

PENERAPAN METODE SVM UNTUK KLASIFIKASI RESIKO KREDIT KEPEMILIKAN KENDARAAN (MOTOR)

Hendra Supendar ¹ Henny Leidiyana ²

Abstract—Given Credit Risk in a lease is always there , So credit analysis should be done carefully . if Consumer troubled in paying installment payments , so would be detrimental to the lease . Therefore , quantitative and qualitative Credit analyzing will provide clarity for decision makers . To achieve this goal , credit preparations should be done by collecting information and data. The quality of the result of the analysis depends on the quality of human resources , the data, and analysis techniques . In this research discussed regarding the application using support vector machine (SVM) method for determination of the Credit Risk of Motor Vehicle ownership . The resulting model is then evaluated its performance by calculating accuracy model prediction using Confusion Matrix and ROC curve, obtained 79.78 % accuracy value and AUC value is 0.894. The classification is Good Enough because have AUC values between from 0.8 to 0.9.

Keywords : Support Vector Machines , Confusion Matrix , ROC curve

Intisari— Mengingat resiko tidak kembalinya kredit pada sebuah leasing, selalu ada, maka analisis kredit dengan cermat perlu dilakukan. Jika konsumen bermasalah dalam pembayaran angsuran, maka akan merugikan pihak leasing. Oleh karena itu, penilaian kuantitatif dan kualitatif dalam menganalisa kredit akan memberikan kejelasan bagi pembuat keputusan. Untuk mewujudkan hal tersebut, perlu dilakukan persiapan kredit, yaitu dengan mengumpulkan informasi dan data untuk bahan analisis. Kualitas hasil analisis tergantung pada kualitas SDM, data yang diperoleh, dan teknik analisis.

Dalam penelitian ini dibahas mengenai penerapan metode Support vector machines (SVM) untuk penentuan resiko kredit kepemilikan kendaraan bermotor. Model yang dihasilkan kemudian dievaluasi performansinya dengan cara menghitung akurasi model dalam memprediksi menggunakan metode Confusion Matrix dan Kurva ROC, didapat nilai akurasi 79.78% dan termasuk klasifikasi cukup baik karena memiliki nilai AUC antara 0.8-0.9, yaitu sebesar 0.894.

Kata kunci: Support Vector Machines, Confusion Matrix, Kurva ROC

¹ Program Studi Teknik Komputer AMIK BSI Jakarta, Jl. RS Fatmawati No. 24 Pondok Labu Jakarta Selatan DKI Jakarta (telp:021-7500282/021-7500680 fax: 021-7513790

² Program Studi Manajemen Informatika AMIK BSI Jakarta, Jl. Kramat Raya No. 18 Jakarta Pusat, Jakarta, Indonesia; email : hendra.hds@bsi.ac.id , henny.hnl@bsi.ac.id

I. PENDAHULUAN

Lembaga keuangan adalah badan usaha yang kekayaannya terutama berbentuk aset keuangan (*financial asset*) atau tagihan (*claims*) dibandingkan dengan aset non keuangan (*non financial asset*). Lembaga keuangan terutama memberikan kredit dan menanamkan dananya dalam surat-surat berharga. Dalam kegiatannya, lembaga keuangan banyak diterpa masalah. Penyebab munculnya permasalahan bukan karena sistem dan perangkat peraturan yang disiapkan oleh Bank Indonesia yang kurang memadai, tetapi lebih banyak dipengaruhi oleh kualitas SDM yang mengelola perkreditan pada lembaga keuangan tersebut [7].

Masalah yang sering terjadi banyak disebabkan karena ulah konsumen, seperti konsumen yang dianggap layak ternyata menunggak angsuran, ada juga yang sampai akhirnya motor harus ditarik oleh perusahaan pemberi kredit, bahkan ada konsumen yang melarikan diri bersama dengan motor kreditannya. Munculnya permasalahan itu salah satunya diakibatkan karena ketidaktajaman analisis *account officer* (analiskredit). *Account officer* tidak melakukan monitor secara berkala serta lemahnya pengawasan kredit setelah konsumen mendapatkan fasilitas, baik pengawasan langsung maupun administratif. Agar kredit yang diberikan mencapai sasaran, yaitu aman, maka analisis kredit dilakukan. Analisis kredit adalah kajian yang dilakukan untuk mengetahui kelayakan dari suatu permasalahan kredit [7].

Untuk membantu mengurangi permasalahan tersebut, maka dalam penelitian ini akan dibuat sebuah model untuk memprediksi apakah calon konsumen akan lancar atau bermasalah dalam pembayaran angsurannya.

