

KLASIFIKASI PASIEN DIABETES MELLITUS MENGGUNAKAN METODE *SMOOTH SUPPORT VECTOR MACHINE (SSVM)*

Rizky Adhi Nugroho¹, Tarno², Alan Prahutama³

¹Mahasiswa Departemen Statistika FSM Universitas Diponegoro

^{2,3}Staf Pengajar Departemen Statistika FSM Universitas Diponegoro

ABSTRACT

Diabetes Mellitus (DM) is a high-risk metabolic diseases. Laboratory tests are needed to determine if the patients suffer from a Diabetes Mellitus. Therefore, it needs a classification methods that can precisely classify data according to the classes criteria. SVM is one of commonly used methods of classification. The basic concept of SVM is to find the best separator function (hyperplane) that separates the data according its class. SVM uses a kernel trick for nonlinear problems, which transforms data into high-dimensional space using kernel functions, so it can be classified linearly. This research will use a developed methods of SVM called SSVM, that adds smoothing function using Newton-Armijo method. The smoothing methods is needed to correct the effectiveness of SVM in big data classifying. The result is indicating tha SSVM classification prediction using Gaussian RBF kernel function, can classify 98 out of 110 patient data of Diabetes Mellitus correctly according the original class.

Keywords : Diabetes Mellitus, Classification, Support Vector Machine (SVM), Smooth Support Vector Machine (SSVM), Kernel Gaussian RBF.

1. PENDAHULUAN

Kesehatan merupakan aspek penting dalam kehidupan, banyak permasalahan yang terjadi dalam peningkatan taraf kesehatan masyarakat yang berhubungan dengan gaya hidup. Gaya hidup seseorang yang kurang sehat dapat mengakibatkan munculnya gangguan kesehatan. Gangguan kesehatan yang kerap terjadi pada masyarakat salah satunya adalah Diabetes Mellitus (DM) yang merupakan penyakit yang disebabkan kadar gula darah yang tinggi. Hal ini menjadi tantangan yang berat pada sistem pelayanan kesehatan di Indonesia (Zahtamal, 2007).

Dari berbagai penelitian di Indonesia yang dilakukan oleh pusat-pusat diabetes, sekitar tahun 1980-an prevalensi DM pada penduduk usia 15 tahun ke atas sebesar 1,5-2,3% dengan prevalensi di daerah rural/pedesaan lebih rendah dibandingkan perkotaan. Pada Survei Kesehatan Rumah Tangga (SKRT) 2001 mendapatkan prevalensi DM pada penduduk usia 25-64 tahun di Jawa dan Bali sebesar 7,5%. Prevalensi dari penderita DM cenderung meningkat pada perempuan dibandingkan dengan laki-laki dan terjadi peningkatan prevalensi penyakit DM sesuai dengan pertambahan umur namun mulai umur 65 tahun cenderung menurun.

Untuk dapat mengidentifikasi penyakit DM tersebut, perlu diketahui ciri pasien penyakit DM melalui berbagai hasil uji laboratorium. Hasil pengecekan tersebut memiliki nilai diskret yang dapat dikategorikan. Dibutuhkan sebuah metode klasifikasi yang nantinya digunakan untuk mengetahui apakah terjadi penyakit DM pada data-data hasil uji laboratorium. Salah satu metode klasifikasi yang dikenal adalah metode *Smooth Support Vector Machine (SSVM)*.

Penelitian mengenai klasifikasi dilakukan untuk mengetahui apakah suatu metode dapat melakukan klasifikasi dengan baik. Beberapa penelitian yang sebelumnya telah dilakukan, SVM dan SSVM diketahui dapat menghasilkan nilai akurasi klasifikasi yang tinggi. Seiring dengan berkembangnya metode-metode untuk melakukan klasifikasi, maka metode klasifikasi yang baik dan efisien akan menjadi lebih sering digunakan. Metode *smoothing* yang ditambahkan pada SVM non-linier biasa diharapkan mampu melakukan klasifikasi dengan lebih baik.

