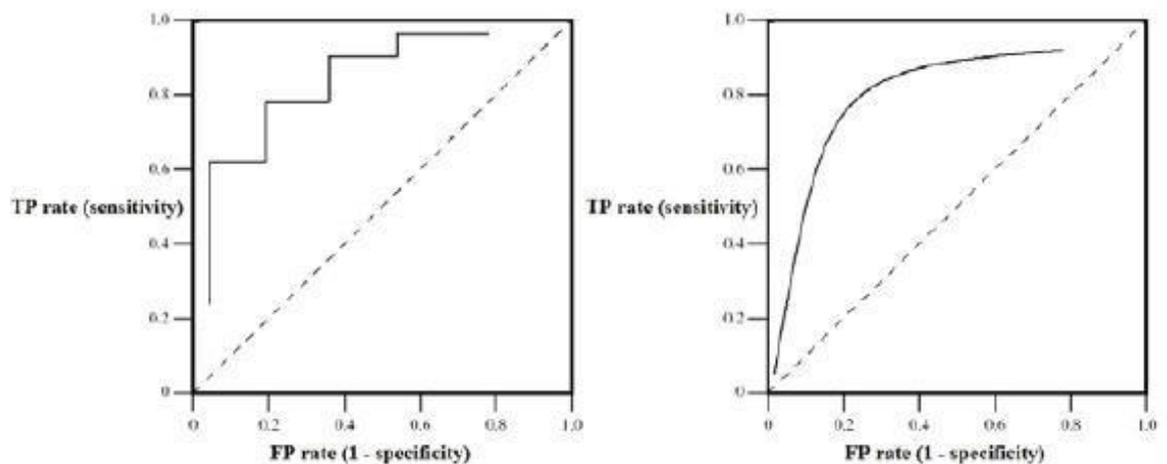
- f. Presisi (*p*) adalah proporsi prediksi kasus positif yang benar, yang dihitung dengan menggunakan persamaan:

$$\text{prc} = \frac{v}{q + v} \quad (2.10)$$

### 2.1.6.3 Kurva ROC (Receiver Operating Characteristic)

Grafik kurva ROC (Receiver operasi karakteristik) digunakan untuk mengevaluasi akurasi classifier dan untuk membandingkan klasifikasi yang berbeda model (Vercellis, 2009). Sebuah grafik ROC adalah grafik dua dimensi dengan proporsi negatif pada sumbu horisontal dan proporsi positif yang benar di sumbu vertikal (Vercellis, 2009). Kegunaan kurva ROC adalah (Gorunescu, 2011) adalah untuk radar selama Perang Dunia II untuk mendeteksi benda-benda musuh di medan perang, teori deteksi sinyal, dalam psikologi ke rekening untuk deteksi sinyal persepsi, penelitian medis dan dalam mesin pembelajaran dan penelitian data mining, serta masalah klasifikasi.

Dalam masalah klasifikasi menggunakan kelas keputusan dua (klasifikasi biner), masing-masing objek dikelompokkan dalam (P, N), yaitu positif atau negatif. Sementara model klasifikasi beberapa (misalnya, pohon keputusan) menghasilkan label kelas diskrit (menunjukkan hanya kelas diprediksi objek), pengklasifikasi lainnya (misalnya, naif Bayes, jaringan saraf) menghasilkan output yang berkesinambungan, yang ambang batas yang berbeda mungkin diterapkan untuk memprediksi keanggotaan kelas, secara teknis, ROC kurva, juga dikenal sebagai grafik ROC, dua-dimensi grafik di mana tingkat TP diplot pada sumbu Y-dan tingkat FP diplot pada X-sumbu (Gorunescu, 2011).



Gambar 2.5 Grafik ROC (*discrete/continuous case*)

(Gorunescu,2011)

Pada Gambar 2.5 garis diagonal membagi ruang ROC, yaitu:

1. (a) poin diatas garis diagonal merupakan hasil klasifikasi yang baik.
2. (b) point dibawah garis diagonal merupakan hasil klasifikasi yang buruk.

Dapat disimpulkan bahwa, satu point pada kurva ROC adalah lebih baik dari pada yang lainnya jika arah garis melintang dari kiri bawah ke kanan atas didalam grafik.

Untuk keakuransian nilai AUC dalam klasifikasi *data mining* dibagi menjadi lima kelompok (Gorunescu, 2011), yaitu:

- a. 0.90 - 1.00 = klasifikasi sangat baik (*excellent classification*)
- b. 0.80 - 0.90 = klasifikasi baik (*good classification*)
- c. 0.70 - 0.80 = klasifikasi cukup (*fair classification*)
- d. 0.60 - 0.70 = klasifikasi buruk (*poor classification*)
- e. 0.50 - 0.60 = klasifikasi salah (*failure*)

## 2.2 Tinjauan Studi

Literatur mengenai pembahasan prediksi penyakit liver telah banyak dilakukan dengan beberapa metode. Penelitian yang dilakukan adalah menggunakan data set yang bersifat *public*, dimana data set yang diambil dari repositori publik yang disepakati oleh para peneliti data mining yaitu UCI Repository. Berikut metode- metode yang pernah digunakan untuk menyelesaikan masalah prediksi penyakit liver:

Dalam penelitian ini algoritma klasifikasi populer dianggap dapat mengevaluasi kinerja dalam hal akurasi, presisi, sensitivitas dan spesififikasi dalam mengklasifikasi dataset pasien liver AP dibandingkan UCLA dataset liver dengan semua algoritma yang dipilih, dengan memakai atribut umum untuk dataset liver seperti Umur, jenis kelamin, SGOT, SGPT, ALP, Total Bilirubin, Direct Bilirubin, Total Protein dan Albumin sangat penting dalam memutuskan status penyakit pada kasus dataset liver. Dengan dataset yang dipilih, algoritma KNN, Back Propagation dan SVM memberikan hasil yang lebih baik dengan semua kombinasi set fitur. (Ramana, Babu & Venkateswarlu, 2011).

Dalam penelitian ini atribut umum dari data dua set yaitu Alkaphos, SGPT dan SGOT diambil untuk menganalisa atribut yang ada pada dua dataset yang berbeda yaitu dengan cara oneway Analysis Variance (ANOVA) dan Multivariate Analysis of Variance (MANOVA), penelitian dilakukan 3 tahap, yaitu percobaan ke-1 menunjukkan bahwa

terdapat perbedaan yang lebih signifikan dalam kelompok dengan semua kombinasi atribut yang memungkinkan, percobaan ke-2 juga menunjukkan bahwa ada lebih signifikan perbedaan dalam kelompok dengan semua kemungkinan atribut kombinasi, dan percobaan yang ke-3 menunjukkan ada lebih perbedaan yang signifikan dengan semua kombinasi atribut yang mungkin, kecuali analisis SGPT antara pasien yang tidak terkena liver pada UCI dataset dan INDIA dataset. (Ramana, Babu & Venkateswarlu, 2012).

