

## Prediksi Calon Nasabah Gadai Potensial pada PT. Pegadaian (Persero) dengan Menggunakan Metode Suport Vector Machine-Sequential Minimal Optimization (SVM-SMO)

### *Predicion of Potnetial Pawn Customer's in PT. Pegadaian Persero by Using Suport Vector Machine-Sequetnial Minimal Optimization (SVM-SMO) Method*

Ramadhan Wahyu Pratama<sup>1</sup>, Arie Ardyanti Suryani, ST., MT.<sup>2</sup>, Siti Sa'adah, ST., MT.<sup>3</sup>

<sup>1,2,3</sup> Jurusan Teknik Informatika, Institut Teknologi Telkom  
Jl. Telekomunikasi, Dayeuh Kolot Bandung 40257 Indonesia

[ramadhannow@gmail.com](mailto:ramadhannow@gmail.com), [arie006@yahoo.com](mailto:arie006@yahoo.com), [sitisaadah@gmail.com](mailto:sitisaadah@gmail.com)

#### ABSTRAKSI

Salah satu tantangan yang dihadapi oleh perusahaan jasa gadai terutama PT. Pegadaian adalah usaha menurunkan barang lelang milik nasabah akibat nasabah tersebut tidak melakukan pencicilan pinjaman sebelum jatuh tempo berakhir. Jenis nasabah yang tidak melakukan usaha untuk menebus barang gadai mereka dapat dikatakan sebagai seorang nasabah yang tidak potensial bagi perusahaan. Untuk itu maka perusahaan perlu mengetahui jenis nasabah mereka apakah termasuk dalam kategori potensial atau tidak, informasi ini berguna untuk mengantisipasi kemungkinan kemunculan nasabah lelang yang akan merugikan perusahaan. Penentuan nasabah potensial dan tidak potensial ini dapat dilakukan pada tahap klasifikasi dengan menggunakan metode *Support Vector Machine Sequential Minimal Optimization*.

Tugas akhir ini bertujuan untuk menerapkan Menerapkan *Support Vector Machine Sequential Minimal Optimization* (SVM-SMO) untuk memprediksi nasabah potensial dan tidak potensial. Pemilihan metode SVM dalam tugas akhir ini karena SVM telah terbukti kehandalannya dalam melakukan klasifikasi data dalam jumlah yang besar dan memiliki atribut data yang kompleks. Sebelum melalui tahap klasifikasi terlebih dahulu data akan *dipreprocessing* dan dinormalisasi dengan menggunakan normalisasi linear. Terdapat tiga parameter uji yang digunakan sebagai evaluasi sistem yaitu *Precision*, *Recall* dan *F-measure*, dengan hasil rata-rata setiap parameter uji bernilai diatas 75%.

**Kata Kunci:** Nasabah Gadai, *Support Vector Machine Sequential Minimal Optimization*

#### ABSTRACT

Nowadays, The challenges Ade by PT. Pegadaian Company is do decrease The Action intens belonging to the customers as a Renault of these customers do not the pay the duty dua date expires. Types of customers who do not make an effort to redeem their pledges can be considered as a potential customer who is not for the company. For those reasons, companies need to know the type of customer they are included in the category of potential or not, this information is useful to anticipate the possibility of the emergence of the auction customers will hurt the company. Determination of potential customers and potential can be performed on stage classification using Support Vector Machine Sequential Minimal Optimization.

This final project aims to applying Support Vector Machine Sequential Minimal Optimization (SMO-SVM) to predict potential customers and potential. SVM method has been selected in this final project because SVM has proven reliability in the classification of data in large quantities and data with a complex attributes. Before going through the first stage of classification, the data will be preprocessing and normalized by using linear normalization. There are three parameters that are used as an evaluation test of the system, namely precision, recall and F-measure, with the average yield of each parameter is above 75%.

**Key Word :** *Pawn Customer, Suport Vector Machine Sequential Minimal Optimization*

#### 1. Pendahuluan

##### 1.1 Latar Belakang

PT. Pegadaian (Persero) adalah sebuah badan usaha milik negara yang menyediakan layanan jasa gadai yang berbadan hukum dengan 90% dari jumlah nasabah keseluruhan berupa nasabah gadai emas[6]. Seiring Dengan semakin banyaknya jumlah nasabah gadai emas yang dimiliki oleh perusahaan, maka semakin banyak pula tantangan yang harus dihadapi oleh perusahaan. Salah satu tantangan yang ada hingga kini dan belum memiliki solusi yang nyata adalah masih banyaknya nasabah yang tidak memenuhi kewajiban mereka untuk melakukan pencicilan kredit ataupun bunga di tiap

periode waktu jatuh tempo. Dari data yang diperoleh pada tahun 2010, jumlah nasabah yang tidak melakukan kewajibannya sebanyak 179 orang nasabah[9], kemudian pada tahun 2011, terdapat 124 orang nasabah yang tidak melakukan kewajibannya melakukan pencicilan kredit[9], dan terakhir pada tahun 2012, jumlah nasabah yang merugikan perusahaan ini bertambah menjadi 199 orang[9]. Hal ini tentu saja ini berdampak bagi laba perusahaan, karena semakin banyak kredit yang tidak dicicil oleh nasabah maka semakin banyak barang lelang yang menumpuk di dalam kas lelang. Oleh karena itu diperlukan sebuah cara untuk dapat mengetahui nasabah jenis yang potensial bagi perusahaan dan tidak potensial

atau merugikan perusahaan. Metode klasifikasi dapat digunakan untuk memprediksi masalah ini, yaitu dengan mencari sekumpulan model atau fungsi yang membedakan masing-masing kelas nasabah.

