

Klasifikasi Hasil *Pap Smear Test* Sebagai Upaya Pencegahan Sekunder Penyakit Kanker Serviks di Rumah Sakit “X” Surabaya Menggunakan *Piecewise Polynomial Smooth Support Vector Machine* (PPSSVM)

Mukti Ratna Dewi dan Santi Wulan Purnami
Jurusan Statistika, FMIPA, Institut Teknologi Sepuluh Nopember (ITS)
Jl. Arief Rahman Hakim, Surabaya 60111 Indonesia
e-mail: santi_wp@statistika.its.ac.id

Abstrak — Kanker serviks merupakan kanker yang menyerang leher rahim dan ditandai dengan pertumbuhan abnormal pada sel leher rahim. Berdasarkan data Riset Kesehatan Dasar tahun 2013, kanker serviks merupakan salah satu kanker yang paling sering menyerang perempuan di Indonesia. Oleh karena itu pencegahan sekunder melalui deteksi dini penting dilakukan. Penelitian ini mengangkat permasalahan klasifikasi hasil *pap smear test* di Rumah Sakit “X” Surabaya tahun 2010 yang didasarkan pada ketujuh faktor risiko, yaitu usia, usia pertama kali menstruasi, usia pertama kali melahirkan, siklus menstruasi, penggunaan alat kontrasepsi, paritas, dan riwayat keguguran menggunakan metode SSVM dengan *piecewise polynomial function 1* (PPSSVM1) dan *piecewise polynomial function 2* (PPSSVM2). Hasil penelitian menunjukkan bahwa dari 3586 pasien yang melakukan *pap smear test*, 1172 di antaranya menunjukkan hasil abnormal. Pada hasil klasifikasi *pap smear test*, metode PPSSVM1 memiliki rata-rata tingkat sensitivitas tertinggi, yaitu 91,22% pada data *training* dan 94,66% pada data *testing*. Sementara metode PPSSVM2 memiliki rata-rata nilai akurasi dan spesifisitas tertinggi, yaitu 89,22% dan 88,53% pada data *training* serta 92,84% dan 91,12% untuk data *testing*. Berdasarkan kurva ROC dan nilai AUC, kedua metode PPSSVM sama-sama baik dalam kasus pengklasifikasian hasil *pap smear test* di Rumah Sakit “X” Surabaya. Model PPSSVM1 lebih baik digunakan bila tingkat sensitivitas lebih diutamakan. Sementara bila tingkat spesifisitas lebih diutamakan maka model PPSSVM2 lebih baik untuk digunakan.

Kata Kunci — kanker serviks, *pap smear test*, klasifikasi, SSVM, PPSSVM

I. PENDAHULUAN

KANKER serviks adalah salah satu kanker yang paling sering diderita oleh perempuan di seluruh dunia dan menduduki peringkat keempat paling umum pada wanita dan secara keseluruhan ketujuh, dengan perkiraan 528.000 kasus baru pada tahun 2012 [1]. Di Indonesia, prevalensi penyakit kanker juga cukup tinggi. Berdasarkan data Riset Kesehatan Dasar tahun 2013, kanker yang paling sering menyerang perempuan adalah kanker payudara dan kanker serviks dengan 5.349 kasus (12,8%) untuk kanker serviks [2].

Berdasarkan penelitian yang telah berkembang didapatkan beberapa faktor risiko yang dapat memicu munculnya penyakit kanker serviks. Faktor risiko yang paling penting dari kanker serviks adalah infeksi Human Papilloma Virus (HPV) [3]. Di samping itu, faktor usia [4], usia saat pertama kali melahirkan [5], usia saat menstruasi pertama kali [6], jumlah anak [7], penggunaan alat kontrasepsi [8] [9] [10], dan riwayat keguguran juga merupakan beberapa faktor risiko lain dari kanker serviks [11] [10].

Kanker serviks sangat mungkin untuk dicegah dan sebuah program deteksi dini yang efektif dapat mengurangi angka kejadian dan kematian. Salah satu kegiatan deteksi dini kanker serviks yang paling umum di Indonesia adalah menggunakan metode *pap smear*. Suatu prosedur *pap smear* konvensional yang dilakukan dengan sangat baik dapat menghasilkan akurasi hasil sebesar 76% sampai dengan 94% [4]. Oleh karena itu diperlukan suatu penelitian untuk mendapatkan suatu model yang mampu mengklasifikasikan hasil *pap smear* dengan tingkat kesalahan klasifikasi yang minimal.

Salah satu metode klasifikasi yang akhir-akhir ini dikembangkan untuk meningkatkan performansi akurasi adalah SVM (*Support Vektor Machine*) yang digagas oleh Vapnik pada tahun 1995 [12]. SVM memanfaatkan optimasi dengan *quadratic programming* yang apabila digunakan untuk data berdimensi tinggi dengan jumlah besar menjadi kurang efisien [13]. Salah satu teknik yang dikembangkan untuk mengatasi permasalahan ini adalah *smoothing technique* sehingga melahirkan metode baru yang disebut *Smooth Support Vector Machine* (SSVM).

Pengembangan metode SSVM terus dilakukan dengan memodifikasi *smoothing function*. Salah satunya adalah mengubah *smoothing function* standar ke dalam bentuk *polynomial* yang memiliki beberapa formulasi, yaitu *quadratic polynomial function* dan *fourth polynomial function* yang digagas oleh Yuan dan Hang [14], *piecewise polynomial function* oleh Luo, dkk. [15], dan *spline function* oleh Yuan, dkk. [16]. Dari keempat fungsi ini, *piecewise polynomial function* memiliki pendekatan paling optimum terhadap *plus function* [17] serta memiliki performansi akurasi yang lebih baik [18]. Pada tahun 2013, Wu dan Wang mengembangkan *piecewise polynomial function* yang memiliki formulasi berbeda dari *piecewise polynomial function* yang telah ditemukan oleh Luo, dkk [19]. Dari kedua fungsi *piecewise polynomial* ini, *piecewise polynomial* yang digagas oleh Wu dan Wang secara teoritis memiliki performansi klasifikasi lebih baik [20].

Berdasarkan keunggulan kedua fungsi *piecewise polynomial* tersebut maka penelitian ini akan mencoba mendapatkan model terbaik dari metode SSVM menggunakan kedua fungsi *piecewise polynomial function* untuk klasifikasi hasil *pap smear* pada Rumah Sakit “X” Surabaya. Selanjutnya, *piecewise polynomial function* yang digagas oleh Luo, dkk. akan disebut sebagai *piecewise polynomial function 1* (PPSSVM1), sementara *piecewise polynomial function* yang ditemukan oleh Wu dan Wang disebut sebagai *piecewise polynomial function 2* (PPSSVM2).