Dalam penelitian ini algoritma Classification and regression tree (CART) dan Case base-reasoning (CBR) dapat digunakan untuk mendiagnosa penyakit liver. Metodenya berdasarkan saran dan bantuan dari dokter dan spesialis medis dari kondisi hati, 510 pengunjung rawat jalan menggunakan ICD-9 (international Classification of Diseases, edisi 9) kode disebuah pusat medis di Taiwan dari tahun 2005 sampai 2006 sebagai kasus untuk diagnosis penyakit hati, dan didapat kan hasil pada tahap pertama CART pada pemeriksaan kesehatan menunjukkan pasien menderita penyakit hati akurasinya sebesar 92,94% dan pada tahap kedua menggunakan CBR untuk mengambil kasus yang serupadari kasus dasar untuk mendukung pengobatan penyakit hati hasil akurasinya sebesar 90%.(Lin, 2009).

Dalam penelitian ini Algoritma Berbasis rule dapat digunakan untuk mengklasifikasi pasien penyakit Liver, beberapa algoritma berbasis rule yang digunakan adalah algoritma C.45, ZeroR dan PART. Namun dalam penelitian ini algoritma C.45 memiliki tingkat akurasi yang paling tinggi. (Ramana, Babu, Venkateswarlu, 2012).

Dalam penelitian ini algoritma klasifikasi bayesian dengan bagging dan boosting digunakan untuk mendiagnosis penyakit hati, akurasi bagging yang dikombinasikan dengan akurasi boosting dapat meningkatkan kinerja algoritma klasifikasi bayesian, percobaan dilakukan dengan memisahkan atribut menjadi dua kelompok yaitu perbedaan jenis kelamin pasien penderita liver. Hasil dari kombinasi akurasi bagging dan boosting untuk algoritma klasifikasi bayesian dapat lebih baik jika jumlah datanya lebih besar lagi. (Ramana, Babu, Venkateswarlu, 2011).

Berdasarkan tinjauan studi diatas dapat disimpulkan bahwa ada beberapa peneliti yang sudah menggunakan metode *Support Vector Machine* (SVM) namun tidak ada penelitian sebelumnya yang menggunakan optimasi dalam hal pemilihan fitur yang sesuai. Dalam penelitian ini akan menggunakan algoritma *Support Vector Machine* (SVM) berbasis

*particle swarm optimization* untuk menentukan fitur terbaik pada bobot atribut yang sesuai dan optimal sehingga hasil prediksi lebih akurat.

### 2.3 Obyek Penelitian

Sumber data yang terhimpun langsung oleh peneliti disebut dengan sumber primer, sedangkan apabila melalui tangan kedua disebut sumber sekunder (Riduwan, 2008). Data yang diperoleh adalah data sekunder karena diperoleh dari Indian Liver Patient dataset dalam UCI (singkatan dari ILPD) url :[http://archive.ics.uci.edu/ml/datasets/ILPD+\(Indian+Liver+Patient+Dataset\)](http://archive.ics.uci.edu/ml/datasets/ILPD+(Indian+Liver+Patient+Dataset)). Masalah yang harus dipecahkan di sini adalah prediksi terjadinya penyakit Liver berisi 583 orang yang diperiksa, dan sebanyak 167 pasien tidak terdeteksi terkena penyakit liver, sehingga 416 pasien terdeteksi penyakit liver. Dengan atribut dari penyakit liver adalah usia, Jenis Kelamin, Total Bilirubin (TB) , Direct Bilirubin (DB), Alkaline phosphatase (Alkhapus), Sgpt Alamine (sgpt), Sgot Proteins (Sgot), Total Proteins (TP), Albumin (Alb), rasio albumin dan Globulin (A/G) dan hasil sebagai label yang terdiri atas sakit dan sehat. Data pasien penyakit liver bisa di lihat pada tabel 1. berikut.

**Tabel 1.** Atribut dan data penyakit liver (UCI Repository)

Usia	Jenis Kelamin	TB	DB	Alkhapus	Sgpt	Sgot	TP	ALB	Rasio (A/G)	Hasil
65	Wanita	0.7	0.1	187	16	18	6.8	3.3	0.9	Sakit
62	Pria	10.9	5.5	699	64	100	7.5	3.2	0.74	sakit
62	Pria	7.3	4.1	490	60	68	7	3.3	0.89	Sakit
58	Pria	1	0.4	182	14	20	6.8	3.4	1	Sakit
72	Pria	3.9	2	195	27	59	7.3	2.4	0.4	Sakit
46	Pria	1.8	0.7	208	19	14	7.6	4.4	1.3	Sakit
26	Wanita	0.9	0.2	154	16	12	7	3.5	1	Sakit
29	Wanita	0.9	0.3	202	14	11	6.7	3.6	1.1	sakit
17	Pria	0.9	0.3	202	22	19	7.4	4.1	1.2	Sehat
55	Pria	0.7	0.2	290	53	58	6.8	3.4	1	Sakit
57	Pria	0.6	0.1	210	51	59	5.9	2.7	0.8	sakit
72	Pria	2.4	1.3	260	31	56	7.4	3	0.6	Sakit
64	Pria	0.9	0.3	310	61	58	7	3.4	0.9	sehat
74	Wanita	1.1	0.4	214	22	30	8.1	4.1	1	Sakit
61	Pria	0.7	0.2	145	53	41	5.8	2.7	0.87	Sakit
25	Pria	0.6	0.1	183	91	53	5.5	2.3	0.7	sehat
38	Pria	1.8	0.8	342	168	441	7.6	4.4	1.3	Sakit
33	10pria	1.6	0.5	165	15	23	7.3	3.5	0.92	sehat
40	Wanita	0.9	0.3	293	232	245	6.8	3.1	0.8	Sakit
40	wanita	0.9	0.3	293	232	245	6.8	3.1	0.8	Sakit

## BAB III

### METODOLOGI PENELITIAN

#### 3.1 Desain Penelitian

Pengertian penelitian dalam akademik yaitu digunakan untuk mengacu pada aktivitas yang rajin dan menyelidiki sistematis atau investigasi di suatu daerah, dengan tujuan menemukan atau merevisi fakta, teori, aplikasi dan tujuannya adalah untuk menemukan dan menyebarkan pengetahuan baru (Berndtsson, Olsson, & Lundell, 2008). Menurut (Dawson, 2009) ada empat metode penelitian yang umum digunakan yaitu tindakan penelitian, eksperimen, studi kasus dan survey. Dalam konteks penelitian, metode yang dilakukan mengacu kepada pemecahan masalah yang meliputi mengumpulkan data, merumuskan hipotesis atau proposisi, pengujian hipotesis, menafsirkan hasil, dan kesimpulan (Berndtsson, Hansson, Olsson, & Lundell, 2008).