Pada penelitian tugas akhir ini metode klasifikasi yang digunakan adalah *Support Vector Machine Sequential Minimal Optimization*. Metode ini dipilih karena, jika dibandingkan dengan metode k-mean, *artificial network* dan naive bayes, dalam proses pembelajarannya dapat menghasilkan hasil klasifikasi yang berbeda-beda. Sedangkan pada algoritma yang terdapat *Support Vector Machine* hasil klasifikasi yang dihasilkan selalu tetap, karena pendekatan klasifikasi yang dilakukan oleh SVM tidak mengukur jarak dari masing-masing *Vector*, melainkan mencari beberapa *Support Vector* terbaik, lalu mencari margin antar *Support Vector*nya. Selain itu juga pemilihan *Support Vector Machine Sequential Minimal Optimization* ini dikarenakan dalam penanganan *quadratic programming*, SMO jauh lebih baik karena mampu meminimalkan waktu komputasi untuk data yang sangat besar[1].

Dengan menggunakan metode ini maka prediksi dapat dilakukan dengan mempelajari model dataset yang telah dikategorikan dalam proses pelatihan kedalam dua kelas berbeda yaitu kelas nasabah yang potensial dan nasabah yang tidak potensial[1].

## 1.2 Perumusan Masalah

Permasalahan yang akan diangkat adalah bagaimana melakukan prediksi nasabah yang terdiri dari nasabah potensial dan tidak dengan menggunakan *Support Vector Machine Sequential Minimal Optimization*, serta mengevaluasi performansi dari sistem dalam memberikan hasil prediksi yang diperoleh dari evaluasi parameter-parameter pada metode tersebut.

## 1.3 Tujuan

Tujuan yang ingin dicapai dalam pengerjaan Tugas Akhir ini adalah sebagai berikut:

1. Menerapkan *Support Vector Machine Sequential Minimal Optimization* (SVM-SMO) untuk memprediksi nasabah potensial dan tidak potensial.
2. Mengevaluasi *Precision*, *Recall* dan *F-measure* yang dihasilkan oleh *Support Vector Machine Sequential Minimal Optimization* dalam memberikan hasil prediksi terhadap studi kasus ini.

## 1.4 Batasan Masalah

Adapun batasan masalah dalam pengerjaan tugas akhir ini adalah:

1. Data yang digunakan sebagai bahan Analisis yaitu dokumen surat bukti kredit dan database nasabah yang berasal sebuah kantor cabang di PT. Pegadaian (Persero) Unit Pelayanan Cabang Pegadaian Seririt-Bali.
2. Data surat bukti kredit yang digunakan adalah data kredit dalam kurun waktu 2010-2012.

## 2. Landasan Teori

### 2.1 Nasabah Gadai

Nasabah Gadai merupakan orang yang menjaminkan barang mereka kepada perusahaan untuk mendapatkan uang tunai sesuai dengan taksiran barang yang akan mereka gadai[10]. Dalam studi kasus ini, jenis nasabah

gadai yang digunakan adalah nasabah yang menggadaikan barang mereka ke perusahaan dalam wujud emas. Siapapun dapat menjadi nasabah gadai pegadaian, selama orang tersebut memiliki kartu tanda penduduk dan barang yang akan digadaikan. Seorang nasabah gadai dikatakan potensial apabila nasabah tersebut terus memberikan kontribusi terhadap perusahaan, kontribusi tersebut bisa berupa banyaknya barang jaminan yang digadaikan ke perusahaan, nasabah tersebut selalu mencicil uang kredit pinjaman yang mereka terima tiap bulan atau empat bulan sekali. Sedangkan nasabah yang tidak melunasi uang pinjaman mereka sehingga barang yang dijaminkan dilelang dapat dikategorikan sebagai nasabah tidak potensial. Pihak perusahaan belum memiliki aturan atau sistem yang pasti dalam menentukan nasabah potensial atau tidak tersebut, sehingga hal ini yang mendasari penulis untuk mengangkat kasus ini menjadi bahan penelitiannya[10].

Setiap nasabah gadai yang memperoleh pinjaman akan mendapatkan suatu Surat Bukti Kredit atau SBK, dimana kredit gadai yang diberikan oleh pegadaian disini adalah suatu pemberian pinjaman (kredit) dalam jangka waktu tertentu kepada nasabah atas dasar hukum gadai dan persyaratan tertentu yang telah ditetapkan oleh perusahaan. Nasabah wajib menyelesaikan pinjamannya kepada perusahaan (Pegadaian) sebagai pemberi pinjaman (kreditur), dengan cara mengembalikan uang pinjaman dan membayar sewa modalnya berdasarkan ketentuan yang berlaku. Pelunasan kredit yang dapat dilakukan oleh pihak nasabah antara lain melalui, melunasi dengan membayar pokok pinjaman ditambah dengan sewa modal pada saat jatuh tempo atau pelunasan. Serta dapat juga melalui hasil penjualan lelang barang jaminannya. Dimana sisa hasil penjualan lelang barang jaminan (harga lelang dikurangi pokok pinjaman dan sewa modal) dikembalikan kepada nasabah. Apabila hasil penjualan lelang tidak mencukupi pokok pinjaman dan sewa modalnya, maka kekurangannya tetap menjadi kewajiban nasabah.[10]

Jenis barang-barang yang diterima sebagai barang jaminan gadai antara lain, kain seperti bahan pakaian, kain permadani, barang perhiasaan seperti emas, perak, berlian dan batu mulia, kendaraan seperti mobil, sepeda motor, dan sepeda, dan barang rumah tangga, seperti perabotan rumah tangga, dan elektronik.[10].