II. TINJAUAN PUSTAKA

A. Support Vector Machine (SVM)

Misal diberikan sebuah matriks \mathbf{A} berukuran $m \times n$ yang merepresentasikan permasalahan klasifikasi sebanyak m poin dalam R^n . Keanggotaan setiap poin \mathbf{A}_i diberikan oleh matriks \mathbf{D} berukuran $m \times m$ dengan nilai 1 atau -1 pada diagonal utamanya.

Formulasi SVM untuk permasalahan linier diberikan oleh.

$$\min_{(\mathbf{w}, \gamma, \xi) \in R^{n+1+m}} v\mathbf{e}^T \xi + \frac{1}{2} \|\mathbf{w}\|_2^2 \quad (1)$$

$$s.t \quad \mathbf{D}(\mathbf{A}\mathbf{w} - \mathbf{e}\gamma) + \xi \geq \mathbf{e}, \quad \xi \geq \mathbf{0}$$

Keterangan :

- ξ : vektor *slack* nonnegatif berukuran $m \times 1$
- \mathbf{w} : vektor normal berukuran $n \times 1$
- \mathbf{e} : vektor satu berukuran $m \times 1$
- γ : parameter penentu lokasi bidang pemisah terhadap titik asal
- v : parameter positif yang menyeimbangkan bobot dari *training error* dan *margin maximization term*.

Ketika kedua kelas dapat terpisah linier secara sempurna (*linearly separable*) maka variabel *slack* bernilai nol dan bidang pemisah linier diberikan oleh.

$$\mathbf{x}^T \mathbf{w} - \gamma = 0 \quad (2)$$

Kedua kelas dipisah dengan sepasang *paralel bounding planes*, yaitu.

$$\mathbf{x}^T \mathbf{w} - \gamma = +1 \quad (3)$$

$$\mathbf{x}^T \mathbf{w} - \gamma = -1$$

Maximum margin dapat ditemukan dengan memaksimalkan jarak antara kedua *bounding planes*, yaitu $2/\|\mathbf{w}\|_2$.

Ketika kedua kelas tidak terpisah linier secara sempurna (*linearly non separable*) maka digunakan *soft margin* yang ditentukan oleh variabel *slack* ξ_i , sehingga.

$$\mathbf{x}^T \mathbf{w} - \gamma + \xi_i \geq +1 \text{ untuk } \mathbf{x}^T = \mathbf{A}_i \text{ dan } D_{ii} = +1$$

$$\mathbf{x}^T \mathbf{w} - \gamma - \xi_i \leq -1 \text{ untuk } \mathbf{x}^T = \mathbf{A}_i \text{ dan } D_{ii} = -1 \quad (4)$$

$$(i = 1, 2, \dots, m)$$

Pada umumnya banyak data yang tidak dapat dipisahkan menggunakan bidang pemisah linier namun dapat dipisahkan secara linier bila dipetakan pada ruang dimensi yang lebih tinggi menggunakan sebuah pemetaan nonlinier. Pemetaan nonlinier ini dapat diperoleh dengan memanfaatkan sebuah fungsi kernel yang memenuhi kondisi Mercer, yaitu semidefinit positif. Formulasi SVM pada permasalahan nonlinier adalah.

$$\min_{\mathbf{u} \in R^m} \frac{1}{2} \mathbf{u}^T \mathbf{D} \mathbf{K}(\mathbf{A}, \mathbf{A}^T) \mathbf{D} \mathbf{u} - \mathbf{e}^T \mathbf{u} \quad (5)$$

$$s.t. \quad \mathbf{e}^T \mathbf{D} \mathbf{u} = 0, \quad 0 \leq \mathbf{u} \leq v\mathbf{e}$$

Pemanfaatan fungsi kernel $K(\mathbf{A}, \mathbf{A}^T)$ yang bebas dari kondisi Mercer melahirkan *Generalized Support Vector Machine* (GSVM) dengan fungsi objektif sebagai berikut.

$$\min_{(\mathbf{u}, \gamma, \xi) \in R^{m+1+m}} v\mathbf{e}^T \xi + f(\mathbf{u}) \quad (6)$$

Pada persamaan (6), $f(\mathbf{u})$ adalah fungsi konveks pada R^m yang menekan parameter \mathbf{u} untuk memaksimalkan jarak antara *bounding planes*. Sementara parameter $v > 0$ memboboti *error* klasifikasi sejalan dengan proses

suppression dari parameter \mathbf{u} . Sebuah solusi dari program matematika untuk \mathbf{u} dan γ menghasilkan bidang pemisah nonlinier, yaitu.

$$K(\mathbf{x}^T, \mathbf{A}^T) \mathbf{D} \mathbf{u} = \gamma \quad (7)$$

B. Smooth Support Vector Machine (SSVM)

Pada pendekatan *smoothing* yang digagas oleh Lee dan Mangasarian [13], kuadrat 2-norm dari vektor variabel *slack* ξ diminimalkan dengan bobot $v/2$ menggantikan 1-norm dari vektor variabel *slack* ξ pada persamaan (1). Sebagai tambahan, jarak antara *bounding planes* diukur pada ruang dimensi $(n+1)$ dari $(\mathbf{w}, \gamma) \in R^{n+1}$, yaitu $2/\|(\mathbf{w}, \gamma)\|_2$ sehingga hasil modifikasi dari fungsi objektif SVM linier standar adalah sebagai berikut.

$$\min_{(\mathbf{w}, \gamma, \xi) \in R^{n+1+m}} \frac{v}{2} \xi^T \xi + \frac{1}{2} (\mathbf{w}^T \mathbf{w} + \gamma^2)$$

$$s.t \quad \mathbf{D}(\mathbf{A}\mathbf{w} - \mathbf{e}\gamma) + \xi \geq \mathbf{e}, \quad \xi \geq \mathbf{0} \quad (8)$$

Kendala pada persamaan (8) dapat ditulis menjadi

$$\xi = (\mathbf{e} - \mathbf{D}(\mathbf{A}\mathbf{w} - \mathbf{e}\gamma))_+ \quad (9)$$

di mana dalam hal ini $x_+ = \max\{0, x\}$. Bila persamaan (9) dimasukkan ke dalam persamaan (8) maka akan diperoleh fungsi objektif bebas kendala, yaitu.

