

# PENINGKATAN AKURASI KELAYAKAN KREDIT MENGGUNAKAN PARTICLE SWARM OPTIMIZATION

Suamanda Ika Novichasari

Program Studi SI Teknik Informatika, Universitas Ngudi Waluyo

vichareal0311@gmail.com

**Abstrak**— Penentuan kelayakan kredit adalah sebuah proses untuk menentukan apakah seorang nasabah termasuk kredit baik atau kredit buruk. Dengan demikian teknik data mining yang tepat digunakan adalah klasifikasi. Naive Bayes Classifier (NBC), Decision Trees (DT), dan Support Vector Machines (SVM) adalah algoritma yang digunakan untuk klasifikasi, namun akurasi kurang maksimal. Particle Swarm Optimization (PSO) yang digunakan untuk pembobotan atribut dapat meningkatkan kinerja dari algoritma klasifikasi. Penelitian ini membandingkan algoritma NBC-PSO, DT-PSO dan SVM-PSO. Dataset yang digunakan adalah German Credit Data. Proses validasi menggunakan tenfold-cross validation, sedangkan pengujian modelnya menggunakan confusion matrix dan kurva ROC. Hasil eksperimen menunjukkan kinerja dari masing-masing algoritma meningkat ketika digabungkan dengan PSO, namun membutuhkan waktu eksekusi yang relatif lebih lama.

**Kata kunci**— Kelayakan kredit, klasifikasi, data mining, NBC, Decision trees, SVM, PSO

## I. PENDAHULUAN

Kredit dalam perbankan sangat mempengaruhi pembiayaan perekonomian nasional dan berfungsi sebagai penggerak pertumbuhan ekonomi [1]. Proses penilaian kredit bagi bank tidak dapat diremehkan, karena bagi bank sangat penting untuk membedakan antara UKM baik dan buruk dalam hal kredit melalui proses penilaian kredit tersebut [2]. Dalam kondisi keuangan yang sangat lemah dengan jumlah kredit bermasalah yang semakin besar, serta likuiditas yang semakin rendah dapat menyebabkan bank kesulitan untuk membiayai kegiatan usahanya [3]. Seperti yang tercantum dalam Gatra News, Lembaga Penjamin Simpanan (LPS) menyatakan bahwa sejak delapan tahun terakhir ini, dari 2006 hingga 2013, sebanyak 55 bank dilikuidasi, yang terdiri dari Bank Perkreditan Rakyat (BPR) dan 1 bank umum, serta menyelamatkan Bank Century yang kini bersulih nama menjadi Bank Mutiara [4]. Oleh karena itu analisa kelayakan kredit sangat penting dilakukan untuk mencegah jumlah kredit bermasalah yang semakin besar guna menjaga kelangsungan hidup sebuah bank.

Untuk menyelesaikan masalah analisa resiko kredit data akan diklasifikasikan menjadi dua kelas, yaitu kredit baik dan kredit buruk. Sehingga tepat menggunakan

teknik klasifikasi data mining. Metode yang paling populer digunakan untuk teknik klasifikasi adalah Decision Trees (DS), Naive Bayes Classifiers (NBC), Support Vector Machines (SVM), dan lain lain [5]. Hasil akurasi dari teknik klasifikasi tersebut kurang maksimal, maka dibutuhkan teknik optimasi seperti PSO untuk meningkatkan akurasi. PSO adalah algoritma pencarian berbasis populasi partikel yang digunakan untuk memecahkan masalah optimasi. Setiap partikel memiliki kecenderungan untuk terbang menuju daerah pencarian yang lebih baik dan lebih baik selama proses pencarian [6], [7]. Beberapa peneliti telah menggunakan algoritma PSO untuk meningkatkan akurasi dari suatu algoritma klasifikasi data mining [8],[9],[10],[11],[12],[13].