### 2.1.1 Perolehan Kredit Gadai oleh Nasabah

Untuk apa memperoleh kredit pinjaman dari pihak perusahaan terdapat syarat-syarat yang harus dipenuhi oleh pihak nasabah, diantaranya:

1. Nasabah gadai wajib mengisi Formulir Permintaan Kredit (FPK) dan menandatangani.
2. Nasabah gadai wajib menyerahkan foto copy kartu identitas (KTP, SIM, paspor atau sejenisnya) yang masih berlaku, dengan menunjukkan aslinya.
3. Menyerahkan barang jaminan
4. Menandatangani perjanjian hutang piutang dengan jaminan gadai pada Surat Bukti Kredit (SBK)
5. Membayar biaya Administrasi (BA)
6. Khusus untuk Barang Jaminan (BJ) kendaraan bermotor dilengkapi dengan persyaratan lainnya yang diatur dalam peraturan direksi.

Setelah memenuhi persyaratan diatas nasabah dapat dinyatakan layak untuk memperoleh kredit dari pihak perusahaan. Jangka waktu kredit yang diberikan adalah jangka waktu kredit yang tetap dengan maksimal waktu pelunasan kredit adalah 120 hari dari waktu awal pengajuan, tetapi nasabah dapat melunasi kreditnya sewaktu-waktu sebelum jangka waktu kredit berakhir. Apabila setelah jangka waktu kredit tersebut berakhir belum dapat dilunasi, maka jangka waktu kreditnya dapat diperpanjang dengan cara nasabah tersebut melakukan pembayaran bunga kredit berdasarkan tipe golongan kreditnya. Dimana, apabila golongan kredit A, maka bunga yang harus dibayar yaitu 0.75 % per 15 hari dari jumlah nominal kredit, golongan kredit B dan C sebesar 1.2 % per 15 hari, dan golongan kredit D 1.02 % per 15 hari.

## 2.2 Klasifikasi

Klasifikasi adalah suatu proses untuk menyatakan suatu objek ke salah satu kategori yang sudah didefinisikan sebelumnya. Tujuannya adalah untuk mengetahui kelas yang seakurat mungkin dari record-record data yang sebelumnya yang tidak terlihat. Model klasifikasi digunakan untuk pemodelan deskriptif sebagai sebuah perangkat untuk penggambaran perbedaan objek-objek dari kelas yang berbeda. Selain itu juga klasifikasi dapat digunakan sebagai sebuah model prediktif yang dimana dapat digunakan untuk memprediksi label kelas untuk record yang tidak diketahui atau tidak dikenal. Beberapa contoh teknik dalam klasifikasi adalah decision tree, rute-based, neural Networks, support vector machine dan naive bayes

### 2.2.1 Preprocessing

Preprocessing merupakan teknik pengolahan data menjadi data yang lebih terstruktur. Data set yang digunakan berupa data mentah nasabah dari pihak perusahaan. Data set ini akan melalui tahap preprocessing agar data sesuai dengan algoritma untuk memperoleh hasil klasifikasi.

### 2.2.2 Transformasi Data

Didalam sebuah data, tidak seluruh data yang harus digunakan, setelah data didapatkan maka data tersebut harus mengalami proses menghilangkan noise dan data yang tidak konsisten atau data tidak relevan. Umumnya data yang diperoleh, baik dari database suatu perusahaan maupun hasil eksperimen, memiliki isian-isian yang tidak sempurna seperti data yang hilang, data yang tidak valid atau juga hanya sekedar salah ketik. Data yang ada pada database sering kali tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk dianalisis yang akan diambil dari database. Sebagai contoh, sebuah kasus yang meneliti faktor kecenderungan orang membeli dalam kasus Market basket analysis, tidak perlu mengambil nama pelanggan, cukup dengan di pelanggan saja dan data-data transaksi pelanggan tersebut.

Setelah suatu data dipilih, maka selanjutnya data tersebut diubah atau digabungkan ke dalam format yang sesuai untuk diproses dalam data mining. Beberapa metode data mining membutuhkan format data yang khusus sebelum bisa diaplikasikan. Sebagai contoh beberapa metode standar seperti analisis asosiasi dan clustering hanya bisa menerima input data kategorial.

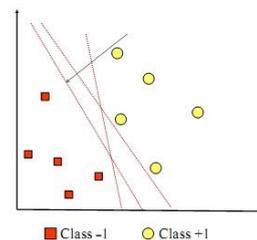
Karenanya data berupa angka numerik yang berlanjut perlu dibagi-bagi menjadi beberapa interval. Proses ini sering disebut sebagai transformasi data.