$$\min_{(\mathbf{w}, \gamma) \in R^{n+1}} \frac{v}{2} \|(\mathbf{e} - \mathbf{D}(\mathbf{A}\mathbf{w} - \mathbf{e}\gamma))_+\|_2^2 + \frac{1}{2} (\mathbf{w}^T \mathbf{w} + \gamma^2) \quad (10)$$

Fungsi objektif (10) tidak memiliki turunan kedua sehingga metode optimasi konvensional tidak bisa digunakan. Oleh karena itu, Lee dan Mangasarian mengusulkan *smoothing technique* yang mengganti *plus function* x_+ pada fungsi objektif (10) dengan integral dari fungsi *sigmoid neural networks*, yaitu

$$p(x, \alpha) = x + \frac{1}{\alpha} \ln(1 + \exp(-\alpha x)) \quad (11)$$

dengan $\alpha > 0$ adalah *smooth parameter* sehingga fungsi objektif (10) menjadi [13].

$$\min_{(\mathbf{w}, \gamma) \in R^{n+1}} \frac{v}{2} \|p(\mathbf{e} - \mathbf{D}(\mathbf{A}\mathbf{w} - \mathbf{e}\gamma), \alpha)\|_2^2 + \frac{1}{2} (\mathbf{w}^T \mathbf{w} + \gamma^2) \quad (12)$$

Sementara pada kasus nonlinier, formulasi SSVM nonlinier dikembangkan dari formulasi GSVM (6) yang menggunakan fungsi kernel $K(\mathbf{A}, \mathbf{A}^T)$ untuk menentukan bidang pemisah nonlinier. Fungsi objektif dari SSVM nonlinier adalah sebagai berikut [13].

$$\min_{(\mathbf{u}, \gamma) \in R^{m+1}} \frac{v}{2} \|p(\mathbf{e} - \mathbf{D}(K(\mathbf{A}, \mathbf{A}^T) \mathbf{D} \mathbf{u} - \mathbf{e}\gamma), \alpha)\|_2^2 + \frac{1}{2} (\mathbf{u}^T \mathbf{u} + \gamma^2) \quad (13)$$

Pada penelitian ini, fungsi kernel yang digunakan adalah fungsi kernel Gaussian yang memiliki formulasi

$$K(\mathbf{A}, \mathbf{A}^T)_{ij} = \exp\left(-\mu \|\mathbf{A}_i^T - \mathbf{A}_j^T\|_2^2\right) \quad (14)$$

di mana μ merupakan parameter dari fungsi kernel dengan $i, j = 1, 2, \dots, m$.

Bidang pemisah nonlinier ditentukan oleh solusi optimum dari \mathbf{u} dan γ dari persamaan (13) yang dapat diselesaikan dengan algoritma Newton Armijo sehingga *decision function* $f(\mathbf{x})$ yang terbentuk untuk SSVM nonlinier adalah sebagai berikut [13].

$$f(\mathbf{x}) = \text{sign}(K(\mathbf{x}^T, \mathbf{A}^T) \mathbf{D} \mathbf{u} - \gamma) \quad (15)$$

C. Polynomial Function

Salah satu hasil modifikasi *smooth function* yang dikembangkan oleh beberapa peneliti adalah *polynomial smoothing function*. Beberapa jenis dari *polynomial smoothing function* adalah.

- Quadratic Polynomial Function* [14].
- Fourth Polynomial Function*, [21].
- Spline Function* [16].
- Piecewise Polynomial Function* [15] [19].

Pada penelitian ini, *polynomial function* yang digunakan adalah *piecewise polynomial function* yang memiliki dua formulasi berbeda. Formulasi pertama ditemukan oleh Luo, dkk. pada tahun 2006 [15].

$$f_1(x, \alpha) = \begin{cases} x, & x \geq \frac{1}{\alpha} \\ \frac{1}{32}(\alpha x + 1)^3, & -\frac{1}{\alpha} < x < \frac{1}{\alpha} \\ \left(\alpha^2 x^3 - 3\alpha x^2 + x + \frac{5}{\alpha}\right), & \\ 0, & x \leq -\frac{1}{\alpha} \end{cases} \quad (16)$$

Sementara formulasi kedua ditemukan oleh Wu dan Wang pada tahun 2013 [19].

$$f_2(x, \alpha) = \begin{cases} 0, & x < -\frac{1}{3\alpha} \\ \frac{3}{2}\alpha^2 \left(x + \frac{1}{3\alpha}\right)^3, & -\frac{1}{3\alpha} \leq x < 0 \\ x + \frac{3}{2}\alpha^2 \left(\frac{1}{3\alpha} - x\right), & 0 \leq x \leq \frac{1}{3\alpha} \\ x, & x > \frac{1}{3\alpha} \end{cases} \quad (17)$$

Pada penelitian ini, *smoothing parameter* yang digunakan adalah $\alpha = 10$ dengan acuan penelitian Wu dan Wang pada tahun 2013 [19].

D. Seleksi Parameter

Pada tahun 2007, Huang, dkk. mengembangkan suatu metode untuk *model selection* pada SVM yang memungkinkan pengguna menemukan kombinasi parameter yang baik secara efisien dan otomatis, disebut *Uniform Design* (UD).

Pada SVM, seleksi parameter dilakukan terhadap parameter fungsi kernel dan parameter penyeimbang SVM, yaitu ν . Fungsi kernel yang paling sering digunakan dalam pemetaan nonlinier adalah kernel gaussian yang memiliki parameter μ . Huang, dkk. [22] menyatakan bahwa nilai dari sebuah parameter kernel gaussian yang baik berada dalam rentang

$$\frac{10^{-3}}{\rho} < \mu < \frac{1,9}{\rho}$$

di mana $\rho = \min_{A_i \neq A_j} \|A_i^T - A_j\|$. Ketika $\mu < 10^{-3}/\rho$ maka akan

terjadi kondisi *underfitting*, sementara bila $\mu > 1,9/\rho$ maka akan terjadi kondisi *overfitting*. Di sisi lain, sebuah parameter penyeimbang ν yang baik berada dalam rentang

$$10^{-2} \leq \nu \leq 10^4.$$

E. Performasi Klasifikasi

Performasi klasifikasi dapat dilihat melalui tingkat akurasi, sensitivitas, dan spesifisitas. Akurasi menunjukkan performasi teknik klasifikasi secara keseluruhan. Sensitivitas adalah tingkat akurasi kelas positif sementara spesifisitas adalah tingkat akurasi kelas negatif [23].

