

# Klasifikasi Pasien Hasil *Pap Smear Test* sebagai Pendeteksi Awal Upaya Penanganan Dini pada Penyakit Kanker Serviks di RS. “X” Surabaya dengan Metode *Bagging Logistic Regression*

Ida Ayu Sevita Intansari, Santi Wulan Purnami, dan Sri Pingit Wulandari

Jurusan Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Institut Teknologi Sepuluh Nopember (ITS)

Jl. Arief Rahman Hakim, Surabaya 60111

E-mail: santi\_wp@statistika.its.ac.id, sri\_pingit@statistika.its.ac.id

**Abstrak**— Kanker serviks merupakan pertumbuhan dari suatu kelompok sel yang tidak normal pada serviks (mulut rahim). Menurut WHO (2003), Indonesia merupakan negara dengan penderita kanker mulut rahim nomor satu di dunia. Kanker serviks adalah kanker yang paling sering dijumpai pada wanita setelah kanker payudara dan dapat menyebabkan kematian. Oleh karena itu, pasien hasil pap smear test sebagai pendeteksi awal upaya penanganan dini pada penyakit kanker serviks di RS. “X” Surabaya guna melihat karakteristik dan faktor resiko yang paling berpengaruh terhadap pendeteksi awal kanker serviks. Berdasarkan hasil penelitian diperoleh analisis deskriptif yang menunjukkan bahwa lebih banyak pasien yang memiliki hasil test abnormal dibandingkan dengan yang normal menggunakan Pap Smear Test. Pada analisis regresi logistik diperoleh tiga variabel prediktor yang signifikan berpengaruh terhadap hasil test, yaitu usia pasien, jumlah anak (paritas) dan penggunaan kontrasepsi, dengan nilai ketepatan klasifikasi sebesar 66,7%. Nilai ketepatan klasifikasi tersebut dapat ditingkatkan dengan pendekatan bagging regresi logistik (bagging logistic regression). Hasil analisis bagging regresi logistik pada 2 jenis kombinasi data, menunjukkan bahwa pada 80 kali replikasi bootstrap diperoleh nilai ketepatan klasifikasi meningkat sebesar 1,11%..

**Kata Kunci**— bagging, kanker serviks, ketepatan klasifikasi, regresi logistik.

## I. PENDAHULUAN

KANKER serviks merupakan pertumbuhan dari suatu kelompok sel yang tidak normal pada serviks (mulut rahim). Perubahan ini biasanya memakan waktu beberapa tahun sebelum berkembang menjadi kanker. Menurut WHO (2003) Indonesia merupakan negara dengan penderita kanker mulut rahim nomor satu di dunia.

Penelitian terhadap kanker serviks pernah dilakukan oleh Nurdin[1], Melva[2] dan Yastuti[3] untuk mengetahui faktor-faktor yang mempengaruhi kejadian kanker serviks. Penelitian menggunakan metode *Bagging* Regresi Logistik pernah dilakukan oleh Ningrum[4] untuk mengklasifikasikan kesejahteraan rumah tangga di Kota Malang, hasil kesimpulan penelitian tersebut menunjukkan bahwa nilai ketepatan klasifikasinya sebesar 97,8% menjadi 98% setelah dilakukan *bagging*. Oleh karena itu penelitian kali ini dilakukan pada

data di RS. X Surabaya pada tahun 2010 menggunakan *Bagging* Regresi Logistik dengan harapan hasil penelitian ini mendapatkan ketepatan klasifikasi yang lebih tinggi dibanding metode pada penelitian sebelumnya.

## II. LANDASAN TEORI

### 1. Statistika Deskriptif

Statistika deskriptif adalah metode yang digunakan untuk mendeskripsikan data, meliputi pengumpulan, pengorganisasian, serta penyajian data [5].

### 2. Analisis Regresi Logistik Biner (*Binary Logistic Regression*)

Dalam model statistika dengan dua kategori, dengan variabel respon mengandung unsur “sukses” atau “gagal”. Data biner ini merupakan bentuk paling sederhana dari data kategori. Model yang paling sering digunakan untuk data dua kategori adalah regresi logistik biner [6].

### Model Regresi Logistik Biner (*Dikotomus*)

Regresi logistik merupakan suatu metode regresi yang menggambarkan hubungan antara suatu variabel respon (*dependent*) dan satu atau lebih variabel prediktor (*independent*). Bentuk model regresi logistik dengan variabel prediktor  $i$  adalah sebagai berikut

$$\pi(x) = \frac{e^{(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_i x_i)}}{1 + e^{(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_i x_i)}} \quad (1)$$

Dengan menggunakan transformasi logit dari  $\pi(x)$ , maka model regresi fungsi logit dapat didefinisikan sebagai berikut

$$g(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_i x_i \quad (2)$$

Bentuk logit  $g(x)$  ini merupakan model logit, fungsi linear dalam parameter-parameternya, dan berada dalam jarak antara  $-\infty$  sampai  $+\infty$  tergantung dari variabel  $X$  [7].