2. TINJAUAN PUSTAKA

2.1. Diabetes Mellitus

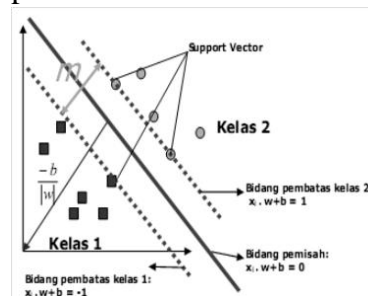
Diabetes Mellitus (DM) atau di Indonesia lebih dikenal dengan istilah kencing manis adalah sebuah penyakit metabolik yang disebabkan oleh kurangnya hormon insulin atau ketidakmampuan tubuh dalam memanfaatkan insulin, sehingga kadar glukosa atau kadar gula dalam darah tidak terkendali. Insulin adalah suatu hormon yang dihasilkan oleh kelompok sel beta di pankreas. Insulin memberi sinyal kepada sel tubuh agar menyerap glukosa.

Menurut laporan Riset Kesehatan Dasar (Riskesdas) pada tahun 2007 menunjukkan bahwa prevalensi nasional penyakit DM adalah 1,1% berdasarkan diagnosis tenaga kesehatan dan gejala. Hasil Riskesdas tahun 2007 itu juga menunjukkan bahwa prevalensi nasional DM (berdasarkan pengukuran gula darah pada penduduk umur diatas 15 tahun di daerah perkotaan) adalah 5,7%. Sebanyak 13 provinsi di Indonesia mempunyai prevalensi DM diatas prevalensi nasional.

Ada dua tipe DM yaitu DM tipe I (childhood-onset diabetes, insulin-dependent diabetes mellitus, IDDM) adalah diabetes yang terjadi karena berkurangnya rasio insulin dalam sirkulasi darah akibat hilangnya sel beta penghasil insulin pada pulau-pulau Langerhans pankreas. IDDM dapat diderita oleh anak-anak maupun orang dewasa. Diabetes tipe ini hanya dapat diobati dengan menggunakan insulin. Kemudian diabetes tipe 2 adalah kondisi dimana tubuh pasien tidak cukup menerima insulin, sehingga menyebabkan kadar gula dalam darah menjadi tinggi.

2.2. Smooth Support Vector Machine (SSVM)

Support Vector Machines (SVM) adalah suatu teknik untuk melakukan prediksi, baik dalam kasus klasifikasi maupun regresi (Santosa, 2007). Pada ruang berdimensi tinggi, akan dicari hyperplane (fungsi pemisah) yang dapat memaksimalkan jarak (margin) antara dua kelas data. Klasifikasi menggunakan SVM dapat dijelaskan secara sederhana yaitu usaha untuk mendapatkan garis sebagai fungsi pemisah terbaik yang dapat memisahkan dua kelas yang berbeda (+1,-1) pada ruang input.



Gambar 1. Konsep Fungsi Pemisah pada SVM

Pada gambar 1 memperlihatkan bahwa beberapa data yang merupakan anggota dari kelas -1 dan +1. Data yang disimbolkan dengan kotak adalah anggota data -1 sedangkan yang bulat menyimbolkan anggota data +1. Fungsi pemisah terbaik adalah fungsi yang mampu memisahkan data dengan nilai margin (m) yang terbesar, dan tepat berada di antara kedua kelas data. Margin merupakan jarak antara fungsi pemisah dengan data terdekat dari masing-masing kelas (Prasetyo, 2012). Pada gambar 1, fungsi pemisah terbaik ditunjukkan dengan garis tebal yang memisahkan kedua kelas. Data yang berada pada bidang pembatas dan terdekat dengan fungsi pemisah terbaiklah yang disebut dengan support vector. Hanya data support vector yang digunakan selama proses mendapatkan fungsi pemisah terbaik (Prasetyo, 2012). SVM merupakan metode berbasis *machine learning* yang berpotensi untuk dikembangkan lebih jauh karena memiliki performansi tinggi dan dapat diaplikasikan secara luas untuk klasifikasi dan estimasi.