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

##### 1. Pengumpulan data

Pada tahap ini ditentukan data yang akan di proses. Mencari data yang tersedia, memperoleh

Tabel 3.1. Atribut dan data penyakit Liver (UCI Repository)

Usia	Jenis Kelamin	TB	DB	Alkhapus	Sgpt	Sgot	TP	ALB	Rasio (A/G)	Hasil
65	Wanita	0.7	0.1	187	16	18	6.8	3.3	0.9	Sakit
62	Pria	10.9	5.5	699	64	100	7.5	3.2	0.74	sakit
62	Pria	7.3	4.1	490	60	68	7	3.3	0.89	Sakit
58	Pria	1	0.4	182	14	20	6.8	3.4	1	Sakit
72	Pria	3.9	2	195	27	59	7.3	2.4	0.4	Sakit
46	Pria	1.8	0.7	208	19	14	7.6	4.4	1.3	Sakit
26	Wanita	0.9	0.2	154	16	12	7	3.5	1	Sakit
29	Wanita	0.9	0.3	202	14	11	6.7	3.6	1.1	sakit
17	Pria	0.9	0.3	202	22	19	7.4	4.1	1.2	Sehat
55	Pria	0.7	0.2	290	53	58	6.8	3.4	1	Sakit
57	Pria	0.6	0.1	210	51	59	5.9	2.7	0.8	sakit
72	Pria	2.4	1.3	260	31	56	7.4	3	0.6	Sakit
64	Pria	0.9	0.3	310	61	58	7	3.4	0.9	sehat
74	Wanita	1.1	0.4	214	22	30	8.1	4.1	1	Sakit
61	Pria	0.7	0.2	145	53	41	5.8	2.7	0.87	Sakit
25	Pria	0.6	0.1	183	91	53	5.5	2.3	0.7	sehat
38	Pria	1.8	0.8	342	168	441	7.6	4.4	1.3	Sakit
33	10pria	1.6	0.5	165	15	23	7.3	3.5	0.92	sehat
40	Wanita	0.9	0.3	293	232	245	6.8	3.1	0.8	Sakit
40	wanita	0.9	0.3	293	232	245	6.8	3.1	0.8	Sakit

### 3.3 Pengolahan data awal

Jumlah data awal yang diperoleh dari pengumpulan data yaitu sebanyak 583 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 (vecellis, 2009):

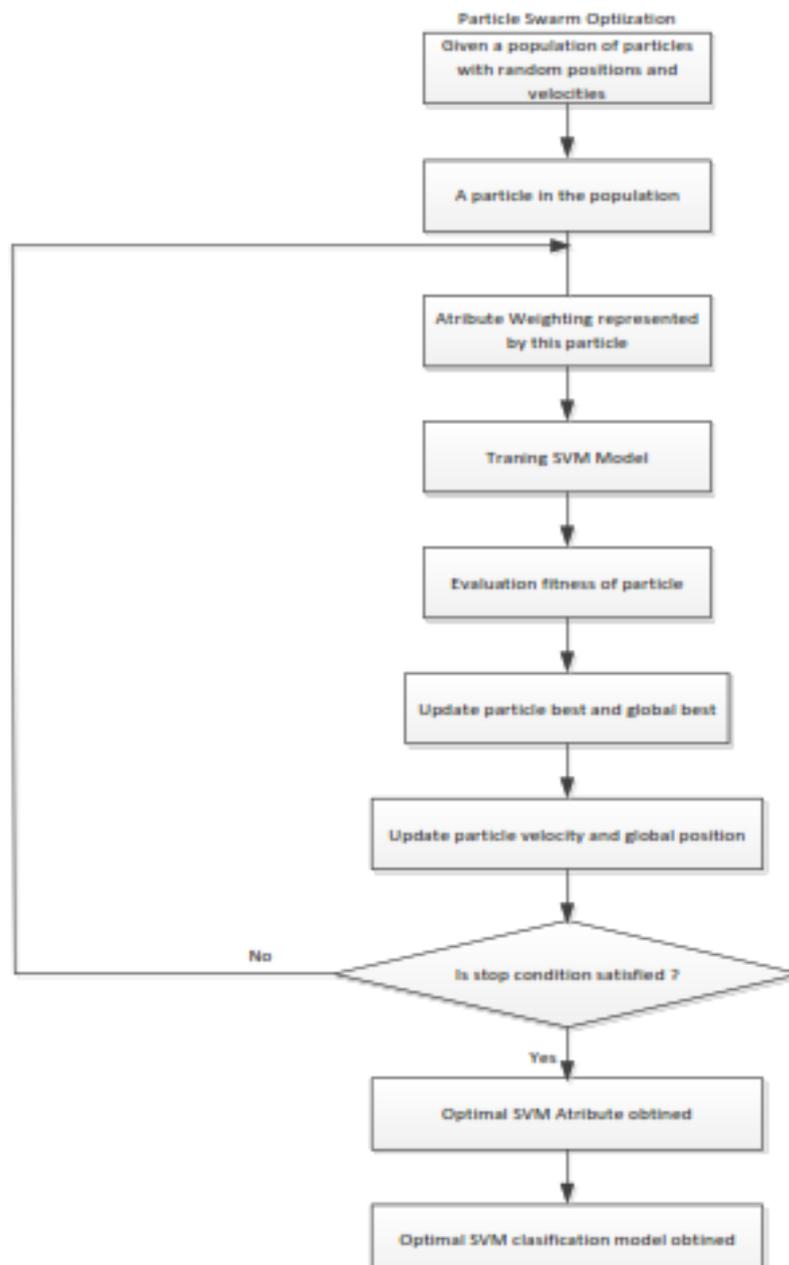
1. Data validation, untuk mengidentifikasi dan menghapus data yang ganjil (outlier/noise), data yang tidak konsisten, dan data yang tidak lengkap (missing value).
2. Data integration and transformation, untuk meningkatkan akurasi dan efisiensi algoritma. Data yang digunakan dalam penulisan ini bernilai kategorikal. Data ditransformasikan kedalam software Rapidminer. Tabel kategorikal atribut terlihat pada table 3.2.
3. Data size reduction and discretization, untuk memperoleh data set dengan jumlah atribut dan record yang lebih sedikit tetapi bersifat informative.

Tabel 3.2 Tabel Atribut yang digunakan (UCI Repository)

No	Atribut	Nilai
1	Usia	Usia (Tahun)
2	Jenis Kelamin	Jenis kelamin (female/male)
3	Total Bilirubin (TB)	Sel darah merah dihati (mg / dl)
4	Direct Bilirubin (DB)	Bilirubin bebas (mg / dl)
5	Alkaline Phosphotase (alkphos)	Enzim dari usus
6	SgptAlamine (Sgpt)	Enzim didalam hati
No	Atribut	Nilai
7	Sgot Protiens (Sgot)	Enzim didalam hati
8	Total Protiens (TP)	Serum protein (g /dl)
9	Albumin	Sintesa protein
10	Rasio Albumin dan Globulin (Rasio A/G)	Perbandingan albumin dan globulin

#### 3.4 Metode yang disulkan

Pada tahap modeling ini dilakukan pemrosesan data traning sehingga akan membahas metode algoritma yang diuji dengan memasukan data penyakit liver kemudian di analisa dan dikomparasi. Berikut ini bentuk gambaran metode algoritma yang akan diuji.



