

KOMPARASI ALGORITMA C4.5 DAN SVM BERBASIS PARTICLE SWARM OPTIMAZATION DALAM PENENTUAN KREDIT

Syaifur Rahmatullah

Program Studi Teknik Informatika
STMIK Nusa Mandiri JAKARTA
Jl. Damai no. 8 Warung Jati Barat
syaifur.rahmatullah@gmail.com

ABSTRACT

Bad credit is one of the credit risk faced by the financial and banking industry. Bad credit happens if in the long run, financial institutions or banks can not attract loans within a predetermined time. Bad credit has a negative effect on credit providers in the form of risk of loss course this should not be allowed to drag on and had to find a way out. However, to ensure accuracy in the determination of credit worthiness required an accurate algorithm. Therefore, there should be a study that aims to find an algorithm that accurately by means mengkomparasi some of them C4.5 algorithm, and SVM. To further improve the accuracy of the algorithms are in Particle Swarm Optimization with Optimazation. Berupan confusion matrix research results prove the accuracy of Support Vector Machine-based Particle Swarm Optimazation exists at the level of accuracy of 96.20% and the AUC by 0989.

Keywords: *Data Mining, SVM, C4.5, Particle Swarm Optimazation*

ABSTRAK

Kredit macet merupakan 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 Tentu saja hal ini tidak boleh dibiarkan berlarut-larut dan harus dicari jalan keluarnya. Namun untuk menjamin akurasi dalam penentuan kelayakan kredit diperlukan suatu algoritma yang akurat. Oleh karena itu perlu diadakan sebuah penelitian yang bertujuan untuk mencari algoritma yang akurat dengan cara mengkomparasi beberapa algoritma diantaranya C4.5, dan SVM. Untuk lebih meningkatkan akurasi kedua algoritma tersebut di optimasi dengan Particle Swarm Optimazation. Hasil penelitian berupa confusion matrix membuktikan akurasi Support Vector Machine berbasis Particle Swarm Optimazation memiliki tingkat akurasi sebesar 96,20% dan hasil AUC sebesar 0.989.

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

I. PENDAHULUAN

Kredit macet merupakan 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 (Jianguo & Tao, 2008). Kredit macet memiliki dampak yang buruk bagi penyedia kredit yaitu berupa resiko kerugian (Hian, Wei & Chwee, 2006). Tentu saja hal ini tidak boleh dibiarkan berlarut-larut dan harus dicari jalan keluarnya, karena masalah ini dapat mengakibatkan penghentian operasi dan mengarah kepada kebangkrutan lembaga penyedia kredit tersebut (Jianguo & Tao, 2008).

Kredit macet dapat dihindari dengan cara melakukan analisa kredit yang akurat terhadap calon debitur (Bellotti & 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 & Hongbing, 2011).

Penilaian kredit mengelompokkan para calon debitur menjadi dua jenis yaitu debitur baik dan debitur buruk. Debitur baik memiliki kemungkinan besar akan membayar kewajiban keuangannya dengan lancar, sedangkan debitur buruk memiliki kemungkinan besar mengalami kredit macet (Gang, Jinxing, Jian & Hongbing, 2011). Penilaian kredit sangat penting karena banyak keuntungan yang diperoleh yaitu mengurangi biaya analisa kredit, pengambilan keputusan lebih cepat, pemantauan lebih dekat dengan data yang ada dan memungkinkan untuk

menetapkan calon debitur prioritas (Ping, 2009). 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 (Hian, Wei & Chwee, 2006).