II. KAJIAN LITERATUR

Dalam Surat Keputusan Bersama Menteri Keuangan, Perindustrian dan Perdagangan No.1169/KMK.01/1991 tanggal 21 Nopember 1991 tentang kegiatan Sewa Guna Usaha, *Leasing* (Noerlina, 2007) disebutkan setiap kegiatan pembiayaan perusahaan dalam bentuk penyediaan barang-barang modal untuk digunakan oleh suatu perusahaan untuk jangka waktu tertentu, berdasarkan pembayaran-pembayaran berkala disertai dengan hak pilih (opsi) bagi perusahaan tersebut untuk membeli barang-barang modal yang

bersangkutan atau memperpanjang jangka waktu *leasing* berdasarkan nilai sisa yang telah disepakati.

A. Kredit

Kredit adalah penyerahan barang, jasa, atau uang dari satu pihak (kreditur/pemberi pinjaman) atas dasar kepercayaan kepada pihak lain (nasabah atau penguang/*borrower*) dengan janji membayar dari penerima kredit kepada pemberi kredit pada tanggal yang telah disepakati kedua belah pihak [7].

B. Data Mining

Data Mining didefinisikan sebagai proses penemuan pola dalam data [10]. Berdasarkan tugasnya, *data mining* dikelompokkan menjadi Deskripsi, Estimasi, Prediksi, Klasifikasi, klustering, Asosiasi [4]. Proses dalam tahap *data mining* terdiri dari tiga langkah utama, yaitu [8]:

1. Data Preparation

Pada langkah ini, data dipilih, dibersihkan, dan dilakukan *preprocessed* mengikuti pedoman dan pengetahuan (*knowledge*) dari ahli domain yang menangkap dan mengintegrasikan data internal dan eksternal ke dalam tinjauan organisasi secara menyeluruh.

2. Algoritma data mining

Penggunaan algoritma data mining dilakukan pada langkah ini untuk menggali data yang terintegrasi untuk memudahkan identifikasi informasi bernilai.

3. Fase analisa data

Keluaran dari data mining dievaluasi untuk melihat apakah ranah pengetahuan (*knowledge domain*) ditemukan dalam bentuk *rule* yang telah diekstrak dari jaringan.

C. Klasifikasi

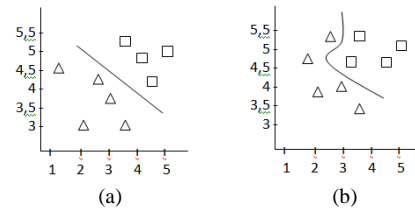
Klasifikasi adalah proses penemuan model (atau fungsi) yang menggambarkan dan membedakan kelas data atau konsep yang bertujuan agar bisa digunakan untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui [3]. Algoritma klasifikasi yang banyak digunakan secara luas, yaitu *Decision/classification trees*, *Bayesian classifiers/ Naïve Bayes classifiers*, *Neural networks*, Analisa Statistik, Algoritma Genetika, *Rough sets*, *k-nearest neighbor*, Metode *Rule Based*, *Memory based reasoning*, dan *Support vector machines (SVM)*.

Klasifikasi data terdiri dari 2 langkah proses. Pertama adalah *learning* (fase *training*), dimana algoritma klasifikasi dibuat untuk menganalisa data *training* lalu direpresentasikan dalam bentuk *rule* klasifikasi. Proses kedua adalah klasifikasi, dimana data tes digunakan untuk memperkirakan akurasi dari *rule* klasifikasi [3].

D. Support Vector Machines (SVM)

Support Vector Machines (SVM) adalah algoritma *supervised learning* yang mengklasifikasikan data yang linier maupun non linier dengan memaksimalkan margin antara *support points* dan sebuah pemetaan non linier untuk

mentransformasikan data training ke dimensi yang lebih tinggi. SVM pertama kali diperkenalkan oleh Vapnik pada tahun 1992.



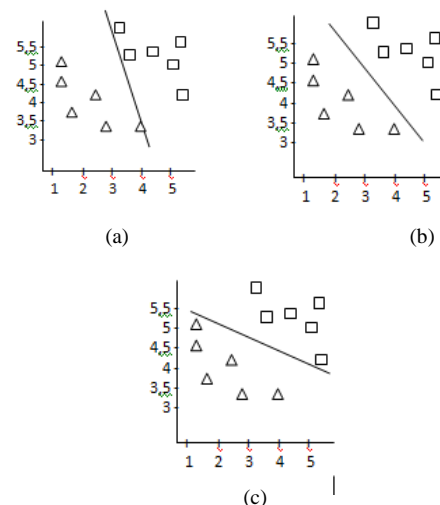
Sumber: Witten et al. (2011)

(a) data terpisah secara linier (b) data terpisah secara non linier

Gambar 1. Ilustrasi kasus algoritma SVM.

Gambar 1 (a) adalah contoh data yang digambarkan dengan grafik dua dimensi yang merepresentasikan dua atribut prediktor. Pada gambar tersebut, garis lurus memisahkan atribut kelas secara linier (*linearly separable*). Sedangkan pada Gambar 1(b) adalah contoh dimana data terpisah tidak linier (*non-linearly separable*).