### 2.2.3 Normalisasi

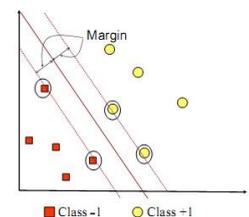
Selain mengubah nilai atribut, normalisasi atribut yang bernilai bilangan real umumnya juga diperlukan agar proses perhitungan lebih mudah dan atribut yang rentang nilainya besar tidak mendominasi atribut yang rentang nilainya lebih kecil. Normalisasi umumnya memberikan hasil yang lebih baik [9]. Rentang nilai atribut yang dianjurkan untuk digunakan adalah [0,1] atau [-1,+1] [9]. Normalisasi dalam rentang nilai [0,1] menghasilkan akurasi yang sama dengan [-1,+1] jika menggunakan fungsi kernel RBF, tetapi waktu komputasi yang dibutuhkan mungkin berbeda. Karena nilai atribut data pada KDD CUP 99 banyak yang bernilai 0, maka digunakan rentang normalisasi [0,1] karena waktu yang akan dibutuhkan menjadi lebih singkat dibandingkan dengan penggunaan rentang nilai [-1,+1]. Alasannya, jika dinormalisasi ke dalam rentang nilai [-1,+1] banyak nilai 0 akan berubah menjadi -1 sehingga perhitungannya menjadi lebih lambat. Adapun rumus normalisasi ke dalam rentang [0,1] adalah persamaan (3.1) dimana  $\min x$  dan  $\max x$  adalah nilai minimum dan nilai maksimum atribut  $x$ .

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (2.1)$$

## 2.3 Support Vector Machine



Gambar 2.1 Hyperplane



Gambar 2.2 Support Vector

Support Vector Machine (SVM) adalah sebuah algoritma sistem pembelajaran yang berupa fungsi-fungsi linear yang dilatih. Support Vector Machine ini telah diperkenalkan sejak tahun 1992 oleh Vapnik, Guyon dan

Pada Gambar 2.1 memperlihatkan dimana svm berusaha untuk menemukan hyperplane terbaik yang memisahkan kedua kelas yaitu +1 dan -1, dimana kelas +1 disimbolkan dengan lingkaran berwarna kuning dan -1 disimbolkan dengan persegi berwarna merah.

Pada Gambar 2.2 *Hyperplane* pemisah terbaik antara kedua class dapat ditemukan dengan mengukur margin hyperplane yang ada 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*. Garis solid pada gambar 2.2 menunjukkan hyperplane yang terbaik, yaitu yang terletak tepat pada tengah-tengah kedua class, sedangkan titik merah dan kuning yang berada dalam lingkaran hitam adalah support vector. Usaha untuk mencari lokasi hyperplane ini merupakan inti dari proses pembelajaran pada SVM.[2]

Data yang tersedia dinotasikan sebagai  $\vec{x}_i \in \mathbb{R}^{d \times 2,3}$  sedangkan label masing-masing dinotasikan sebagai  $y_i \in \{-1,+1\}$  untuk  $i = 1,2,\dots,l$ , dimana  $l$  adalah banyaknya data. Diasumsikan kedua kelas  $-1$  dan  $+1$  dapat terpisah secara sempurna oleh hyperplane berdimensi  $d$ , yang didefinisikan[2]:

$$\vec{w} \cdot \vec{x} + b = 0 \tag{2.1}$$

Apabila pattern  $\vec{x}_i$  termasuk kedalam kelas  $-1$  (sampel negatif) maka dirumuskan sebagai pattern yang memenuhi pertidaksamaan:

$$\vec{w} \cdot \vec{x}_i + b \leq -1 \tag{2.2}$$

sedangkan pattern  $\vec{x}_i$  yang termasuk kelas  $+1$  (sampel positif) :

$$\vec{w} \cdot \vec{x}_i + b \geq +1 \tag{2.3}$$

Margin terbesar dapat ditemukan dengan memaksimalkan nilai jarak antara hyperplane dan titik terdekatnya, yaitu  $1/\|\vec{w}\|$ . Hal ini dapat dirumuskan sebagai *Quadratic Programming (QP) problem*, yaitu mencari titik minimal persamaan (2.4), dengan memperhatikan constraint persamaan (2.5).[2]

$$\min_w \tau(w) = \frac{1}{2} \|\vec{w}\|_2 \tag{2.4}$$

$$y_i(\vec{x}_i \cdot \vec{w} + b) - 1 \geq 0 \tag{2.5}$$

Pemecahan problem dapat dilakukan dengan teknik *Lagrange Multiplier*, sesuai persamaan persamaan berikut (2.6).

$$L(\vec{w}, b, \alpha) = \frac{1}{2} \|\vec{w}\|^2 - \sum_{i=1}^l \alpha_i (y_i (\vec{x}_i \cdot \vec{w} + b) - 1) \quad (i = 1, 2, \dots, l) \tag{2.6}$$

$\alpha_i$  merupakan Lagrange multipliers yang dapat bernilai positif atau nol. Sesuai dengan sifat bahwa apabila  $L$  bernilai no sebagai titik optimal gradient, maka persamaan (2.6) dapat dimodifikasi menjadi:

$$\sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j=1}^l \alpha_i \alpha_j y_i y_j \vec{x}_i \cdot \vec{x}_j \tag{2.7}$$

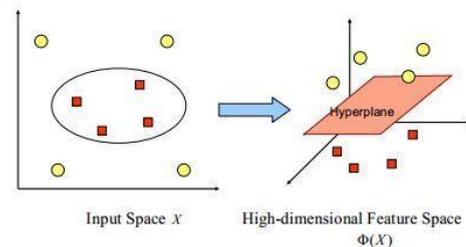
Harus memenuhi:

$$\alpha_i > 0, \quad (i = 1, 2, \dots, l) \quad \sum_{i=1}^l \alpha_i y_i = 0 \tag{2.8}$$

Dari hasil perhitungan di atas didapatkanlah  $\alpha_i$  yang kebanyakan bernilai positif. Data yang berkorelasi dengan  $\alpha_i$  yang positif inilah yang disebut sebagai *support vector*. Setelah menemukan *support vector*. [2]

### 2.3.1 Data Yang Tidak Terpisah Secara Linear dan Kernel Trik

Menurut teori Cover yang menyatakan “Jika suatu transformasi bersifat non linear dan dimensi dari feature space cukup tinggi, maka data pada input space dapat dipetakan ke feature space yang baru, dimana pattern-pattern tersebut pada probabilitas tinggi dapat dipisahkan secara linear”. [2], hal ini dapat digunakan untuk menyelesaikan kasus non linear, maka SVM dimodifikasi kedalam fungsi kernel.