Elon S. Correa, Alex A. Freitas, dan Colin G. Johnson (2008), mengukur kinerja Diskrit Particle Swarm Optimization (DPSO) dan PSO dalam seleksi atribut yang diterapkan pada Bayesian network dan Naive Bayes. Hasil dari penelitian ini adalah dalam hal seleksi atribut pada satu data set DPSO lebih baik dari PSO dan pada 6 data set DPSO terbukti lebih efektif dan meningkatkan efisiensi komputasi dari classifier. Hasil lainnya adalah bahwa pendekatan Bayesian Network melebihi pendekatan dari Naive Bayes pada semua percobaan [8].

S.Devi, A.K. Jagadev, S.Dehuri, R.Mall (2009) menggunakan kombinasi antara Bayesian classifier dan PSO untuk menghilangkan atribut yang palsu dan yang tidak relevan dari dataset bio-medis. Penelitian ini membuktikan bahwa fungsi diskriminan non-linier yang dikombinasikan dengan seleksi fitur PSO terbukti menjadi seleksi atribut terbaik tanpa menurunkan tingkat akurasi [9].

Henny Leidiyana pada tahun (2011) membandingkan algoritma C4.5, Naive Bayes, dan Neural Network yang digunakan untuk penentuan kelayakan kredit motor. Hasilnya dengan metode pengujian Cross Validation, Confusion Matrix dan Kurva ROC terbukti C4.5 memiliki akurasi tertinggi [10].

Siti Masriyah pada tahun 2011 membandingkan dua metode yaitu metode C4.5 dan metode C4.5 berbasis PSO. Hasilnya, terbukti metode C4.5 berbasis PSO lebih akurat dengan akurasi 94% dan Area Under Curve (AUC) sebesar 0,955 [11].

Friska Handayanna pada tahun 2012 menerapkan metode PSO untuk seleksi atribut pada metode Support Vector Machines (SVM) karena Atribut dalam Metode SVM memiliki pengaruh penting dalam akurasi klasifikasi. Hasil penelitian menunjukkan bahwa PSO yang diterapkan untuk seleksi atribut pada metode SVM untuk prediksi penyakit dengan data set Pima Diabetes, terbukti dapat meningkatkan nilai akurasi dan AUC dibanding prediksi hanya dengan menggunakan metode SVM saja [12].

Suamanda Ika N pada tahun 2013 melakukan penelitian dengan menerapkan PSO pada pembobotan atribut pada Naive Bayes Classifier (NBC) untuk meningkatkan akurasi NBC. Penelitian ini menggunakan dataset publik German credit data yang di ambil dari University of California, Irvine (UCI) repository. Hasil penelitian menunjukkan bahwa akurasi dari model NBC berbasis PSO (NBC-PSO) sebesar 78,00% dan AUC sebesar 0,778. Nilai tersebut lebih tinggi dibanding dengan model NBC tanpa PSO yang hanya 73,70% dan 0,774. Dengan demikian terbukti bahwa penerapan PSO pada pembobotan atribut NBC dapat meningkatkan akurasi dari NBC [13].

## II. DATA SET

Data yang digunakan pada penelitian ini berasal dari University of California, Irvine (UCI) Machine Learning dengan judul German Credit data. Data yang disediakan oleh Professor Dr. Hans Hofmann ini berjumlah 1000 record dan terdiri dari 20 atribut bebas, dengan 7 atribut bertipe numerik dan 13 bertipe kategorikal dan 1 atribut terikat bertipe biner (0-1) seperti berikut ini :

- Salary assignments for at least 1 year, merupakan atribut bebas bertipe kategorikal
- Duration in month, merupakan atribut bebas bertipe numerik
- Credit history, merupakan atribut bebas bertipe kategorikal
- Purpose, merupakan atribut bebas bertipe kategorikal
- Credit amount, merupakan atribut bebas bertipe numerik
- Savings account, merupakan atribut bebas bertipe kategorikal
- Present employment since, merupakan atribut bebas bertipe kategorikal
- Installment rate in percentage of disposable income, merupakan atribut bebas bertipe numerik
- Personal status and sex, merupakan atribut bebas bertipe kategorikal
- Other debtors / guarantors, merupakan atribut bebas bertipe kategorikal
- Present residence since, merupakan atribut bebas bertipe numerik
- Property, merupakan atribut bebas bertipe kategorikal