Penelitian terdahulu mengenai topik penilaian kredit telah banyak dilakukan seperti penelitian yang dilakukan diantaranya adalah 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). Peneliti selanjutnya yang membahas tentang risiko pemberian kredit adalah Yun Li dengan judul penelitian *Application of the PSO - SVM Model for Credit Scoring*, penelitian kali ini menggunakan algoritma SVM dengan metode penelitian membandingkan beberapa fitur seleksi untuk meningkatkan akurasi SVM dan menghasilkan bahwa PSO dengan SVM memiliki akurasi yang tinggi dalam pemecahan masalah risiko penentuan kredit (Yun Li, Qiu Yancoo, Hua Zhang, 2011). Penelitian selanjutnya yang membahas tentang kredit adalah Tony Bellotti dan Jonathan Crook yang berjudul *Support vector machines for credit scoring and discovery of significant features*, menggunakan model *Support Vector Machine (SVM)*, *Logistic Regression (LR)*, *Linear Discriminant Analysis (LDA)* dan *k-Nearest Neighbours (kNN)* untuk penentuan kelayakan pemberian kredit dan menentukan fitur yang berpengaruh. Hasil menunjukkan bahwa metode *Support Vector Machine (SVM)* mengungguli ketiga metode lainnya dan dapat digunakan sebagai metode yang baik dalam seleksi fitur yang berpengaruh secara signifikan terhadap dasar keputusan kelayakan pemberian kartu kredit dan juga sangat tepat dalam pengolahan data dengan jumlah besar (Bellotti & Crook, 2007). Selanjutnya penelitian yang dilakukan Zurada dengan judul *Could Decision Trees Improve the Classification Accuracy and Interpretability of Loan Granting Decisions ?* (Zurada, 2010). Membuat penelitian untuk membandingkan beberapa algoritma seperti Regresi Linier, Neural Network, Support Vector Machine, Case Base Reasoning, Rule Based Fuzzy Neural Network dan Decision Tree. Semua model algoritma tadi digunakan untuk

menganalisa persetujuan pinjaman dalam bentuk kredit. Dari hasil penelitian didapatkan bahwa Decision Tree terbukti mempunyai akurasi tertinggi dalam menentukan keputusan dibandingkan algoritma lain, dan penelitian yang dilakukan Satchidananda & Simha, 2006. *Comparing decision trees with logistic regression for credit risk analysis*. Penelitian ini membandingkan dua model algoritma untuk analisa resiko kredit, yaitu Pohon Keputusan dan Regresi Logistik. Data diambil dari dua bank yang berbeda, kemudian untuk mengelompokkan kasus positif dan negatif maka dilakukan klustering data dengan menggunakan k-means. Hasil analisa dari masing-masing model dikomparasi dan kemudian diukur kemudian didapatkan bahwa algoritma pohon keputusan mempunyai tingkat akurasi yang tinggi dibandingkan algoritma regresi logistik.

Particle swarm optimization (PSO) merupakan teknik komputasi evolusioner yang mampu menghasilkan solusi optimal secara global dalam ruang pencarian melalui interaksi individu dalam segerombolan partikel. Setiap partikel menyampaikan informasi berupa posisi terbaiknya kepada partikel yang lain dan menyesuaikan posisi dan kecepatan masing-masing berdasarkan informasi yang diterima mengenai posisi yang terbaik tersebut (Shuzhou & Bo, 2011). Particle swarm optimization dapat digunakan sebagai teknik optimasi untuk mengoptimalkan subset fitur dan parameter secara bersamaan (Yun, Qiu-yan & Hua, 2011).

Dari beberapa penelitian diatas di temukan beberapa metode yang dapat digunakan untuk penentuan kelayakan kredit pemilikan rumah dan pada kesempatan ini penelitian yang akan dilakukan adalah menganalisis komparasi algoritma klasifikasi data mining, diantaranya adalah algoritma, algoritma Decision Tree, dan algoritma Support Vector Machine serta penerapan metode Particle swarm optimization (PSO) dalam penyeleksian atribut dengan tujuan untuk mendapatkan algoritma yang paling akurat dalam penentuan kelayakan pemberian kredit.

