

PENERAPAN PARTICLE SWARM OPTIMAZATION UNTUK MENENTUKAN KREDIT KEPEMILIKAN RUMAH DENGAN MENGGUNAKAN ALGORITMA C4.5

Mulkan Syarif

Program Studi Manajemen Informatika, AMIK BSI Jakarta
Jl. Kamal Raya No.18, Ringroad, Cengkareng, Jakarta barat
mulkan.mlk@bsi.ac.id

Abstract

In studies that have been done previously to determine ownership loan home. One the methods of the most widely used method with a high degree of accuracy is C4.5 algorithm. In conducting this study also used a method algorithm C4.5 and to improve the accuracy will be performed using the addition of particle swarm optimization method for the determination of credit ratings. Home ownership after testing the results obtained is a support vector machine produces a value of 91.93% accuracy and AUC value of 0.860 was then performed using particle swarm optimization method in which the attributes which originally totaled 8 predictor variables selected from eight attributes used. The results showed higher accuracy value that is equal to 94.15%, and AUC value of 0.941. So as to achieve the increased accuracy of 2.22%, and an increase in AUC of 0.081. By looking at the accuracy and AUC values, the algorithm of support vector machines based on particle swarm optimization and therefore is in the category of classification is very good.

Kata Kunci: Data Mining, C4.5, Particle Swarm Optimazation

PENDAHULUAN

Dalam pertumbuhan perekonomian semakin meningkat, salah satu dari pertumbuhan tersebut adalah pada sektor properti, untuk memenuhi kebutuhan kepemilikan rumah yang semakin meningkat, bank sebagai debitur memberikan kredit kepemilikan rumah. Selain itu terdapat juga terdapat faktor-faktor lainnya yang dapat menyebabkan pertumbuhan kredit pemilikan rumah meningkat, antara lain :

1. Penurunan BI rate sebesar 50 basis point dari 11,75 % menjadi 11,25% sehingga menyebabkan banyak bank yang menurunkan suku bunga kredit pemilikan rumah (KPR). (www.btn.co.id)
2. Dengan adanya penurunan suku bunga KPR, banyak masyarakat tertarik dengan kredit pemilikan

rumah ini sehingga daya beli meningkat.

3. Rendahnya resiko dari kredit pemilikan rumah ini sehingga menjadi alternatif untuk berinvestasi.

Kredit macet adalah salah satu resiko kredit yang dihadapi oleh pelaku industri keuangan dan perbankan. Kredit macet terjadi apabila dalam jangka panjang, lembaga keuangan atau perbankan tidak dapat menarik pinjaman kredit dalam waktu yang telah ditentukan. Kredit macet memiliki dampak yang buruk bagi penyedia kredit yaitu berupa resiko kerugian. Kredit macet dapat di minimasilir dengan cara melakukan analisa kredit yang akurat terhadap calon debitur (Bellotti dan Crook, 2007). Tujuan utamanya adalah untuk memperoleh keyakinan bahwa

debitur akan dapat memenuhi setiap kewajibannya dan memastikan kualitas kredit tetap lancar sampai berakhirnya perjanjian kredit. Keakuratan penilaian kredit sangat penting untuk profitabilitas lembaga keuangan (Gang, Jinxing, Jian dan Hongbing, 2011).

Penilaian kredit juga bermanfaat bagi penyedia kredit untuk mengukur dan mengelola risiko keuangan dalam memberikan kredit sehingga mereka dapat membuat keputusan yang lebih baik, lebih cepat dan lebih obyektif. Penelitian terdahulu mengenai topik penilaian kredit telah banyak dilakukan seperti penelitian yang dilakukan diantaranya dari peneliti Yi Jian yang mengambil judul *A Bank Customer Credit Evaluation Based on the Decision Tree and the Simulated Annealing Algorithm* yang bertujuan untuk penentuan nasabah baik dan buruk dalam pemberian kredit, dan terbukti algoritma C4.5 merupakan algoritma yang memiliki keandalan, efisiensi dan kesederhanaan dalam hal pengklasifikasian jenis nasabah (Yi Jian, Ya Chen, & Xiangjian Dia, 2007). Dalam penelitian ini penulis mengusulkan dengan menggunakan metode berbasis *decision tree* C4.5 kredit untuk menentukan skor yang lebih sederhana untuk dapat berfungsi sebagai pengganti untuk model yang rumit saat digunakan pada tujuan umum. Hasil penulisan diharapkan untuk memberikan informasi yang berharga untuk di evaluasi. Dari penelitian sebelumnya hasil yang baik dalam menggunakan algoritma dibandingkan algoritma data mining lainnya. Algoritma C4.5 memiliki kecepatan dalam membentuk sebuah dan tingkat akurasi yang baik serta model pembelajaran dari data, selain itu