SSVM adalah pengembangan SVM dengan menggunakan teknik smoothing. Metode ini pertama kali diperkenalkan oleh Lee pada tahun 2001. SVM memanfaatkan optimasi dengan quadratic programming, sehingga untuk data berdimensi tinggi dan data jumlah besar SVM menjadi kurang efisien. Oleh karena itu dikembangkan smoothing technique yang menggantikan plus function SVM dengan integral dari fungsi sigmoid neural network yang selanjutnya dikenal dengan Smooth Support Vector Machine (SSVM).

Diberikan masalah klasifikasi dari n objek dalam ruang dimensi R^p sehingga susunan data berupa matriks A berukuran $n \times p$ dan keanggotaan tiap titik terhadap kelas $\{+1\}$ atau $\{-1\}$ yang didefinisikan pada diagonal matriks D berukuran $m \times n$, problem optimasinya adalah :

$$\min_{w,b,\xi} \frac{c}{2} \xi' \xi + \frac{1}{2} (w'w + b^2), \quad (1)$$

dengan kendala $D(Aw + eb) + \xi \geq e, \xi \geq 0$.

Solusi problem (1) adalah

$$\xi = (e - D(Aw + eb)), \quad (2)$$

dimana ξ adalah variabel *slack* yang mengukur kesalahan klasifikasi.

Kemudian dilakukan substitusi dan konversi, sehingga persamaan (2) dapat ditulis sebagai berikut :

$$\min_{w,b} \frac{c}{2} \|(e - D(Aw - eb))\|_2^2 + \frac{1}{2} (w'w + b^2), \quad (3)$$

Fungsi objektif dalam persamaan (3) tidak memiliki turunan kedua. Teknik *smoothing* yang diusulkan dilakukan dengan mengganti fungsi plus dengan $p(x, \alpha)$ yaitu integral dari fungsi sigmoid *neural network* $(1 + \xi^{-\alpha x})^{-1}$ atau dapat dituliskan sebagai berikut :

$$p(x, \alpha) = x + \frac{1}{\alpha} \log(1 + \varepsilon^{-\alpha x}), \alpha > 0 \quad (4)$$

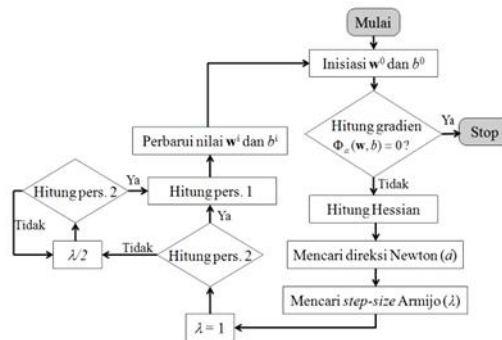
dimana α adalah parameter *smoothing*. Dengan menggantikan fungsi plus dengan $p(x, \alpha)$ maka diperoleh model SSVM sebagai berikut :

$$\min_{(w,b) \in R^{p+1}} \Phi_\alpha(w, b) = \min_{(w,b) \in R^{p+1}} \frac{c}{2} \|p(e - D(Aw - eb))\|_2^2 + \frac{1}{2} (w'w + b^2), \quad (5)$$

Secara umum, problem optimasi SSVM dapat ditulis sebagai berikut :

$$\min_{(w,b) \in R^{p+1}} \Phi_\alpha(w, b) = \min_{(w,b) \in R^{p+1}} \frac{c}{2} \|p(e - D(K(x_i, x_j)Dw - eb), \alpha)\|_2^2 + \frac{1}{2} (w'w + b^2), \quad (6)$$

Yang diselesaikan dengan iterasi Newton Armijo (Gambar 2) dan $K(x_i, x_j)$ merupakan fungsi kernel yang dalam penelitian ini digunakan kernel Gaussian atau bisa dirumuskan berikut $K(x_i, x_j) = \exp(-\gamma(\|x_i - x_j\|^2))$ dengan parameter kernel γ .