Gambar 3.1 Metode yang diusulkan

### 3.5 Eksperimen dan Pengujian Metode

Tahap modeling untuk menyelesaikan prediksi penyakit diabetes dengan menggunakan dua metode yaitu algoritma support vector machine dan algoritma particle swarm optimization.

1. *Support vector machine* yaitu suatu metode sebuah metode seleksi fitur, dan mengambil salah satu yang memiliki akurasi klasifikasi terbaik.

2. *Particle Swarm Optimization* (PSO) yaitu metode optimasi yang melakukan pencarian menggunakan populasi (swarm) dari individu (partikel) yang diperbaharui dari iterasi untuk iterasi dengan menyeleksi atribut yang ada.

Pada penelitian kali ini yang digunakan adalah penelitian Experiment. Penelitian eksperimen melibatkan penyelidikan hubungan kausal menggunakan tes dikendalikan oleh si peneliti itu sendiri. Dalam penelitian eksperimen digunakan spesifikasi software dan hardware sebagai alat bantu dalam penelitian pada Tabel 3.3:

Tabel 3.3 Spesifikasi hardware dan software

Software	Hardware
Sistem Operasi: Windows 10	CPU: Intel Core i3 Gen 2
Data Mining: Rapid Miner versi 5.2	Memory: 6 GB
	Hardisk: 640 GB

### 3.6 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. Sedangkan penerapan algoritma support vector machine berbasis Particle swarm optimization beracuan pada nilai weight pada algoritma tersebut. Setelah ditemukan nilai akurasi yang paling ideal dari parameter tersebut langkah selanjutnya adalah menentukan nilai weight. Setelah ditemukan nilai akurasi yang paling ideal dari parameter tersebut langkah selanjutnya adalah menentukan weight sehingga terbentuk struktur algoritma yang ideal untuk pemecahan masalah tersebut.

## BAB IV

### HASIL PENELITIAN DAN PEMBAHASAN

#### 4.1 Hasil Eksperimen dan Pengujian Metode

##### 4.1.1 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 4.1. Eksperiment penentuan nilai *training cycle SVM*

C	epsiolon	SVM	
		accury	AUC
0.0	0.0	71,36%	0,500
1.0	1.0	65%	0,500
1.0	1.0	65%	0,500
1.0	0.0	71,36%	0,500
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	71,36%	0,500
1.0	0.0	71,36%	0,500

Hasil terbaik pada *eksperiment SVM* diatas adalah dengan C=0.0 dan Epsilon=0.0 dihasilkan accuracy 71,36 dan AUCnya 0.500 untuk SVM dengan C=1.0 dan Epsilon=0.0 dihasilkan accuracy 77,36 dan AUCnya 0.661.

##### 4.1.2 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 4.2. Eksperiment penentuan nilai *training cycle SVM berbasis PSO*

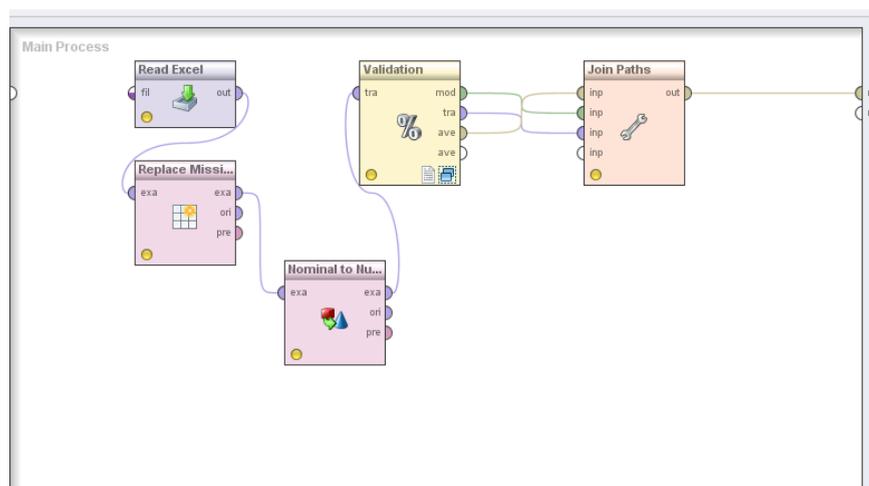
C	epsiolon	SVM		population size	SVM-PSO	
		accury	AUC		accury	AUC
0.0	0.0	71,36%	0,500	5	75.79%	0,572
1.0	1.0	65%	0,500	5	65,11%	0,5
1.0	1.0	65%	0,500	5	65,11%	0,5
1.0	0.0	71,36%	0,500	5	77,36%	0,661

C	epsiolon	SVM		population size	SVM-PSO	
		accury	AUC		accury	AUC
1.0	1.0	65%	0,500	5	65,11%	0,5
1.0	1.0	65%	0,500	5	65,11%	0,5
1.0	1.0	65%	0,500	5	65,11%	0,5
1.0	0.0	71,36%	0,500	100	77,36%	0,661
1.0	0.0	71,36%	0,500	10	77,21%	0,773

Hasil terbaik pada *eksperiment* SVM berbasis PSO diatas adalah dengan C=0.0 dan Epsilon=0 serta population size=5 yang dihasilkan accuracy 75,79% dan AUCnya 0.764 dan dengan C=1.0 dan Epsilon=0 serta population size=5 untuk SVM berbasis PSO dihasilkan accuracy 77,36 dan AUCnya 0.775 .

#### 4.2 Evaluasi dan Validasi Hasil

Hasil dari pengujian model yang dilakukan adalah memprediksi penyakit diabetes 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 algoritma *neural network* berbasis *particle swarm optimization*. Metode pengujiannya menggunakan *cross validation* dengan desain modelnya sebagai berikut.