### Penaksiran Parameter Model Regresi Logistik

*Maximum Likelihood* (kemungkinan maksimum) merupakan salah satu metode yang digunakan untuk menduga parameter-parameter yang terdapat dalam persamaan regresi

logistik [7]. Fungsi *likelihood*-nya adalah sebagai berikut.

$$l(\beta) = \prod_{i=1}^n \pi(x_i)^{y_i} (1 - \pi(x_i))^{1-y_i}$$

dengan

$$\pi(x_j) = \frac{\exp(\sum_{j=0}^p \beta_j x_{ij})}{1 + \exp(\sum_{j=0}^p \beta_j x_{ij})}$$

$$l(\beta) = \ln(l(\beta))$$

$$= \sum_{j=0}^p \left[ \sum_{i=1}^n y_i x_{ij} \right] \beta_j - \sum_{i=1}^n \ln \left[ 1 + \exp \left( \sum_{j=0}^p \beta_j x_{ij} \right) \right] \quad (3)$$

Persamaan (3) didiferensialkan terhadap  $\beta$

$$\frac{\partial l(\beta)}{\partial \beta_j} = \sum_{i=1}^n y_i x_{ij} - \sum_{i=1}^n x_{ij} \pi(x_{ij}) \quad (4)$$

dengan  $j=1,2,\dots,p$

Untuk mencari turunan dari persamaan (4) yang disamadengankan nol, seringkali tidak mendapatkan hasil yang eksplisit sehingga digunakan metode iterasi *Newton Raphson* untuk mengatasinya. Algoritma metode iterasi *Newton Raphson* menurut Agresti [6] adalah sebagai berikut.

1. Menentukan nilai awal estimasi parameter  $\beta^{(0)}$  dengan *Ordinary Least Square* (OLS), yaitu:
2. Membentuk vektor gradien  $g$ .
3. Membentuk matriks Hessian  $H$ .
4. Memasukkan nilai ke vektor  $g$  dan matriks  $H$  sehingga diperoleh vektor dan matriks .
5. Melakukan iterasi pada persamaan dimulai dari  $t = 0$ , dimana  $\beta^{(t)}$  adalah estimasi parameter yang konvergen pada iterasi ke- $t$ .
6. Apabila belum diperoleh estimasi parameter yang konvergen, langkah (5) diulang kembali hingga iterasi ke  $t = t + 1$  dan iterasi akan berhenti saat  $|\beta^{(t+1)} - \beta^{(t)}| \leq \epsilon$ , dimana  $\epsilon$  adalah bilangan yang sangat kecil. Hasil estimasi yang diperoleh adalah  $\beta^{(t+1)}$  pada iterasi terakhir..

### Pengujian Parameter Model Regresi Logistik

#### Uji Serentak

Uji serentak ini dilakukan yaitu untuk mengetahui signifikansi parameter  $\beta$  terhadap variabel respon secara keseluruhan. Pengujian signifikansi parameter tersebut menggunakan statistik uji G, dimana statistik uji G mengikuti distribusi *Chi-Square* dengan derajat bebas satu [7].

Hipotesis yang digunakan:

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0$$

$H_1$ : paling sedikit ada satu  $\beta_j \neq 0$ , dengan  $j = 1, 2, \dots, p$

Taraf Signifikan : 0,05

Statistik Uji :

$$G = -2 \ln \frac{\binom{n_1}{n} \binom{n_0}{n}}{\sum_{i=1}^n \hat{\pi}_i^{y_i} (1 - \hat{\pi}_i)^{(1-y_i)}} \quad (5)$$

$$n_1 = \sum_{i=1}^n y_i \quad n_0 = \sum_{i=1}^n (1 - y_i) \quad \text{dan} \quad n = n_1 + n_0$$

Daerah penolakan : tolak  $H_0$  jika  $G > \chi_{(db)}^2$  dengan db adalah derajat bebas (I-1)(J-1).

#### Uji Individu

Untuk menguji pengaruh setiap  $\beta_j$  secara individual. Hasil pengujian secara individual akan menunjukkan apakah suatu variabel prediktor layak untuk masuk dalam model atau tidak.