Gambar 2.3 Pemetaan input space dengan pemetaan ke dimensi tinggi[2]

Gambar 2.3 pada Input Space X terlihat bahwa antara class lingkaran(kuning) dan persegi(merah) saling terpisah karena berbeda dimensi. Sedangkan pada sisi kanan yaitu  $\Phi(X)$  terlihat bahwa terjadi pemetaan tiap data pada input space tersebut ke ruang vektor baru yang berdimensi lebih tinggi. Notasi matematika dari proses ini adalah:

$$\Phi : \mathbb{R}^d \rightarrow \mathbb{R}^q \quad d < q \tag{2.9}$$

Proses pembelajaran pada SVM umumnya menggunakan dot product dari data yang sudah ditransformasikan pada ruang baru yang berdimensi lebih tinggi, berupa  $\Phi(\vec{x}_i) \cdot \Phi(\vec{x}_j)$ .

Akibat Transformasi  $\Phi$  sulit untuk dipahami, maka menurut teori Mercer perhitungan dot product dapat digantikan dengan fungsi kernel  $K(\vec{x}_i, \vec{x}_j)$  yang mendefinisikan secara implicit transformasi  $\Phi$ . Hal ini disebut sebagai Kernel Trick, yang dirumuskan sebagai berikut:

$$K(\vec{x}_i, \vec{x}_j) = \Phi(\vec{x}_i) \cdot \Phi(\vec{x}_j) \tag{2.10}$$

Kernel trick mudah dalam pengaplikasiannya karena dalam proses pembelajaran SVM, untuk menentukan support vektornya kita hanya perlu mengetahui fungsi kernel yang akan dipakai, dan tidak perlu mengetahui wujud dari fungsi non linear  $\Phi$ . Berbagai jenis fungsi kernel, sebagaimana sebagai berikut:

Polynomial:  $K(\vec{x}_i, \vec{x}_j) = (\vec{x}_i \cdot \vec{x}_j + 1)^p \tag{2.11}$

Gaussian:  $K(\vec{x}_i, \vec{x}_j) = \exp(-\gamma \|\vec{x}_i - \vec{x}_j\|^2) \tag{2.12}$

Sigmoid:  $K(\vec{x}_i, \vec{x}_j) = \tanh(\alpha \vec{x}_i \cdot \vec{x}_j + \beta) \tag{2.13}$

Pada pengerjaan penelitian ini, jenis kernel yang digunakan adalah Gaussian Kernel. Gaussian Kernel adalah pilihan kernel dimana penggunaan kernelnya ini mampu secara *nonlinear* memetakan sampel ke dalam ruang dimensi yang lebih tinggi, sehingga tidak seperti kernel *linear*, kernel ini dapat menangani kasus ketika hubungan antara label kelas dan atributnya tidak

*linear*[3] dengan persamaan:

$$K(\vec{x}_i, \vec{x}_j) = \exp(-\gamma \|x - x'\|^2) \tag{2.14}$$

Dimana nilai  $\gamma \geq 0$  adalah sebuah parameter yang mengontrol besarnya fleksibilitas persamaan Gaussian.

### 2.3 Sequential Minimal Optimization

Sequential Minimal Optimization merupakan algoritma pembelajaran yang menggunakan aturan Quadratic Programming (QP), dimana algoritma ini bertujuan untuk memecahkan masalah optimasi yang muncul selama pelatihan SVM. Sequential Minimal Optimization ditemukan pertama kali pada tahun 1998 oleh John Platt di sebuah research microsoft. SMO ini banyak diaplikasikan algoritma pelatihan mesin menggunakan LIBSVM tools.

Masalah optimasi pada SMO mempertimbangkan masalah klasifikasi biner dengan dataset  $(x_1, y_1), \dots, (x_n, y_n)$ , dimana  $x_i$  adalah sebuah vektor inputan dan  $y_i \in \{-1, 1\}$  adalah label biner yang sesuai.

Soft margin SVM yang dilatih menggunakan masalah Quadratic Programming, dapat ditunjukkan dengan formula berikut:

Bergantung pada:

$$\begin{aligned} \max_{\alpha} & \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n y_i y_j K(x_i, x_j) \alpha_i \alpha_j, \\ & 0 \leq \alpha_i \leq C, \text{ for } i = 1, 2, \dots, n, \\ & \sum_{i=1}^n y_i \alpha_i = 0 \end{aligned} \tag{2.15}$$

SMO mengatasi masalah optimasi dengan cara membagi masalah menjadi serangkaian sub-problem yang mungkin, yang kemudian akan diselesaikan secara analitis.

Dengan algoritma yang terjadi sebagai berikut:

1. Temukan Lagrange multiplier yang melanggar kondisi Karush – Kuhn – Tucker (KKT) dalam optimasi.
2. Pilih Multiplier kedua dan optimalkan pasangannya.
3. Ulangi langkah 1 dan 2 hingga konvergen. Ketika semua Lagrange Multipliers memenuhi kondisi KKT, maka masalah telah terselesaikan.