model yang terbentuk mudah dipahami karena digambarkan dalam bentuk diagram pohon keputusan. Namun data yang tidak relevan juga dapat menurunkan sebuah akurasi algoritma C4.5 (Zhang & etc, 2010).

Untuk menyelesaikan suatu permasalahan di atas, maka penulis menggunakan model pohon keputusan algoritma C4.5. Model ini akan digunakan untuk memprediksi apakah peminjam nantinya akan bermasalah dalam pembayaran kredit atau tidak. Model ini dipilih karena proses *learning* dan klasifikasi pada algoritma C4.5 sederhana dan cepat. Secara umum, model algoritma C4.5 mempunyai tingkat akurasi yang tinggi (Han & Kamber, 2006).

BAHAN DAN METODE

A. Data Mining

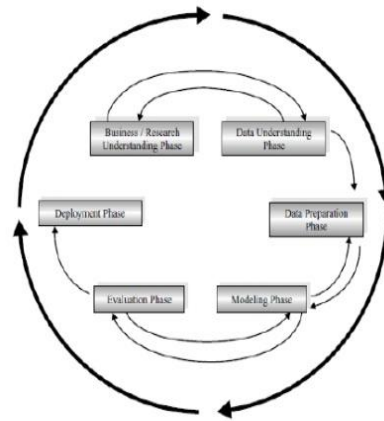
Data mining adalah aplikasi algoritma spesifik untuk mengekstrak pola dari data (Abraham, Grosan dan Ramos, 2006). Data Mining didefinisikan sebagai proses penemuan pola dalam data (Witten, 2011). Data mining sering juga disebut analisis data eksploratif. Data dalam jumlah besar yang diperoleh dari mesin kasir, pemindaian barcode dan dari berbagai basis data dalam perusahaan, kemudian ditelaah, dianalisis, dihapus dan dipakai ulang. Pencarian dilakukan pada model yang berbeda untuk memprediksi penjualan, respon pasar, keuntungan dan lain-lain (Olson dan Shi, 2008)

Kemajuan dalam bidang data mining didorong oleh beberapa faktor, antara lain (Larose, 2005):

1. Pertumbuhan yang cepat dalam kumpulan data.
2. Penyimpanan data dalam data *warehouse*, sehingga seluruh perusahaan memiliki akses ke dalam *database* yang andal.

3. Adanya peningkatan akses data melalui navigasi web dan intranet.
4. Tekanan kompetisi bisnis untuk meningkatkan penguasaan pasar dalam globalisasi ekonomi.
5. Perkembangan teknologi perangkat lunak untuk data mining (ketersediaan teknologi).
6. Perkembangan yang hebat dalam kemampuan komputasi dan pengembangan kapasitas media penyimpanan.

Sebuah standar lintas industri diperlukan dalam perusahaan untuk penggalan data yang dimiliki. Cross-Industry Standard *Process for Data Mining* (CRISP-DM) dikembangkan tahun 1996 oleh analis dari beberapa industri seperti Daimler Chrysler, SPSS dan NCR. CRISP-DM menyediakan standar proses data mining sebagai strategi pemecahan masalah secara umum dari bisnis atau unit penelitian. Dalam CRISP-DM, Sebuah proyek data mining memiliki siklus hidup yang terbagi dalam enam fase (Gambar 1). Keseluruhan fase berurutan yang ada tersebut bersifat adaptif. Fase berikutnya dalam urutan bergantung kepada keluaran dari fase sebelumnya. Hubungan penting antar fase digambarkan dengan panah. Sebagai contoh, jika proses berada pada fase modeling, maka berdasar pada perilaku dan karakteristik model, proses mungkin harus kembali kepada fase data preparation untuk perbaikan lebih lanjut terhadap data atau berpindah maju kepada fase *evaluation* (Larose, 2005).