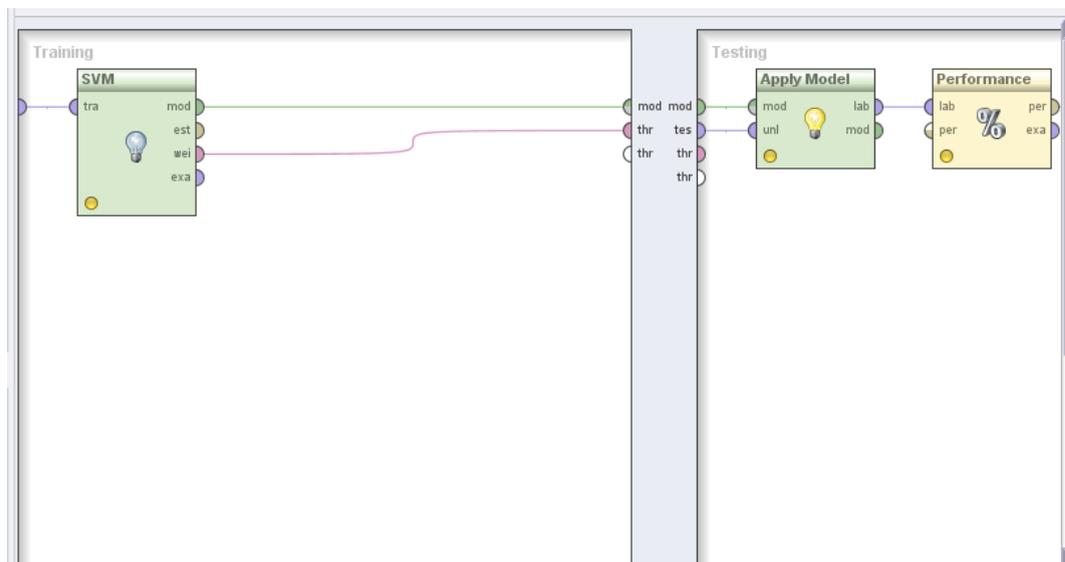


Gambar 4.1. Desain model Validasi

### 4.2.1 Hasil Pengujian Model Support Vector Machine

Pada penelitian penentuan hasil penyakit diabetes menggunakan algoritma

*Support Vector Machine* berbasis pada framework RapidMiner sebagai berikut



Gambar 4.2. 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 4.1

#### 1. *Confusion Matrix*

Tabel 4.2 diketahui dari 583 data, 416 diklasifikasikan *sakit* sesuai dengan prediksi yang dilakukan dengan metode svm, lalu 167 data diprediksi *sakit* tetapi ternyata hasilnya prediksi sehat, 0 data *class tidak* diprediksi sesuai, dan 0 data diprediksi tidak ternyata hasil prediksinya y a

Tabel 4.3 Model *Confusion Matrix* untuk Metode *Support Vector Machine*

accuracy:71.36% +/-0.97% (mikro: 71.36%)			
	True Sakit	True Sehat	Class precision
pred. Sakit	416	167	71,36%

pred. Sehat	0	0	0.00%
class recall	100.00%	0.00%	

Berdasarkan Tabel 4.1 tersebut menunjukkan bahwa, tingkat akurasi dengan menggunakan algoritma SVM adalah sebesar 71,36%, dan dapat dihitung untuk mencari nilai *accuracy*, *sensitivity*, *specificity*, *ppv*, dan *npv* hasilnya dan dapat dihitung untuk mencari nilai *accuracy*, *sensitivity*, *specificity*, *ppv*, dan *npv* pada persamaan dibawah ini:646456

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} = \frac{118 + 452}{118 + 452 + 150 + 48} = 0.7421$$

$$\frac{TP + TN + FP + FN}{118 + 452 + 150 + 48}$$

$$Sensitivity = \frac{TP}{TP + FN} = \frac{118}{118 + 48} = 0.7108$$

$$\frac{TP}{TP + FN} = \frac{118}{118 + 48}$$

$$Specificity = \frac{TN}{N + FP} = \frac{452}{452 + 150} = 0.7508$$

$$\frac{TN}{N + FP} = \frac{452}{452 + 150}$$

Hasil perhitungan terlihat pada tabel 4.4 dibawah ini:

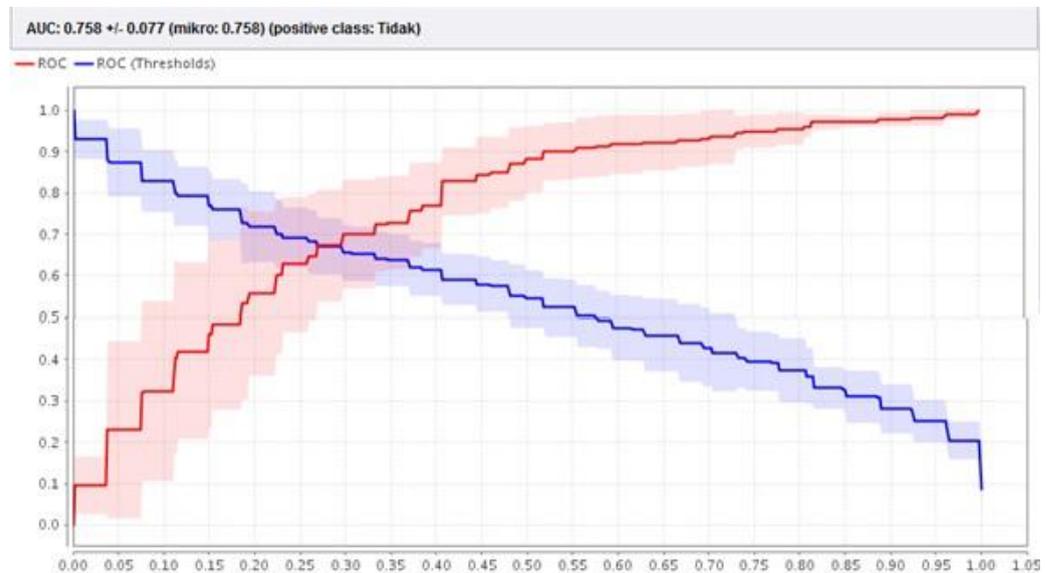
Tabel 4.4 Nilai *accuracy*, *sensitivity*, *specificity*, *ppv*, dan *npv* untuk Metode *Support Vector Machine*

	Nilai (%)
<b><i>Accuracy</i></b>	74,21
<b><i>Sensitivity</i></b>	72.43
<b><i>Specificity</i></b>	42.16

<b>PPV</b>	74.67
<b>NPV</b>	91.4

## 2. Kurva ROC

Hasil perhitungan divisualisasikan dengan kurva ROC. Perbandingan kedua metode komparasi bisa dilihat pada Gambar 4.2 yang merupakan kurva ROC untuk algoritma *Support Vector Machines*. Kurva ROC pada gambar 4.2 mengekspresikan *confusion matrix* dari Tabel 4.1. Garis horizontal adalah *false positives* dan garis vertikal *true positives*.