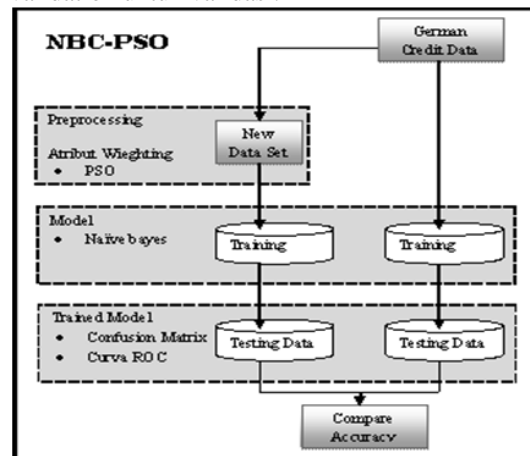
- Age in years, merupakan atribut bebas bertipe numerik
- Other installment plans, merupakan atribut bebas bertipe kategorikal
- Housing, merupakan atribut bebas bertipe kategorikal
- Number of existing credits at this bank, merupakan atribut bebas bertipe numerik
- Job, merupakan atribut bebas bertipe kategorikal
- Number of people being liable to provide maintenance for, merupakan atribut bebas bertipe numerik
- Telephone, merupakan atribut bebas bertipe kategorikal
- Foreign worker, merupakan atribut bebas bertipe kategorikal
- Credit Status, merupakan atribut terikat bertipe biner (0-1)

Data ini juga pernah digunakan sebagai dataset penelitian oleh beberapa peneliti diantaranya adalah Jeroen Eggermont, Joost N. Kok dan Walter A. Kusters pada tahun 2004 dan Ke Wang, Shiyu Zhou, Ada Wai-Chee Fu dan Jeffrey Xu Yu pada tahun 2003 [14].

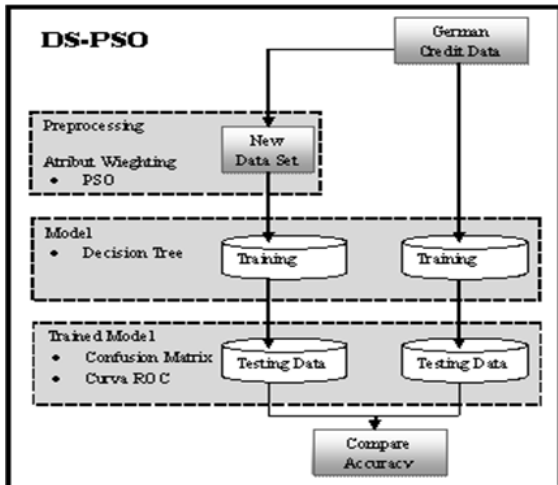
Keterbatasan hardware yang digunakan memaksa peneliti hanya menggunakan 20% data dari German Credit Data yaitu 200 record. Dataset yang diambil adalah 100 record pertama dan 100 record terakhir.

## III. METODE

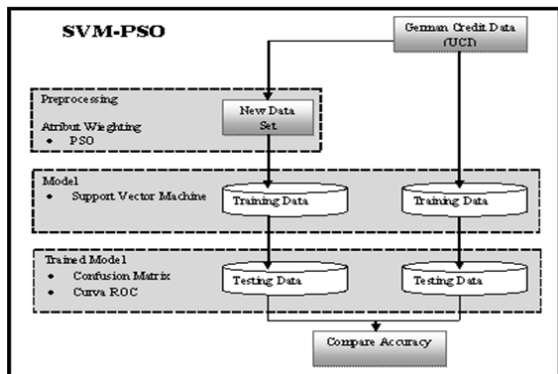
Secara keseluruhan penelitian ini menggunakan 6 model algoritma yaitu NBC, DT, SVM, NBC berbasis PSO (NBC-PSO), DT berbasis PSO (DT-PSO) dan SVM berbasis PSO (SVM-PSO). PSO diterapkan pada saat preprossesing untuk pembobotan atribut. Untuk membandingkan atau mengkomparasi dalam penelitian ini akan menggunakan framework RapidMiner versi 5.3.015 sehingga akan ditemukan algoritma mana yang paling akurat. Pengukuran keakuratan hasil yang dicapai oleh model menggunakan confusion matrix dan kurva ROC, cross-validation untuk validasi.