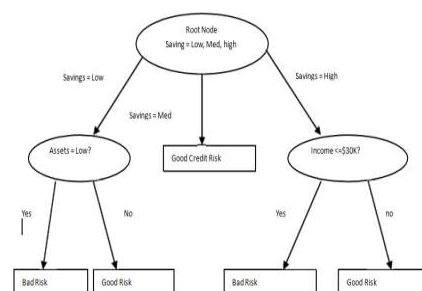


Sumber : Larose (2005)

Gambar 1. Proses Data Mining CRISP – DM

B. Algoritma C4.5

Salah satu metode klasifikasi menarik yang melibatkan konstruksi pohon keputusan, koleksi node keputusan, terhubung oleh cabang-cabang, memperpanjang bawah dari simpul akar sampai berakhir di node daun. Dimulai dari node root, yang oleh konvensi ditempatkan dibagian atas dari diagram pohon keputusan, atribut diuji pada node keputusan, dengan setiap hasil yang mungkin menghasilkan cabang. Setiap cabang kemudian mengarah ke node lain baik keputusan atau ke node daun untuk mengakhiri (Larose, 2005)



Sumber: (Larose, 2005)

Gambar 2 Contoh konsep pohon keputusan sederhana

Algoritma C4.5 merupakan bagian dari kelompok algoritma decision trees dan merupakan kategori 10 algoritma yang paling populer. Diakhir tahun 1970 hingga awal

tahun 1980-an, J.Ross Quinlan seorang peneliti dibidang mesin pembelajaran mengembangkan sebuah model keputusan yang dinamakan ID3(*Iterative Dichotomiser*), walaupun sebelumnya proyek ini telah dibuat oleh E.B Tahapan dalam membuat sebuah pohon keputusan dengan algoritma C4.5(Larose, 2005) yaitu:

1. Mempersiapkan data *training*, data *training* biasanya diambil dari data histori yang pernah terjadi sebelumnya atau disebut data masa lalu dan sudah dikelompokan dalam kelas-kelas tertentu.
2. Menghitung Total *Entropy* sebelum dicari masing-masing Entropy class

$$H(T) = -\sum P_j \log_2(P_j)$$

Keterangan :

H = Himpunan Kasus

T = Atribut

P_j = Proporsi dari H_j terhadap H

3. Hitung Nilai Gain dengan information gain dengan rata-rata Gain *average* = H(T)-H_{saving}(T)

Keterangan:

H(T) = Total Entropy

H_{saving}(T) = Total Gain information untuk masing-masing Atribut

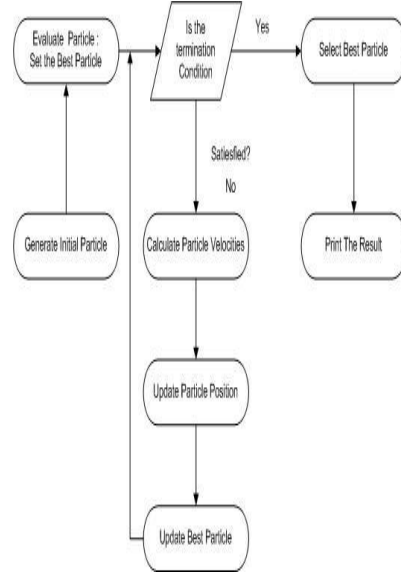
4. Ulangi langkah ke-2 dan ke-3 hingga semua tupel terpartisi
Proses partisi pohon keputusan akan berhenti disaat:

- a. Semua tupel dalam node N mendapatkan kelas yang sama
- b. Tidak ada atribut didalam tupel yang dipartisi lagi
- c. Tidak ada tupel didalam cabang yang kosong