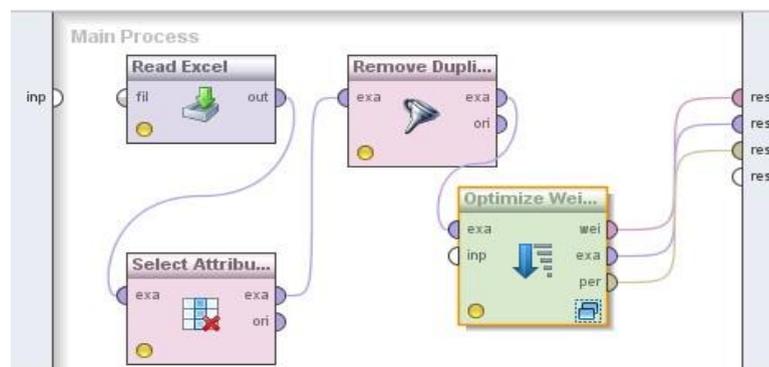


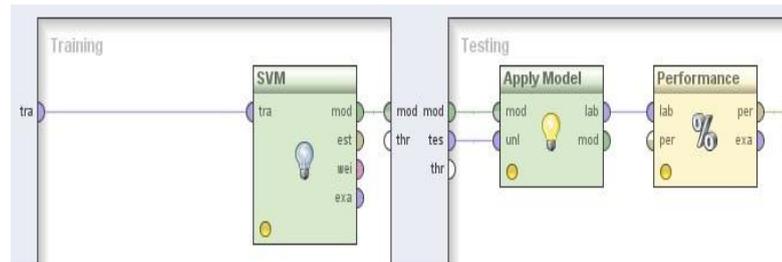
Gambar 4.3. Kurva ROC dengan Metode *Support Vector Machines*

Dari Gambar 4.3 terdapat grafik ROC dengan nilai AUC (*Area Under Curve*) sebesar 0.758 dimana diagnosa hasilnya *Fair classification*.

#### 4.2.2 Hasil Pengujian Model *Support Vector Machine* berbasis Algoritma *Particle Swarm Optimization* (PSO)

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





Gambar 4.4. Model pengujian validasi *support vector machine* 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* didapatkan hasil pada tabel 4.2

#### 1. *Confusion Matrix*

Tabel 4.5 diketahui dari 583 data, 295 diklasifikasikan *sakit* sesuai dengan prediksi yang dilakukan dengan metode SVM berbasis *Particle Swarm Optimization* (PSO), lalu 112 data diprediksi *sakit* tetapi ternyata hasilnya prediksi sehat, 8 data *class sehat* diprediksi sesuai, dan 1 data diprediksi sehat ternyata hasil prediksinya sakit.

Tabel 4.5 Model *Confusion Matrix* untuk Metode *Support Vector Machine* berbasis *Particle Swarm Optimization*

<b>accuracy:77.36% +/-4.06% (mikro: 77.34%)</b>			
	True sakit	True Sehat	Class precision
pred. Ya	295	112	72,48%
pred. Tidak	1	8	88.89%
class recall	99.66%	6.67%	

berdasarkan Tabel 4.5 tersebut menunjukkan bahwa, tingkat akurasi dengan menggunakan algoritma SVM berbasis *Particle Swarm Optimization* (PSO) adalah sebesar 77,36%, dan dapat dihitung untuk mencari nilai *accuracy*, *sensitivity*, *specificity*, *ppv*, dan *npv* hasilnya dan dapat dihitung untuk mencari nilai *accuracy*, *sensitivity*, *specificity*, *ppv*, dan *npv* pada persamaan dibawah ini:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} = \frac{129 + 465}{129 + 465 + 139 + 35} = 0.7736$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} = \frac{129}{129 + 35} = 0.7866$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} = \frac{465}{465 + 139} = 0.7699$$

$$\text{PPV} = \frac{\text{TP}}{\text{TP} + \text{FP}} = \frac{129}{129 + 139} = 0.4813$$

$$\text{NPV} = \frac{\text{TN}}{\text{TN} + \text{FN}} = \frac{465}{465 + 35} = 0.933$$

Hasil perhitungan terlihat pada tabel 4.6 dibawah ini:

Tabel 4.6 Nilai *accuracy*, *sensitivity*, *specificity*, *ppv*, dan *npv* Metode *Support Vector Machine* berbasis *Particle Swarm Optimization*

	Nilai (%)
<i>Accuracy</i>	77.36
<i>Sensitivity</i>	78.86
<i>Specificity</i>	76.99
<i>PPV</i>	48.13
<i>NPV</i>	93.3

## 2. Kurva ROC

Hasil perhitungan divisualisasikan dengan kurva ROC. Perbandingan kedua metode komparasi bisa dilihat pada Gambar 4.4 yang merupakan kurva ROC untuk algoritma *Support Vector Machines* berbasis *Particle Swarm Optimization* (PSO). Kurva ROC pada gambar 4.4 mengekspresikan *confusion matrix* dari Tabel 4.3. Garis horizontal adalah *false positives* dan garis vertikal *true positives*.



Gambar 4.5. Kurva ROC dengan Metode *Support Vector Machines* berbasis *Particle Swarm Optimization* (PSO)

Dari Gambar 4.5 terdapat grafik ROC dengan nilai AUC (*Area Under Curve*) sebesar 0.775 dimana diagnosa hasilnya *Fair classification*.

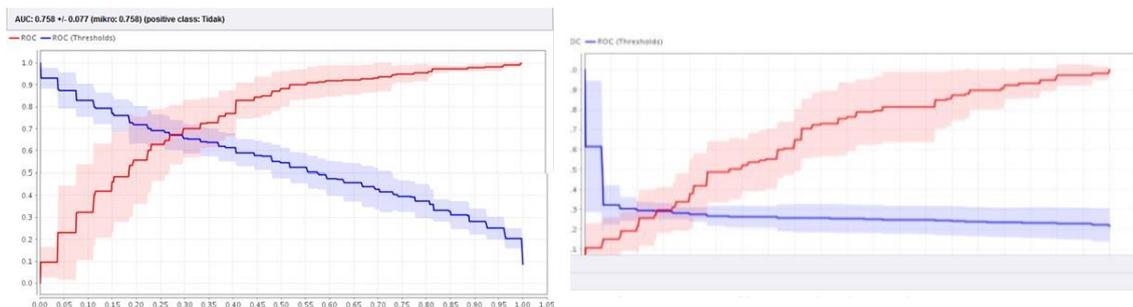
### 4.2.3 Analisis Evaluasi dan Validasi Model

Dari hasil pengujian diatas, baik evaluasi menggunakan *confusion 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 71.36% dan nilai akurasi untuk model algoritma SVM berbasis PSO sebesar 77.36 % dengan selisih akurasi 3.15%, dapat dilihat pada Tabel 4.5 dibawah ini:

Tabel 4.7 Pengujian algoritma SVM dan SVM berbasis PSO

	<i>Accuracy</i>	<i>AUC</i>
<b>SVM</b>	71.36%	0.500
<b>SVM berbasis PSO</b>	77.36 %	0.661

Untuk evaluasi menggunakan ROC *curve* sehingga menghasilkan nilai AUC (*Area Under Curve*) untuk model algoritma SVM menghasilkan nilai 0.500 dengan nilai diagnosa *Fair Classification*, sedangkan untuk algoritma SVM berbasis PSO (*Particle Swarm Optimization*) menghasilkan nilai 0.661 dengan nilai diagnose *Fair Classification*, dan selisih nilai keduanya sebesar 0.161. Dapat dilihat pada Gambar 4.5 dibawah ini.