II. LANDASAN TEORI

A. Data Mining

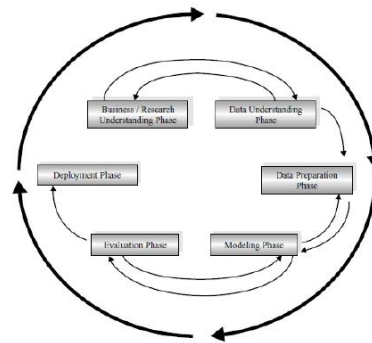
Data mining adalah aplikasi algoritma spesifik untuk mengekstrak poladari data (Abraham, Grosan & Ramos, 2006). Data Mining didefinisikan sebagai proses penemuan pola dalam data (Witten, 2011). Data mining sering juga disebut analisis data eksploratif. Data dalam jumlah besar yang diperoleh dari mesin

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 & 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 analisis dari beberapa industri seperti DaimlerChrysler, 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 2.1). 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).

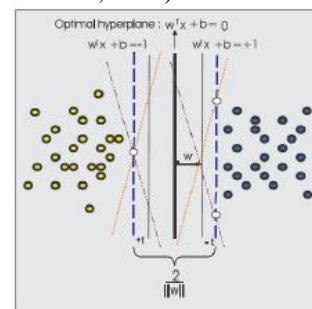


Sumber: (Larose, 2005)

Gambar 1. Proses Data Mining CRISP – DM

B. Algoritma Support Vector Machine

Support Vector Machine (SVM) adalah metode learning machine yang bekerja atas prinsip Structural Risk Minimization (SRM) dengan tujuan menemukan hyperplane terbaik yang memisahkan dua buah class pada input space (Bellotti & Crook, 2007). Hyperplane terbaik adalah hyperplane yang terletak ditengah-tengah antara dua set obyek dari dua class. 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 (Aydin, Karakose & Akin, 2011).



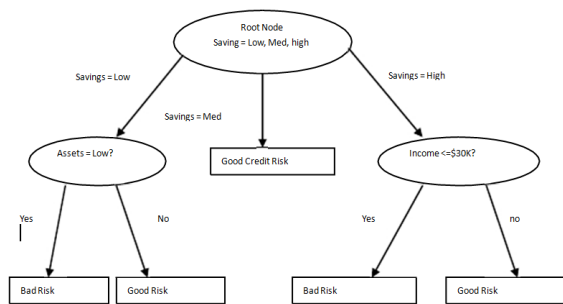
Sumber: (Larose, 2005)

Gambar 2. Konsep SVM untuk mencari hyperplane terbaik

C. 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(Larose, 2005). Dimulaidari 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 3. Contoh konsep pohon keputusan sederhana

Algoritma C4.5 merupakan bagian dari kelompok algoritma decision trees dan merupakan katerogi 10 algoritma yang paling populer. Diakhir tahun 1970 hingga awal tahun 1980-an, J.Rosss Quinlan seorang peneliti dibidang mesin pembelajaran mengembangkan sebuah model keputusan yang dinamakan ID3 (Interative 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 dikelompokkan 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
 Pj = Proporsi dari Hj terhadap H
3. Hitung Nilai Gain dengan information gain dengan rata-rata

$$\text{Gain average} = H(T) - H_{\text{saving}}(T)$$

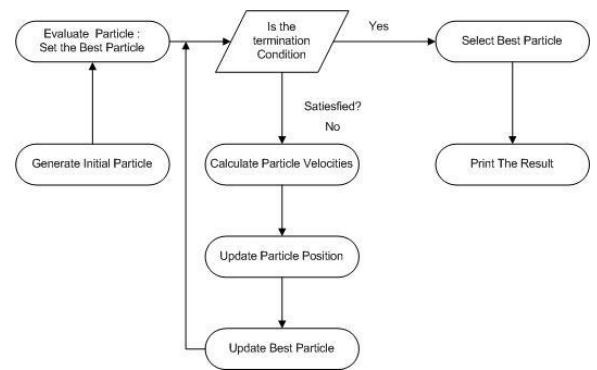
Keterangan:
 H(T) = Total Entropy
 Hsaving(T) = Total Gain information untuk masing-masing Atribut
4. Ulangi langkah ke-2 dan ke-3 hingga semua tupel terpartisi Proses partisi pohon keputusan akan berhenti disaat:
 - a. Semua tupel dalam node N mendapatkan kelas yang sama
 - b. Tidak ada atribut didalam tupel yang dipartisi lagi

- c. Tidak ada tupel didalam cabang yang kosong

D. Particle Swarm Optimization

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).