Hipotesis yang digunakan adalah

$$H_0 : \beta_j = 0$$

$H_1 : \beta_j \neq 0$  dengan  $j = 1, 2, 3, \dots, p$

Taraf Signifikan :  $\alpha$

Statistik Uji : Statistik Uji *Wald*

$$W = \frac{\beta_j^2}{SE(\beta_j)} \quad (6)$$

Daerah penolakan : tolak  $H_0$  jika  $W^2 > Z_{\alpha}^2$  atau  $W^2 > \chi_{(db)}^2$  dimana db adalah derajat bebas dengan nilai sama dengan 1

#### Uji Kesesuaian Model

Adapun hipotesis yang digunakan adalah :

$H_0$ : model sesuai

$H_1$ : model tidak sesuai

Statistik uji yang digunakan untuk menguji kesesuaian model adalah *Chi-Square* dari Hosmer dan Lemeshow *test*.

$$\hat{C} = \sum_{k=1}^g \frac{(O_k - n'_k \bar{\pi}_k)^2}{n'_k \bar{\pi}_k (1 - \bar{\pi}_k)} \quad (7)$$

dengan

$$O_k = \sum_{j=1}^{n'_k} y_j, \text{ jumlah variabel respon pada group ke-}k$$

$$\bar{\pi}_k = \frac{\sum_{j=1}^{C_k} m_j \hat{\pi}_j}{n'_k}, \text{ rata-rata taksiran probabilitas}$$

$g$  = Jumlah grup (kombinasi kategori model serentak)

$n'_k$  = banyaknya observasi pada group ke- $k$

Daerah Penolakan : Tolak  $H_0$  jika  $\chi^2_{hitung} > \chi^2_{(db,\alpha)}$

#### Interpretasi Model

Untuk regresi logistik dengan variabel prediktor bersifat dikotomis, nilai x dikategorikan 0 atau 1. Tabel 1 menyatakan variabel prediktor dikotomis.

Tabel 1.

Nilai-Nilai  $\pi(x)$  dan  $1 - \pi(x)$  Untuk Variabel Prediktor Biner (Dikotomos)

Variabel respon	Variabel Prediktor	
	x = 1	x = 0
y = 1	$\pi(1) = \frac{\exp(\beta_0 + \beta_1)}{1 + \exp(\beta_0 + \beta_1)}$	$\pi(0) = \frac{\exp(\beta_0)}{1 + \exp(\beta_0)}$
y = 0	$1 - \pi(1) = \frac{1}{1 + \exp(\beta_0 + \beta_1)}$	$1 - \pi(0) = \frac{1}{1 + \exp(\beta_0)}$
Total	1	1

Sumber: Hosmer and Lemeshow[5]

Odds ratio, dinotasikan  $\psi$ , didefinisikan sebagai *ratio odds* untuk  $x = 1$  terhadap *odds* untuk  $x = 0$ , yang dapat dituliskan dalam persamaan (8) berikut [7].

$$\psi = \frac{\pi(1)/1 - \pi(1)}{\pi(0)/1 - \pi(0)} \tag{8}$$

Berdasarkan Tabel 2.1, nilai *odds ratio* adalah

$$\begin{aligned} \psi &= \frac{\left(\frac{\exp(\beta_0 + \beta_1)}{1 + \exp(\beta_0 + \beta_1)}\right) \left(\frac{1}{1 + \exp(\beta_0)}\right)}{\left(\frac{\exp(\beta_0)}{1 + \exp(\beta_0)}\right) \left(\frac{1}{1 + \exp(\beta_0 + \beta_1)}\right)} \\ &= \frac{\exp(\beta_0 + \beta_1)}{\exp(\beta_0)} \\ &= \exp(\beta_1) \end{aligned} \tag{9}$$

3. Evaluasi Ketepatan Klasifikasi

Evaluasi ketepatan klasifikasi adalah suatu evaluasi yang melihat peluang kesalahan klasifikasi yang dilakukan oleh suatu fungsi klasifikasi [8].

Tabel 2. Tabel Klasifikasi Regresi Logistik

Hasil Observasi	Taksiran	
	$y_1$	$y_2$
$y_1$	$n_{11}$	$n_{12}$
$y_2$	$n_{21}$	$n_{22}$

Keterangan :

- $n_{11}$  = Jumlah subjek dari  $y_1$  tepat diklasifikasikan sebagai  $y_1$
- $n_{12}$  = Jumlah subjek dari  $y_1$  salah diklasifikasikan sebagai  $y_2$
- $n_{21}$  = Jumlah subjek dari  $y_2$  salah diklasifikasikan sebagai  $y_1$
- $n_{22}$  = Jumlah subjek dari  $y_2$  tepat diklasifikasikan sebagai  $y_2$

4. Bootstrap Aggregating (Bagging)

Metode *bagging* pertama kali digunakan oleh Breiman[9]. *Bagging* prediktor adalah metode untuk membangkitkan *multiple version* dari prediktor dan menggunakannya untuk *aggregate* prediktor. *Aggregate classifier* atau metode klasifikasi agregat  $\mu_A$  diberikan secara umum dalam :

$$\mu_A(y) = E_F [ (y, k) ]$$

dimana ekspektasi sampel  $D_k$  berdasarkan distribusi F (fungsi distribusi empirik). Breiman mengenalkan prosedur *bagging* dimana syarat  $\mu_A(y)$  diperoleh dengan membangkitkan B dari  $D_k$  dengan metode *bootstrap*. *Bootstrap* sampel ditentukan dari sampling dengan pengembalian data observasi. Untuk setiap resample *bootstrap* dari dihitung [10]:

$$^{(*)}(y, k^{(*)}), b = 1, \dots, B$$

dan kemudian menaksir *classifier* sebagai berikut.

