

# Komparasi Algoritma Naive Bayes Dengan *Support Vector Machine* Berbasis *Particle Swarm Optimization* untuk Prediksi Kesuburan

Ela Nurelasari<sup>1,\*</sup>

<sup>1</sup> Manajemen Informatika; AMIK BSI Jakarta; Jl. R.S. Fatmawati No.24, Pondok Labu, Jakarta Selatan, telp/fax (021)7500282/(021)7513790; e-mail: ela.eur@bsi.ac.id.

\* Korespondensi: e-mail: ela.eur@bsi.ac.id

Diterima: 15 Mei 2018; Review: 29 Mei 2018; Disetujui: 12 Juni 2018

Cara sitasi: Nurelasari E. 2018. Komparasi Algoritma Naive Bayes Dengan *Support Vector Machine* Berbasis *Particle Swarm Optimization* untuk Prediksi Kesuburan. Bina Insani ICT Journal. 5(1): 61-70.

---

**Abstrak:** Berdasarkan penelitian bahwa kesuburan pria telah mengalami penurunan, hal ini dapat disebabkan oleh beberapa faktor lingkungan dan gaya hidup, diantaranya seperti pecandu alkohol, rokok, usia, faktor genetik dan musim dapat berpengaruh pada sperma yang berkualitas. Penelitian ini menguji kemampuan antara metode algoritma Naive Bayes dengan Support Vector Machine berbasis Particle Swarm Optimization, dimana dataset yang digunakan diambil dari dataset fertilitas UCI Machine Learning Repositori. Dataset terdiri dari 100 sample dan 10 field/atribut. Hasil dari komparasi kedua metode tersebut dimana klasifikasi Support Vector Machine berbasis Particle Swarm Optimization memperoleh nilai accuracy lebih tinggi 88.00% dibandingkan dengan algoritma naive bayes dengan nilai accuracy 85.00%.

**Kata kunci:** Kesuburan, *Naive Bayes*, *Support Vector Machine*, *Particle Swarm Optimization*.

**Abstract:** Previous research proved that man fertility has decreased. This is caused by some environmental factors and life style. Alcohol, cigarette, age, genetic factor, and season may cause the quality of sperm. This research analyzes the capability between Naive Bayes Algorithm method and Support Vector Machine based on Particle Swarm Organisation. The dataset was taken from fertility dataset in UCI Machine Learning Repositori. It consists of 100 samples and 10 fields. The result of the both methods showed that Particle Swarm Organisation has 88.00% accuration score compared to Naive Bayes Algorithm has 85.00%.

**Keywords:** Fertility, *Support Vector Machine*, *Naive Bayes*, *Particle Swarm Optimization*.

## 1. Pendahuluan

Tingkat kesuburan terutama pada pria telah menurun secara drastis. Yang mempengaruhi kualitas sperma adalah faktor lingkungan serta kebiasaan hidup. Dimana yang mempengaruhi potensi kesuburan pria yaitu adanya peningkatan penyakit reproduksi laki-laki [Irvine, 2000]. Tetapi juga dapat dipengaruhi oleh faktor lingkungan atau pekerjaan [Giwerzman and Giwerzman, 2011]. Kualitas kesuburan pria dipengaruhi juga oleh konsumsi alkohol dan rokok. Ada beberapa faktor yang dapat mempengaruhi kesuburan pria diantaranya musim/cuaca, usia, penyakit bawaan dari kecil, adanya trauma, bedah, demam tinggi, alkohol, rokok, dan jam lama duduk [Gil et al., 2012].

Data mining merupakan cabang ilmu yang termasuk masih baru tetapi telah menghasilkan keuntungan yang cukup besar pada saat ini. Manfaat yang dirasakan dengan penggunaan data mining sudah tidak diragukan lagi, beberapa kasus telah banyak diselesaikan dengan menggunakan teknik data mining [Widodo et al., 2013]. Algoritma naive bayes terbukti

efektif dalam banyak aplikasi praktis. Termasuk klasifikasi teks, medical diagnosis dan manajemen performa sistem [Domingos and Pazzani, 1997]. Klasifikasi naive bayes menggunakan teori bayesian dengan mengasumsikan tidak ada hubungan antar atribut. Model ini mudah digunakan serta sederhana sehingga dianggap tepat untuk database yang besar. Hasil dari naive bayes dianggap baik karena mampu melakukan klasifikasi dengan baik [Khasanah].

Naive Bayes adalah klasifikasi statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu class [Kusrini and Lutfi, 2009]. Dengan diaplikasikan ke dalam database dengan dataset yang besar Naive Bayes terbukti memiliki akurasi tinggi. Dan dikenal dengan teorema bayes karena dapat memprediksi peluang dimasa depan dengan melihat pengalaman di masa sebelumnya, Metode ini ditemukan oleh ilmuwan bernama Thomas Bayes dari Inggris. Menurut penelitian sebelumnya *Teorema bayes* dapat dipadukan dengan naive bayes yang diasumsikan dengan kondisi antar atribut yang saling bebas [Hand, 2007]. Naive Bayes juga termasuk metode klasifikasi yang sangat terkenal dan masuk dalam sepuluh algoritma terbaik dalam data mining [Kusrini and Lutfi, 2009]. Berikut bentuk umum dari teorema bayes:

The diagram illustrates Bayes' Theorem with the following components and labels:

- Likelihood:**  $P(x|c)$
- Class Prior Probability:**  $P(c)$
- Posterior Probability:**  $P(c|x)$
- Predictor Prior Probability:**  $P(x)$

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)}$$

$$P(c|X) = P(x_1|c) \times P(x_2|c) \times \dots \times P(x_n|c) \times P(c)$$

Sumber: [Kusrini and Lutfi, 2009]

Gambar 1. Teorema bayes

Keterangan:

X= Data dimana classnya belum diketahui

C= Hipotesis data X merupakan suatu class spesifik

$P(C|X)$  = Probabilitas hipotesis H didasarkan dari kondisi X (*posteriori probability*)

$P(C)$  = Probabilitas Hipotesis H (*Prior probability*)

$P(X|C)$  = *Probabilitas* X berdasar kondisi Hipotesis H

$P(X)$  = Probabilitas X

Metode bayes merupakan penyederhanaan dari Metode algoritma naive bayes. Untuk mempermudah pemahaman, maka teorema Bayes disederhanakan menjadi:

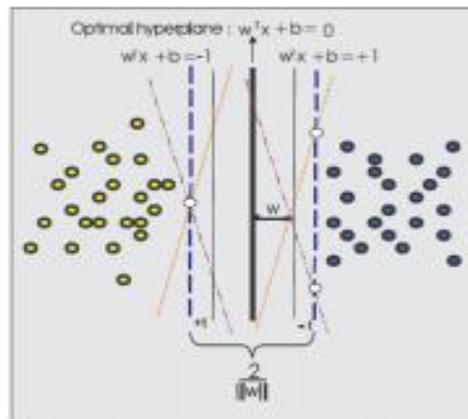
$P(H|X) P(X)$

Namun menurut bahwa naive bayes mempunyai kekurangan dimana attribut atau fitur independen sering salah dan hasil estimasi probabilitas tidak dapat berjalan optimal.

*Support Vector Machine* (SVM) adalah suatu metode learning machine yang bekerja dengan prinsip *Structural Risk Minimization* (SRM) dan mempunyai fungsi menemukan *hyperplane* terbaik dilakukan dengan cara memisahkan dua buah class pada input space [Bellotti and Crook, 2009]. *Hyperplane* yang berada ditengah antara dua set obyek dari dua class adalah *Hyperplane* terbaik. Ada dua cara untuk mencari titik maksimalnya *Hyperplane* pemisah terbaik antara kedua class yaitu dengan mengukur margin *hyperplane* dan mencari titik tersebut. *Pattern* yang paling dekat ini disebut sebagai *support vector* [Aydin et al., 2011] Karakteristik dari metode *Support Vector Machine* (SVM) adalah: 1). *Linear classifier* merupakan Prinsip SVM. 2). *Pattern recognition* dilakukan dengan memodifikasi data pada input space ke ruang yang berdimensi lebih tinggi, dan optimisasi dilakukan pada ruang vector yang baru tersebut. Dimana hal tersebut dapat membedakan SVM dari solusi *pattern*

*recognition* pada umumnya, yang melakukan optimisasi parameter pada ruang hasil perubahan yang berdimensi lebih rendah dibanding dimensi *input space*. 3). *Structural Risk Minimization* (SRM) merupakan strategi yang digunakan. 4). Prinsip kerja SVM pada dasarnya hanya mampu menangani klasifikasi dua class.

Secara sederhana konsep SVM adalah sebagai usaha mencari *hyperlane* terbaik yang berfungsi sebagai pemisah dua buah class pada input space, dimana dapat dilihat pada gambar 2:



Sumber: [Santosa, 2007]

Gambar 2. Konsep SVM untuk mencari *hyperlane* terbaik

Pada gambar 2 memperlihatkan beberapa pattern yang merupakan anggota dari dua buah class: +1 dan -1. Pattern yang tergabung pada class -1 disimbolkan dengan warna kuning. Sedangkan pattern pada class +1, disimbolkan dengan warna biru. Problem klasifikasi dapat diterjemahkan dengan usaha menemukan garis (*hyperplane*) yang memisahkan antara kedua kelompok tersebut. Berbagai alternatif garis pemisah (*discrimination boundaries*) ditunjukkan garis berwarna orange. *Hyperplane* pemisah terbaik antara kedua class dapat ditemukan dengan mengukur margin *hyperplane* tersebut dan mencari titik maksimalnya.

Margin adalah jarak antara *hyperplane* tersebut dengan pattern terdekat dari masing-masing class. Pattern yang paling dekat ini disebut sebagai support vector. *Hyperplane* yang terbaik yaitu yang terletak tepat pada tengah-tengah kedua class, sedangkan titik putih yang berada dalam garis bidang pembatas class adalah support vector. Usaha untuk mencari lokasi *hyperplane* ini merupakan inti dari proses pembelajaran pada SVM.

Algoritma PSO pertama kali diusulkan oleh Kennedy dan Eberhart pada tahun 1995. *Particle swarm optimization* (PSO) adalah jenis algoritma kecerdasan yang berasal dari perilaku kawanan burung mencari makan [Yun et al.] *Particle swarm optimization* (PSO) adalah teknik komputasi yang mampu menghasilkan solusi secara global optimal dalam ruang pencarian melalui interaksi perseorangan dalam kawanan partikel. Setiap partikel mengungkapkan informasi berupa posisi terbaiknya kepada partikel yang lain dan menyesuaikan posisi serta kecepatan masing-masing berdasarkan informasi yang diterima terkait posisi yang paling baik tersebut [Wang and Meng, 2011].