C. Particle Swarm Optimazation

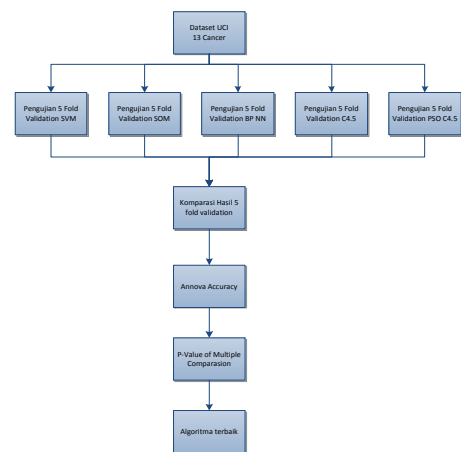
Particle swarm optimization (PSO) dapat diasumsikan dengan sekelompok burung yang secara acak

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, Grosan & Ramos, 2006).



Sumber: Abraham, Grosan & Ramos (2006)

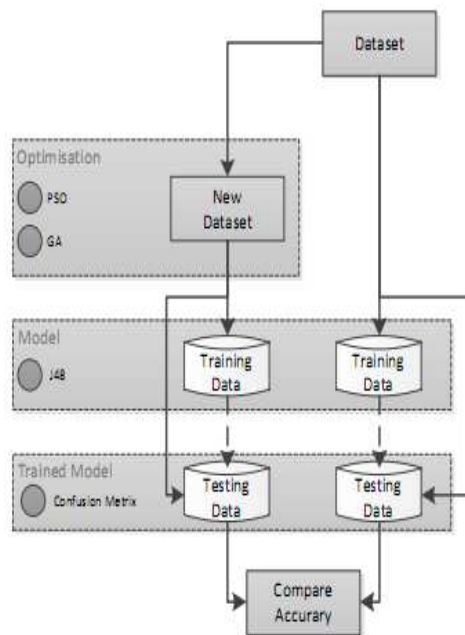
Gambar 3 Struktur dasar PSO



Sumber: Sausa (2004)

Gambar 4 Model Applying particle swarm optimization-based decision tree classifier for cancer classification on gene expression Data

Model penelitian yang dilakukan (Sausa & etc, 2004) Berdasarkan hasil akurasi 98% yang ada maka bisa disimpulkan bahwa PSO dapat mengandung hasil yang sangat baik pada dataset yang digunakan.

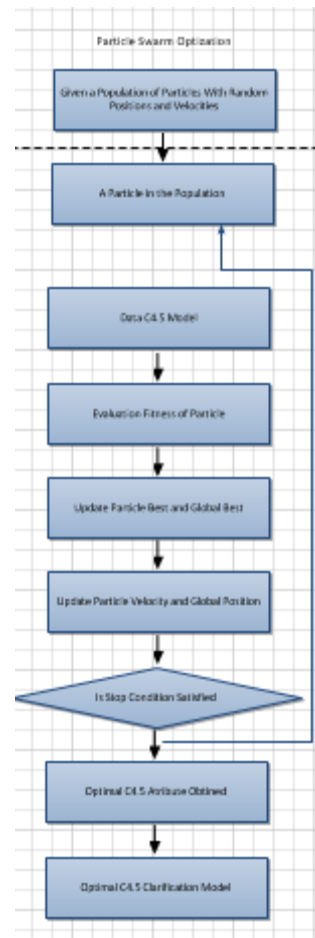


Sumber :Sausa &etc (2004)

Gambar 5 Application of the PSO model Scoring

HASIL DAN PEMBAHASAN

Dari beberapa penelitian sebelumnya maka peneliti menggunakan metode penelitian yang akan digunakan dalam eksperimen ini adalah algoritma C4.5 dengan *Particle Swarm Optimazation* ini dengan menggunakan model penelitian sebagai berikut:



Gambar 6 Model yang di usulkan

A. Data Preparation

Data kredit diambil dari data konsumen KPR Bank Syariah Mandiri 2013 sampai 2014, dimana dari 632 data debitur, ditemukan 97 debitur yang mengalami masalah dalam pembayaran angsuran. Ada beberapa atribut yang digunakan dalam data debitur yaitu akun bank, lama pinjaman, jaminan, jumlah kredit, jumlah angsuran, umur, pekerjaan, masa kerja, gaji, persentase pemotongan gaji, jenis kelamin, status, pinjaman sebelumnya, response. Nilai dari atribut tersebut ada yang merupakan nilai kategorikal diantaranya akun bank, jaminan, pekerjaan, masa

kerja, jenis kelamin, status, pinjaman sebelumnya

B. Data Integration

untuk meningkatkan akurasi dan efisiensi algoritma. Data yang digunakan dalam penulisan ini bernilai kategorikal.