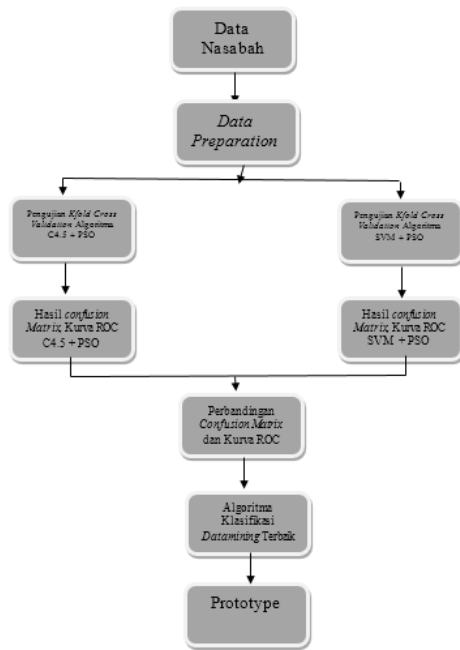


Sumber: (Abraham, Grosan & Ramos, 2006)

Gambar 4. Struktur dasar PSO

III. Metode Penelitian

peneliti menggunakan metode penelitian yang akan digunakan dalam eksperimen komparasi algoritma Support Vector Machine, dan Decision Tree berbasis Particle Swarm Optimazation ini dengan menggunakan model penelitian sebagai berikut :



Sumber : Olahan penulis (2015)

Gambar 5. Model yang di usulkan

A. Data Preparation

Data kredit diambil dari data konsumen KPR Bank DKI 2012 sampai 2013, dimana dari 632 data debitur, ditemukan 97 debitur yang mengalami masalah dalam pembayaran angsuran. Jadi ada sekitar 15,35% debitur yang bermasalah. 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. Untuk model SVM data ditransformasi ke dalam angka.

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	Tidak Memiliki akun bank
		2	Memiliki akun bank
2	Jaminan	1	Surat - Surat

			Berharga
		2	Rumah
		3	Tanah
		4	Lain lain
3	Pekerjaan	1	Pensiunan
		2	PNS
		3	Karyawan Swasta
		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	Bercerai
7	pinjaman sebelumnya	1	tidak ada pinjaman
		2	ada pinjaman (belum lunas)
		3	ada pinjaman (sudah lunas)
8	Response	1	Lancar
		2	Bermasalah

Sumber : Olahan penulis (2015)

C. Pengujian Algoritma

Dalam penelitian ini akan dilakukan analisis komparasi menggunakan tiga metode klasifikasi data mining. Metode yang diusulkan untuk pengolahan data debitur kredit adalah penggunaan algoritma C4.5, dan *Support Vector Machine* 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 data yaitu seperti berikut:

Hasil pengujian dengan akurasi yang paling tinggi adalah metode yang akan digunakan untuk penentuan kelayakan

pemberian kredit ini. Berikut gambaran kateristik dari masing-masing metode:

1. Algoritma C4.5 yaitu salah satu algoritma dalam metode decision tree yang merubah data menjadi pohon keputusan menggunakan rumus perhitungan entropi
2. Support Vector Machine yaitu model untuk menemukan hyperlane terbaik yang memisahkan dua buah class.