Proses data mining membutuhkan biaya komputasi yang tinggi ketika berhadapan dengan kumpulan data dalam jumlah besar. Mengurangi dimensi yaitu jumlah atribut set data atau kelompok atribut, secara efektif dapat memotong biaya tersebut. Pengurangan dimensi tersebut dilakukan dengan menekan seminimal mungkin kerugian yang dapat terjadi akibat kehilangan sebagian informasi. Tujuan pengurangan dimensi dalam domain data mining adalah untuk mengidentifikasi biaya terkecil di mana algoritma data mining dapat menjaga tingkat kesalahan di bawah perbatasan garis efisiensi. Yang dimaksud dengan biaya adalah fungsi dari kompleksitas teoritis dari algoritma data mining yang berasal dari model, dan berkorelasi dengan waktu yang dibutuhkan algoritma tersebut dalam menjalankan model, serta ukuran dari kumpulan data [Dahan et al., 2014].

Seleksi atribut adalah masalah terkait erat dengan pengurangan dimensi. Tujuan seleksi atribut adalah untuk mengidentifikasi tingkat kepentingan atribut dalam kumpulan data,

dan membuang semua atribut lain seperti informasi yang tidak relevan dan berlebihan. Karena seleksi atribut mengurangi dimensi dari data, maka hal ini akan memungkinkan operasi algoritma data mining dapat berjalan lebih efektif dan lebih cepat. Dalam beberapa kasus dengan dilakukannya seleksi atribut dihasilkan peningkatan tingkat akurasi klasifikasi [Dahan et al., 2014].

Di sisi lain, seleksi atribut adalah proses yang mahal, dan juga bertentangan dengan asumsi awal yaitu bahwa semua informasi atau atribut diperlukan dalam rangka mencapai akurasi maksimal. Ada empat alasan utama untuk melakukan pengurangan dimensi yaitu [Dahan et al., 2014]: 1). Penurunan biaya model pembelajaran, 2). Meningkatkan kinerja model pembelajaran, 3). Mengurangi dimensi yang tidak relevan 4). Mengurangi dimensi yang berlebihan.

Tujuan seleksi atribut adalah untuk pengurangan atribut dari dataset untuk menghilangkan variabel yang dianggap tidak relevan. Metode seleksi fitur dapat diklasifikasikan ke dalam tiga kategori utama [Vercellis, 2011]: 1). 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 menilai sekelompok variabel dengan menggunakan klasifikasi yang sama atau algoritma regresi digunakan untuk memprediksi nilai dari variabel target. 3). Untuk metode embedded, proses seleksi atribut terletak di dalam algoritma pembelajaran, sehingga pemilihan set optimal atribut secara langsung dibuat selama fase generasi model.

*Particle swarm optimization* (PSO) dapat diandaikan dengan segerombolan burung yang secara random mencari makanan di suatu daerah. Hanya ada satu potong makanan di daerah yang dicari tersebut. Burung-burung tidak tahu di mana makanan tersebut. Tapi mereka tahu seberapa jauh makanan tersebut dan posisi rekan-rekan mereka. Jadi strategi terbaik untuk menemukan makanan adalah dengan mengikuti burung yang terdekat dari makanan [Abraham et al., 2006].

Untuk menemukan solusi yang optimal, maka setiap partikel akan bergerak ke arah posisi yang terbaik sebelumnya ( $pbest$ ) dan posisi terbaik secara global ( $gbest$ ). Sebagai contoh, partikel ke-  $i$  dinyatakan sebagai:  $x_i = (x_{i,1}, x_{i,2}, \dots, x_{i,d})$  dalam ruang  $d$ -dimensi. Posisi terbaik sebelumnya dari partikel ke-  $i$  disimpan dan dinyatakan sebagai  $pbest_i = (pbest_{i,1}, pbest_{i,2}, \dots, pbest_{i,d})$ . Modifikasi kecepatan dan posisi tiap partikel dapat dihitung menggunakan kecepatan saat ini dan jarak  $pbest_{i,d}$  ke  $pbest_d$  seperti ditunjukkan oleh persamaan berikut:

$$V_{i,m} = w \cdot V_{i,m} + c_1 \cdot R \cdot (pbest_{i,m} - x_{i,m}) + c_2 \cdot R \cdot (gbest_m - x_{i,m})$$

$$x_{i,d} = x_{i,m} + V_{i,m}$$

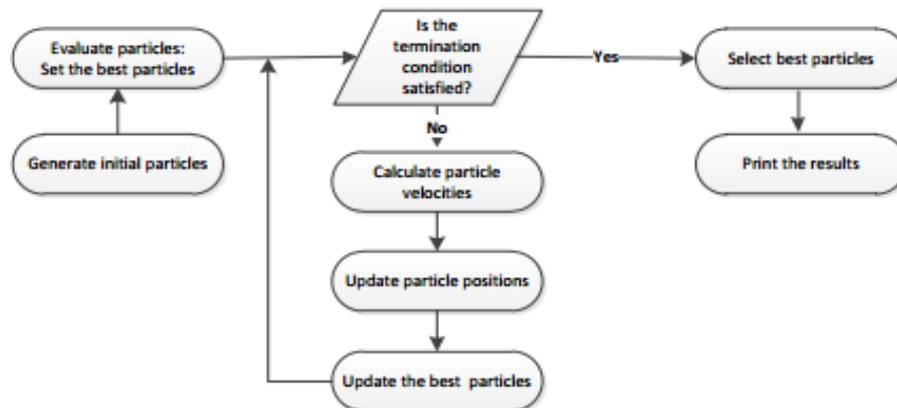
Dimana:

- $n$  : jumlah partikel dalam kelompok
- $d$  : dimensi
- $V_{i,m}$  : kecepatan partikel ke- $i$  pada iterasi ke-  $i$
- $w$  : faktor bobot inersia
- $c_1, c_2$  : konstanta akselerasi (learning rate)
- $R$  : bilangan random (0-1)
- $x_{i,d}$  : posisi saat ini dari partikel ke-  $i$  pada iterasi ke-  $i$
- $pbest$  : posisi terbaik sebelumnya dari partikel ke-  $i$
- $gbest$  : partikel terbaik diantara semua partikel dalam satu kelompok atau populasi

Persamaan ( $V_{i,m}$ ) menghitung kecepatan baru untuk tiap partikel (solusi potensial) berdasarkan pada kecepatan sebelumnya ( $V_{i,m}$ ), lokasi partikel dimana nilai fitness terbaik telah dicapai ( $pbest$ ), dan lokasi populasi global ( $gbest$  untuk versi global,  $lbest$  untuk versi local) atau *local neighborhood* pada algoritma versi local dimana nilai fitness terbaik telah dicapai.

Persamaan ( $x_{i,d}$ ) memperbaharui posisi tiap partikel pada ruang solusi. Dua bilangan acak  $c_1$  dan  $c_2$  dibangkitkan sendiri. Penggunaan berat inersia  $w$  telah memberikan performa yang meningkat pada sejumlah aplikasi.

Secara garis besar, struktur dasar dari *Particle Swarm Optimization* dapat digambarkan dalam Gambar 3:



Sumber: [Abraham et al, 2006]

Gambar 3. Struktur dasar *Particle Swarm Optimization*

## 2. Metode Penelitian

Penelitian yang telah dilakukan adalah eksperimen dalam bentuk sistem penunjang keputusan dalam memprediksi tingkat kesuburan dengan mengkomparasi antara Naive Bayes dan Support Vector Machine dengan optimasi Particle Swarm Optimization. Data yang digunakan dalam penelitian ini diambil dari dataset sekunder fertilitas sperma yang diambil dari UCI Machine Learning Repositori, dimana dataset terdiri dari 100 instance dan 10 atribut dengan 1 atribut label, dimana atribut terdiri dari analisa cuaca, usia, penyakit pada kanak-kanak, trauma, bedah, konsumsi alkohol, merokok, jumlah jam duduk dan prediktor. Penelitian dilakukan dengan menggunakan tools aplikasi rapidminer 5. Langkah-langkah dalam penelitian ini diantaranya: 1). Pengumpulan Data pada tahap ini dikumpulkannya dataset yang akan diproses dan diolah. 2). Pengolahan data awal pada tahap ini dilakukan penyeleksian data, membersihkan data yang redundansi dan ditransformasikan sehingga dapat dilakukan dalam pembuatan model. 3). Metode yang diusulkan tahap ini data dianalisis, mengelompokan variabel yang akan digunakan. 4). Pengujian Metode tahap ini mengkomparasi antara algoritma Naive Bayes dan Support Vector Machine dengan optimasi Particle Swarm Optimization. 5). Evaluasi dan validasi. Tahap ini dilakukan penilaian terhadap dataset untuk mengetahui tingkat akurasi model.

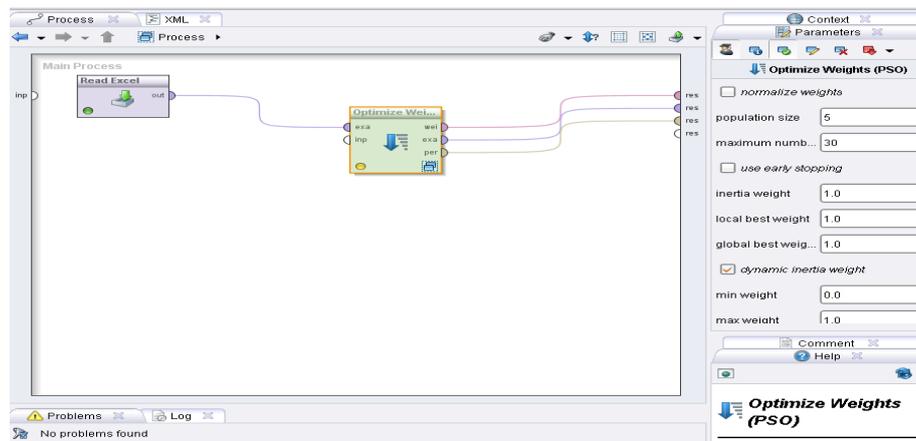
## 3. Hasil dan Pembahasan

Tujuan dari penelitian ini adalah mengkomparasi antara dua algoritma antara naive bayes dan support vector machine dan menerapkan *Particle swarm optimization* (PSO) dengan melakukan seleksi atribut untuk meningkatkan keakuratan yang berbeda.