Tabel 1 dibawah ini menampilkan nama atribut, kategori, dan nilai angka(range), berikut rule nilainya:

Tabel 1 Kategori Atribut

No	Atribut	Nilai	Keterangan
1	akun_bank	1	TidakMemilikiakun bank
		2	Memilikiakun bank
2	Jaminan	1	Surat - SuratBerharga
		2	Rumah
		3	Tanah
		4	Lain lain
3	pekerjaan	1	Pensiunan
		2	PNS
		3	KaryawanSwasta
		4	Wiraswasta
		5	Guru
4	masa_kerja	1	< 1 tahun
		2	1 - 4 tahun
		3	4 - 7 tahun
		4	> 7tahun
5	Jenis_kelamin	1	Pria
		2	Wanita
6	Status	1	Single
		2	Menikah
		3	Berceraai
7	pinjamansebelumnya	1	tidakadapinjaman
		2	adapinjaman (belumlunas)
		3	adapinjaman (sudahlunas)
8	Response	1	Lancar
		2	Bermasalah

Sumber :Bank Syariah Mandiri (2013-2014)

C. Pengujian Algoritma

Dalam penelitian ini akan dilakukan analisis menggunakan

metode klasifikasi data mining. Metode yang diusulkan untuk pengolahan data debitur kredit adalah penggunaan algoritma C4.5, serta di tingkatkan dalam segi seleksi atribut dengan *Particle Swarm Optimazation*, setelah diolah dan menghasilkan model, maka terhadap model yang dihasilkan tersebut dilakukan pengujian menggunakan 90% data pada k-fold cross validation, sedangkan 10 % dari data akan di pakai untuk testing. Kemudian dilakukan evaluasi dan validasi hasil dengan confusion matrix dan kurva ROC. Tahap selanjutnya adalah memperbandingkan hasil akurasi dan AUC dari setiap model, sehingga diperoleh model dari metode klasifikasi penentuan kelayakan pemberian kredit dengan nilai akurasi dan AUC tertinggi. Dalam tahapan ini akan dilakukan beberapa langkah-langkah metode yang diusulkan yaitu:

1. Algoritma C4.5 yaitu salah satu algoritma dalam metode decision tree yang merubah data menjadi pohon keputusan menggunakan rumus perhitungan entropi.
2. Particle Swarm Optimazation yaitu pencarian solusi optimal secara global dalam ruang pencarian melalui interaksi individu dalam segerombolan partikel dengan cara melakukan seleksi atribut yang ada.

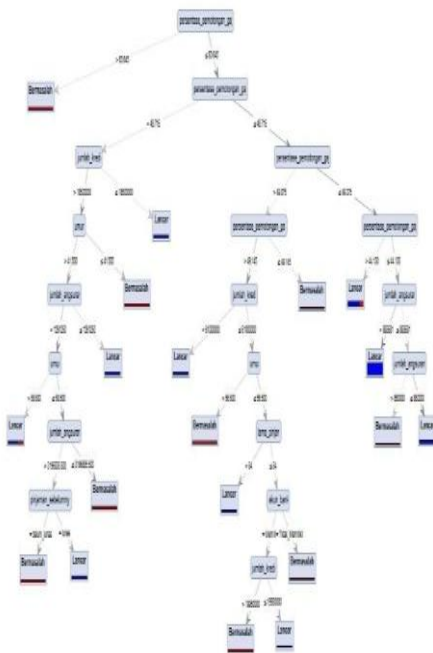
Dari hasil pengujian algoritma c4.5 didapatkan pohon keputusan sebagai berikut:

Tabel 2 Nilai Accuracy, Sensitivity, Specificity, ppv, dan npv C4.5

Nilai (%)	
Accuracy	91,93
Sensitivity	94,11
Specificity	75,00
PPV	96,70
NPV	62,07