### 2.4 Pengukuran Evaluasi

Pengujian keefektifan atau kualitas hasil suatu klasifikasi terdapat 3 buah parameter pengujian performansi yang dapat dilakukan, yaitu precision, recall dan f-measure. Precision adalah tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh sistem. Recall adalah tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi. Dan f-measure didefinisikan sebagai tingkat kedekatan antara nilai prediksi dengan nilai aktual.

$$\tag{2.16}$$

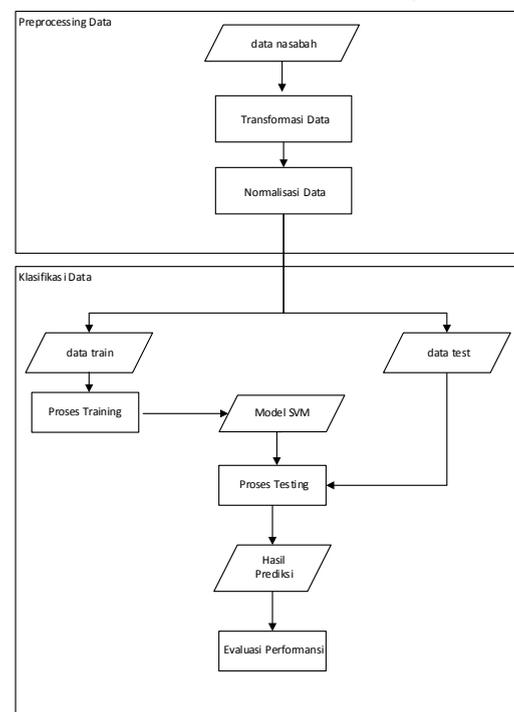
$$Recall = \frac{TP}{TP + FN} \tag{2.17}$$

$$F - measure = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{2.18}$$

### 3. Deskripsi Umum Sistem

Secara umum, sistem ini akan melakukan prediksi melalui data nasabah pegadaian yang terdapat di PT. Pegadaian (Persero) kantor cabang Seririt Bali, apakah nasabah tersebut termasuk kedalam kategori nasabah yang potensial atau tidak dengan menggunakan metode support vector machines sequential minimal optimization. Proses prediksi ini meliputi tahapan preprocessing data nasabah dan klasifikasi data nasabah. Yang dimana tahapan preprocessing data ini terdiri dari tahapan cleaning data, transformasi data, dan normalisasi data. Setelah tahapan preprocessing data ini dilakukan maka akan dihasilkan sebuah data dalam bentuk data train dan data test, yang dimana kedua data ini akan digunakan untuk proses klasifikasi nantinya.

Gambar 3.1 Gambaran sistem secara global



Setelah melalui preprocessing data, tahap selanjutnya adalah tahap klasifikasi data, pada tahap ini, sistem akan melakukan pelatihan data train dengan menggunakan metode support vector machine sequential minimal optimization, setelah melakukan training data, sistem akan menghasilkan model svm, dimana model svm inilah yang nantinya akan diuji dengan menggunakan data testing pada tahap pengujian model.

### 4. Pengujian dan Analisis

#### 4.1 Uji Parameter Gamma

Skenario pengujian pertama ini adalah pengujian performansi dari sistem yang diamati dari perubahan parameter gamma selama proses pelatihan. Dari 3 data

yang diuji, pertama pada data kredit 2010 nilai *precision* dan *recall* terbaik didapatkan ketika nilai gamma bernilai

4 dan 0.1, dengan nilai *precision* sebesar 0.6791 dan 0.9200. Sedangkan untuk data kedua, pada kredit 2011 nilai *precision* dan *recall* terbaik didapatkan ketika nilai *gamma* bernilai 0.6 dan [0, 0.1] dengan nilai *precision* dan *recall* sebesar 0.9578 dan 0.8575. Dan terakhir untuk data ketiga, yaitu data kredit 2012, nilai *precision* dan *recall* terbaik didapatkan ketika nilai *gamma* bernilai 4 dan 0.9 dengan nilai *precision* yang dihasilkan yaitu 0.8731 dan nilai *recall*nya 0.7325.

Sedangkan untuk nilai *f-measure* pada ketiga data uji, untuk data pertama pada kredit 2010 nilai *f-measure* terbaik didapatkan ketika nilai *gamma* 0.8 yaitu dengan nilai 0.7889, sedangkan pada data kredit 2011 nilai *f-measure* terbaik didapatkan ketika nilai *gamma* 0.6 yaitu dengan nilai 0.8821, dan terakhir untuk data kredit 2012 nilai *f-measure* terbaik didapatkan ketika nilai *gamma* bernilai 0.8 yaitu dengan nilai 0.7951.

Secara keseluruhan dari hasil yang didapatkan ketika nilai *gamma* bernilai pada rentang [0.4..0.9], nilai *precision*, *recall*, dan *f-measure* yang dihasilkan lebih baik dan stabil, hal ini dikarenakan parameter *gamma* merupakan parameter yang mengontrol lebar gaussian, yang mengatur fleksibilitas *decision boundary* dalam memberikan hasil dari suatu klasifikasi. Semakin besar *gamma*, maka hasil klasifikasi akan semakin fleksibel yang dapat mengakibatkan terjadinya *overfitting*[7], tetapi apabila nilai *gamma* semakin kecil, maka hasil klasifikasi tersebut akan semakin mendekati kernel linear.