### a. Metode Support Vector Machine berbasis Particle Swarm Optimization

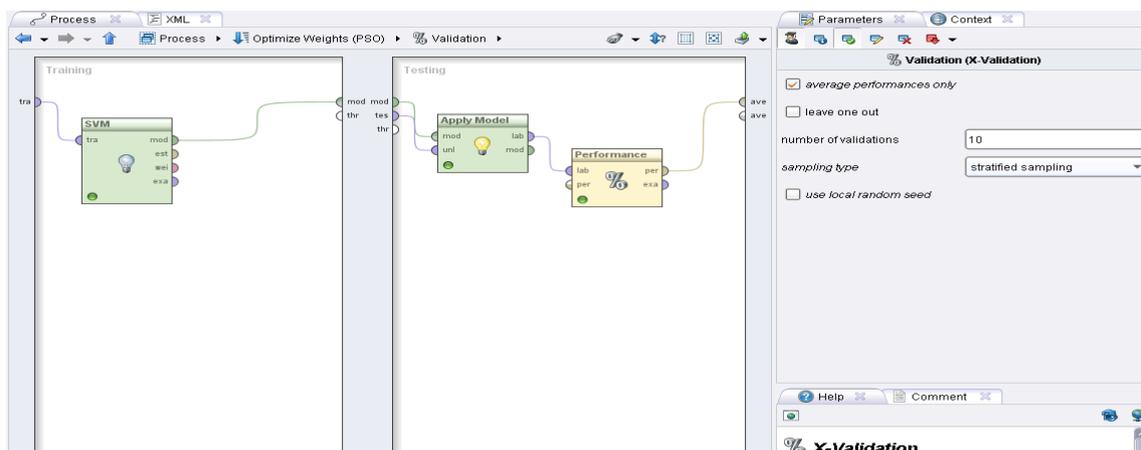
Gambar 4 menunjukkan pengujian algoritma Support Vector Machine berbasis Particle swarm optimization menggunakan metode K-Fold Cross Validation dengan menggunakan RapidMiner.

Nilai *training cycles* dalam penelitian ini ditentukan dengan cara melakukan uji coba memasukkan C dan epsilon. Selanjutnya dilakukan observasi terhadap variabel C dan  $\epsilon$  dari *Support Vector Machine* dan *Particle swarm optimization*. Gambar 5 adalah percobaan yang telah dilakukan.



Sumber: Hasil Pengolahan Data (2017)

Gambar 4. Pengujian Algoritma Support Vector Machine dengan optimasi Particle Swarm Optimization menggunakan Tools Rapid Miner

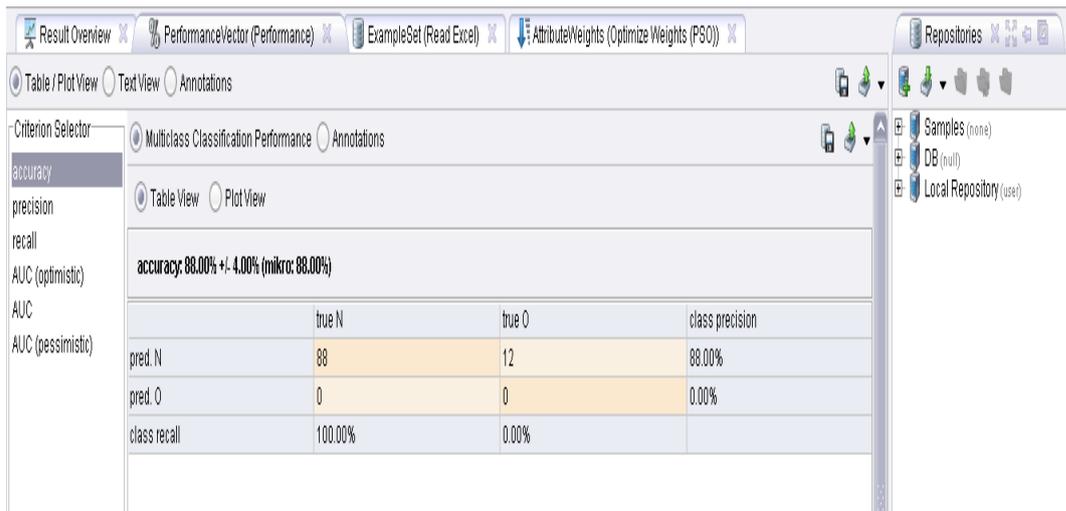


Sumber: Hasil Pengolahan Data (2017)

Gambar 5. Pengujian Algoritma Support Vector Machine dengan optimasi Particle Swarm Optimization menggunakan Tools Rapid Miner

Hasil pengujian model adalah untuk mengukur tingkat akurasi dan AUC (*Area Under Curve*) dari penentuan prediksi kesuburan dengan metode *cross validation*. Model klasifikasi bisa dievaluasi berdasarkan kriteria seperti tingkat akurasi, kecepatan, kehandalan, skabilitas dan interpretabilitas (Vecellis, 2009). Setelah data diolah maka dapat diuji tingkat akurasinya untuk melihat kinerja dari masing-masing model.