Sumber: Hasil Pengujian Dengan Software Rapid Miner

Grafik ROC dengan nilai AUC (Area Under Curve) untuk algoritma C4.5 sebesar 0.860 dengan tingkat diagnosa *Good Classification*



Sumber: Hasil Pengujian Dengan Software Rapid Miner

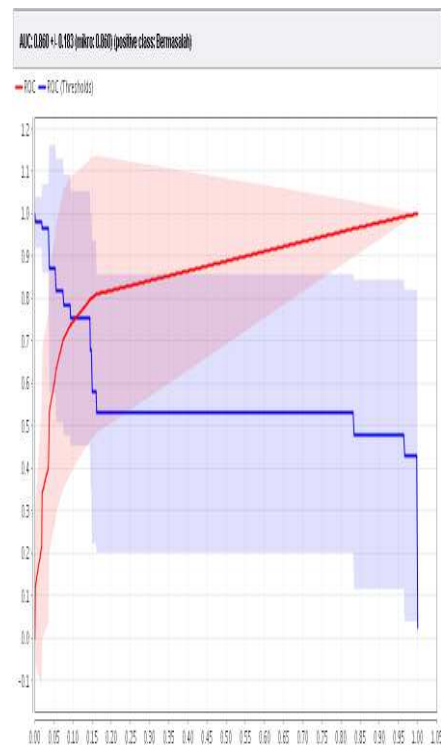
Gambar 7 Pohon Keputusan C4.5

Confusion matrix membentuk matriks yang terdiri dari true positif atau tupel positif dan true negatif atau tupel negatif. Pada gambar 7, akurasi C4.5 sebesar 91,93%. dimana Jumlah *True Positive* (TP) adalah 527, Untuk *False Negative* (FN) adalah 33, untuk *False Positive* (FP) adalah 18 dan Untuk *True Negative* (TN) adalah 54

accuracy: 91.93% +/- 3.77% (mikro: 91.93%)			
	true Lancar	true Bermasalah	class precision
pred Lancar	527	33	94.11%
pred Bermasalah	18	54	75.00%
class recall	96.70%	62.07%	

Sumber: Hasil Pengujian Dengan Software Rapid Miner

Gambar 8 Hasil Akurasi Algoritma C4.5

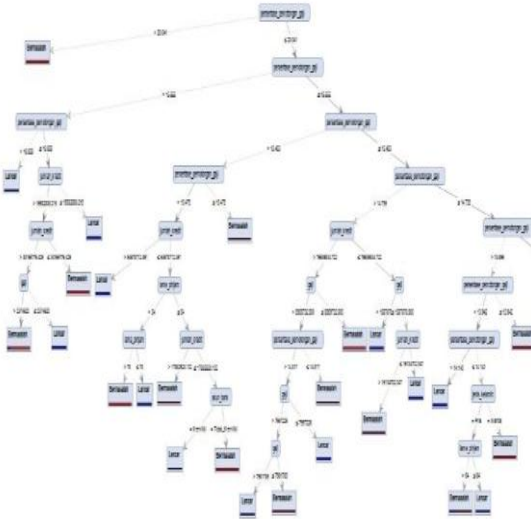


Sumber: Hasil Pengujian Dengan Software Rapid Miner

Gambar 9 Nilai AUC C4.5 dalam ROC Curve

A. Evaluasi model C4.5 berbasis PSO dengan Confusion Matrix

Dari hasil pengujian algoritma model C4.5 berbasis *particle swarm optimazation* diperoleh pohon keputusan sebagai berikut :



Sumber: Hasil Pengujian Dengan Software Rapid Miner

Gambar 10 Pohon keputusan C4.5 Berbasis PSO

Confusion matrix membentuk matriks yang terdiri dari true positif atau tupel positif dan true negatif atau tupel negatif. Pada gambar 10, akurasi C4.5 berbasis PSO sebesar 94.15 %. Dimana Jumlah True Positive (TP) adalah 523 , Untuk False Negative (FN) adalah 15 , untuk False Positive (FP) adalah 22 dan Untuk True Negative (TN) adalah 72.