Gambar 2. Diagram Alir Algoritma Newton-Armijo

Persamaan 1 :

$$\Phi_\alpha(w_i, b_i) - \Phi_\alpha((w_i, b_i) + \lambda_i d_i) \geq -\delta \lambda_i \nabla \Phi_\alpha(w_i, b_i) d_i$$

Persamaan 2 :

$$(w_{i+1}, b_{i+1}) = (w_i, b_i) + \lambda_i d_i$$

Saat iterasi pada algoritma Newton-Armijo berhenti, diperoleh nilai w dan b yang konvergen. Dengan demikian fungsi pemisah yang diperoleh untuk kasus klasifikasi linier adalah

$$f(x) = \text{sign}(w'x + b), \quad (7)$$

Sedangkan fungsi pemisah untuk kasus klasifikasi nonlinier adalah sebagai berikut

$$f(x) = \text{sign}(w'x + b) = \text{sign}(u'D'K(x_i, x_j) + b)$$

(8)

Perumusan program linier SVM 1-norm adalah salah satu cara untuk memilih atribut (*feature selection*) di antara varian-varian norm SVM, problem linier tersebut adalah sebagai berikut

$$\min_{(w,b,s,\xi) \in R^{(2p)+1+n}} C e' \xi + e' s$$

Dengan kendala $D(Aw + eb) + \xi \geq e$

(8)

$$\begin{aligned} -s &\leq w \leq s \\ \xi &\geq 0 \end{aligned}$$

Solusi dari w mampu menghasilkan model yang parsimoni dan bersifat *sparsity*. Jika nilai dari elemen vektor $w_p = 0$, maka variabel p tidak berkontribusi dalam penentuan kelas. Kontribusi atribut atau variabel prediktor dapat dinilai dari besarnya nilai w_i untuk masing-masing atribut, dengan $l=1,2, \dots, p$.

2.3. Seleksi Parameter

2.3.1. K-Fold Cross Validation

Menurut Santosa (2007), cross-validation adalah salah satu cara menentukan parameter terbaik dari satu model dengan cara menguji besarnya error pada data testing. Dalam cross-validation, data dibagi sebanyak k bagian dengan ukuran sampel yang sama. Kemudian $k-1$ bagian digunakan sebagai data untuk training sistem dan sisanya untuk testing. Pengujian ini dilakukan sebanyak k kali dengan mengganti-ganti partisi yang berfungsi sebagai data training dan data testing.

2.3.2. Algoritma Grid Search

Menurut Yao, et al.(2014), algoritma grid search membagi jangkauan pencarian parameter yang akan dioptimalkan ke dalam grid dan melintasi semua titik grid untuk mendapatkan nilai optimal. Jika parameter c dan g masing-masing memiliki taksiran nilai sebanyak m dan n, algoritma grid search akan melatih dan menguji m x n model SSVM untuk menemukan parameter yang optimal.

2.4. Pengukuran Kinerja Klasifikasi

Berkaitan dengan evaluasi performansi klasifikasi, sensitivity dan specificity merupakan statistik yang mengukur performansi klasifikasi biner. Sensitivity mengukur proporsi dari kondisi yang positif, dan specificity mengukur proporsi dari kondisi yang negatif. Pengukuran kinerja klasifikasi pada data asli dan data hasil dari model klasifikasi dilakukan dengan menggunakan tabulasi silang (matriks konfusi) yang berisi informasi tentang kelas data asli yang direpresentasikan pada baris matriks dan kelas data hasil prediksi suatu algoritma direpresentasikan pada kolom klasifikasi (Prasetyo, 2012).