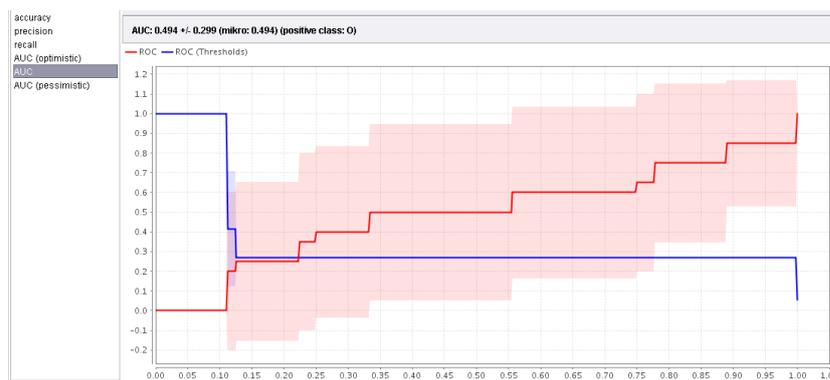
Gambar 6 merupakan hasil dari *Confusion Matrix* dengan Metode *Support Vector Machine* dan *Particle Swarm Optimization* untuk seleksi atributnya. Dimana hasil percobaan tersebut memiliki *accuracy* 88.00%. Jumlah True Positive (TP) adalah 88 diklasifikasikan sebagai 1 sesuai dengan prediksi yang dilakukan dengan metode *Support Vector Machine*, lalu False Negative (FN) sebanyak 12 data diprediksi sebagai 1 tetapi ternyata 2, kemudian True Negative (TN) sebanyak 0 data sebagai 2 sesuai dengan prediksi, dan False Positive (FP) sebanyak 0 data diprediksi 2 ternyata 1.



Sumber: Hasil pengolahan data(2017)

Gambar 6. Pengujian Algoritma *Support Vector Machine* dengan optimasi *Particle Swarm Optimization* menggunakan Tools Rapid Miner

Evaluasi dengan kurva ROC, berikut ini adalah kurva AUC dengan menggunakan algoritma *Support Vector Machine* dengan optimasi *Particle Swarm Optimization*.

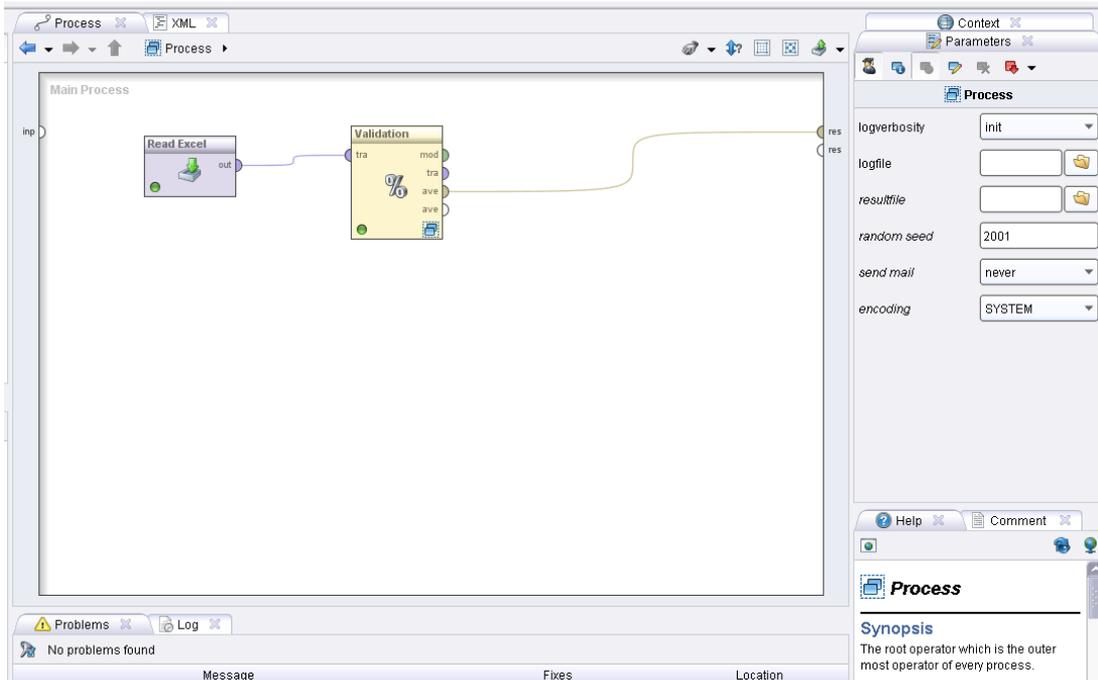


Gambar 7. Kurva AUC

Pada gambar 7 menunjukkan grafik ROC dengan nilai AUC (*Area Under Curve*) sebesar 0.49 dengan tingkat diagnosa *fair Classification* untuk model *Support Vector Machine* dengan optimasi *Particle Swarm Optimization*.