accuracy: 94.15% (+/- 3.54%) (nilai: 94.15%)			
	true Lancar	true Bermasalah	class prevision
pred. Lanar	523	15	97.21%
pred. Bermasalah	22	72	76.60%
class recall	95.88%	92.78%	

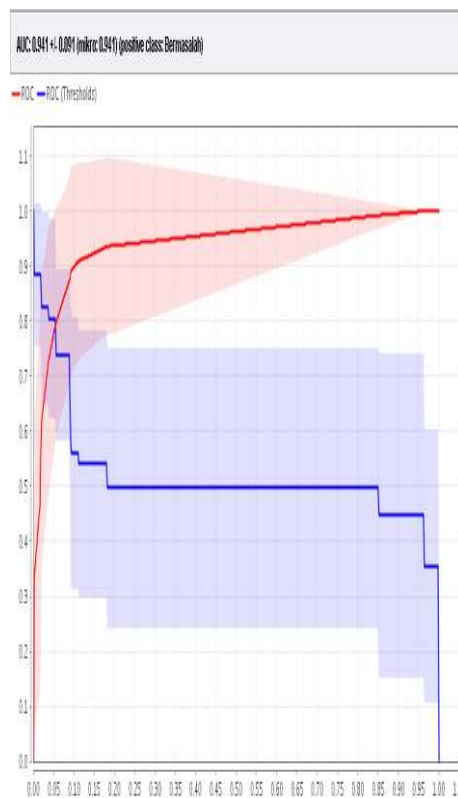
Gambar 11 Hasil Akurasi Algoritmat C4.5

berbasis PSO
Tabel 3 Nilai Accuracy, Sensitivity, Specificity, ppv, dan npv C4.5 berbasis PSO

Nilai (%)	
Accuracy	94,15
Sensitivity	97,21
Specificity	76,60
PPV	95,98
NPV	82,76

Sumber: Hasil Pengujian Dengan Software Rapid Miner

Grafik ROC dengan nilai AUC (Area Under Curve) untuk algoritma C4.5 berbasis PSO sebesar 0.941 dengan tingkat diagnosa *Excellent Classification*



Sumber: Hasil Pengujian Dengan Software Rapid Miner

Gambar 12 Nilai AUC C4.5 berbasis PSO dalam ROC Curve

Berdasarkan dari analisa pengujian algoritma diatas, maka dapat dirangkumkan hasilnya sebagai berikut :

Tabel 4 Perbandingan Performance Mode

	C4.5	C4.5 + PSO
<i>Accuracy</i>	91,93%	94,15%
AUC	0,860	0,941

Melihat hasil perbandingan dari seluruh pengujian yang telah dilakukan didapatkan hasil dengan nilai akurasi untuk C4.5 *Accuracy* 91,93 % dan AUC 0,860. Sedangkan algoritma C4.5 berbasis PSO dengan nilai Akurasi 94,15% dan AUC 0,941. Dilihat dari dari hasil Akurasi dan nilai AUC ke dua algoritma tersebut, maka dipilihlah algoritma C4.5 berbasis *Particle Swarm Optimazation* sebagai algoritma yang paling akurat karena memiliki nilai akurasi dan ROC paling tinggi diantara algoritma yang lain.

B. Graphical user interface (GUI)

Berdasarkan hasil akurasi yang telah di terapkan, diketahui bahwa algoritma C4.5 yang dioptimaskan dengan Particle Swarm Optimization (PSO) memiliki akurasi dan performa terbaik, sehingga *rule* yang dihasilkan oleh algoritma C4.5 dioptimasi dengan *Particle Swarm Optimization* (PSO) dijadikan sebagai *rule* untuk pembuatan interface yang dapat memudahkan dalam penentuan kelayakan kredit pemilikan rumah. *Interface* yang digunakan dalam penelitian ini dibuat dengan menggunakan bahasa PHP berbasis web. Tampilan untuk

Graphical user interface (GUI) system penunjang keputusan. Dapat dilihat pada gambar 13

Id	Account	Bank	Collateral	Job	Work Duration	Gender	Status	Previous Loan	Response	Action
3	Rumah	Wibakarata	5 Tahun	Wanita	Sangat Baik	Sangat Baik	LUNCAIR	<input type="button" value="Edit"/>		
2	Surat Berharta	Sweeta	4 Tahun	Wanita	Buruk	Buruk	TIDAK LUNCAIR	<input type="button" value="Edit"/>		
1	Tanah	PNB	7 Tahun	Pria	Baik	Baik	LUNCAIR	<input type="button" value="Edit"/>		