$$A(y) = \hat{A} [ (y, k^*) ]$$

Estimasi Bagging Class Probability

Perbandingan dilakukan dengan membandingkan prediksi error dari hasil model tunggal sebagai berikut.

$$e_S = | (j|x) - p^*(j|x) | \tag{10}$$

dimana  $(j|x)$  adalah penaksir peluang dari model tunggal dan  $p^*(j|x)$  adalah peluang sebenarnya. Sedangkan untuk model hasil *bagging*, pada setiap iterasi *bootstrap* dilakukan perhitungan prediksi error.

$$e_B = | {}_B(j|x) - p^*(j|x) | \tag{11}$$

dimana  ${}_B(j|x)$  penaksir dari peluang pada setiap replikasi, sehingga prediksi error dari model *bagging* merupakan hasil rata-rata prediksi error pada setiap pengambilan sampel pada setiap B replikasi *bootstrap* [9].

Algoritma *bagging* untuk regresi logistik biner adalah sebagai berikut :

1. Mengambil sampel *bootstrap* sebanyak n dari data set dengan pengulangan sebanyak n.
2. Menghitung peluang respon untuk setiap observasi dan menghitung ketepatan klasifikasi. Kesalahan klasifikasi pada langkah ini disebut  $e_B$ .
3. Mengulang langkah 1 sampai langkah 3 sebanyak B kali (replikasi *bootstrap*).
4. Memperoleh ketepatan klasifikasi *bagging* dari pengambilan sampel sampai B, sehingga kesalahan klasifikasi *bagging* untuk replikasi B kali adalah  ${}_B$ .
5. Membentuk model *bagging* regresi logistik biner dari setiap parameter pada pengambilan sampel sampai B.
6. Untuk memperoleh hasil yang lebih baik, maka replikasi *bootstrap* dilakukan sebanyak mungkin [11]. Replikasi *bootstrap* yang biasa digunakan adalah 50 sampai 200.

5. Kanker Serviks

Kanker mulut rahim (serviks) atau karsinoma serviks uterus merupakan kanker pembunuh wanita nomor dua di dunia setelah kanker payudara. Kanker ini 99,7% disebabkan oleh *Human Papilloma Virus* (HPV) [12].

Pap Smear Test

Hasil *Pap test* sesuai klasifikasi *Papanicolaou* terdiri dari lima klas yaitu [13]-[15].

1. Normal, dimana tidak ditemukan sel abnormal
2. Tampak sel abnormal tetapi tidak ada bukti keganasan
3. Tampak sel-sel atypik yang meragukan untuk keganasan
4. Tampak sel-sel yang mencurigakan keganasan
5. Tampak adanya sel-sel ganas.

Hasil *Pap test* pada RS “X” Surabaya disajikan dalam Tabel 3.

Tabel 3. Gambaran Sitologi Hasil *Pap Test*.

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

III. METODOLOGI PENELITIAN

1. Sumber Data Penelitian

Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari RS “X” Surabaya tahun 2010.

2. Variabel Penelitian

Variabel penelitian yang digunakan dalam penelitian ini terdiri atas variabel respon dan variabel prediktor. Variabel

penelitian yang digunakan dalam penelitian ini adalah :

1. Variabel respon yang digunakan dalam penelitian ini adalah hasil sitologi *Pap Smear Test* (Y) yang merupakan variabel dengan kategori :  
 1 = "Normal"  
 2 = "Abnormal"

Variabel penelitian disajikan pada Tabel 4.

Tabel 4.  
Variabel Penelitian

Kode	Variabel	Definisi Operasional	Skala Pengukuran
Variabel respon :			
Y	Hasil <i>Pap Smear Test</i>	1. Normal (1) 2. Abnormal (2)	Nominal
Variabel prediktor :			
X <sub>1</sub>	Usia	Usia pasien	Rasio
X <sub>2</sub>	Usia melahirkan pertama kali	Usia pertama kali melahirkan	Rasio
X <sub>3</sub>	Usia menstruasi pertama kali	Usia pertama kali mengalami menstruasi	Rasio
X <sub>4</sub>	Siklus menstruasi	1. teratur (1) 2. tidak teratur (2)	Nominal
X <sub>5</sub>	Frekuensi melahirkan	1. pernah melahirkan ≤2 anak (1) 2. pernah melahirkan >2 anak (2)	Nominal
X <sub>6</sub>	Penggunaan kontrasepsi	1. tidak menggunakan kontrasepsi (1) 2. menggunakan kontrasepsi (2)	Nominal
X <sub>7</sub>	Riwayat keguguran	1. tidak pernah keguguran (1) 2. pernah keguguran (2)	Nominal

IV. ANALISIS DAN PEMBAHASAN

1. Karakteristik Pasien *Pap Smear Test* Berdasarkan Faktor Resiko

Tabel 4.