b. Metode Naive Bayes

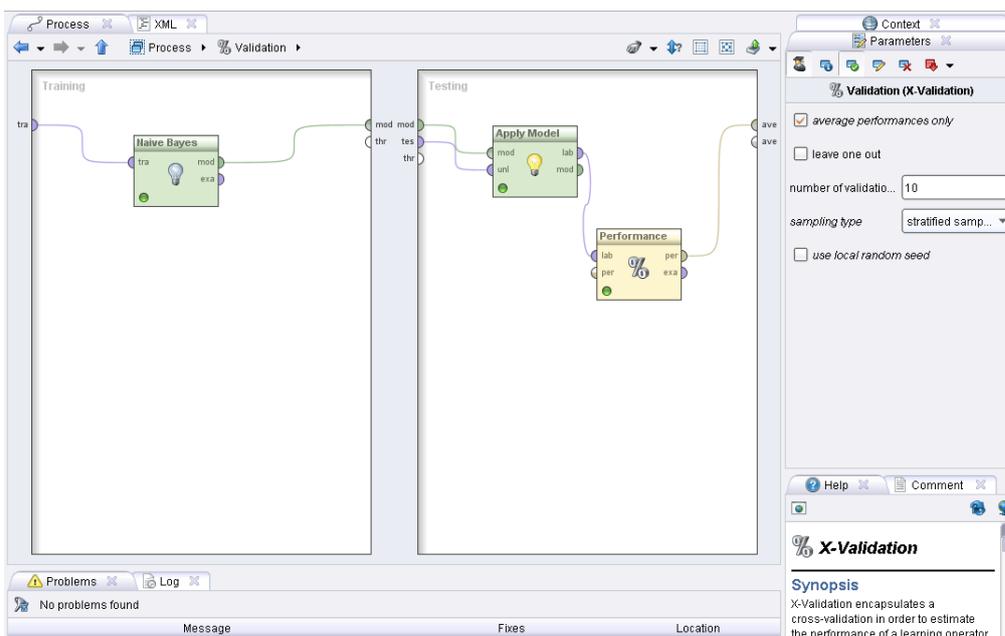
Berikut adalah gambar pengujian algoritma Naive Bayes menggunakan metode *K-Fold Cross Validation* dimana data excel akan diimport ke dalam Rapid Miner setelah itu di validasi. Database Diagnosa Tingkat Kesuburan dihubungkan dengan operator *optimize selection (evolutionary)* untuk pemilihan atribut-atribut yang sesuai dengan proses prediksi hasil tingkat kesuburan. Di dalam *optimize selection (evolutionary)* terdapat proses *cross validation* yang terlihat pada gambar 8.



Sumber: Hasil pengolahan data(2017)

Gambar 8. Pengujian Operator *Cross Validation* menggunakan Tools Rapid Miner

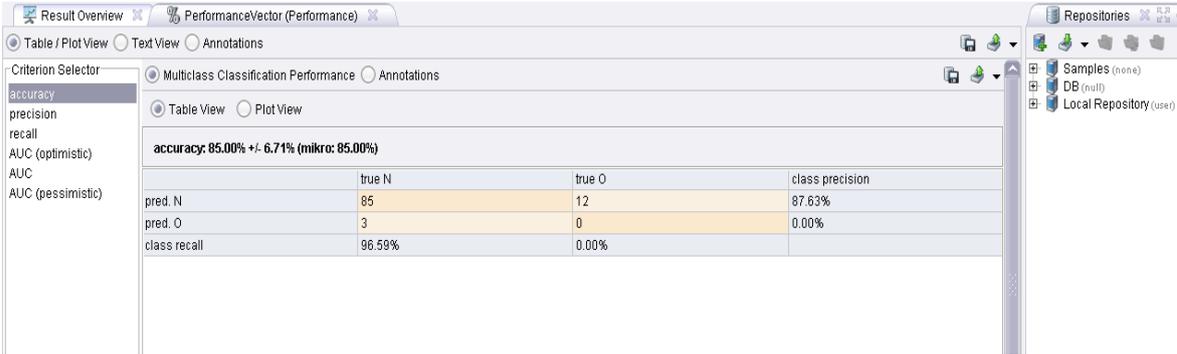
*Cross Validation* yang digunakan dalam penelitian ini adalah *10-fold validation*. Dataset yang berjumlah 160 data dengan 9 atribut akan dibagi menjadi 10 bagian. Dimana setiap bagian akan dibentuk secara acak. Prinsip dari 10 fold validation adalah 1:9 dimana 1 bagian menjadi data testing dan 9 bagian menjadi data trainin, sehingga 10 bagian tersebut dapat berkesempatan menjadi data testing. Setelah dilakukan training dan testing maka dapat diukur tingkat akurasi [Buani, 2016]. Berikut proses penerapan algoritma naive bayes Gambar 9.



Sumber: Hasil pengolahan data(2017)

Gambar 9. Pengujian Algoritma Naive Bayes menggunakan Tools Rapid Miner

Setelah dilakukan penerapan model algoritma naive bayes maka dihasilkan akurasi sesuai gambar 10. Gambar 10 merupakan hasil dari *Confusion Matrix* dengan Metode *Naive Bayes* untuk seleksi atributnya. Dimana hasil percobaan tersebut memiliki *accuracy* 85.00%. Jumlah True Positive (TP) adalah 85 diklasifikasikan sebagai 1 sesuai dengan prediksi yang dilakukan dengan metode *Support Vector Machine*, lalu False Negative (FN) sebanyak 12 data diprediksi sebagai 1 tetapi ternyata 2, kemudian True Negative (TN) sebanyak 3 data sebagai 2 sesuai dengan prediksi, dan False Positive (FP) sebanyak 0 data diprediksi 2 ternyata 1.