Particle Swarm Optimazation yaitu pencarian solusi optimal secara global dalam ruang pencarian melalui interaksi individu dalam segerombolan partikel dengan cara melakukan seleksi terhadap atribut yang ada

IV. Hasil Penelitian Dan Pembahasan
A. Evaluasi Model C4.5 dengan Confusion Matrix dan Grafik ROC

Confusion matrix membentuk matriks yang terdiri dari true positif atau tupel positif dan true negatif atau tupel negatif. Pada gambar IV.10, akurasi C4.5 sebesar 93.81 % . dimana Jumlah True Positive (TP) adalah 525 , Untuk False Negative (FN) adalah 29 , untuk False Positive (FP) adalah 10 dan Untuk True Negative (TN) adalah 67

	true Lancar	true Bermasalah	class precision
pred Lancar	525	29	94.77%
pred Bermasalah	10	67	87.01%
class recall	93.15%	69.79%	

Sumber : Olahan Penulis (2015)
Gambar 6. Hasil Akurasi Algoritma C4.5

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

	Nilai (%)
Accuracy	93,81
Sensitivity	94,76
Specificity	87,01
PPV	98,13
NPV	69,79

Sumber : Olahan Penulis (2015)

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



Sumber : Olahan Penulis (2015)
Gambar 7. Nilai AUC C4.5 dalam ROC Curve

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

Confusion matrix membentuk matriks yang terdiri dari true positif atau tupel positif dan true negatif atau tupel negatif. Pada gambar IV.11 , akurasi C4.5 berbasis PSO sebesar 94.29 % . dimana Jumlah True Positive (TP) adalah 529 , Untuk False Negative (FN) adalah 30 , untuk False Positive (FP) adalah 6 dan Untuk True Negative (TN) adalah 66

	true Lancar	true Bermasalah	class precision
pred Lancar	529	30	94.67%
pred Bermasalah	6	66	91.67%
class recall	94.68%	69.79%	

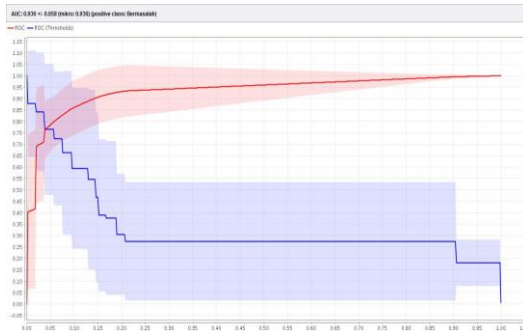
Sumber : Olahan Penulis (2015)
Gambar 8. Hasil Akurasi Algoritma C4.5 berbasis PSO

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

	Nilai (%)
Accuracy	94.29
Sensitivity	94.63
Specificity	91.66
PPV	98.87
NPV	68.75

Sumber : Olahan Penulis (2015)

grafik ROC dengan nilai AUC (Area Under Curve) untuk algoritma C4.5 berbasis PSO sebesar 0.936 dengan tingkat diagnosa *Excellent Classification*



Sumber : Olahan Penulis (2015)
Gambar 8. Nilai AUC C4.5 berbasis PSO dalam ROC Curve

C. Evaluasi model Support Vector Machine dengan Confusion Matrix

Confusion matrix membentuk matriks yang terdiri dari true positif atau tupel positif dan true negatif atau tupel negatif. Pada gambar IV.12 , akurasi *Support Vector Machine* sebesar 89,85 % . dimana Jumlah True Positive (TP) adalah 524 , Untuk False Negative (FN) adalah 53 , untuk False Positive (FP) adalah 11 dan Untuk True Negative (TN) adalah 43

accuracy: 89.85% +/- 2.69% (mikro: 89.85%)			
	true Lancar	true Bermasalah	class precision
pred. Lancar	524	53	90.81%
pred. Bermasalah	11	43	79.62%
class recall	97.94%	44.79%	