Karakteristik Pasien *Pap Smear Test* Berdasarkan Variabel Usia Pasien, Usia Melahirkan dan Usia Menstruasi Pertama

Variable	N	N*	Mean	SE Mean	StDev	Minimum	Maximum
X1	900	0	42,280	0,308	9,244	23,000	80,000
X2	900	0	25,867	0,137	4,103	16,000	45,000
X3	900	0	12,936	0,045	1,354	8,000	16,000

Tabel 4.1 menunjukkan bahwa usia pasien yang melakukan tes terkecil adalah usia 23 tahun sedangkan yang terbesar adalah usia 80 tahun. Untuk variabel usia melahirkan pertama kali ditunjukkan bahwa usia melahirkan pertama kali termuda adalah usia 16 tahun dan usia melahirkan tertua adalah 45 tahun. Variabel usia menstruasi pertama termuda adalah usia 8 tahun dan usia menstruasi pertama tertua adalah usia 16 tahun.

Karakteristik Pasien *Pap Smear Test* Berdasarkan Faktor Resiko Siklus Menstruasi

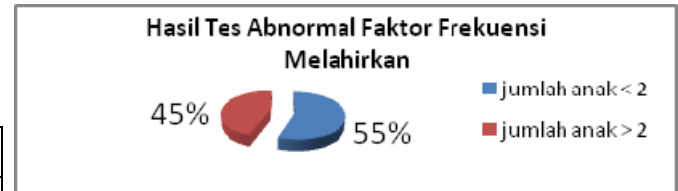


Gambar. 1. Hasil Tes Abnormal Faktor Siklus Menstruasi.

Gambar 1. menjelaskan pasien memiliki hasil test abnormal paling banyak adalah pasien dengan siklus menstruasi teratur

yaitu sebesar 67%, sedangkan pasien yang siklus menstruasinya tidak teratur sebesar 33%.

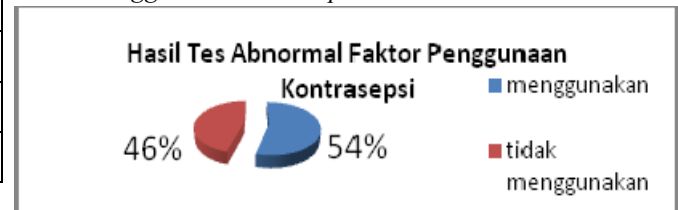
Karakteristik Pasien *Pap Smear Test* Berdasarkan Faktor Resiko Frekuensi Melahirkan



Gambar. 2. Hasil Tes Abnormal Faktor Frekuensi Melahirkan.

Gambar 2. menjelaskan pasien memiliki hasil test abnormal paling banyak adalah pasien dengan jumlah anak kurang atau sama dengan 2 yaitu sebesar 55%, sedangkan pasien yang memiliki anak lebih dari 2 sebesar 45%.

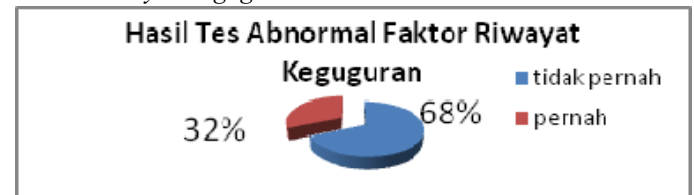
Karakteristik Pasien *Pap Smear Test* Berdasarkan Faktor Resiko Penggunaan Kontrasepsi



Gambar. 3. Hasil Tes Abnormal Faktor Penggunaan Kontrasepsi.

Gambar 3. menjelaskan pasien memiliki hasil test abnormal paling banyak adalah pasien yang menggunakan kontrasepsi yaitu sebesar 54%, sedangkan pasien yang tidak menggunakan kontrasepsi sebesar 46%.

Karakteristik Pasien *Pap Smear Test* Berdasarkan Faktor Resiko Riwayat Keguguran



Gambar. 4. Hasil Tes Abnormal Fakotr Riwayat Keguguran.

Gambar 4. menjelaskan pasien memiliki hasil test abnormal paling banyak adalah pasien yang tidak pernah keguguran yaitu sebesar 68%, sedangkan pasien yang pernah keguguran sebesar 32%.

2. Regresi Logistik Biner

Regresi Logistik Biner Secara Individu

Pengujian ini digunakan untuk mengetahui pengaruh masing-masing variabel prediktor terhadap variabel respon secara individu. Hipotesis yang digunakan adalah sebagai berikut.

$$H_0 : \beta_j = 0$$

$$H_1 : \beta_j \neq 0, \text{ dengan } i = 1, 2, \dots, 7$$

Tingkat signifikansi yang digunakan yaitu  $\alpha = 10\%$ .