Gambar 1. Model NBC yang diusulkan



Gambar 2 Model DS yang diusulkan



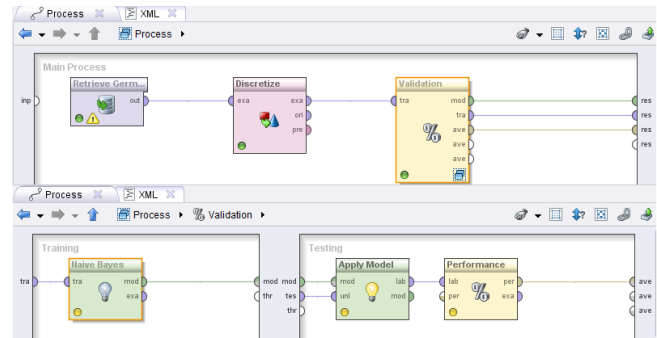
Gambar 3 Model SVM yang diusulkan

#### IV. HASIL DAN PEMBAHASAN

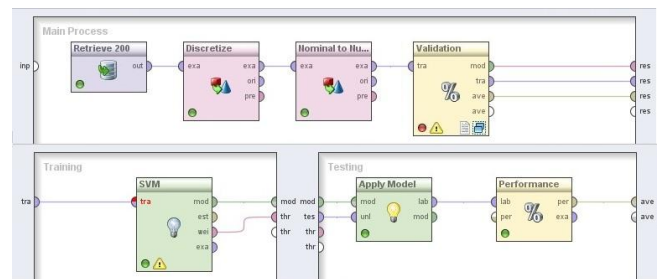
Tujuan utama penelitian ini adalah untuk mengetahui nilai akurasi dari algoritma NBC-PSO, DT-PSO dan SVM-PSO menggunakan German credit dataset. Kemudian membandingkan ketiga algoritma tersebut sehingga dapat diperoleh salah satu algoritma yang terbaik.

Desain model NBC dapat dilihat pada gambar 4. Retrieve berfungsi untuk memasukkan dataset ke dalam RapidMiner. Discretize untuk mengubah isi dari atribut numerik yang kontinu menjadi nominal. Discretize menggunakan *Discretize by frequency* yaitu transformasi atribut numerik kontinu menjadi atribut nominal oleh operator frekuensi yang menciptakan kelompok rentang terbaik sehingga jumlah nilai yang unik dalam kelompok rentang hampir sama. *Validation* menggunakan *tenfold cross-validation*. Didalam *validation* terdapat dua kolom, *training* dan *testing*. Didalam kolom *training* terdapat algoritma klasifikasi yang diterapkan yaitu *Naive Bayes*, sedangkan di dalam kolom *testing* terdapat *Apply Model* untuk menjalankan algoritma/model *Naive Bayes* dan *Performance* untuk mengukur performa dari model *Naive Bayes* tersebut. Desain model DT sama seperti model NBC namun pada kolom *training* didalam *validation* berisi *Decision trees*. Desain model SVM berbeda dengan desain model NBC dan DT. Seperti yang terlihat pada

gambar 5, setelah *Discretize* terdapat *Nominal to numeric* untuk mengubah nilai dari atribut yang bertipe nominal menjadi numerik. Hal ini dilakukan karena algoritma SVM tidak dapat menangani atribut yang bernilai nominal.