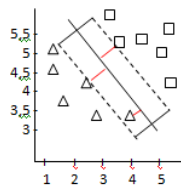
Gambar 2 menunjukkan bagaimana dua kelas terpisah oleh garis pemisah kelas untuk data yang terpisah secara linier. Gambar 2(a), (b), dan (c) menunjukkan bahwa sebuah kasus dapat memiliki beberapa garis pemisah.



Sumber: Witten et al. (2011)

Gambar 2. Garis pemisah kelas untuk data yang terpisah secara linier

Dari sekian garis pemisah kelas yang terbaik adalah garis pemisah dimana ketika ada data baru ditambahkan ke dalam data training maka garis pemisah tersebut tidak perlu diubah. Pemodelan dengan SVM bertujuan memprediksi data baru akan termasuk pada kelas yang mana berdasarkan garis pemisah. Jika garis berubah maka akurasi prediksi bisa berkurang.



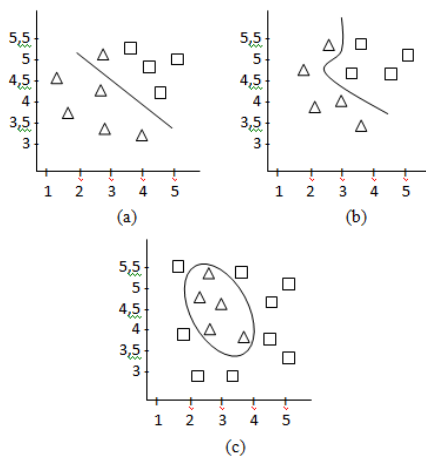
Sumber: Witten et al. (2011)
Gambar 3. Perluasan Garis Pemisah

Karena tidak diketahui garis mana yang paling baik maka yang perlu dilakukan adalah meminimalkan resiko terjadinya perubahan garis. Garis pemisah yang ada diperluas sampai menyentuh titik data terdekat. Lebar perluasan itu disebut dengan *margin*. Garis pemisah dengan *margin* terbesar adalah garis pemisah yang terbaik karena meminimalkan resiko terjadi perubahan garis. Jadi kita berusaha untuk memaksimalkan *margin*. Titik yang pertama kali tersentuh oleh margin disebut *support vector* atau *support point*. Rumus untuk menentukan garis pemisah pada data yang terpisah secara linier adalah:

$$\min_{w,b} \frac{1}{2} w^T w$$

$$y_i (w^T x_i + b) \geq 1$$

Pada Gambar 4 terlihat tiga contoh kasus data non linier. Pada Gambar 4(a) terdapat titik yang berada di kelas yang salah. Pada Gambar (b) dan (c) garis pemisah tidak linier.



Sumber: Witten et al. (2011)
Gambar 4. Data terpisah secara non linier

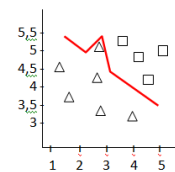
Gambar 4(a) hampir terpisah secara linier tetapi ada beberapa poin yang berada di sisi kelas atau margin yang salah. Untuk kasus ini perbaikan bisa dilakukan dengan variabel *slack*, menggunakan rumus:

$$\min_{w,b} \frac{1}{2} w^T w + \lambda \sum_{i=1}^n s_i$$

$$y_i (w^T x_i + b) + s_i \geq 1 \text{ untuk } \forall i$$

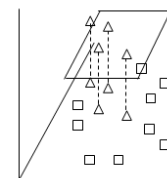
$$s_i \geq 0$$

Variabel *slack* akan memperbaiki pemisah akibat dari adanya poin yang masuk ke kelas yang salah (*misclassification*). Variabel *slack* bernilai seminimal mungkin sementara lebar margin semaksimal mungkin. Hal ini merupakan kontradiksi dan untuk mengatasinya yaitu dengan regularisasi menggunakan parameter λ . Gambar 5 merupakan perbaikan margin dengan variabel *slack*.



Sumber: Witten et al. (2011)
Gambar 5. Setelah perbaikan dengan variabel *slack*

Gambar 4(b) dan Gambar 4(c) jelas terpisah secara tidak linier. Untuk kedua kasus ini perbaikannya menggunakan *kernel trick* untuk mengubah pemisah kelas yang tidak linier menjadi linier pada dimensi yang lebih tinggi. Untuk data yang tidak terpisah secara linier digunakan kernel. Kernel dan fungsi pemetaan (*mapping function*) akan memproyeksikan data x ke dimensi yang lebih tinggi sehingga hasilnya tidak lagi non linier seperti pada Gambar 5.