Statistik uji yang digunakan adalah statistik uji Wald. Keputusan penolakan  $H_0$  jika  $p\text{-value} < \alpha$  atau nilai  $W^2$  lebih besar dari  $(1;0,05) = 3,841$ .

Tabel 5.  
Uji Signifikansi Parameter Secara Individu

Variabel Prediktor	Deskripsi	Wald	p-value
X <sub>1</sub>	Usia	18,772	0,000*
X <sub>2</sub>	Usia melahirkan pertama	2,799	0,094*
X <sub>3</sub>	Usia menstruasi pertama	0,118	0,665
X <sub>4</sub>	Siklus menstruasi	23,532	0,000*
X <sub>5</sub>	Paritas (Jumlah anak)	0,035	0,852
X <sub>6</sub>	Penggunaan kontrasepsi	13,484	0,000*
X <sub>7</sub>	Riwayat keguguran	0,475	0,490

**Keterangan :**

\*parameter signifikan berpengaruh pada  $\alpha = 5\%$

Tabel 5. menunjukkan bahwa variabel prediktor signifikan berpengaruh pada taraf  $\alpha = 10\%$  adalah variabel X<sub>1</sub>, X<sub>2</sub>, X<sub>4</sub> dan X<sub>6</sub>.

*Regresi Logistik Biner Secara Serentak*

Teknik yang digunakan dalam uji serentak adalah teknik *backward wald* dengan hipotesis sebagai berikut.

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_7 = 0$$

H<sub>1</sub> : paling sedikit ada satu  $\beta_i \neq 0$ , dengan  $i = 1, 2, \dots, 7$

Tingkat signifikansi yang digunakan yaitu  $\alpha = 5\%$ . Keputusan penolakan  $H_0$  jika  $p\text{-value} < \alpha$  atau nilai G lebih besar  $(4;0,05) = 9,488$ .

Tabel 6.

Uji Signifikansi Parameter Secara Serentak

		Chi-square	df	Sig.
Step 1	Step	57.517	4	0,000
	Block	57.517	4	0,000
	Model	57.517	4	0,000

**Keterangan :**

\*parameter signifikan berpengaruh pada  $\alpha = 5\%$

Nilai statistik uji G yang dihasilkan adalah sebesar 57,517 dimana nilai ini lebih besar dari nilai  $(4;0,05)$  sehingga keputusan yang diambil adalah tolak  $H_0$ , atau dengan kata lain paling sedikit ada satu  $\beta_i \neq 0$ . Jadi, model regresi logistik biner yang diperoleh adalah sebagai berikut.

$$\pi(x) = \frac{\exp(1,209 + 0,095X_1 - 0,090X_2 - 0,792X_4(1) - 0,681X_6(1))}{1 + \exp(1,209 + 0,095X_1 - 0,090X_2 - 0,792X_4(1) - 0,681X_6(1))}$$

*Uji Kesesuaian Model*

Uji kesesuaian model digunakan untuk mengetahui apakah terdapat perbedaan antara hasil observasi dengan kemungkinan hasil prediksi. Hipotesis yang digunakan adalah sebagai berikut.

H<sub>0</sub> : model sesuai

H<sub>1</sub> : model tidak sesuai

Tingkat signifikansi yang digunakan adalah  $\alpha = 5\%$ . Pada *Hosmer and Lemeshow test* diperoleh nilai p-value sebesar 0,175 dimana nilai p-value  $> \alpha$  sehingga keputusannya adalah Gagal tolak  $H_0$  yang berarti model sesuai, sehingga model yang digunakan adalah

$$\pi(x) = \frac{\exp(1,209 + 0,095X_1 - 0,090X_2 - 0,792X_4(1) - 0,681X_6(1))}{1 + \exp(1,209 + 0,095X_1 - 0,090X_2 - 0,792X_4(1) - 0,681X_6(1))}$$

dengan model logit :

$$g(x) = 1,209 + 0,095X_1 - 0,090X_2 - 0,792X_4(1) - 0,681X_6(1)$$

*Evaluasi Ketepatan Klasifikasi*

Tabel 7.

Tabel Klasifikasi Model Regresi Logistik

Observasi	Prediksi		Ketepatan Klasifikasi(%)
	Normal	Abnormal	
Normal	82	179	31,4
Abnormal	121	518	81,1
Ketepatan Klasifikasi Total (%)			66,7

Tabel 7. diketahui bahwa persentase seluruh observasi terklasifikasi dengan benar adalah 66,7%.

*3. Bagging Regresi Logistik*

Data dibagi menjadi data training dan testing dengan 2 kombinasi yaitu training-testing 80%-20% dan training-testing 70%-30% kemudian masing-masing kombinasi data direplikasi bootstrap sebanyak 50 hingga 100 kali.

*Estimasi Bagging Class Probability Kombinasi Data I*

Tabel 8.

Ketepatan Klasifikasi dengan Regresi Logistik Biner pada data testing pada data kombinasi 1

Regresi Logistik	Ketepatan Klasifikasi	$\bar{e}_B$
Biner	67,78%	32,22%

Tabel 9.