Gambar 4.7 Kurva ROC dengan *support vector machine* dan *support vector machine* berbasis *Particle Swarm Optimization*)

Dengan demikian algoritma SVM berbasis PSO dapat memberikan solusi untuk permasalahan dalam prediksi hasil prediksi penyakit diabetes. Untuk rinciannya dapat dilihat pada Tabel 4.5. dan Gambar 4.7.

### 4.3 Pembahasan

Tingkat populasi penyakit liver semakin lama mengalami peningkatan, menunjukkan terjadinya permasalahan dalam diagnosa penyakit diabetes. Untuk menganalisa penyakit diabetes yang akurat diperlukan metode yang paling tepat. Pada penelitian ini *Support Vector Machine (SVM)* digunakan karena diketahui dari hasil penelitian, sebelumnya bahwa *Support Vector Machine (SVM)* memiliki kemampuan dapat meminimalkan kesalahan melalui

**Program Pascasarjana Magister Ilmu Komputer STMIK Nusa Mandiri**

memaksimalkan margin dengan misahkan antara *hyper-lane* dan satu set data bahkan dengan jumlah sample yang kecil. 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 71.36 % dan mempunyai nilai AUC sebesar 0.500. Setelah dilakukan penyesuaian pada parameter C dan epsilon didapat nilai akurasi terbaik untuk algoritma *support vector machine* yaitu mempunyai akurasi sebesar 71.36 % dan nilai AUCnya sebesar 0.500. Dari hasil tersebut diketahui bahwa keberhasilan dari *Support Vector Machine (SVM)* sangat dipengaruhi oleh pemilihan atribut yang tepat. Semakin banyak atribut dan informasi yang digunakan akan mengakibatkan banyaknya waktu dan biaya yang dikorbankan bahkan akan mengurangi tingkat akurasi dan kompleksitas yang lebih tinggi. kompleksitas yang lebih tinggi.

Mengingat pentingnya seleksi atribut dalam *Support Vector Machine (SVM)* maka diterapkan *Particle swarm optimization (PSO)* untuk melakukan tugas tersebut. *Particle swarm optimization (PSO)* diketahui dapat digunakan sebagai teknik optimasi untuk mengoptimalkan subset fitur. Algoritma PSO sederhana dan memiliki kompleksitas yang lebih rendah, sehingga dapat memastikan solusi optimal dengan menyesuaikan pencarian global dan lokal, sehingga kinerja klasifikasi *Support Vector Machine (SVM)* dapat ditingkatkan. *Eksperiment* dilakukan kembali dengan menerapkan *Particle swarm optimization (PSO)* untuk seleksi atribut dalam *Support Vector Machine (SVM)* dan dilakukan penyesuaian pada parameter C,  $\epsilon$  dan *population*. Dari 8 atribut yang digunakan kemudian diseleksi atribut sehingga terpilih 5 atribut yang digunakan dengan hasil akurasi 77.36 % dan nilai AUC sebesar 0.661. Setelah dilakukan penyesuaian pada parameter C,  $\epsilon$  dan *population* didapat nilai akurasi terbaik untuk algoritma *support vector machine* berbasis *particle swarm optimization* yaitu mempunyai akurasi sebesar 77.36 % dan nilai AUC sebesar 0.661.

Hasil dari *eksperiment* tersebut menunjukkan pengujian data kredit German UCI data set menggunakan *Support Vector Machine (SVM)* dan penerapan *Particle Swarm Optimization* dalam pemilihan atribut menunjukkan hasil

yang lebih akurat dalam prediksi penyakit diabetes dengan peningkatan nilai akurasi sebesar 3.15% dan perbedaan nilai AUC sebesar 0,161.

#### 4.4 Implikasi Penelitian

Implikasi penelitian pada tesis ini peneliti arahkan pada tiga aspek, yaitu:

##### 1. Aspek Sistem

Penerapan kebijakan dalam penentuan prediksi penyakit liver di rumah sakit akan membawa pengaruh pada sistem, di mana melalui data-data yang ada rumah sakit akan mempermudah dalam menentukan penyakit liver.

Dengan data set yang ada serta atribut yang telah ditentukan diharapkan penentuan penyakit diabetes dapat lebih akurat sehingga penyakit diabetes yang semakin meningkat dari tahun ketahun dapat berkurang jumlah penderitanya. Dan dengan atribut yang sudah ada dan faktor penyebab yang telah diinformasikan diharapkan penderita penyakit diabetes maupun yang belum terkena serta mempunyai gejala penyakit diabetes tersebut dapat lebih memerhatikannya sehingga tidak meningkat lagi jumlah penderita penyakit diabetes. Dalam penerapannya dengan data serta atribut yang telah ada dipenyakit diabetes, maka data tersebut dapat diuji dengan menggunakan *software RapidMiner*, *software* ini merupakan *open source*, sehingga bisa didapatkan dengan gratis. Sehingga dapat diketahui tingkat akurasi dari data tersebut. Dengan memanfaatkan perangkat yang ada, seperti *hardware* dan jaringan, system akan berjalan dengan lebih baik dalam memecahkan permasalahan yang ada.

##### 2. Aspek Manajerial

Secara manajerial kinerja dokter akan lebih membawa dampak yang signifikan dalam memprediksi penyakit liver. Dengan penelitian ini diharapkan penderita penyakit liver lebih meningkatkan kepeduliannya mengenai kesehatan sehingga tidak mengalami komplikasi yang bisa menyebabkan kematian. Dan dengan adanya penelitian ini dapat mengatasi masalah penyakit liver yang semakin meningkat dengan atribut yang ada dapat diprediksi penyakit

liver dengan lebih akurat dan meningkatkan kesadaran manusia mengenai bahayanya penyakit liver. Selain itu penerapannya dapat dilakukan untuk memprediksi penyakit liver berikutnya kepada penderita penyakit liver dalam hal untuk meningkatkan hasil prediksi penyakit liver.

### 3. Penelitian Lanjutan

Penelitian semacam ini dapat dikembangkan pada bidang kesehatan. Penelitian ini juga dapat dikembangkan dengan algoritma aturan klasifikasi yang lain yang terdapat dalam data mining, seperti algoritma *Naive Bayes*, *KNN*, dan dapat ditingkatkan akurasinya dengan optimasi lainnya seperti ACO dan Adabost.

## BAB V

### PENUTUP

#### 5.1 Kesimpulan

Dalam penelitian ini dilakukan pengujian model dengan menggunakan *Support Vector Machines* dan *Support Vector Machines* berbasis *Particle Swarm Optimization* dengan menggunakan data penyakit liver yang terkena penyakit atau tidak. Model yang dihasilkan diuji untuk mendapatkan nilai *accuracy*, *precision*, *recall* dan AUC dari setiap algoritma sehingga didapat pengujian dengan menggunakan *support vector machines* didapat nilai *accuracy* adalah 71.36 % dengan nilai dan nilai AUC adalah 0.500. sedangkan pengujian dengan menggunakan *support vector machines* berbasis *Particle Swarm Optimization* (PSO) didapatkan nilai *accuracy* 77.36% dengan nilai dan nilai AUC adalah 0.661. Maka dapat disimpulkan pengujian pengujian data liver UCI data set menggunakan *support vector support vector machines* berbasis *Particle Swarm Optimization* (PSO) didapat bahwa pengujian *support vector machines* berbasis algoritma berbasis *Particle Swarm Optimization* (PSO) lebih baik dari pada *support vector machines* sendiri. Dengan demikian dari hasil pengujian model diatas dapat disimpulkan bawa *support vector machines* berbasis *Particle Swarm Optimization* memberikan pemecahan untuk permasalahan prediksi penyakit diabetes lebih akurat.