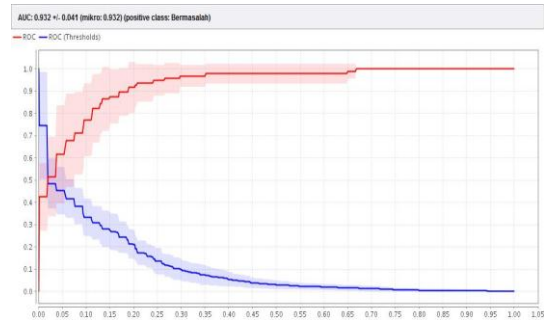
Sumber : Olahan Penulis (2015)
Gambar 9. Hasil Akurasi Algoritma Support Vector Machine

Tabel 4. Nilai Accuracy, Sensitivity, Specificity, ppv, dan npv Support Vector Machine

	Nilai (%)
Accuracy	89,85
Sensitivity	90,81
Specificity	79,62
PPV	97,94
NPV	44,79

Sumber : Olahan Penulis (2015)

grafik ROC dengan nilai AUC (Area Under Curve) untuk algoritma *Support Vector Machine* sebesar 0.932 dengan tingkat diagnosa *Excellent Classification*



Sumber : Olahan Penulis (2015)
Gambar 10. Nilai AUC Support Vector Machine dalam ROC Curve

D. Evaluasi model Support Vector Machine berbasis PSO dengan Confusion Matrix

Confusion matrix membentuk matriks yang terdiri dari true positif atau tupel positif dan true negatif atau tupel negatif. Pada gambar IV.13 , akurasi *Support Vector Machine* berbasis PSO sebesar 96.20 % . dimana Jumlah True Positive (TP) adalah 531 , Untuk False Negative (FN) adalah 20 , untuk False Positive (FP) adalah 4 dan Untuk True Negative (TN) adalah 76

accuracy: 96.20% +/- 1.91% (mikro: 96.20%)			
	true Lancar	true Bermasalah	class precision
pred. Lancar	531	20	96.37%
pred. Bermasalah	4	76	95.00%
class recall	99.25%	79.17%	

Sumber : Olahan Penulis (2015)
Gambar 11. Hasil Akurasi Algoritma Support Vector Machine berbasis PSO

Tabel 5 Nilai Accuracy, Sensitivity, Specificity, ppv, dan npv Support Vector Machine berbasis PSO

	Nilai (%)
Accuracy	89,85
Sensitivity	90,81
Specificity	79,62
PPV	97,94
NPV	44,79

Sumber : Olahan Penulis (2015)

grafik ROC dengan nilai AUC (Area Under Curve) untuk algoritma *Support Vector Machine* berbasis PSO sebesar 0.989 dengan tingkat diagnosa *Excellent Classification*



Sumber : Olahan Penulis (2015)

Gambar 11. Nilai AUC Support Vector Machine

berbasis PSO dalam *ROC Curve* Berdasarkan dari analisa pengujian masing-masing algoritma di atas maka dapat dirangkumkan hasilnya sebagai berikut :

Tabel 6. Perbandingan Performance Mode

	C4.5	C4.5 + PSO	SVM	SVM + PSO
Accuracy	93,81%	94,29%	89,85%	96,20%
AUC	0,941	0,936	0,932	0,989

Sumber : Olahan Penulis (2015)

Melihat hasil perbandingan dari seluruh pengujian yang telah dilakukan didapatkan hasil dengan nilai akurasi untuk C4.5 *Accuracy* 93,81 % dan AUC 0,941. Algoritma *Support Vector Machine* dengan nilai akurasi 89,85% dan AUC 0,932, algoritma C4.5 berbasis PSO dengan nilai Akurasi 94,29% dan AUC 0,939 serta Algoritma *Support Vector Machine* berbasis PSO dengan nilai akurasi 96,20% dan AUC 0,989. Dilihat dari dari hasil Akurasi dan nilai AUC ke empat algoritma tersebut, maka dipilihlah algoritma *Support Vector Machine* berbasis Particle Swarm Optimazation sebagai algoritma yang paling akurat karena memiliki nilai akurasi dan ROC paling tinggi diantara algoritma yang lain.