Hasil *Bagging* Regresi Logistik Data Training data kombinasi 1

Replikasi Bootstrap	Ketepatan Klasifikasi	$\bar{e}_B$
50 kali	71,29%	28,71%
60 kali	71,07%	28,93%
70 kali	71,15%	28,85%
<b>80 kali</b>	<b>71,67%</b>	<b>28,33%</b>
90 kali	71,48%	28,52%
100 kali	71,61%	28,39%

Tabel 10.

Hasil *Bagging* Regresi Logistik Data Testing data kombinasi 1

Replikasi Bootstrap	Ketepatan Klasifikasi	$\bar{e}_B$
50 kali	68,89%	31,11%
60 kali	68,89%	31,11%
70 kali	68,89%	31,11%
<b>80 kali</b>	<b>68,89%</b>	<b>31,11%</b>
90 kali	68,89%	31,11%
100 kali	68,33%	31,67%

Tabel 8., Tabel 9. dan Tabel 10. memberikan informasi bahwa dengan 80 kali replikasi bootstrap yang dilakukan pada data training kombinasi 1, diperoleh rata-rata ketepatan klasifikasi terbesar, maka estimasi parameter pada bagging dengan 80 replikasi digunakan pada data testing dan menghasilkan ketepatan klasifikasi sebesar 68,89%. Dapat disimpulkan bahwa diperoleh *bagging* prediktor terbaik adalah pada replikasi bootstrap sebanyak 80 kali. Prediksi errornya yang mengalami penurunan sebesar 1,11%, yaitu sebesar 32,22% menjadi 31,11% dan mengalami peningkatan klasifikasi dari 67,78% menjadi 68,89%.

Estimasi Bagging Class Probability Kombinasi Data 2

Tabel 11.

Ketepatan Klasifikasi dengan Regresi Logistik Biner pada data testing data kombinasi 2

Regresi Logistik	Ketepatan Klasifikasi	$\epsilon_s$
Biner	69,63%	30,37%

Tabel 12.

Hasil Bagging Regresi Logistik Data Training pada data kombinasi 2

Replikasi Bootstrap	Ketepatan Klasifikasi	$\bar{\epsilon}_B$
50 kali	70,88%	29,12%
60 kali	71,16%	28,84%
70 kali	70,45%	29,55%
<b>80 kali</b>	<b>71,49%</b>	<b>28,51%</b>
90 kali	70,97%	29,03%
100 kali	70,67%	28,33%

Tabel 13.

Hasil Bagging Regresi Logistik Data Testing pada data kombinasi 2

Replikasi Bootstrap	Ketepatan Klasifikasi	$\bar{\epsilon}_B$
50 kali	70,74%	29,26%
60 kali	70,00%	30,00%
70 kali	70,74%	29,26%
<b>80 kali</b>	<b>70,74%</b>	<b>29,26%</b>
90 kali	70,37%	29,63%
100 kali	70,37%	29,63%

Tabel 11. Tabel 12. dan Tabel 13. memberikan informasi bahwa dengan 80 kali replikasi bootstrap yang dilakukan pada data kombinasi 2, diperoleh rata-rata ketepatan klasifikasi terbesar, yaitu sebesar 71,49%. Dapat disimpulkan bahwa diperoleh bagging prediktor terbaik adalah pada replikasi bootstrap sebanyak 80 kali. Prediksi errornya yang mengalami penurunan sebesar 1,11%, yaitu sebesar 30,37% menjadi 29,63% dan mengalami peningkatan klasifikasi dari 69,63% menjadi 70,74%.

Tabel 14.

Hasil Bagging Regresi Logistik

Replikasi Bootstrap	Ketepatan Klasifikasi	
	Kombinasi 1 80%-20%	Kombinasi 2 70%-30%
50 kali	68,89%	70,74%
60 kali	68,89%	70,00%
70 kali	68,89%	70,74%
<b>80 kali</b>	<b>68,89%</b>	<b>70,74%</b>
90 kali	68,89%	70,37%
100 kali	68,33%	70,37%

Hasil perbandingan pada Tabel 14. didapatkan hasil bootstrap data kombinasi kedua dengan data kombinasi pertama, data kombinasi pertama dan kedua memiliki peningkatan ketepatan klasifikasi yang sama.