Tabel 1. Matriks Konfusi

F_{gh}		Kelas Hasil Prediksi (h)	
		Kelas = 1	Kelas = 0
Kelas Asli (g)	Kelas = 1	F_{11}	F_{01}
	Kelas = 0	F_{10}	F_{00}

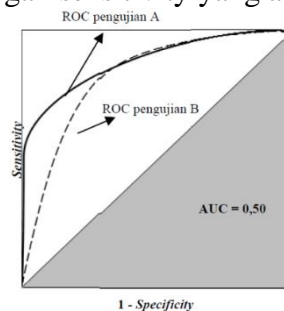
Ketepatan klasifikasi dapat dilihat dari akurasi klasifikasi. Kedua ukuran sensitivity-specificity menjelaskan akurasi diagnosis dengan lebih bermakna daripada indeks persentase akurasi tunggal

$$\text{Akurasi klasifikasi (\%)} = \frac{F_{11} + F_{00}}{F_{11} + F_{10} + F_{01} + F_{00}}$$

$$\text{Sensitivity (\%)} = \frac{F_{11}}{F_{11} + F_{10}}$$

$$\text{Specificity (\%)} = \frac{F_{00}}{F_{10} + F_{00}}$$

Membuat plot ROC (Receiver Operating Characteristic) merupakan cara yang digunakan untuk menggambarkan akurasi diskriminasi dari suatu pengujian diagnosis untuk menentukan apakah seseorang menderita suatu penyakit tertentu atau tidak. Kurva ROC adalah plot dari sensitivity terhadap 1-specificity untuk beberapa nilai threshold yang digunakan untuk menerangkan ketepatan uji dalam berbagai tingkatan titik potong dalam membaca specificity yang sesuai dengan sensitivity yang ada.



Gambar 3. Kurva ROC dari dua pengujian diagnosis (A dan B), masing-masing uji paling sedikit memiliki AUC seluas 0,50

Prinsip dasar yang harus diperhatikan adalah semakin besar nilai AUC maka semakin baik.

3. METODOLOGI PENELITIAN

3.1. Data dan Variabel Penelitian

Jenis data yang digunakan dalam penelitian ini merupakan data sekunder. Data tersebut adalah data rekam medis pegawai Kementerian Perindustrian yang melakukan pengobatan di Balai Kesehatan Kementerian Perindustrian mulai bulan Juli 2014 sampai dengan September 2014 dengan jumlah data sebanyak 496 data. Data bersumber dari penelitian tugas akhir Tyasti (2015).

Tabel 2. Kriteria Jenis Kelamin dan Usia Pasien
Sumber (Depkes, 2009)

Atribut	Keterangan
Diabetes	Positif Negatif
Jenis Kelamin	Perempuan Laki-laki
Usia Pasien	26-35 = dewasa awal 36-45 = dewasa akhir 46-55 = lansia awal 56-65 = lansia akhir

Tabel 3. Kriteria Diabetes Mellitus
Sumber (Toruan, 2012)

Atribut	Keterangan
Glukosa darah puasa (mg/dL)	80-109 = baik 110-125 = baik ≥ 126 = buruk
Glukosa darah 2 jam (mg/dL)	80-144 = baik 145-179 = sedang ≥ 180 = buruk
HDL (mg/dL)	>45 = baik ≤ 45 = buruk
LDL (mg/dL)	<100 = baik 100-129 = sedang ≥ 130 = buruk
Trigliserida (mg/dL)	<150 = baik 150-199 = sedang ≥ 200 = buruk
hbA1c	$<6,5$ = baik 6,5-8 = sedang >8 = buruk

3.2. Metode Analisis Data

Metode analisis data yang digunakan dalam penelitian ini adalah sebagai berikut:

1. Membuat deskripsi data
2. Membagi data sampel pelatihan dan sampel pengujian dengan melakukan beberapa kali percobaan dengan melihat hasil akurasi yang paling tinggi.
3. Menentukan parameter SVM dengan menerapkan metode grid search untuk mendapatkan model SSVM dengan kombinasi parameter C dan γ terseleksi.
4. Menentukan confusion matrix untuk menghitung specificity, sensitivity dan akurasi klasifikasi bagi model SSVM.
5. Membuat kurva ROC.
6. Menghitung akurasi klasifikasi SSVM.