Sumber: Witten et al. (2011)
Gambar 5. Memproyeksikan data ke dimensi yang lebih tinggi

Fungsi untuk memproyeksikan data ke dimensi yang lebih tinggi disebut fungsi pemetaan (*mapping function*). *Kernel trick* artinya fungsi pemetaan tidak perlu diketahui, yang penting adalah cara untuk menghitung fungsi pemetaan dari kernel. Kernel adalah pangkat dari fungsi pemetaan $\phi(x)$ atau dapat ditulis

$$K(x,x) = \phi^T(x) \cdot \phi(x)$$

Secara umum kernel dari vektor u dan v ditulis sebagai

$$K(\mathbf{u}, \mathbf{v}) = \phi^T(\mathbf{u}) \cdot \phi(\mathbf{v})$$

Kernel dapat dibuat dengan menciptakan sebuah fungsi Kernel dengan perkalian, penjumlahan, atau penskalaan dengan sebuah konstanta. Contoh membuat fungsi Kernel:

$$p = aK(\mathbf{u}, \mathbf{v}) + bK(\mathbf{u}, \mathbf{v}) \text{ dan} \\ q = cK(\mathbf{u}, \mathbf{v}) + dK(\mathbf{u}, \mathbf{v})$$

Rumus kernel yang umum dipakai dalam SVM adalah:

Tabel 1. Rumus Kernel

Nama Kernel	Rumus	Parameter
Polynomial	$K(\mathbf{u}, \mathbf{v}) = (a\mathbf{u}^T \mathbf{v} + b)^d$	$a, b, d >= 2$
Gaussian	$K(\mathbf{u}, \mathbf{v}) = \exp(-a/\ \mathbf{u}-\mathbf{v}\ ^2)$	a
Sigmoid	$K(\mathbf{u}, \mathbf{v}) = \tanh(a\mathbf{u}^T \mathbf{v} + b)$	a, b

Sumber: Witten et al. (2011)

E. Pengujian Model

Untuk mengukur akurasi model maka dilakukan evaluasi menggunakan teknik [10]:

1. Cross Validation

Dalam penelitian ini digunakan *10 fold-cross validation* dimana data training dibagi secara random kedalam 10 bagian dengan perbandingan yang sama kemudian *error rate* dihitung bagian demi bagian, selanjutnya hitung rata-rata seluruh *error rate* untuk mendapatkan *error rate* secara keseluruhan.

2. Confusion Matrix

Metode ini menggunakan table matriks seperti pada Tabel 2, jika data set hanya terdiri dari dua kelas, kelas yang satu dianggap sebagai positif dan yang lainnya negatif [1].

Tabel 2 Model confusion matrix

Klasifikasi yang benar	Diklasifikasikan sebagai	
	+	-
+	true positives	false negatives
-	false positives	true negatives

Sumber: Bramer (2006)

True positives adalah jumlah *record* positif yang diklasifikasikan sebagai positif, *false positives* adalah jumlah *record* negatif yang diklasifikasikan sebagai positif, *false negatives* adalah jumlah *record* positif yang diklasifikasikan sebagai negatif, *true negatives* adalah jumlah *record* negatif yang diklasifikasikan sebagai negatif. Kemudian masukkan data uji kedalam model *confusion matrix* maka akan

didapatkan hasil. Setelah data uji dimasukkan kedalam *confusion matrix* lalu dihitung nilai-nilai yang telah dimasukkan tersebut untuk dihitung jumlah *sensitivity (recall)*, *specificity*, *precision* dan *accuracy*. *Sensitivity* digunakan untuk membandingkan jumlah TP terhadap jumlah *record* yang positif sedangkan *specificity* adalah perbandingan jumlah TN terhadap jumlah *record* yang negatif. Untuk menghitung digunakan persamaan di bawah ini: (Han & Kamber, 2006)

$$sensitivity = \frac{TP}{P}$$

$$specificity = \frac{TN}{N}$$

$$precision = \frac{TP}{TP + FP}$$

$$accuracy = sensitivity \frac{P}{(P + N)} + specificity \frac{N}{(P + N)}$$

Keterangan:

- TP = jumlah *true positives*
- TN = jumlah *true negatives*
- P = jumlah *record* positif
- N = jumlah *tupel* negatif
- FP = jumlah *false positives*

3. Kurva ROC

Kurva ROC menunjukkan akurasi dan membandingkan klasifikasi secara visual. ROC mengekspresikan *confusion matrix*. ROC adalah grafik dua dimensi dengan *false positives* sebagai garis horizontal dan *true positives* sebagai garis vertikal [9]. *The area under curve (AUC)* dihitung untuk mengukur perbedaan performansi metode yang digunakan. AUC dihitung menggunakan rumus: [5]

$$\theta^r = \frac{1}{mn} \sum_{j=1}^n \sum_{i=1}^m \psi(x_i^r, x_j^r)$$

Dimana

$$\psi(X, Y) = \begin{cases} 1 & Y < X \\ \frac{1}{2} & Y = X \\ 0 & Y > X \end{cases}$$

III. METODE PENELITIAN

Metode penelitian yang digunakan yaitu model cross standard industry for data mining (CRISP DM) yang terdiri dari enam fase yaitu 1). Business understanding, 2). Data Understanding, 3). Data Preparation, 4). Modeling, 5). Evaluation, 6). Deployment.