#### 5.2 Saran

Agar penelitian ini bisa ditingkatkan, berikut adalah saran-saran yang diusulkan:

1. Penelitian ini diharapkan dapat digunakan pihak medis sebagai bahan pertimbangan memprediksi penyakit liver, sehingga dapat meningkatkan akurasi dalam prediksi penyakit liver.
2. Penelitian ini dapat dikembangkan dengan metode optimasi lainnya seperti *Ant Colony Optimization* (ACO), *Genetic Algorithm* (GA), dan lainnya.
3. Penelitian ini dapat dikembangkan dengan metode klasifikasi data mining lainnya seperti *Naive Bayes*, *KNN* dan lainnya untuk melakukan perbandingan.

## DAFTAR PUSTAKA

- Bellazzi, R., & Zupanb, B. (2008). *Predictive Data Mining In Clinical Medicine: Current Issues And And Guidelines*. International Journal Of Medical Informatics 77 , 81–97.
- Berndtssom, M., Hansson, J., Olsson, B., & Lundell, B. (2008). *A Guide For Students In Computer Science And Information Systems*. London: Springer.
- Burges, C. J. (1998). *A Tutorial On Support Vector Machines For Pattern Recognition*. Boston: Kluwer Academic Publishers.
- Dong, Y., Xia, Z., Tu, M., & Xing, G. (2007). *An Optimization Method For Selecting Parameters In Support Vector Machines*. Sixth International Conference On Machine Learning And Applications , 1-6.
- 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.
- Huang, K., Yang, H., King, I., & Lyu, M. (2008). *Machine Learning Modeling*
- Kononenko, I. (2001). *Machine Learning For Medical Diagnosis: History, State Of The Art and Perspective*. *Artificial Intelligence In Medicine*, 23, 89-109.
- Lin Rong-Ho. 2009. *An intelligent model for liver disease diagnosis*. *Artificial Intelligence in medicine*
- Lin, S.-W., Shiue, Y.-R., Chen, S.-C., & Cheng, H.-M. (2009). *Applying enhanced data mining approaches in predicting bank performance: A case of Taiwanese commercial banks*. *Expert Systems with Applications* 36 , 11543–11551.
- 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.
- Lumongga, F. (2009, Juli). *Struktur Liver*. Retrieved From USU Institutional Repository: <http://repository.usu.ac.id/handle/123456789/2052>.
- Maimon, O. (2010). *Data Mining And Knowledge Discovery Handbook*. New York Dordrecht Heidelberg London: Springer.
- Moertini, V. S. (2002). *Data Mining Sebagai Solusi Bisnis*. *Integral*, Vol. 7 No. 1, April 2002 , 4.

- Parimala, R., & Nallaswamy, R. (2012). *Feature Selection using a Novel Particle Swarm Optimization and It's Variants*. *I.J. Information Technology and Computer Science*, , 16-24.
- Praetorius, G. (2006). Retrieved from Linkoping University Department Of Computer and Information Science: <http://www.ida.liu.se/~729G50/studentpapper-06/gesa-praetorius.pdf>.
- Ramana, B. V., Babu, S.P., & Venkateswarlu, N. B. 2011. *A Critical Study Of Selected Classification Algorithms For Liver Disease Diagnosis*. *International Journal Of Database Management Systems*, 3 (2), 101-114.
- Ramana, B. V., Babu, S.P., & Venkateswarlu, N. B. 2012. *A Critical Comparative Study of Liver Patients from USA and INDIA: An Exploratory Analysis*. *International Journal Of Computer Science Issues*, Vol. 9, Issues 3, No.2.
- Ramana, B. V., Babu, S.P., & Venkateswarlu, N. B. 2012. *Liver Classification Using Modified Rotation Forest*. *International Journal Of Engineering Research And Development*, 1 (6), 17-24.
- Ramana, B. V., Prabu, S. P., & Venkateswarlu, N. B. 2012, Mei 21. *ILPD (Indian Liver Patient Dataset)*. Retrieved Juni 17, 2013, from MachineLearningRepository: <https://archive.ics.uci.edu/ml/datasets/ILPD+%28Indian+Liver+Patient+Dataset%29>
- Ramana, B. V., Babu, S.P., & Venkateswarlu, N. B. 2011. *A Critical Evaluation of Bayesian Classifier For Liver Diagnosis Using Bagging and Boosting Methods*. *International Journal Of Engineering Science and Technology (IJEST)*, 3 (4), 3422-3426.
- Riduwan. (2008) *Metode dan Teknik Menyusun Tesis*. Alfabeta. Bandung).
- Salindeho, E. (2013). Penyakit Liver. [http://www.bmodtcenter.com/artikel\\_kesehatan.html](http://www.bmodtcenter.com/artikel_kesehatan.html). Diakses 20 November 2013.
- Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data Mining Practical Machine Learning Tools And Techniques*. Burlington, Usa: Morgan Kaufmann Publishers.



	<b>LEMBAR KONSULTASI BIMBINGAN TESIS</b>
	SEKOLAH TINGGI MANAJEMEN INFORMATIKA & KOMPUTER <b>NUSA MANDIRI</b>

- NIM : 14000959
- Nama Lengkap : Nu'man Musyaffa
- Dosen Pembimbing : Dr. Mochamad Wahyudi, MM, M.Kom, M.Pd
- Judul Tesis : "Penerapan *Particle Swarm Optimization* Untuk Seleksi Atribut Pada Metode *Support Vector Machine* Untuk Prediksi Penyakit Liver"

Foto  
2x3

No	Tanggal Bimbingan	Pokok Bahasan	Paraf Dosen Pembimbing
1	26 Okt '15	Bimbingan perdana & Pengajuan Bab I	
2	28 Okt '15	Pengajuan Bab II	
3	31 Okt '15	Pengajuan Bab III	
4	5 Nop '15	Pengajuan Bab IV	
5	21 Nop '15	Pengajuan Bab V	
6	30 Nop '15	Periksa keseluruhan & Acc Ujian Sidang	

Catatan untuk dosen pembimbing

Bimbingan Tesis

- Dimulai pada tanggal : 26 Oktober 2015
- Diakhiri pada tanggal : 30 Nopember 2015
- Jumlah pertemuan bimbingan : 6 Kali Bimbingan

Disetujui Oleh,  
Dosen Pembimbing



Dr. Mochamad Wahyudi, MM, M.Kom, M.Pd