4. PEMBAHASAN

4.1. Analisis Deskriptif

Data yang digunakan dalam penelitian ini adalah status pasien Diabetes Mellitus yang berobat di Balai Kesehatan Kementerian Perindustrian Jakarta Periode Juli 2014 sampai dengan September 2014. Terdapat 496 orang yang berobat di Balai Kesehatan Kementerian Perindustrian. Berikut merupakan jumlah pasien Diabetes Mellitus yang positif maupun negatif yang berobat di Balai Kesehatan Kementerian Perindustrian.

Tabel 4. Status Pasien Diabetes Mellitus

Status Pasien	Jumlah	Persentase
Positif	261	52,62 %
Negatif	235	47,38 %
Total	496	100%

Tabel 5. Status Pasien Diabetes Mellitus Berdasarkan Jenis Kelamin

Jenis Kelamin	Positif	Negatif	Total
Laki-laki	160	139	299
Perempuan	101	96	197
Total	261	235	496

4.2. Klasifikasi dengan *Smooth Support Vector Machine (SSVM)*

Klasifikasi pasien Diabetes Mellitus (DM) pada Balai Kesehatan Kementerian Perindustrian menggunakan metode Smooth Support Vector Machine (SSVM). Pada penelitian ini kernel yang digunakan adalah kernel Radial Basis Function (RBF) dalam menentukan hyperplane terbaik untuk klasifikasi data, dengan menghitung nilai akurasi yang didapatkan dari matriks konfusi untuk setiap percobaan.

Klasifikasi dengan metode SSVM pada penelitian ini menggunakan software Matlab R2015b dengan Toolbox SSVM yang didapatkan dari http://dmlab8.csie.ntust.edu.tw/downloads/SSVM_Code.htm.

Untuk memperoleh hasil klasifikasi pasien penyakit Diabetes Mellitus di Balai Kesehatan Kementerian Perindustrian Jakarta, maka dilakukan beberapa tahapan:

1. Penelitian ini membagi data menjadi dua yaitu data training dan testing, dengan pembagian proporsi klasifikasi 60:40, 70:30, dan 80:20.
2. Fungsi kernel yang digunakan adalah fungsi kernel Radial Basis Function (RBF). Parameter yang dibutuhkan dalam RBF adalah parameter cost (C) dan parameter gamma (γ). Parameter C (cost) yang digunakan adalah 10⁻³, 10⁻², 10⁻¹, dan 100. Sedangkan parameter γ (gamma) yang digunakan adalah 0,01, 0,05, 0,1, 0,5, 1, 5 dan 10.
3. Penentuan parameter terbaik dilakukan dengan metode grid search sebagai pengukur kinerja metode klasifikasi SSVM pada data training dengan program Matlab. Berikut adalah hasil uji coba parameter C dan γ yang telah ditentukan menurut masing-masing proporsi :
 - a. Proporsi data 60:40

Pada proporsi data training dan data testing sebesar 60:40, jumlah data training sebanyak 297 data dan sisanya sebanyak 199 data sebagai data testing. Data tersebut diuji dengan parameter C dan γ yang telah ditentukan. Hasil dari proporsi data 60:40 adalah nilai error terkecil didapatkan pada nilai parameter C = 100 atau C = 1 dan parameter $\gamma = 1$, dengan nilai error sebesar 0.0603.

b. Proporsi data 70:30

Pada proporsi data training dan data testing sebesar 70:30, jumlah data training sebanyak 347 data dan sisanya sebanyak 149 data sebagai data testing. Data tersebut diuji dengan parameter C dan γ yang telah ditentukan. Nilai error terkecil ada pada dua kombinasi parameter C dan γ . Yang pertama adalah pada nilai C = 100 dengan nilai $\gamma = 0.1$. Yang kedua adalah pada nilai C = 100 dengan nilai $\gamma = 1$. Keduanya sama-sama berasal dari nilai C = 100 atau 1. Dengan nilai error sebesar 0,470.

c. Proporsi data 80:20

Pada proporsi data training dan data testing sebesar 80:20, jumlah data training sebanyak 395 data dan sisanya sebanyak 101 data sebagai data testing. Data tersebut diuji dengan parameter C dan γ yang telah ditentukan. nilai error terkecil didapatkan pada nilai parameter C = 100 atau C = 1 dan parameter $\gamma = 0,5$, dengan nilai error sebesar 0.0297.