IV. HASIL DAN PEMBAHASAN

Untuk penelitian ini data yang digunakan sebanyak 487 *record* transaksi kredit baik yang bermasalah maupun yang tidak bermasalah. Data tersebut diperoleh dari sebuah *leasing* yang berlokasi di Cikarang. Semua atribut pada data *training* bernilai kategori yang terdiri dari 14 atribut, dimana 13 atribut merupakan prediktor dan 1 atribut label, yaitu remark. Informasi mengenai atribut dan nilainya dapat dilihat pada Tabel 3. Untuk atribut remark bernilai *good* atau *bad*. Nilai *good* artinya konsumen tersebut sudah pernah mendapat kredit motor dan pembayarannya lancar sampai lunas. Sedangkan *bad* artinya artinya konsumen tersebut sudah pernah mendapat kredit motor dan pembayarannya tidak lancar. Sehingga data ini merupakan klasifikasi dengan dua kelas yaitu *good* atau *bad*.

Tabel 3. Daftar Atribut dan Nilainya

no	Atribut	Nilai atribut
1	status perkawinan	Menikah belum menikah janda/duda
2	jumlah tanggungan	tidak ada 1 orang 2-3 orang > 3 orang
3	pendidikan terakhir	>S1 S1 Diploma SLTA SLTP SD tidak sekolah
4	Usia	< 21 tahun atau > 60 tahun 21 - 55 tahun 55 - 60 tahun
5	kepemilikan rumah	milik sendiri (PBB/srtfkt/AJB/rek listrik) milik sendiri (PBB a/n orang lain) KPR orang tua Keluarga dinas sewa/kontrak >= tenor

Tabel 3. Daftar Atribut dan Nilainya (lanjutan)

no	Atribut	Nilai atribut
		sewa/kontrak < tenor Kost
6	lama tinggal	> 5 tahun 3 - 5 tahun 1 - 3 tahun < 1 tahun
7	kondisi rumah	Permanen non permanen
8	jenis pekerjaan	PNS TNI/POLRI Jaksa Karyawan wiraswasta kecil wiraswasta menengah wiraswasta besar Konsultan Dokter Dosen Guru Pengacara Pensiunan
9	status perusahaan	BUMN/D swasta besar swasta menengah swasta kecil perorangan lembaga pendidikan lembaga negara
10	status kepegawaian	tetap kontrak harian buruh pabrik harian buruh non pabrik pemilik
11	Masa kerja/usaha	> 5 tahun 2 - 5 tahun < 2 tahun
12	penghasilan perbulan	> 3x angsuran dengan slip > 3x angsuran tanpa slip

Tabel 3. Daftar Atribut dan Nilainya (lanjutan)

Tabel 4. Konversi Nilai Atribut

no	Atribut	Nilai atribut
		> 2x angsuran dengan slip
		> 2x angsuran tanpa slip
		> 1x angsuran dengan slip
		> 1x angsuran tanpa slip
		< 1x angsuran
13	pembayaran pertama	> 30 %
		20 - 30%
		10 - 20 %
		< 10 %
14	Remark	bad cust
		Good cust

Sumber: Perusahaan Leasing Cabang Cikarang

Berikut adalah sebuah sampel data konsumen yang pernah mengambil kredit motor:

1. Status perkawinan : menikah
2. Jumlah tanggungan : tidak ada
3. Pendidikan terakhir : SLTA
4. Usia : 21-55
5. Kepemilikan rumah : KPR
6. Lama tinggal : 3-5 tahun
7. Kondisi rumah : permanen
8. Jenis pekerjaan : karyawan
9. Status perusahaan : swasta
10. Status kepegawaian : tetap
11. Masa kerja : >5 tahun
12. Penghasilan perbulan : 10-20%
13. Total pembayaran pertama : > 3x ang slip
14. Remark : good

Data yang diperoleh dijadikan data training untuk membuat model menggunakan algoritma *Support Vector Machines* (SVM). Model yang dibuat adalah model untuk memprediksi apakah seorang pemohon kredit layak mendapatkan kredit dan tidak akan bermasalah dalam pembayaran kreditnya.

Untuk mendapatkan data yang berkualitas, dilakukan *preprocessing*. Setelah dilakukan *preprocessing*, 487 record direduksi dengan menghilangkan duplikasi.