Sumber: Hasil Perhitungan GUI berdasarkan Java Netbean 6.5

Gambar 13 Graphical user interface (GUI)

KESIMPULAN

Dari penelitian yang telah dilakukan pembuatan menggunakan algoritma C4.5, dan *Support Vector Machine* dengan berbasis *Particle Swarm Optimazation* dan menggunakan data perkreditan Bank Syariah Mandiri. Model yang dihasilkan dikomparasi untuk dapat diketahui metode terbaik dalam penentuan kelayakan Kredit Pemilikan Rumah.

Dari hasil pengujian dengan mengukur kinerja ketiga metode tersebut menggunakan *confusion matrix*, kurva *ROC* diketahui bahwa C4.5 menghasilkan nilai akurasi 91,93 % dan nilai AUC 0,860. Metode C4.5 berbasis PSO menghasilkan nilai akurasi 94,15 % dan nilai AUC 0,941.

Dengan demikian metode C4.5 berbasis *Particle Swarm Optimazation* adalah metode yang terbaik untuk pemecahan masalah

kelayakan penentuan kredit Pemilikan Rumah.

Untuk keperluan penelitian lebih lanjut mengenai metode klasifikasi data mining dengan menggunakan data dibidang kredit perbankan dapat dilakukan dengan menggunakan *dataset* lebih dari satu institusi untuk mengetahui keakuratan dari algoritma yang akan diuji. Melakukan *training* kepada seluruh karyawan agar dapat dengan baik menjalankan sistem untuk mendapatkan keputusan dengan cepat dan tepat.

Melakukan pengujian *dataset* kredit dengan algoritma lain seperti metode *Naive Bayes*, *K-Nearest Neighbor* dan sebagainya, serta melakukan optimasi untuk peningkatan akurasi dengan algoritma *Genetic Algorithm*.

UCAPAN TERIMA KASIH

Dalam penyelesaian penelitian ini penulis tidak lupa mengucapkan terimakasih kepada semua pihak yang telah membantu untuk terselesaikannya penelitian ini kepada:

1. Ketua STMIK Nusa Mandiri
2. Rekan-rekan di STMIK Nusa Mandiri yang telah memberikan motivasi dalam pengerjaan penelitian ini.
3. Kapala Cabang Bank Syariah Mandiri KCU Tangerang

DAFTAR PUSTAKA

Abraham, A., Grosan, C., Ramos, V., (2006). *Swarm Intelligence in Data Mining*. Springer-Verlag Berlin Heidelberg.

Gang, W., Jinxing, H., Jian, M., & Hongbing, J. (2011). A comparative assessment of ensemble learning for credit scoring. *Expert Systems with Applications: An International Journal*. 38, 223-230.

Han, J., & Kamber, M. (2006). *Data Mining Concept and Tehniques*. San Fransisco: Morgan Kauffman.

Larose, D. T. (2005). *Discovering Knowledge in Data*. New Jersey: John Willey & Sons, Inc.

Olson, D, & Shi, Y. (2008). *Pengantar Ilmu Penggalan Data Bisnis*. Jakarta: Penerbit Salemba Empat.

Sausa & etc, (2004). *Particle swarm based Data Mining Algorithms for classification tasks*. *ACM Digital Library*.

Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data Mining: Practical Machine Learning and Tools*. Burlington: Morgan Kaufmann Publisher.

Yi Jiang, Yan Chen, Zhiming Zeng, & Xiangjian He. (2009). A Bank Customer Credit Evaluation Based on the Decision Tree and the Simulated Annealing Algorithm. *World Congress on Computer Science and Information Engineering*, 18-22.

Zhang, & etc., (2010). *Vertical bagging decision trees model for credit scoring*. Elsevier Ltd.

Bellotti, T., & Crook, J. (2007). Support vector machines for credit scoring and discovery of significant features. *Expert System with Application: An International Journal*, 36, 3302-3308.