Sesuai dengan parameter C dan γ yang telah diuji cobakan menurut masing-masing proporsi data, didapatkan nilai error yang terkecil dari 4 percobaan yang dapat dirangkum menjadi tabel berikut :

Tabel 6. Error minimal yang didapat dari masing-masing percobaan

Data	Error	C	γ
60:40	0.0603	1	1
70:30	0.0470	1	0,1
		1	1
80:20	0.0297	1	0,5

Dari hasil percobaan yang telah dilakukan, model terbaik ada pada proporsi data 80:20 menggunakan kernel RBF dengan parameter C = 1 dan $\gamma = 0,5$.

4. Tahapan selanjutnya adalah menentukan hyperplane untuk model terbaik dengan parameter-parameter tersebut. Untuk nilai b yang diperoleh dari proses optimasi untuk hyperplane SSVM dengan fungsi kernel RBF adalah $b = 0,0879$. Sehingga persamaan untuk hyperplane (fungsi pemisah) untuk SSVM dapat ditulis sebagai berikut :

$$D(x) = w^T \cdot x + b$$

$$D(x) = \sum_{i=1}^n a_i y_i K(x, x_i) + 0,0879$$

dengan

$$K(x, x_i) = \exp(-0.5\|x - x_i\|^2)$$

5. Evaluasi Hasil Klasifikasi

Parameter yang terbaik didapatkan pada partisi data 80:20 dengan kombinasi parameter C = 1 dan $\gamma = 0,5$. Nilai parameter tersebut kemudian diterapkan pada klasifikasi data testing yang kemudian dievaluasi dengan menghitung ketepatan klasifikasinya.

Matriks konfusi :

		Kelas asli	
		Positif dm	Negatif dm
Output Klasifikasi	Positif dm	59	2
	Negatif dm	1	39

Dari matriks konfusi tersebut didapatkan akurasi sebagai berikut :

$$Akurasi = \frac{59 + 39}{59 + 2 + 1 + 39} = 0,9703$$

Dapat disimpulkan bahwa klasifikasi dengan metode SSVM yang terbaik didapatkan dengan nilai akurasi sebesar 0,9703 atau 97,03 %. Selain akurasi, tingkat ketepatan akurasi dapat diketahui dari nilai sensitivity yaitu keakuratan klasifikasi pada kelas positif dan specificity yaitu keakuratan klasifikasi pada kelas negatif.

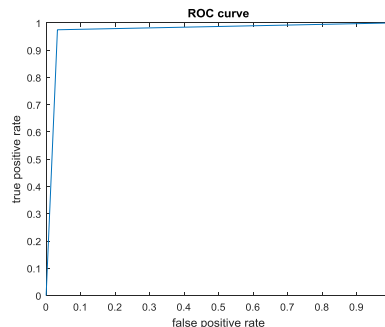
$$Sensitivity = \frac{59}{59 + 1} = 0,98333$$

$$Specificity = \frac{39}{2 + 39} = 0,95121$$

Nilai sensitivity atau akurasi klasifikasi menggunakan metode SSVM pada kelas positif adalah sebesar 98,33%, sedangkan specificity yang terbentuk adalah sebesar 95,12%. Artinya adalah akurasi klasifikasi pada masing-masing kelas positif dan negatif bernilai tinggi.

6. Kurva ROC

Kurva ROC dibentuk untuk mengetahui gambaran akurasi diskriminasi dari suatu pengujian. Ketepatan uji diterangkan pada area di bawah kurva ROC yang disebut Area Under the Curve (AUC). Nilai AUC berkisar antara 50% - 100% yang menunjukkan semakin besar nilainya maka semakin besar nilai AUC maka klasifikasi semakin baik.