Pada metode SVM data harus berupa numerik sementara data training adalah nominal maka data nominal tersebut diubah ke dalam numerik. Tabel 4 berisi konversi nilai atribut yang berupa nominal yang diubah menjadi numerik. Setiap nilai diberikan angka numerik yang berbeda berupa bilangan bulat mulai dari nol.

no	Atribut	Nilai atribut	Nilai Numerik
1	status perkawinan	Menikah	0
		belum menikah	1
		janda/duda	2
2	jumlah tanggungan	tidak ada	0
		1 orang	1
		2-3 orang	2
		> 3 orang	3
3	pendidikan terakhir	>S1	0
		S1	1
		Diploma	2
		SLTA	3
		SLTP	4
		SD	5
		tidak sekolah	6
4	Usia	< 21 tahun atau > 60 tahun	0
		21 - 55 tahun	1
		55 - 60 tahun	2
5	kepemilikan rumah	milik sendiri (PBB/srtfkt/AJB/rek listrik)	0
		milik sendiri (PBB a/n orang lain)	1
		KPR	2
		orang tua	3
		Keluarga	4
		dinas	5
		sewa/kontrak >= tenor	6
		sewa/kontrak < tenor	7
		Kost	8
6	lama tinggal	> 5 tahun	0
		3 - 5 tahun	1
		1 - 3 tahun	2
		< 1 tahun	3
7	kondisi rumah	Permanen	0
		non permanen	1
8	jenis pekerjaan	PNS	0
		TNI/POLRI	1
		Jaksa	2
		Karyawan	3
		wiraswasta kecil	4
		wiraswasta menengah	5

Tabel 4. Konversi Nilai Atribut (lanjutan)

no	Atribut	Nilai atribut	Nilai Numerik
		Konsultan	7
		Dokter	8
		Dosen	9
		Guru	10
		Pengacara	11
		Pensiunan	12
9	status perusahaan	BUMN/D	0
		swasta besar	1
		swasta menengah	2
		swasta kecil	3
		perorangan	4
		lembaga pendidikan	5
		lembaga negara	6
10	status kepegawaian	tetap	0
		kontrak	1
		harian buruh pabrik	2
		harian buruh non pabrik	3
		pemilik	4
11	Masa kerja/usaha	> 5 tahun	0
		2 - 5 tahun	1
		< 2 tahun	2
12	penghasilan perbulan	> 3x angsuran dengan slip	0
		> 3x angsuran tanpa slip	1
		> 2x angsuran dengan slip	2
		> 2x angsuran tanpa slip	3
		> 1x angsuran dengan slip	4
		> 1x angsuran tanpa slip	5
		< 1x angsuran	6
13	pembayaran pertama	> 30 %	0
		20 - 30%	1
		10 - 20 %	2
		< 10 %	3
14	Remark	bad cust	-1
		Good cust	1

Sumber: Perusahaan Leasing Cabang Cikarang

Karena data training adalah terpisah secara non linier maka (*non-linearly separable*) maka pembuatan model menggunakan metode SVM dengan *kernel trick*. Seperti yang

telah dijelaskan pada bagian sebelumnya ada beberapa metode kernel yang umum digunakan diantaranya polinomial, *gaussian*, dan *sigmoid*. Kernel yang digunakan pada penelitian ini adalah polinomial dengan rumus:

$$K(\mathbf{u}, \mathbf{v}) = (a\mathbf{u}^T \mathbf{v} + b)^d$$

Mula-mula data dimasukkan ke dalam kernel. Misalkan untuk kasus satu dimensi (data dengan satu variabel atribut) maka vektor $\mathbf{u} = \mathbf{v} = [x]$ menggunakan kernel polinomial derajat 2. Perkalian titik adalah pangkat dari variabel tersebut:

$$K(x, x) = (x \cdot x + 1)^2$$

Fungsi kernel diperluas menjadi

$$K(x, x) = (x^2 + 1)^2 = 1 + 2x^2 + x^4 = \phi^T(x) \cdot \phi(x)$$

Kemudian ambil akar pangkat untuk tiap suku kernel untuk mendapatkan fungsi pemetaan

$$\begin{aligned} \sqrt{1} &= 1 \\ \sqrt{2x^2} &= x\sqrt{2} \\ \sqrt{x^4} &= x^2 \end{aligned}$$

Sehingga fungsi pemetaan adalah vektor dengan tiga elemen

$$\phi^T(x) = [1 \quad x\sqrt{2} \quad x^2]$$

$$\begin{aligned} \phi^T(x) \cdot \phi(x) &= [1 \quad x\sqrt{2} \quad x^2] \begin{pmatrix} 1 \\ x\sqrt{2} \\ x^2 \end{pmatrix} \\ &= 1 + 2x^2 + x^4 \end{aligned}$$

Untuk kasus data dengan dua variabel atribut maka vektor $\mathbf{u}^T = [u_1, u_2]$ dan $\mathbf{v}^T = [v_1, v_2]$ sama seperti contoh kasus satu variabel sebelumnya menggunakan kernel polinomial derajat 2. Perkalian titik adalah pangkat dari variabel tersebut:

$$K(\mathbf{u}, \mathbf{v}) = \left([u_1 \quad u_2] \cdot \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} + 1 \right)^2 = (1 + u_1 \cdot v_1 + u_2 \cdot v_2)^2$$

Fungsi kernel diperluas menjadi

$$K(u, v) = (1 + 2u_1 \cdot v_1 + 2u_2 \cdot v_2 + (u_1 \cdot v_1)^2 + (u_1 \cdot v_1)^2 + 2u_1 v_1 u_2 v_2)$$