Tabel 1. Tabel Klasifikasi

Kondisi yang Sebenarnya	Hasil Prediksi	
	Positif	Negatif
Positif	TP	FN
Negatif	FP	TN

Penghitungan akurasi, sensitivitas, dan spesifisitas adalah sebagai berikut.

$$a) \text{ Akurasi} = \frac{TN+TP}{TN+TP+FN+FP} \quad (18)$$

$$b) \text{ Sensitivitas} = \frac{TP}{TP+FN} \quad (19)$$

$$c) \text{ Spesifisitas} = \frac{TN}{TN+FP} \quad (20)$$

F. Kurva Receiving Operating Characteristics (ROC)

Kurva ROC (*Receiver Operating Characteristics*) adalah representasi grafis dari hubungan timbal balik antara sensitivitas dan spesifisitas [24] di mana kurva ini sering digunakan untuk mengevaluasi pengklasifikasian karena mempunyai kemampuan secara menyeluruh dan cukup baik [25].

Metode yang umum digunakan untuk menghitung performasi klasifikasi adalah dengan menghitung luas daerah di bawah kurva ROC, yaitu *Area Under The ROC Curve* (AUC). Terdapat dua metode dalam mengestimasi luasan AUC, yaitu metode parametrik dan nonparametrik. Metode nonparametrik lebih disukai oleh banyak peneliti karena tidak memerlukan asumsi distribusi pada data penelitian.

Misalkan Y merupakan sebuah variabel random yang mewakili hasil tes diagnosis biner maka diagnosis berdasarkan sembarang *cutoff* c adalah positif bila $Y \geq c$ dan negatif jika $Y < c$. Nilai *cutoff* c dihitung menggunakan rumusan sebagai berikut

$$c_1 = kk - 1; c_2 = \frac{kk + kb}{2}; c_3 = kb + 1 \quad (21)$$

dengan kk adalah kelas terkecil (0) dan kb adalah kelas terbesar (1) pada hasil tes diagnosis.

Penghitungan sensitivitas (TP) dan 1-spesifisitas (FP) pada sembarang *cutoff* c diberikan oleh

$$TP(c) = \frac{s_1(c)}{n_1}; FP(c) = \frac{s_0(c)}{n_0} \quad (22)$$

di mana

$s_1(c)$: jumlah subjek $Y \geq c$ pada kelas 1

$s_0(c)$: jumlah subjek $Y \geq c$ pada kelas 0

n_1 : jumlah kejadian kelas 1

n_0 : jumlah kejadian kelas 0.

Estimasi AUC menggunakan metode nonparametrik dihitung menggunakan *trapezoidal rule* dengan rumusan sebagai berikut [26].

$$AUC = \frac{1}{n_1 n_0} \sum_{i=1}^{n_1} \sum_{j=1}^{n_0} \varphi(Y_{i1}, Y_{j0}) \quad (23)$$

di mana

$$\varphi(Y_{i1}, Y_{j0}) = \begin{cases} 1 & \text{jika } Y_{i1} > Y_{j0} \\ 1/2 & \text{jika } Y_{i1} = Y_{j0} \\ 0 & \text{jika } Y_{i1} < Y_{j0} \end{cases} \quad (24)$$

serta Y_{i1} dan Y_{j0} masing-masing merupakan hasil tes diagnostik untuk kelas positif dan negatif.

Kategori pengklasifikasian model berdasarkan nilai AUC diberikan oleh Tabel 2.

Tabel 2. Kategori Pengklasifikasian Model Berdasarkan Nilai AUC

Nilai AUC	Model Diklasifikasikan Sebagai
0,90-1,00	<i>Excellent</i>
0,80-0,90	<i>Good</i>
0,70-0,80	<i>Fair</i>
0,60-0,70	<i>Poor</i>
0,50-0,60	<i>Fail</i>

G. Kanker Serviks

Kanker serviks merupakan kanker yang meyerang leher rahim. Faktor risiko yang paling penting dari kanker serviks adalah infeksi *Human Papilloma Virus* (HPV) [3]. Di samping itu, faktor usia [4], usia saat pertama kali melahirkan [5], usia saat menstruasi pertama kali [6], jumlah anak [7], penggunaan alat kontrasepsi [8] [9] [10], dan riwayat keguguran juga merupakan beberapa faktor risiko lain dari kanker serviks [11] [10].

Salah satu tes skrining yang berguna untuk mendeteksi kanker serviks adalah *pap smear test*. *Pap smear* merupakan pemeriksaan sitologi yang digunakan untuk mengecek perubahan pada leher rahim di bagian atas vagina. Tes ini merupakan alat *screening* untuk menemukan gejala dini mengenai perkembangan sel kanker di masa mendatang. Secara umum, terdapat dua kemungkinan hasil dari *pap smear test*, yaitu normal dan abnormal. Jika perubahan abnormal ditemukan pada *screening*, tes lebih lanjut akan dilakukan untuk melihat urgensi pengobatan yang dibutuhkan [27].

III. METODOLOGI PENELITIAN

A. Sumber Data dan Variabel Penelitian

Data yang digunakan pada penelitian ini adalah data sekunder yang diperoleh dari Rumah Sakit "X" Surabaya bagian Riset dan Pengembangan tahun 2010. Data mengenai hasil *pap smear test* dan faktor-faktor yang diduga mempengaruhi penyakit kanker serviks berjumlah 3586 kasus dengan rincian jumlah hasil tes yang normal sebanyak 2414 kasus dan abnormal sebanyak 1172 kasus.

Variabel respon merupakan hasil *pap smear test* berdasarkan gambaran sitologi yang terdiri atas dua kelas, yaitu.

Y = 0 : normal

Y = 1 : abnormal.

Tabel 3. Gambaran Sitologi Hasil *Pap Smear Test*

Kelompok	Jenis	
Normal	Normal smear	
	Radang non spesifik	TV
	AS (<i>Atrophic Smear</i>)	HV
	ASC-US	Candida
	ASC-H	Cocccen
	Early as	LSIL
	HSIL	Mild RNS
Abnormal	Herpes Simplek Virus	Metaplasia

Sementara itu variabel prediktor (X) merupakan faktor risiko dari kanker serviks seperti yang disajikan dalam Tabel 4.