V. KESIMPULAN

Dari penelitian yang telah dilakukan pembuatan menggunakan algortima C4.5, dan *Support Vector Machine* dengan berbasis *Particle Swarm Optimazation* dan menggunakan data perkreditan Bank DKI Jakarta Cabang Bendungan Hilir. Model yang dihasilkan dikomparasi untuk dapat diketahui metode terbaik dalam penentuan kelayakan Kredit.

Dari hasil pengujian dengan mengukur kinerja ketiga metode tersebut menggunakan *confusion matrix*, kurva *ROC* diketahui bahwa C4.5 menghasilkan nilai akurasi 93,81 % dan nilai AUC 0,941. Metode C4.5 berbasis PSO

menghasilkan nilai akurasi 94,29 % dan nilai AUC 0,936. Metode *Support Vector Machine* menghasilkan nilai akurasi 89,85 % dan nilai AUC 0,932. Metode *Support Vector Machine* berbasis PSO menghasilkan nilai akurasi 96,20 % dan nilai AUC 0,989. Nilai akurasi dan AUC tertinggi yaitu pada algoritma *Support Vector Machine* berbasis PSO .

Dengan demikian metode *Support Vector Machine* berbasis *Particle Swarm Optimazation* adalah metode yang terbaik untuk pemecahan masalah kelayakan penentuan kredit Pemilikan Rumah.

Untuk keperluan penelitian lebih lanjut mengenai komparasi metode klasifikasi data mining dengan menggunakan data dibidang kredit perbankan dapat dilakukan dengan :

1. Menggunakan dataset lebih dari satu institusi untuk mengetahui keakuratan dari algoritma yang akan diuji.
2. 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*.

DAFTAR PUSTAKA

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

Aydin, I., Karakose, M., & Akin, E. (2011). A multi-objective artificial immune algorithm for parameter optimization in support vector machine. *Journal Applied Soft Computing*, 11, 120-129.

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.

Edward Burst. (2006). *Credit Management Handbook*. USA: Gower Publishing Company.

Gorunescu, Florin (2011). *Data Mining: Concepts, Models, and Techniques*. Verlag Berlin Heidelberg: Springer.

Han, J., & Kamber, M. (2007). *Data Mining Concepts and Technique*. Morgan Kaufmann publisher.

Hian, C.K., Wei, C.T., & Chwee, P.G (2006). A Two-step Method to Construct Credit Scoring Models with Data Mining

- Techniques. *International Journal of Business and Information*, 1, 96-118.
- Jianguo, Z., & Tao, B. (2008). Credit Risk Assessment using Rough Set Theory and GA-based SVM. The 3rd International Conference on Grid and Pervasive Computing, 320-325.
- Larose, D. T. (2005). *Discovering Knowledge in Data*. New Jersey: John Willey & Sons, Inc.
- Maimon, O., & Rokach, L. (2010). *Data Mining and Knowledge Discovery Handbook*. London: Springer.
- Ning Liu, En Jun Xia, & Li. (2010). Research and Application of PSO-BP Neural Network in Credit Risk Assessment. *International Symposium on Computational Intelligence and Design*, 103-106.
- Olson, D., & Shi, Y. (2008). *Pengantar Ilmu Penggalan Data Bisnis*. Jakarta: Penerbit Salemba Empat.
- Shuzhou, W., & Bo, M. (2011). Parameter Selection Algorithm for Support Vector Machine. *Procedia Environmental Sciences*, 11, 538-544.
- UU Perbankan No.10 Tahun 1998.
- 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.
- Yun, L., Qiu-yan, C. & Hua, Z. (2011). Application of the PSO-SVM model for Credit Scoring. *Seventh International Conference on Computational Intelligence and Security*, 47-51.