Pada kurva ROC yang terbentuk, nilai AUC sebesar 97,11%. Menunjukkan bahwa klasifikasi dengan menggunakan metode SSVM dapat dilakukan dengan sangat baik dan mendapatkan nilai akurasi yang tinggi.

5. KESIMPULAN

Berdasarkan analisis yang telah dilakukan, diperoleh kesimpulan yaitu :

1. Pada pengujian klasifikasi pasien diabetes mellitus menggunakan metode SSVM untuk dengan kernel *Gaussian Radial Basis Function* (RBF) nilai akurasi dua kelas yang didapatkan adalah sebesar 97,03%, sementara nilai akurasi untuk kelas positif dan negatif masing-masing yaitu sebesar 98,33% dan 95,12%.
2. Berdasarkan kurva ROC yang terbentuk, nilai AUC yang didapatkan adalah 97,11%, artinya bahwa klasifikasi menggunakan metode SSVM dapat menghasilkan akurasi klasifikasi yang sangat baik.

6. DAFTAR PUSTAKA

- Andari, S., Purnami, S.W., Otok, B.W. (2013). *Smooth Support Vector Machine dan Multivariate Adaptive Regression Spline untuk Mendiagnosis Kanker Payudara*. Jurnal Statistika, Vol. 1, No. 2, 2013. ITS: Surabaya.
- _____. (2008). *Laporan Hasil riset kesehatan dasar nasional tahun 2007*. Departemen Kesehatan RI : Jakarta.
- Faraggi, D., Reiser, B. (2002). *Estimation of the Area Under the ROC Curve*. Journal of Statistics in Medicine. Vol.21, 2002.
- Huang, J., Jingjing, L., & Ling, C. (2003). *Comparing Naive Bayes, Decision Trees and SVM with AUC and Accuracy*. International Conference on Data Mining.
- Gunn, S.R. (1998). *Support Vector Machines for Classification and Regression*. Technical Report. University of Southampton.
- Nugroho, A.S., Arief B.W. dan Handoko, D. (2003). *Support Vector Machine : Teori dan Aplikasinya dalam Bioinformatika*. <http://ilmukomputer.com>. (diakses 10 Januari 2016).
- Prasetyo, E. (2012). *Data Mining Konsep dan Aplikasi Menggunakan MATLAB*. Andi : Yogyakarta.
- Rahman, F., & Purnami, S.W. (2012). *Klasifikasi Tingkat Keganasan Breast Cancer dengan menggunakan Regresi Logistik Ordinal dan Support Vector Machine (SVM)*. Jurnal Sains dan Seni ITS Vol.1 No.1, D130.
- Santosa, B. 2007. *Data Mining, Teori dan Aplikasi*. Jakarta: Graha Ilmu.
- Snealatha, C., dan Ramachandran, A. 2009. *Diabetes mellitus dalam gizi kesehatan masyarakat*. Jakarta: Penerbit Buku Kedokteran EGC.
- Tyasti, A.E., Ispriyanti, D., Hoyyi, A. (2015). *Algoritma Iterative Dichotomiser 3 (ID3) untuk Mengidentifikasi Data Rekam Medis*. Jurnal Gaussian, Vol.4, No.2. 2015.
- Vapnik, V. (1995). *The Nature of Statistical Learning Theory*, Springer-Verlag, NY.
- Wu, Q., and Wang, W. (2013). *Piecewise-Smooth Support Vector Machine For Classification*. Mathematical Problems in Engineering.
- Yao, Y, et.al. (2014). *An Improved Grid Search Algorithm and its Application in PCA and SVM Based Face Recognition*. Journal of Computational Information Systems. 10:3 (2014) 1219-1229. China.
- Zahtamal, Chandra, F., Suyanto, Restuastuti, T. (2007). *Faktor-Faktor Resiko Pasien Diabetes Melitus*. Jurnal Berita Kedokteran Masyarakat Vol.23, No.3, 2007. FK Universitas Riau: Riau.