Tabel 4. Variabel Prediktor

	Variabel	Deskripsi	Skala
X ₁	Usia (tahun)	Usia dari pasien	Rasio
X ₂	Usia pertama kali menstruasi (tahun)	-	Rasio
X ₃	Usia pertama kali melahirkan (tahun)	-	Rasio
X ₄	Siklus menstruasi	1 : Teratur 2 : Tidak Teratur	Nominal
X ₅	Paritas	1 : ≤ 2 anak 2 : > 2 anak	Nominal
X ₆	Penggunaan alat kontrasepsi	1 : Tidak 2 : Ya	Nominal
X ₇	Riwayat keguguran	1 : Pernah 2 : Tidak pernah	Nominal

B. Langkah Analisis

Langkah analisis dalam penelitian ini secara garis besar terdapat enam tahapan, yaitu.

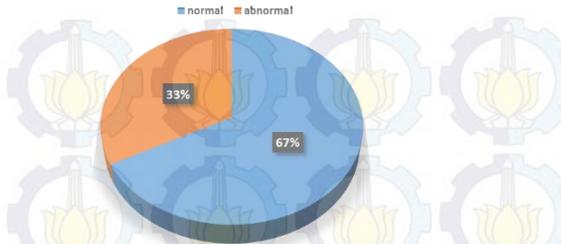
- Menjelaskan karakteristik pasien *pap smear test* berdasarkan faktor risiko yang mempengaruhi kanker serviks.
 - Menghitung ukuran pemusatan dan penyebaran data pada data faktor risiko dengan skala rasio.
 - Membuat tabel kontingensi dan *pie chart*, serta menghitung *odds ratio* pada data faktor risiko berskala nominal.
- Melakukan seleksi parameter pada fungsi kernel Gaussian dan parameter penyeimbang SVM menggunakan *Uniform Design* pada metode SSVM, PPSSVM1, dan PPSSVM2.
- Melakukan optimasi terhadap fungsi objektif SSVM, PPSSVM1, dan PPSSVM2 menggunakan Newton Armidjo.
- Membangun model SSVM, PPSSVM1, dan PPSVM2 untuk klasifikasi deteksi dini kanker serviks.
- Melakukan evaluasi performansi model SSVM dan PPSVM yang telah didapatkan untuk klasifikasi deteksi dini kanker serviks. Langkah-langkahnya adalah sebagai berikut.
 - Membagi data menjadi dua bagian, yaitu data *training* dan data *testing* dengan menggunakan *5-fold cross validation*.
 - Menghitung akurasi, sensitivitas, dan spesifisitas.
 - Membuat kurva ROC.
 - Menghitung nilai AUC.
- Menentukan model terbaik antara SSVM, PPSSVM1, dan PPSSVM2 untuk klasifikasi deteksi dini kanker serviks.

IV. ANALISIS DAN PEMBAHASAN

A. Karakteristik Pasien Berdasarkan Faktor-Faktor Risiko Penyebab Kanker Serviks

Pada pembahasan awal, gambaran karakteristik pasien yang melakukan *pap smear test* akan diberikan berdasarkan ketujuh faktor risiko kanker serviks. Data yang digunakan pada tahapan ini merupakan data asli hasil *pap smear test* sebelum dilakukan metode *balancing* data.

Distribusi kejadian pasien dengan hasil *pap smear test* normal dan abnormal dapat dilihat pada Gambar 1.



Gambar 1. Persentase Hasil *Pap Test* Normal dan Abnormal

Berdasarkan Gambar 1 dapat dilihat bahwa proporsi hasil *pap smear test* normal dan abnormal tidak seimbang. Hasil *pap test* abnormal memiliki persentase sebesar 67% atau setara dengan 2414 kejadian. Sementara sisanya, yaitu 1172 kejadian, merupakan jumlah kejadian dengan hasil *pap smear test* abnormal

Karakteristik pasien berdasarkan faktor risiko kanker serviks berupa usia, usia pertama kali menstruasi, dan usia pertama kali melahirkan ditampilkan pada Tabel 5.

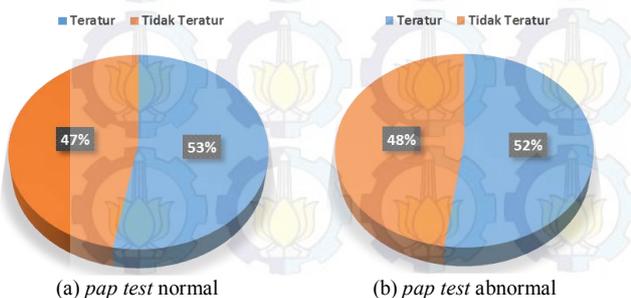
Tabel 5. Statistika Deskriptif Pada Kelompok Hasil *Pap Smear*

	Usia		Usia Pertama Kali Menstruasi		Usia Pertama Kali Melahirkan	
	(N)	(A)	(N)	(A)	(N)	(A)
Minimum	20	14	8	9	12	12
Maksimum	80	71	19	18	49	45
Rata-Rata	43,35	40,38	12,98	12,84	25,64	26,17
Deviasi Std.	10,00	6,78	1,48	1,37	3,99	4,03

*(N) = hasil normal; (A) = hasil abnormal

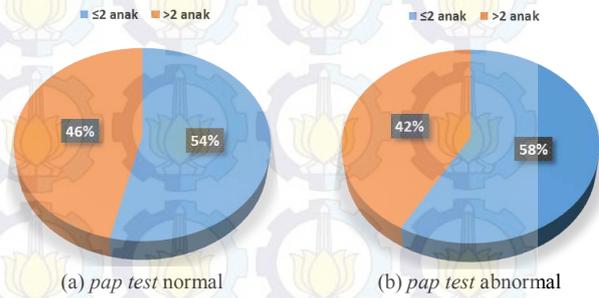
Berdasarkan faktor risiko usia pada Tabel 5, pasien kelompok *pap smear* normal memiliki rata-rata usia dan deviasi standar yang lebih besar dari pasien pada kelompok hasil abnormal. Nilai deviasi standar yang lebih besar ini menunjukkan bahwa usia pasien pada kelompok hasil normal lebih beragam daripada usia pasien pada kelompok hasil abnormal. Sementara itu untuk karakteristik pasien berdasarkan faktor usia pertama kali menstruasi, pasien pada kedua kelompok hasil *pap smear test* memiliki rata-rata yang hampir sama, yaitu sekitar 12 tahun dengan sebaran usia pada kelompok normal yang sedikit lebih tinggi dari pasien pada kelompok abnormal. Usia pertama kali melahirkan pada kelompok hasil *pap smear* normal maupun abnormal juga memiliki yang hampir sama, yaitu sekitar 25 dan 26 tahun, dengan sebaran data pada kelompok abnormal lebih tinggi dari kelompok normal.