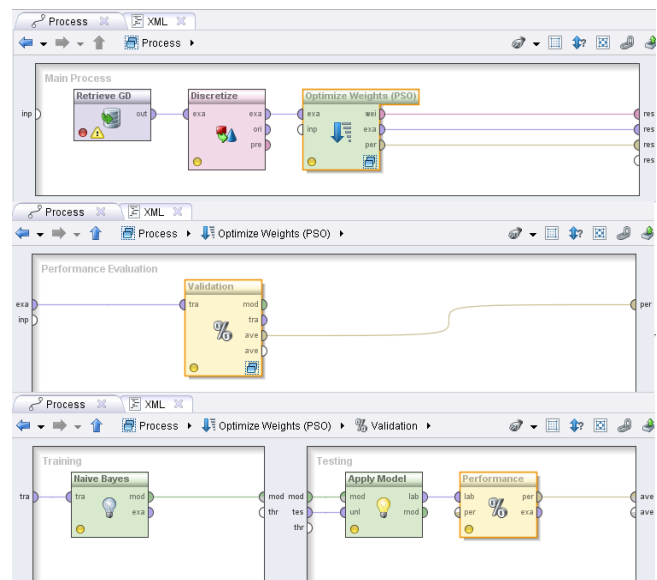


Gambar 4 Model NBC.



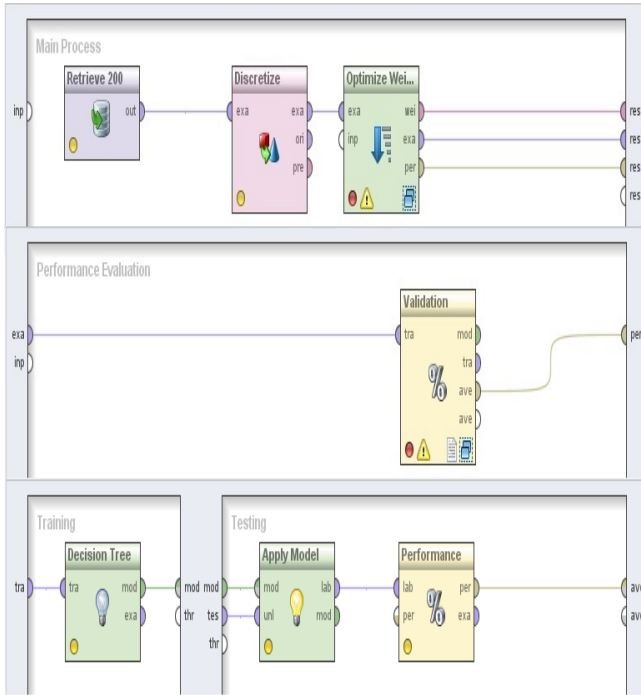
Gambar 5 Model SVM

Sedangkan desain model NBC berbasis PSO pada pembobotan atribut pada Rapidminer dapat dilihat pada gambar 6.



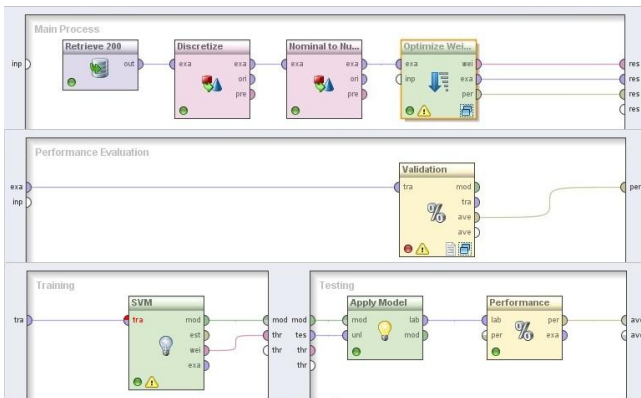
Gambar 6. Desain model NBC-PSO

*Optimize Weights (PSO)* untuk menerapkan algoritma PSO pada pembobotan atribut. Didalam *Optimize Weights (PSO)* terdapat *Validation* yang menggunakan *tenfold cross-validation* yang sama dengan model NBC.



Gambar 7 Desain model DT-PSO

Desain model DT-PSO sama dengan model NBC-PSO, hanya berbeda pada kolom *training* di dalam *validation* menggunakan *Decision Trees*. Sedangkan model SVM-PSO, pada main proses ditambah *Nominal to Numerical* dan kolom *training* digunakan SVM.