#### 4.2 Uji Parameter Tolerance

Skenario pengujian kedua adalah uji parameter *tolerance*. Untuk skenario pengujian yang kedua ini, pengujian yang dilakukan adalah uji parameter *tolerance* terhadap performansi dari sistem yang diamati dari perubahan parameter *tolerance* selama proses pelatihan. Seperti diketahui parameter *tolerance* ini berguna untuk menentukan ukuran maksimum dari kesalahan klasifikasi untuk tetap dapat dianggap masuk ke dalam suatu kelas yang dipisahkan oleh hyperplane. Dari 3 data yang diuji, pertama pada data kredit 2010 nilai *precision* dan *recall* terbaik didapatkan ketika nilai *tolerance* 0.7 dan 0.1, dengan nilai *precision* sebesar 0.6667 dan 0.9200. sedangkan untuk data kedua, pada kredit 2011 nilai *precision* dan *recall* terbaik didapatkan ketika nilai *gamma* bernilai 0.7 dengan nilai *precision* dan *recall* sebesar 0.8321 dan 0.8675. Dan terakhir untuk data ketiga, yaitu data kredit 2012, nilai *precision* dan *recall* terbaik didapatkan ketika nilai *gamma* bernilai 0.8 dan 0.7 dengan nilai *precision* dan *recall* yang dihasilkan yaitu 0.8984 dan 0.7000.

Sedangkan untuk nilai *f-measure* pada ketiga data uji, untuk data pertama pada kredit 2010 nilai *f-measure* terbaik didapatkan ketika nilai *gamma* 0.9 yaitu dengan nilai 0.7665, sedangkan pada data kredit 2011 nilai *f-measure* terbaik didapatkan ketika nilai *gamma* 0.9 yaitu dengan nilai 0.8495, dan terakhir untuk data kredit 2012 nilai *f-measure* terbaik didapatkan ketika nilai *gamma* bernilai 0.7 yaitu dengan nilai 0.7843.

Dari hasil pengujian yang didapatkan, dapat dianalisis bahwa nilai parameter *tolerance* ini jika dibandingkan dengan parameter *gamma*, pengaruh dari parameter *tolerance* tidak berpengaruh signifikan terhadap hasil dari *precision*, *recall* dan *f-measure* yang dihasilkan oleh

sistem. Setelah diamati dari hasil pengujian ini, jika dibandingkan dengan hasil yang dihasilkan ketika melakukan pengujian parameter *gamma*, pada pengujian kali ini, hasil *precision*, *recall* dan *f-measure* yang dihasilkan mengalami perubahan yang tidak terlalu signifikan. Selain itu didapatkan bahwa semakin besar nilai *tolerance* yang diberikan, maka secara umum nilai *precision*, *recall* dan *f-measure* yang dihasilkan akan mengalami penurunan. Dimana apabila diamati, dari 3 data uji ini, dapat dilihat bahwa untuk nilai *precision* secara umum mengalami penambahan nilai pada rentang *gamma* [0.3..0.7], dan nilai *precision* mulai mengalami penurunan ketika nilai *tolerance* yang diberikan melebihi 0.7. Untuk nilai *recall* yang dihasilkan juga mengalami penurunan dan penambahan, dapat dilihat bahwa ketika nilai *tolerance* bernilai [0.1..0.3], nilai *recall* yang dihasilkan cenderung mengalami penambahan, namun setelah nya nilai yang dihasilkan cenderung turun. Berdasarkan dari fungsi parameter *tolerance* ini adalah untuk mentolerir dari tingkat kesalahan klasifikasi, maka yang dapat disimpulkan adalah data uji yang diklasifikasikan memiliki tingkat kesalahan klasifikasi yang rendah, dan sistem mampu memberikan tingkat ketepatan pemberian informasi yang baik. Dimana hal ini dapat dibuktikan dari rata-rata nilai *recall* yang tinggi.

#### 4.3 Uji Komposisi Data Training dan Data Testing

Pada skenario uji terakhir ini adalah skenario uji pembagian komposisi data *training* dan data *testing*. Dari pengujian yang dilakukan dapat dianalisa bahwa kredit tahun 2010, hasil *precision* terbaik didapatkan ketika komposisi pembagian data 80:10 dimana nilai *precision* yang dihasilkan adalah 0.7250, untuk nilai *recall* terbaik yang dihasilkan adalah ketika komposisi pembagian data 80:20 dan 90:10. Sedangkan nilai *f-measure* terbaik didapatkan ketika komposisi data-*training* dan *testing* 80:10 dengan nilai *f-measure* sebesar 0.8108. Selanjutnya untuk kredit tahun 2011, hasil *precision* terbaik didapatkan ketika komposisi pembagian data 70:30 dimana nilai *precision* yang dihasilkan adalah 0.9899, untuk nilai *recall* terbaik yang dihasilkan adalah ketika komposisi pembagian data 90:10 dengan nilai *recall* yang dihasilkan 0.8875. Sedangkan nilai *f-measure* terbaik didapatkan ketika komposisi data-*training* dan *testing* 90:10. untuk kredit tahun 2012, hasil *precision* terbaik didapatkan ketika komposisi pembagian data 70:30 dimana nilai *precision* yang dihasilkan adalah 0.8974, untuk nilai *recall* terbaik yang dihasilkan adalah ketika komposisi pembagian data 70:30 dengan nilai 0.7292. Sedangkan nilai *f-measure* terbaik didapatkan ketika komposisi data-*training* dan *testing* 70:30 dengan nilai 0.8046.