Selanjutnya, karakteristik pasien berdasarkan faktor siklus menstruasi, paritas, penggunaan alat kontrasepsi, dan riwayat keguguran digambarkan secara visual oleh *pie chart* berikut ini.



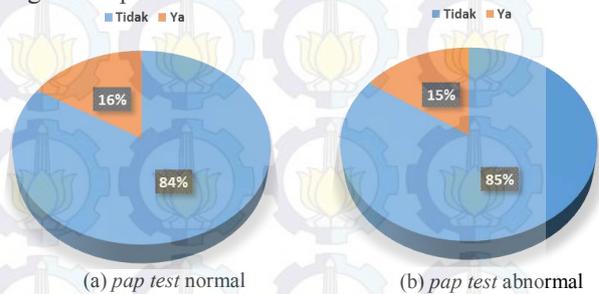
Gambar 2. Karakteristik Berdasarkan Faktor Siklus Menstruasi

Pada Gambar 2 terlihat bahwa pasien cenderung memiliki siklus menstruasi yang teratur baik pada kelompok hasil *pap smear test* normal maupun abnormal dengan distribusi kejadian sebesar 1282 pasien pada kelompok normal dan 606 pada kelompok abnormal.



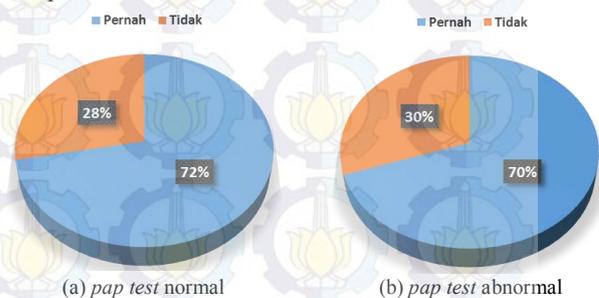
Gambar 3. Karakteristik Berdasarkan Faktor Paritas

Berdasarkan faktor paritas, pasien yang pernah melahirkan lebih dari dua kali adalah sebesar 46% atau setara dengan 1123 pasien untuk hasil *pap smear test* normal, sementara untuk hasil *pap test* abnormal adalah sebesar 42% atau setara dengan 484 pasien.



Gambar 4. Karakteristik Berdasarkan Faktor Penggunaan Alat Kontrasepsi

Pada kedua kelompok hasil *pap smear test*, pasien yang tidak menggunakan alat kontrasepsi memiliki persentase yang lebih tinggi dari pasien yang menggunakan alat kontrasepsi, yaitu 84% pada kelompok normal dan 85% pada kelompok abnormal.



Gambar 5. Karakteristik Berdasarkan Faktor Riwayat Keguguran

Berdasarkan Gambar 5, persentase pasien yang pernah mengalami keguguran lebih besar dari persentase pasien yang tidak pernah mengalami keguguran dengan distribusi kejadian pada kelompok normal adalah 1745 pasien dan 809 pasien pada kelompok abnormal.

Nilai *odds ratio* dari dari masing-masing faktor risiko kanker serviks berskala nominal ditampilkan oleh Tabel 6.

Tabel 6. *Odds Ratio* dari Faktor Risiko

Faktor Risiko	<i>Odss Ratio</i>
Siklus Menstruasi	1,024
Paritas	0,831
Penggunaan Alat Kontrasepsi	0,938
Riwayat Keguguran	0,899

Berdasarkan Tabel 6, faktor risiko siklus menstruasi, penggunaan alat kontrasepsi, dan riwayat keguguran memiliki nilai *odds ratio* mendekati satu. Hal ini menandakan bahwa faktor-faktor tersebut hampir tidak memiliki pengaruh terhadap hasil *pap smear test*.

B. Klasifikasi Hasil Pap Smear Test Menggunakan Metode PPSSVM

Data yang digunakan dalam permasalahan klasifikasi hasil *pap smear test* merupakan data yang telah mengalami *balancing* sehingga proporsi antara kelas positif dan negatif pada data tidak berbeda jauh. *Balancing* pada data dilakukan untuk meningkatkan performansi klasifikasi. Selengkapannya mengenai proses *balancing* pada data hasil *pap smear test* di Rumah Sakit "X" Surabaya tahun 2010 dapat dilihat pada laporan *tesis* Statistika ITS oleh Hartayuni Sain pada tahun 2013 [28]. Setelah mengalami proses *balancing*, jumlah kelompok hasil *pap smear test* normal menjadi 2414 dan 2344 untuk kelompok hasil abnormal.

Langkah pertama dalam pembentukan model klasifikasi menggunakan metode SSVM dan PPSSVM adalah melakukan seleksi parameter μ dan ν menggunakan *Uniform Design* (UD). Hasilnya disajikan dalam Tabel 7.

Tabel 7. Hasil dan Waktu Komputasi Seleksi Parameter

Metode	Parameter		Waktu Komputasi			
	μ	ν	Hari	Jam	Menit	Detik
SSVM	0,2880	10	0	14	37	21
PPSSVM1	0,7897	0,3109	5	17	21	40
PPSSVM2	0,7897	0,3122	3	6	7	30

Terlihat pada Tabel 7 bahwa secara umum metode PPSSVM memakan waktu komputasi paling lama dibandingkan dengan metode SSVM. Perbedaan waktu komputasi ini dipengaruhi oleh jumlah data yang cukup besar serta kerumitan fungsi objektif dari masing-masing metode. Selanjutnya, parameter-parameter ini dimasukkan ke dalam fungsi objektif (13) dan kemudian dilakukan proses optimasi menggunakan algoritma *Newton Armidjo* untuk mendapatkan nilai u dan γ yang optimum. Nilai u dan γ yang optimum ini berguna untuk membentuk *classifier* (15).

Pada proses evaluasi klasifikasi, data akan dipartisi menjadi data *training* dan *testing* menggunakan *5-fold cross validation*. Hasilnya dapat dilihat pada Tabel 8 dan Tabel 9.