Kemudian ambil akar pangkat vektor \mathbf{u} untuk tiap suku kernel untuk mendapatkan fungsi pemetaan

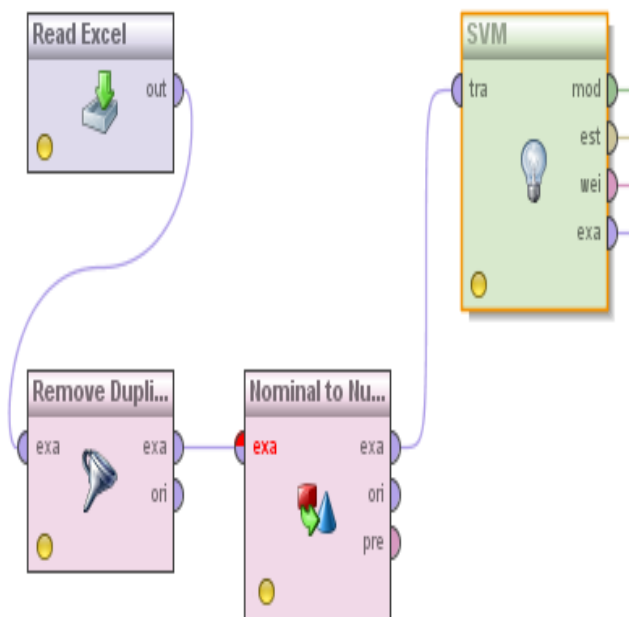
$$\phi^T(\mathbf{u}) = [1 \quad u_1\sqrt{2} \quad u_2\sqrt{2} \quad u_1^2 \quad u_2^2 \quad u_1u_2\sqrt{2}]$$

Dan akar pangkat vektor \mathbf{v} untuk tiap suku kernel untuk mendapatkan fungsi pemetaan

$$\phi^T(\mathbf{v}) = [1 \quad v_1\sqrt{2} \quad v_2\sqrt{2} \quad v_1^2 \quad v_2^2 \quad v_1v_2\sqrt{2}]$$

$$\phi^T(x) \cdot \phi(x) = K(\mathbf{u}, \mathbf{v})$$

Untuk data penelitian yang terdiri dari 13 atribut prediktor perhitungan dilakukan menggunakan aplikasi rapidminer, mulai dari *preprocessing*, modeling, sampai pada pengujian model.



Sumber: Percobaan menggunakan Rapidminer

Gambar 6. Pemodelan dengan rapidminer

Data training dimasukan ke dalam aplikasi menggunakan operator data *Read Excel*, karena data training disimpan dalam aplikasi MS Excel. Untuk menghilangkan duplikasi maka dilakukan penghilangan data duplikasi menggunakan operator *Remove Duplicate*. Karena data training yang diperoleh nilai-nilainya berupa data nominal maka ditransformasikan ke dalam bentuk nilai numerik menggunakan operator konversi tipe *nominal to Numerical*.

Tabel 5. Data setelah dikonversi ke dalam numerik

Row. No	remark	status	tanggungan	pendidikan	usia	...
1	Bad	0	1	1	0	
2	Bad	1	0	0	0	
3	Good	1	1	0	1	
4	Good	0	0	1	0	

Sumber: Percobaan menggunakan Rapidminer

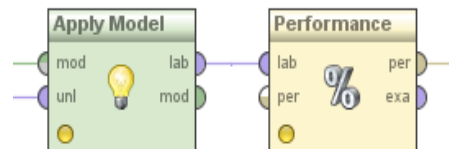
Pembuatan model menggunakan SVM dengan kernel polinomial derajat 2. Nilai bias dan bobot yang diperoleh menggunakan aplikasi rapidminer:

Kernel Model

Total number of Support Vectors: 460
Bias (offset): -0.366

- w[status] = 0.114
- w[tanggungan] = -0.141
- w[pendidikan] = 0.208
- w[usia] = -0.051
- w[kepemilikan rumah] = -0.000
- w[lama tinggal] = -0.000
- w[kondisi rumah] = 0.116
- w[pekerjaan] = 0.471
- w[status perusahaan] = -0.000
- w[kepegawaian] = 0.883
- w[masa kerja] = 1.367
- w[penghasilan] = -0.000
- w[dp] = 0.828

Setelah pembuatan model maka dilakukan pengujian kinerja modelnya menggunakan metode *10-fold cross validation*.