Perubahan komposisi data ini, dapat dilihat pada grafik dibawah ini tingkat perubahan nilai secara umum bernilai maksimal ketika perbandingan data yang digunakan yaitu 80:20, namun apabila dibandingkan dengan data dengan jumlah lainnya perbedaan jumlah nilainya tidak terlalu signifikan.

Sehingga dapat dianalisis bahwa ketika sebuah data *training* memiliki jumlah yang tidak terlalu banyak namun memiliki karakteristik yang baik yang dapat merepresentasikan karakteristik secara keseluruhan dari data, maka model yang dibentuk oleh *Support Vector Machine* akan baik juga. Dari gambar tabel diatas

dapat dilihat bahwa hasil yang dihasilkan relatif stabil dari berbagai macam bentuk komposisi data latih dan uji.

## 5. Penutup

### 5.1 Kesimpulan

Berdasarkan analisis terhadap pengujian yang dilakukan dalam Tugas Akhir ini, dapat disimpulkan bahwa:

1. Metode *Support Vector Machine Sequential Minimal Optimization* dapat diimplementasikan dengan baik untuk mengklasifikasikan nasabah yang potensial dan tidak jika di Analisis dari performasi sistem dalam hal *precision* dan *recall*nya.
2. Nilai gamma jauh lebih berpengaruh jika dibandingkan dengan dengan nilai *tolerance*.
3. Perbedaan komposisi data train dan data *testing* tidak terlalu mempengaruhi performasi sistem, namun secara umum hasil yang baik didapatkan ketika jumlah data *training*nya lebih banyak dibandingkan jumlah data *testing*nya.
4. Hasil performasi *Support Vector Machine Sequential Minimal Optimization* terhadap prediksi nasabah potensial dan tidak potensial adalah baik dimana dengan nilai rata-rata performasi untuk *precision*, *recall*, dan *f-measure* yang dihasilkan diatas 75%.

### 5.2 Saran

Saran yang diperlukan dari Tugas Akhir ini untuk pengembangan sistem lebih lanjut adalah sebagai berikut :

1. Menambahkan sampel data nasabah yang jauh lebih banyak untuk lebih melihat pengaruh dari komposisi data terhadap performasi sistem.
2. Membandingkan performasi yang telah dihasilkan oleh metode klasifikasi SVM SMO ini dengan metode lainnya.
3. Mengkombinasikan SVM SMO dengan model SVM lainnya.
4. Menggunakan satu sumber expert judgment untuk beberapa contoh data.

### Daftar Pustaka

- [1] Abubakar S. Magaji, Auduh Isah, Victor Onomza Waziri, Adeboye K.R. 2013. *LA Conceptual Nigeria Stock Exchange Prediction: Implementation Using Support Vector Machines-SMO Model*
- [2] Cio, Krzystof J., dik. 2007. *"Data Mining A Knowledge Discovery Approach"*. New York: Springer
- [3] Cristiaini Nello & Joh Shawe-Taylor. 2000. *"An Introduction to Support Vector Machines and Other Kernel-Bases Learning Methods"*. UK: Cambrige Press.
- [4] Fisher M.J., Fieldsend J.E., and Everson R.M. 2004. Precision and Recall Optimisation form Information Access Tasks., !st International Workshop, ROCAI-2004
- [5] Nugroho Satriyo Anto, Arief Budi Witarto, dan Dwi Handoko. 2003 *"Support Vector Machine-*

Teori dan Aplikasinya dalam Bioinformatika" dari <http://ilmukomputer.com>

- [6] Masyarakat Ramai Gadai Emas, Pegadaian Raup Omzet Rp 8,5 Triliun . <http://news.detik.com/transisipresiden/read/2012/08/27/114439/1999742/5/masyarakat-ramai-ramai-gadai-emas-pegadaian-raup-omzet-rp-85-triliun>, (Date: 27/ 08/12).
- [7] Hsu, Chih-Wei, Chih-Chung Chang, dan Chih-Jen Lin. 2010. *"A Practical Guide to Support Vector Classification"*. Taiwan: National Taiwan University
- [8] Powers, David M W. 2007. *"Evaluation: From Precision, Recall and F-Factor to ROC, Informedness, Markedness & Correlation"*. AILab, School of Computer Science, Engineering and Mathematics, Flinders University, South Australia, Australia
- [9] PT. Pegadaian. "Data lelang periode 2010-2012 Kantor Cabang Seririt-Bali".
- [10] PT. Pegadaian. "Pedoman Operasional Kantor Cabang".
- [11] Rina Yuliana Siagian. 2011. Klasifikasi Parket Kayu Jati Menggunakan Metode *Support Vector Machine* (SVM)
- [12] Saeed Safari, Mohammad Shojaee, Maysam Eftekhary, Peyman Gholami. 2012. "Ranking Normalization Methods for Improving the Accuracy of SVM Algorithm by DEA Method". Iran: Department of Industrial Engineering, Arak Branch, Islamic Azad University
- [13] Sari Khrisna Dini Yunita. 2006. *Text Categorization with Support Vector Machine (SVM) Classification Method*. Bandung: Institut Teknologi Telkom.
- [14] Simola, B. Scholkopf, C. Burges . *Fast Training of Support Vector Machine pusing Sequential Minimal Optimization, in Advances in Kernel Methods - Support Vector Learning, MIT Press (1998)*.