Model Bagging Regresi Logistik

Estimasi parameter model bagging merupakan nilai rata-rata estimasi parameter B kali replikasi bootstrap. Jika dilakukan replikasi 80 kali, maka estimasi parameter model bagging adalah hasil rata-rata estimasi parameter ke-80 model multiple regresi logistik

Model bagging regresi logistik yang diperoleh pada 80 kali replikasi bootstrap pada data kombinasi 1 adalah :

$$g(x) = -0,0889 + 0,0872x_1 - 0,0868x_2 + 0,827x_3(1) + 0,644x_4(1)$$

Model bagging regresi logistik yang diperoleh pada 80 kali replikasi bootstrap pada data kombinasi 2 adalah :

$$g(x) = -0,1204 + 0,0878x_1 - 0,084x_2 + 0,749x_3(1) + 0,686x_4(1)$$

V. KESIMPULAN

Hasil analisis regresi logistik menunjukkan bahwa pola hubungan hasil *Pap Smear Test* sebagai test awal untuk kanker serviks dengan faktor-faktor yang mem-pengaruhinya digambarkan dalam model regresi logistik berikut.

$$p(x) = \frac{\exp(1,308 + 0,088x_1 - 0,080x_2 - 0,782x_3(1) + 0,681x_4(1))}{1 + \exp(1,308 + 0,088x_1 - 0,080x_2 - 0,782x_3(1) + 0,681x_4(1))}$$

Pada model tersebut terdapat 4 (empat) variabel prediktor yang signifikan berpengaruh terhadap variabel respon, yaitu usia pasien ( $X_1$ ), usia melahirkan pertama ( $X_2$ ), siklus menstruasi ( $X_4$ ) dan penggunaan kontrasepsi ( $X_6$ ), Model tersebut sudah sesuai untuk menjelaskan berapa besar peluang seorang pasien akan positif abnormal pada hasil *Pap Smear Test*, sehingga dapat langsung melakukan tindakan lanjut, dengan ketepatan klasifikasi sebesar 66,7%. Hasil analisis bagging regresi logistik menunjukkan bahwa pada 80 kali pada data kombinasi 1 dan 2, replikasi bootstrap diperoleh peningkatan ketepatan klasifikasi sebesar 1,11%

UCAPAN TERIMA KASIH

Penulis I.A.S.I. mengucapkan terima kasih kepada Jurusan Statistika ITS dan Rumah Sakit "X" Surabaya.

DAFTAR PUSTAKA

- [1] F. I. Nurdin, "Klasifikasi Pasien Hasil Pap Test Penyakit Kanker Leher Rahim dengan Metode Multivariate Adaptive Regression Splines (MARS)". Tugas Akhir, Jurusan Statistika Institut Teknologi Sepuluh Nopember, Surabaya. (2006).
- [2] Melva, "Faktor-faktor yang Mempengaruhi Kejadian Kanker Leher Rahim Pada Penderita Yang Datang Berobat Di RSUP H. Adam Malik Medan". Tesis, Universitas Sumatera Utara, Medan. (2008).
- [3] H. Yastuti, "Bagging Multivariate Adaptive Regression Splines (Mars) Untuk Klasifikasi Pasien Hasil Pap Test Penyakit Kanker Serviks (Studi Kasus Di RS "X" Surabaya)". Tugas Akhir, Jurusan Statistika Institut Teknologi Sepuluh Nopember, Surabaya. (2011).
- [4] E. S. Ningrum, "Klasifikasi Kesejahteraan Rumah Tangga Di Kota Malang Dengan Pendekatan Bagging Regresi Logistik". Tugas Akhir, Jurusan Statistika Institut Teknologi Sepuluh Nopember, Surabaya. (2012).
- [5] R. E. Walpole, *Pengantar Statistika Edisi ke-3*. Jakarta: PT Gramedia Pustaka Utama. (1995).
- [6] A. Agresti, *Categorical Data Analysis*. John Wiley and Sons. New York. (1990).
- [7] D.W. Hosmer dan Lemeshow, *Applied Logistic Regression*. USA : John Wiley and Sons. (2000).
- [8] R. A. Johnson dan D. W. Wichern, *Applied Multivariate Statistical Analysis*. Prentice Hall. New Jersey. (1992).
- [9] L. Breiman, "Bagging Predictor". Technical report No. 421. Departement of statistics University of California. (1994).
- [10] J. G. Dias dan J. K. Vermunt, "A Bootstrap based Aggregate Classifier for Model based Clustering". Journal of Annals Statistics. (2005).
- [11] B. Efron dan R.J. Tibshiran, *An Introduction to the Bootstrap*. Chapman and Hall. New York. (1993).
- [12] Wikipedia. (2012, Februari 9). Kanker Leher Rahim. [http://id.wikipedia.org/wiki/Kanker\\_leher\\_rahim](http://id.wikipedia.org/wiki/Kanker_leher_rahim).
- [13] H. Nurwijaya, Andrijono, Suheimi, H.K., *Cegah dan Deteksi Kanker Serviks*. Gramedia, Jakarta. (2010).
- [14] A. Setiawan, (2010). *Kanker Serviks Penyebab Utama Kematian*, <http://kesehatan.kompas.com/read/2010/05/07/08042334/Kanker.Serviks.Penyebab.Utama.Kematian.htm>.
- [15] D. Wijaya, *Pembunuh Ganas Itu Bernama Kanker Serviks*. Sinar Kejora, Yogyakarta. (2010).