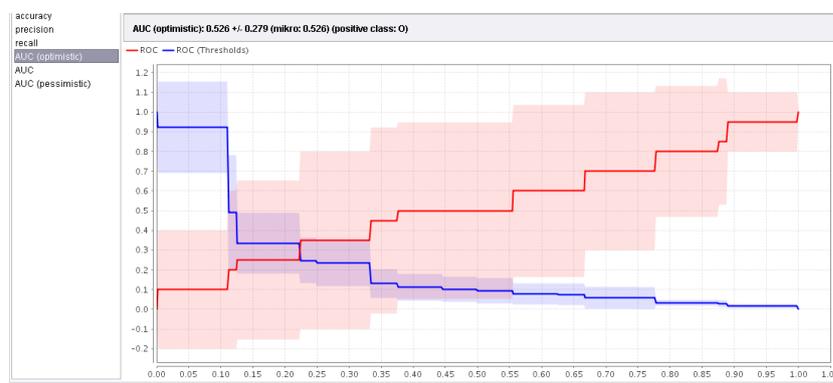


accuracy: 85.00% +/- 6.71% (mikro: 85.00%)			
	true N	true O	class precision
pred. N	85	12	87.63%
pred. O	3	0	0.00%
class recall	96.59%	0.00%	

Sumber: Hasil pengolahan data(2017)

Gambar 10. Hasil dari Pengujian Algoritma Naive Bayes

Evaluasi dengan kurva ROC, berikut ini adalah kurva AUC dengan menggunakan algoritma Naive Bayes.



Gambar 11. Kurva AUC

Pada gambar 11 menunjukkan grafik ROC dengan nilai AUC (*Area Under Curve*) sebesar 0.52 dengan tingkat diagnosa *fair Classification* untuk model Naive Bayes.

#### 4. Kesimpulan

Berdasarkan hasil eksperimen dapat dinyatakan bahwa akurasi dari algoritma support vector machine dengan seleksi atribut diterapkan particle swarm optimization(PSO). Dimana particle swarm optimization diketahui dapat dipakai untuk teknik optimasi dalam memaksimalkan subset fitur. Algoritma PSO mempunyai kerumitan yang lebih rendah, sehingga dapat menghasilkan nilai yang optimal dalam proses pencarian global dan lokal. Tingkat akurasi support vector machine dengan optimasi particle swarm optimization sebesar 88.00% sedangkan dengan menggunakan algoritma naive bayes tingkat akurasi yang dihasilkan sebesar 85.00%.

**Referensi**

- Abraham A, Grosan G, Ramos V. 2006. *Swarm Intelligence in Data Mining*. Berlin.
- Aydin I, Karakose M, Akin E. 2011. A Multi-Objective Artificial Immune Algorithm For Parameter Optimization In Support Vector Machine. *Appl. Soft Comput.* 11: 120–129.
- Bellotti T, Crook J. 2009. Support Vector Machines For Credit Scoring And Discovery Of Significant Features. *Expert Syst. Appl.* 36: 3302–3308.
- Buani DCP. 2016. Optimasi Algoritma Naïve Bayes dengan Menggunakan Algoritma Genetika untuk Prediksi Kesuburan (Fertility). 4.
- Dahan H, Cohen S, Rokach L, Maimon O. 2014. *Proactive Data Mining With Decision Trees*.
- Domingos P, Pazzani M. 1997. On the Optimality of the Simple Bayesian Classifier under Zero-One Loss. *Mach. Learn.* 29: 103–130.
- Gil D, Girela JL, De Juan J, Gomez-Torres MJ, Johnsson M. 2012. Predicting Seminal Quality With Artificial Intelligence Methods. *Expert Syst. Appl.* 39: 12564–12573.
- Giwerzman A, Giwerzman YL. 2011. Environmental Factors And Testicular Function. *Best Pract. Res. Clin. Endocrinol. Metab.* 25: 391–402.
- Hand DJ. 2007. Principles of Data Mining. *Drug Saf.* 30: 621–622.
- Irvine DS. 2000. Male Reproductive Health: Cause For Concern? *Andrologia* 32: 195–208.
- Khasanah FN. 2016. Klasifikasi Proses Penjurusan Siswa Tingkat SMA Menggunakan Data Mining. *Inf. Educ. Prof. [S.I.]*, v. 1, n. 1, p. 65 - 69, dec. 2016. ISSN 2548-3412.
- Kusrini, Taufiq LE. 2009. *Algoritma Data Mining 1st*. Andi.Yogyakarta.
- Santosa B. 2007. *Teknik Pemanfaatan Data Untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu.
- Vercellis C. 2011. *Business Intelligence: Data Mining And Optimization For Decision Making*.
- Wang S, Meng B. 2011. Parameter Selection Algorithm for Support Vector Machine. *Procedia Environ. Sci.* 11: 538–544.
- Widodo PP, Handayanto RT, Herlawati. 2013. *Penerapan Data Mining dengan Matlab*. Bandung: Rekayasa Sains.
- Yun L, Cao Q, and HZ-CI. 2011. Undefined. Application of the PSO-SVM model for Credit Scoring. [ieeexplore.ieee.org](http://ieeexplore.ieee.org).