Sumber: Percobaan menggunakan Rapidminer

Gambar 7. Pengujian Metode SVM

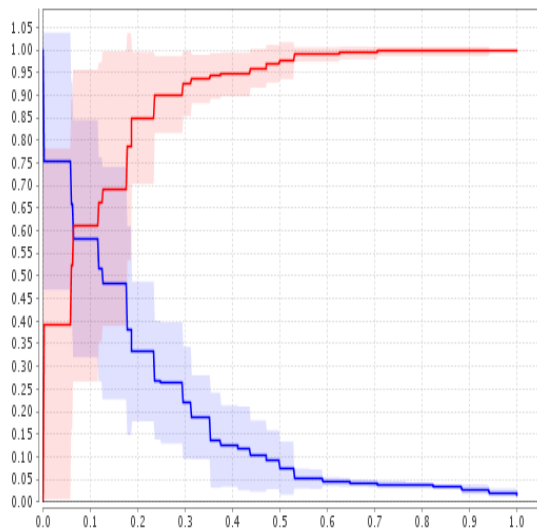
Dari pengolahan didapat nilai akurasi model sebesar 79.78% seperti terlihat pada Tabel 6.

Tabel 6. Model *Confusion Matrix* untuk Metode SVM

Accuracy: 79.78%			
	True Bad	True Good	Class Precission
Pred. Bad	148	74	66.67%
Pred. Good	19	219	92.02%
	88.62%	74.74%	

Sumber: Hasil Pengujian menggunakan Rapidminer

Model yang dihasilkan termasuk klasifikasi cukup baik karena memiliki nilai AUC antara 0.8-0.9, yaitu sebesar 0.894 seperti diperlihatkan pada Gambar 8.



Sumber: Hasil Pengujian menggunakan Rapidminer

Gambar 8. Kurva ROC dengan Metode SVM

V. KESIMPULAN

Dalam penelitian ini dilakukan penerapan algoritma SVM pada data konsumen yang mendapat pembiayaan kredit motor. Agar didapat data yang berkualitas, dilakukan *preprocessing* sebelum diterapkan ke dalam algoritma. Selanjutnya data training yang atributnya bernilai nominal diubah ke dalam numerik. Karena data tidak terpisah secara linier maka pembuatan model SVM menggunakan kernel trick, yaitu polinomial. Untuk mengukur kinerja algoritma tersebut digunakan metode *Cross Validation*, *Confusion Matrix* dan Kurva ROC, didapat nilai akurasi 79.78% dan termasuk klasifikasi cukup baik karena memiliki nilai AUC antara 0.8-0.9, yaitu sebesar 0.894.

REFERENSI

- [1] Bramer, Max. *Principles of Data Mining*. London: Springer. 2007.
- [2] Gorunescu, Florin. *Data Mining: Concepts, Models, and Techniques*. Verlag Berlin Heidelberg: Springer. 2011.
- [3] Han, J., & Kamber, M. *Data Mining Concept and Tehniques*. San Fransisco: Morgan Kauffman. 2006.
- [4] Larose, D. T. *Discovering Knowledge in Data*. New Jersey: John Willey & Sons, Inc. 2005
- [5] Liao. *Recent Advances in Data Mining of Enterprise Data: Algorithms and Application*. Singapore: World Scientific Publishing. 2007.
- [6] Maimon, Oded & Rokach, Lior. *Data Mining and Knowledge Discovey Handbook*. New York: Springer. 2005.

- [7] Rivai, Veithzal, Andria Permata Veithzal. *Credit Management Handbook*. Jakarta: Raja Grafindo Persada. 2006.
- [8] Sumathi, S., Sivanandam, S.N. *Introduction to Data Mining and its Applications*. Berlin Heidelberg New York: Springer. 2006.
- [9] Vercellis, Carlo. *Business Intelligent: Data Mining and Optimization for Decision Making*. Southern Gate, Chichester, West Sussex: John Willey & Sons, Ltd. 2009
- [10] Witten, I. H., Frank, E., Mark A. Hall. *Data Mining: Practical Machine Learning and Tools*. Burlington: Morgan Kaufmann Publisher. 2011.



Hendra Supendar, M.Kom. Tahun 1998 lulus dari Program Strata Satu (S1) Program Studi Teknik Informatika Universitas Persada Indonesia YAI. Tahun 2010 lulus dari Program Strata Dua (S2) Program Studi Magister Ilmu Komputer STMIK Nusa Mandiri Jakarta. Jabatan Fungsional Akademik Lektor di AMIK BSI Jakarta. Aktif mengikuti beberapa seminar Nasional dan telah menulis di beberapa jurnal diantaranya Jurnal Techno STMIK Nusa Mandiri, Jurnal Paradigma AMIK BSI Jakarta, Jurnal Digit STMIK CIC Cirebon dan Jurnal Format Universitas Mercubuana Jakarta



Henny Leidiyana, M.Kom. Tahun 1998 lulus dari Program Strata Satu (S1) Program Studi Teknik Informatika Universitas Persada Indonesia YAI. Tahun 2011 lulus dari Program Strata Dua (S2) Program Studi Magister Ilmu Komputer STMIK Nusa Mandiri Jakarta. Jabatan Fungsional Akademik Asisten Ahli di AMIK BSI Jakarta. Telah menulis beberapa paper di Jurnal Pikel Universitas Islam 45 Bekasi.