Tabel 8. Evaluasi Performansi Klasifikasi Pada Data *Training* (%)

Evaluasi Performansi	Fold	SSVM	PPSSVM1	PPSSVM2
Akurasi	1	82,51	88,05	88,21
	2	82,51	88,16	87,89
	3	82,48	87,74	87,84
	4	82,83	87,74	88,29
	5	88,10	93,88	93,88
Rata-rata		83,69	89,11	89,22
Sensitivitas	1	84,70	90,03	89,07
	2	84,91	90,46	88,65
	3	84,97	89,82	88,49
	4	84,91	89,87	88,43
	5	90,72	95,90	95,04
Rata-rata		86,04	91,22	89,94
Spesifisitas	1	80,38	86,13	87,37
	2	80,18	85,92	87,16
	3	80,07	85,71	87,22
	4	80,80	85,66	88,15
	5	85,56	91,93	92,75
Rata-rata		81,40	87,07	88,53

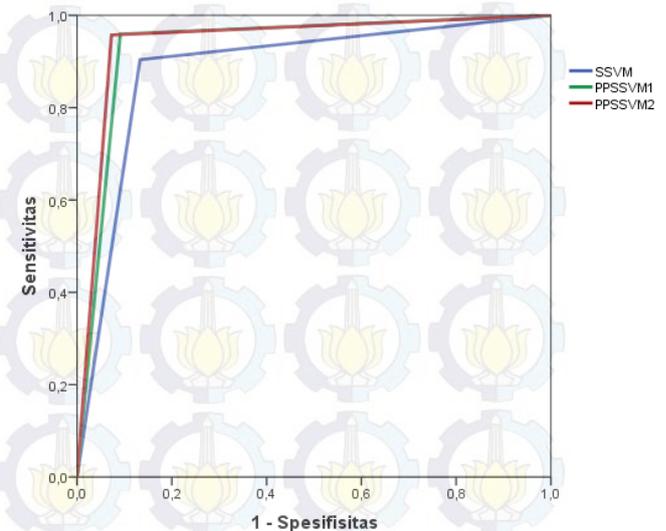
Tabel 9. Evaluasi Performansi Klasifikasi Pada Data *Testing* (%)

Evaluasi Performansi	Fold	SSVM	PPSSVM1	PPSSVM2
Akurasi	1	88,42	93,37	92,74
	2	88,42	92,95	94,00
	3	88,53	94,63	94,21
	4	87,16	94,63	92,42
	5	66,00	88,21	90,84
Rata-rata		83,71	92,76	92,84
Sensitivitas	1	91,45	95,94	93,38
	2	90,60	94,23	95,09
	3	90,38	96,79	95,73
	4	90,60	96,58	95,94
	5	67,31	89,74	92,95
Rata-rata		86,07	94,66	94,62
Spesifisitas	1	85,48	90,87	92,12
	2	86,31	91,70	92,95
	3	86,72	92,53	92,74
	4	83,82	92,74	89,00
	5	64,73	86,72	88,80
Rata-rata		81,41	90,91	91,12

Keterangan: Nilai-nilai yang dicetak tebal menunjukkan akurasi, *sensitivity*, *specificity* tertinggi pada setiap kelompok data.

Secara keseluruhan metode PPSSVM1 dan PPSSVM2 memiliki performansi yang lebih baik dibandingkan dengan metode SSVM. Sebagaimana yang ditampilkan pada Tabel 8 dan Tabel 9, PPSSVM1 memiliki rata-rata sensitivitas yang paling baik pada data *training* maupun *testing* dengan tingkat sensitivitas masing-masing sebesar 91,22% dan 94,66%. Sementara PPSSVM2 memiliki rata-rata nilai akurasi dan spesifisitas tertinggi pada data *training* dan *testing*. Rata-rata nilai akurasi metode PPSSVM2 adalah sebesar 89,22% pada data *training* dan 92,84% pada data *testing*. Sedangkan untuk rata-rata nilai spesifisitas, pada data *training* PPSSVM2 memiliki nilai sebesar 88,53% dan 91,12% untuk data *testing*.

Performansi dari ketiga model juga dapat dilihat menggunakan kurva ROC yang dapat dilihat pada Gambar 6. Sementara nilai AUC pada masing-masing metode dapat dilihat pada Tabel 10.



Gambar 6. Kurva ROC dari SSVM, PPSSVM1, dan PPSSVM2

Tabel 10. Nilai AUC dari metode SSVM, PPSSVM1, dan PPSSVM2

SSVM	PPSSVM1	PPSSVM2
0,886	0,934	0,942

Berdasarkan nilai AUC yang dihasilkan, model PPSSVM1 dan PPSSVM2 masuk dalam kategori *excellent*, sementara model SSVM masuk dalam kategori *good*. Nampak pada

Tabel 10 bahwa metode SSVM memiliki nilai AUC paling kecil bila dibandingkan dengan metode PPSSVM. Hal ini dikarenakan dari ketiga kurva ROC yang terbentuk, model PPSSVM1 dan PPSSVM2 memiliki kurva ROC yang paling mendekati sudut kiri atas dibandingkan dengan model SSVM. Bagian atas kurva ROC dari PPSSVM1 lebih mendekati sudut atas dari bidang plot menunjukkan bahwa PPSSVM1 memiliki nilai sensitivitas yang lebih tinggi dari PPSSVM2.

Berdasarkan analisis kurva ROC dan nilai AUC, dapat dikatakan bahwa kedua metode PPSSVM sama-sama baik dalam kasus pengklasifikasian hasil *pap smear test* di Rumah Sakit "X" Surabaya Tahun 2010. Metode PPSSVM2 mungkin memiliki nilai AUC yang lebih besar dari PPSSVM1. Namun, karena selisih dari kedua nilai AUC tersebut sangat kecil maka hal ini bisa diabaikan. Kedepannya, penggunaan model klasifikasi yang dibangun dengan metode PPSSVM1 dan PPSSVM2 dapat disesuaikan dengan kondisi yang diinginkan. Ketika tingkat sensitivitas yang tinggi lebih diutamakan maka model klasifikasi PPSSVM1 disarankan untuk digunakan. Sementara bila tingkat spesifisitas yang tinggi merupakan instrumen yang lebih penting maka model PPSSVM2 lebih disarankan untuk digunakan.

V. KESIMPULAN / RINGKASAN

Secara keseluruhan metode PPSSVM1 dan PPSSVM2 memiliki performansi yang lebih baik daripada metode SSVM. Metode PPSSVM1 memiliki rata-rata sensitivitas tertinggi, sementara metode PPSSVM2 memiliki rata-rata akurasi dan spesifisitas paling tinggi. Kedua metode PPSSVM sama-sama baik dalam kasus pengklasifikasian hasil *pap smear test* di Rumah Sakit "X" Surabaya Tahun 2010. Model PPSSVM1 lebih baik digunakan bila tingkat sensitivitas lebih diutamakan. Sementara bila tingkat spesifisitas lebih diutamakan maka model PPSSVM2 lebih baik untuk digunakan.

Metode PPSSVM memiliki performansi yang sangat baik, namun sangat tidak efisien dalam waktu komputasi, terutama pada bagian seleksi parameter. Oleh karena itu, pada penelitian selanjutnya dapat dicoba beberapa alternatif metode seleksi parameter yang lebih efisien dalam waktu komputasi.