Gambar 8 Desain model SVM-PSO

Peneliti melakukan 3 tahap *eksperimen*. Tahap pertama melakukan eksperimen dengan model NBC, DT dan SVM. Tahap kedua dengan model NBC-PSO, DT-PSO, dan SVM-PSO dengan parameter default (*population size* bernilai 5 dan *number of generation* bernilai 30). Tahap terakhir menggunakan model NBC-PSO, DT-PSO, dan SVM-PSO dengan *population size* bernilai 350 dan *number of generation* bernilai 500. Nilai *population size* dan *number of generation* didapatkan dari penelitian sebelumnya yang menyebutkan bahwa PSO meningkatkan kinerja secara maksimal pada saat *population size* bernilai 350 dan *number of generation* bernilai 500 [13].

TABEL 1.  
HASIL EKSPERIMEN PERTAMA.

Model	Akurasi	Presisi	Recall	AUC	Waktu
NBC	75,00%	57,48%	48,00%	0,754	0 detik
DT	65,50%	43,71%	31,67%	0,504	0 detik
SVM	79,00%	71,43%	44,33%	0,732	0 detik

TABEL 2.  
HASIL EKSPERIMEN KEDUA.

Model	Akurasi	Presisi	Recall	AUC	Waktu
NBC-PSO	76,60%	64,42%	49,67%	0,780	8 detik
DT-PSO	74,00%	61,67%	33,00%	0,603	46 detik
SVM-PSO	82,00%	90,83%	43,00%	0,772	28 detik

TABEL 3.  
HASIL EKSPERIMEN KETIGA.

Model	Akurasi	Presisi	Recall	AUC	Waktu
NBC-PSO	82,00%	70,67%	56,67%	0,775	35 menit 52 detik
DT-PSO	81,50%	67,38%	63,67%	0,768	6 jam 21 menit 36 detik
SVM-PSO	84,00%	90,17%	51,00%	0,741	8 jam 28 menit 15 detik

V. KESIMPULAN

Penelitian ini menggunakan algoritma NBC, DT, SVM, NBC-PSO, DT-PSO dan SVM-PSO menggunakan data nasabah bank yang akan mengajukan kredit. Penelitian ini berfokus pada penerapan algoritma PSO dalam pembobotan atribut teknik klasifikasi data mining. Validasi model menggunakan 10fold cross-validation dan evaluasi model menggunakan confusion matrix dan kurva ROC.

TABEL 4.

PERBANDINGAN AKURASI DAN PRESISI

METODE	AKURASI			PRESISI		
	Tanpa PSO	PSO-default	PSO-modif	Tanpa PSO	PSO-default	PSO-modif
NBC	75,00%	76,60%	82,00%	57,48%	64,42%	70,67%
DT	65,50%	74,00%	81,50%	43,71%	61,67%	67,38%
SVM	79,00%	82,00%	84,00%	71,43%	90,83%	90,17%

TABEL 5.  
PERBANDINGAN RECALL DAN AUC

METODE	RECALL			AUC		
	Tanpa PSO	PSO-default	PSO-modif	Tanpa PSO	PSO-default	PSO-modif
NBC	48,00%	49,67%	56,67%	0,754	0,78	0,775
DT	31,67%	33,00%	63,67%	0,504	0,768	0,603
SVM	44,33%	43,00%	51,00%	0,732	0,772	0,741

TABEL 6.  
PERBANDINGAN WAKTU EKSEKUSI MODEL

METODE	Waktu eksekusi		
	Tanpa PSO	PSO-default	PSO-modif
NBC	0 detik	8 detik	35 menit 52 detik
DT	0 detik	46 detik	6 jam 21 menit 36 detik
SVM	0 detik	28 detik	8 jam 28 menit 15 detik

Hasil penelitian menunjukkan bahwa akurasi dari masing-masing teknik klasifikasi meningkat ketika digabungkan dengan PSO pada pembobotan atribut. Model SVM-PSO memiliki akurasi dan presisi paling baik diantara teknik lain. Recall dan AUC terbaik dimiliki oleh model NBC-PSO. Waktu eksekusi paling cepat dimiliki oleh NBC-PSO. Sedangkan model DT-PSO memiliki akurasi, presisi, recall dan AUC terendah. Nilai AUC dari ketiga model masuk dalam golongan *fair clasification*. Dengan demikian model NBC-PSO dan SVM-PSO merupakan algoritma yang berkinerja lebih baik dengan kelemahan dan keunggulannya masing-masing dibandingkan model DT-PSO.