DAFTAR PUSTAKA

- [1] Globocan, "GLOBOCAN 2012 : Estimated Cancer Incidence, Mortality and Prevelence Worldwide in 2012," 2012. [Online]. Available: http://globocan.iarc.fr/Pages/fact_sheets_cancer.aspx.
- [2] Dinas Kesehatan, "Hilangkan Mitos Tentang Kanker," Kementerian Kesehatan Republik Indonesia, Jakarta, 2014.
- [3] IARC, "List of Classifications by Cancer Sites with Sufficient or Limited Evidence in Humans," vol. 1, 2014.
- [4] H. Nurwijaya, Andrijono dan H. Suheimi, Cegah dan Deteksi Kanker Serviks, Jakarta: Elex Media, 2010.
- [5] Cancer Research UK, "Cervical Cancer Risk Factors," 8 Agustus 2014. [Online]. Available: <http://www.cancerresearchuk.org/cancer-info/cancerstats/types/cervix/riskfactors/cervical-cancer-risk-factors>.
- [6] I. Yakasai, E. Ugwa dan J. Otubu, "Gynecological Malignancies in Aminu Kano Teaching Hospital Kano : A 3 Years Review," *Nigerian Journal of Clinical Practice*, vol. 16, no. 1, pp. 63-66, 2 Februari 2013.
- [7] R. T. Fatmawati, "Hubungan Paritas dengan Kejadian Kanker Serviks di Ruang Merak RSUD dr. Soetomo Surabaya," Surabaya, 2012.
- [8] S. Sjamsuddin, "Prevention and Early Detection of Cervical Cancer," *Cermin Dunia Kedokteran*, no. 133, pp. 9-14, 2001.
- [9] S. Dalimartha, Deteksi Dini Kanker dan Simplisia Antikanker, Jakarta: Penebar Swadaya, 2004.
- [10] D. S. Tira, "Risiko Jumlah Perkawinan, Riwayat Abortus, dan Pemakaian Alat Kontrasepsi Hormonal Terhadap Kejadian Kanker Serviks di Rumah Sakit Pelamonia Makassar Tahun 2006-2007," 2008.
- [11] A. Abbas, "Beberapa Faktor Risiko Kanker Serviks di Perjan RSUP. dr. Wahidin Sudirohusoso Makassar Tahun 2012," Makassar, 2003.
- [12] C. Cortes dan V. Vapnik, "Support-Vector Networks," *Machine Learning*, vol. 20, no. 3, pp. 273-297, 1 September 1995.
- [13] Y. J. Lee dan O. L. Mangasarian, "SSVM: A Smooth Support Vector Machine for Classification," *Computational Optimization and Applications*, vol. 20, pp. 5-22, 2001.
- [14] Y. Yuan dan T. Huang, "A Polynomial Smooth Support Vector Machine for Classification," *Advanced Data Mining and Applications*, vol. 3584, pp. 157-164, 22-24 July 2005.
- [15] L. Luo, H. Peng dan Q. Zhou, "A Study on Piecewise Polynomial Smooth Approximation to The Plus Function," dalam *International Conference on Control, Automation, Robotics and Vision*, Singapura, 2006.
- [16] Y. Yuan, W. Fan dan D. Pu, "Spline Function Smooth Support Vector Machine for Classification," *Journal of Industrial and Management Optimization (JIMO)*, vol. 3, no. 3, pp. 529-542, Agustus 2007.
- [17] S. Purnami, A. Embong, J. Zain dan S. Rahayu, "A Comparison of Smoothing Function in Smooth Support Vector Machine," dalam *International Conference on Software Engineering and Computer Systems*, 2009a.
- [18] S. Purnami, A. Embong, J. Zain dan S. Rahayu, "Application of Data Mining Technique Using Best Polynomial Smooth Support Vector Machine in Breast Cancer Diagnosis," dalam *International Conference in Robotic, Vision, Signal Symposium and Power Application (Rovsip)*, Langkawi Kedah, Malaysia, 2009b.
- [19] Q. Wu dan W. Wang, "Piecewise-Smooth Support Vector Machine for Classification," *Hindawi Publishing Corporation Mathematical Problems in Engineering*, 2013.
- [20] I. Wulandari, "Piecewise Polynomial Smooth Support Vector Machine Untuk Klasifikasi Desa Tertinggal di Provinsi Kalimantan Timur Tahun 2011," ITS Press, Surabaya, 2014.
- [21] Y. Yuan, J. Yan dan C. Xu, "Polynomial Smooth Support Vector Machine (PSSVM)," *Chinese Journal of Computers*, vol. 28, pp. 9-17, 2005.
- [22] C. M. Huang, Y. J. Lee, D. K. J. Lin dan S. Y. Huang, "Model Selection for Support Vector Machines via Uniform Design," *Computational Statistics & Data Analysis*, vol. 52, no. 1, pp. 335-346, 1 Februari 2007.
- [23] W. Zhu, N. Zeng dan N. Wang, "Sensitivity, Specificity, Accuracy, Associated Confidence Interval and ROC Analysis with Pratical SAS Implementations," dalam *NESUG proceedings : Health Care and Life Sciences*, 2010.
- [24] A. R. V. Erke dan P. M. T. Pattinama, "Receiver operating characteristic (ROC) analysis: Basic principles and applications in radiology," *European Journal of Radiology*, pp. 88-94, 1998.
- [25] S. Chou, J. Shan, Y. Guo dan L. Zhang, "Automated Breast Cancer Detection and Classification Using Ultrasound Image : A Survey, Pattern Recognition," vol. 43, pp. 299-317, 2010.
- [26] E. Colak, F. Mutlu, C. Bal, S. Oner, K. Ozdamar, B. Gok dan Y. Cavusoglu, "Comparison of Semiparametric, Parametric, and Nonparametric ROC Analysis for Continuous Diagnostic Tests Using a Simulation Study and Acute Coronary Syndrome Data," *Computational and Mathematical Methods in Medicine*, vol. 2012, p. 7, 2012.
- [27] Australian Department of Health, "National Cervical Screening Program," 19 November 2013. [Online]. Available: <http://www.cancerscreening.gov.au/internet/screening/publishing.nsf/Content/papsmear#1>. [Diakses 21 Agustus 2014].
- [28] H. Sain, "Combine Sampling Support Vector Machine untuk Klasifikasi Data Imbalanced," ITS Press, Surabaya, 2013.