Untuk penelitian selanjutnya sebaiknya menggunakan lebih banyak algoritma untuk dibandingkan seperti *Neural network*, K-nn, atau yang lainnya. Atau menggunakan banyak dataset yang berbeda dan menambahkan metode pengukuran lain untuk mempermudah dalam menentukan algoritma yang terbaik.

REFERENSI

[1] Utari D, G.A., Arimurti, T., dan Kurniati, I.K (2012). "Pertumbuhan Kredit Optimal," Buletin Ekonomi Moneter dan Perbankan. Oktober 2012. 3-36.

[2] Guidici, P., and Figini, S (2009). *Applied Data Mining for Business and Industry*. 2nd ed. United Kingdom: A John Wiley And Sons, Ltd., Publication.

[3] Triandaru, S., dan Budisantoso, T (2009). *Bank dan Lembaga Keuangan Lain*. Edisi kedua. Jakarta : Salemba Empat.

[4] <http://www.gatra.com/ekonomi-1/44068-lps-selama-8-tahun,-55-bank-dilikuidasi.html>, diakses tanggal 26 Maret 2014.

[5] Gorunescu, F. (2011). *Data Mining Concepts, Models And Techniques*. Verlag Berlin Heidelberg: Springer.

[6] J. Kennedy and R. C. Eberhart (1995). "Particle swarm optimization," In *Proceedings of the 1995 IEEE International Conference on Neural Networks*. IEEE Service Center, Piscataway.

[7] Abraham, A., Grosan, C., & Ramos, V. (2006). *Swarm Intelligence In Data Mining*. Verlag Berlin Heidelberg: Springer.

[8] Correa,E.S., Freitas, A.A., dan Johnson, C.G (2008). "Particle Swarm for Attribute Selection in Bayesian Classification: An Application to Protein Function Prediction," *Journal of Artificial Evolution and Applications*. Volume 2008. Computing Laboratory and Center for Biomedical Informatics, University of Kent.

[9] Devi, S., Jagadev, A.K., Dehuri, S., and R.Mall (2009). "Knowledge Discovery from Bio-medical Data Using a Hybrid PSO/Bayesian Classifier," *International Journal of Computing Science and Communication Technologies*. VOL. 2, NO. 1, July 2009.

[10] Leidiyana, H (2012). *Komparasi Algoritma Klasifikasi Data Mining Dalam Penentuan Resiko Kredit Kepemilikan Kendaraan Bermotor*. Tesis Magister Ilmu Komputer. Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri.

[11] Masripah, S (2011). *Algoritma Klasifikasi C4.5 Berbasis Particle Swarm Optimization Untuk Evaluasi Penentuan Kelayakan Pemberian Kredit Koperasi Syariah*. Tesis Magister Ilmu Komputer. Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri.

[12] Handayanna, F (2012). *Penerapan Particle Swarm Optimization Untuk Seleksi Atribut Pada Metode Support Vector Machine Untuk Prediksi Penyakit Diabetes*. Tesis Magister Ilmu Komputer. Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri.

[13] Ika N, S (2013). *Penentuan Kelayakan Kredit Bank Menggunakan Naïve Bayes Classifier Berbasis Particle Swarm Optimization*. Sripsi Teknik Informatika. Universitas Dian Nuswantoro.

[14] [http://archive.ics.uci.edu/ml/datasets/Statlog+\(German+Credit+Data\)](http://archive.ics.uci.edu/ml/datasets/Statlog+(German+Credit+Data)), di akses pada tanggal 26 Maret 2014.