

# Klasifikasi Kesejahteraan Rumah Tangga di Provinsi Papua dengan Metode Regresi Logistik dan *Support Vector Machine*

Riska Prakasita Sahitayakti dan Kartika Fithriasari.  
Jurusan Statistika, FMIPA, Institut Teknologi Sepuluh Nopember (ITS)  
Jl. Arief Rahman Hakim, Surabaya 60111 Indonesia  
*e-mail*:kartika\_f@statistika.its.ac.id

**Abstrak**—Provinsi Papua dalam lima tahun terakhir ini menunjukkan bahwa presentase penduduk miskin di Provinsi Papua menempati peringkat pertama dibandingkan ke-32 provinsi lainnya di Indonesia. Klasifikasi kesejahteraan rumah tangga ditujukan untuk mengkategorikan suatu rumah tangga, kedalam kategori keluarga miskin ataupun tidak miskin berdasarkan faktor-faktor yang mencirikan suatu kemiskinan. Dalam rangka meningkatkan kesejahteraan masyarakat di Provinsi Papua, salah satu upaya yang dilakukan adalah dijalkannya program pengentasan ke-miskinan, dimana dalam pelaksanaannya perlu dalam mengkategorikan suatu rumah tangga kedalam kategori miskin atau tidak miskin untuk mengetahui layak atau tidaknya mendapatkan bantuan dari program tersebut. Analisis klasifikasi dilakukan menggunakan metode regresi logistik dan *support vector machine*. Karakteristik kesejahteraan rumah tangga di Papua menunjukkan sebanyak 27% rumah tangga dikategorikan rumah tangga miskin dan sebanyak 73% rumah tangga dikategorikan rumah tangga tidak miskin. Klasifikasi menggunakan metode regresi logistik biner menunjukkan bahwa ketepatan klasifikasi tertinggi sebesar 81.20%. Klasifikasi menggunakan metode *support vector machine* menunjukkan bahwa ketepatan klasifikasi tertinggi yakni sebesar 82.05% menggunakan variabel prediktor yang signifikan pada analisis regresi logistik biner. Model klasifikasi yang menghasilkan ketepatan klasifikasi tertinggi tersebut menggunakan fungsi kernel *Radial Basis Function* (RBF) dengan parameter  $\sigma=2$  dan  $C=10$

**Kata Kunci**—Klasifikasi, Kesejahteraan Rumah Tangga, Regresi Logistik, *Support Vector Machine*

## I. PENDAHULUAN

Presentase penduduk miskin di Provinsi Papua dalam lima tahun terakhir ini merupakan yang terbesar dibandingkan ke-32 provinsi lainnya di Indonesia[1], sehingga perlu bagi pemerintah dalam meningkatkan kesejahteraan masyarakat di provinsi tersebut.

Klasifikasi kesejahteraan rumah tangga ini digunakan dalam mengkategorikan suatu rumah tangga, kedalam kategori rumah tangga miskin ataupun tidak miskin berdasarkan faktor-faktor yang mencirikan suatu kemiskinan. Salah satu upaya yang dilakukan dalam rangka meningkatkan kesejahteraan masyarakat di Provinsi Papua, adalah dijalkannya program pengentasan kemiskinan, yang mana dalam pelaksanaan program tersebut perlu untuk mengkategorikan suatu rumah tangga kedalam kategori miskin atau tidak miskin untuk mengetahui layak atau tidaknya suatu rumah tangga mendapatkan bantuan dari

program tersebut. Oleh karena itu, pengkategorian ini dapat digunakan dalam meningkatkan kesejahteraan masyarakat dengan membuat kebijakan atau program pengentasan kemiskinan yang nantinya akan berdampak pada meningkatnya kesejahteraan masyarakat di Provinsi Papua.

Kajian mengenai klasifikasi rumah tangga telah banyak dilakukan oleh beberapa peneliti. Aeni (2010) meneliti klasifikasi kesejahteraan rumah tangga di Provinsi Jawa Tengah dengan pendekatan CART *Arcing*. Tingkat akurasi hasil klasifikasi yang didapatkannya sebesar 79,1%[2]. Ningrum (2011) melakukan penelitian tentang klasifikasi kesejahteraan rumah tangga di kota Malang dengan pendekatan *bagging* regresi logistik. Ningrum menyimpulkan bahwa terdapat empat variabel prediktor berpengaruh signifikan terhadap kesejahteraan rumah tangga, yaitu jumlah anggota rumah tangga, status pekerjaan kepala rumah tangga, pengalaman membeli beras miskin, dan penggunaan telepon seluler pada suatu rumah tangga[3].

Ada beberapa metode klasifikasi, baik untuk kasus klasifikasi *silinear separable* maupun *non-linear separable*. Oleh karena data dalam penelitian ini merupakan kasus *non-linear separable*, maka salah satu metode klasifikasi yang cocok digunakan salah satunya adalah menggunakan *support vector machine* (SVM). Metode SVM merupakan metode klasifikasi non para-metrik yang tidak harus memenuhi asumsi dan distribusi tertentu, serta dapat digunakan pada kasus *non-linear separable*. Selain metode SVM, adapun metode klasik yang sering digunakan dalam pengklasifikasian data adalah dengan menggunakan metode regresi logistik. Oleh karena itu, pada penelitian kali ini dilakukan analisis klasifikasi kesejahteraan rumah tangga di Provinsi Papua dengan membandingkan dua metode, yakni menggunakan metode regresi logistik dan metode *support vector machine*.

## II. TINJAUAN PUSTAKA

### A. Regresi Logistik Biner

Regresi logistik biner merupakan salah satu jenis dari regresi logistik. Regresi logistik adalah suatu metode yang dapat digunakan untuk mencari hubungan antara variabel respon yang bersifat *dichotomous* (dua kategori) atau *polychotomous* (lebih dari dua kategori) dengan satu atau lebih variabel prediktor berskala kategori atau kontinu[4]. Analisis regresi logistik biner digunakan untuk menjelaskan

hubungan antara variabel respon yang hanya memiliki dua kategori saja. Model yang dida-pat dapat dijadikan model dalam mengklasifikasikan variabel prediktor ke dalam variabel respon yang berupa data kategorik. Anggap bahwa sekumpulan  $p$  variabel bebas ditunjukkan se-bagai vektor  $\mathbf{x}' = (x_1, x_2, \dots, x_p)$ . Bentuk logit dari regresi logistik multivariabel adalah sebagai berikut.

$$g(\mathbf{x}) = \ln \left[ \frac{\pi(\mathbf{x})}{1 - \pi(\mathbf{x})} \right] = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p \quad (1)$$

Model regresi logistik dengan variabel independen  $p$  yaitu banyaknya, variabel prediktor ditunjukkan pada persamaan (2).

$$\pi(\mathbf{x}) = \frac{e^{g(\mathbf{x})}}{1 + e^{g(\mathbf{x})}} = \frac{\exp(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)}{1 + \exp(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)} \quad (2)$$

**B. Estimasi Parameter**

Estimasi parameter dari model regresi logistik dapat dilakukan dengan menggunakan metode *Maximum Likelihood Estimation* (MLE). Fungsi probabilitas distribusi *bernoulli* di setiap pengamatan  $(x_i, y_i)$  ditunjukkan pada persamaan berikut

$$f(y_i) = \pi(\mathbf{x}_i)^{y_i} [1 - \pi(\mathbf{x}_i)]^{1 - y_i} \quad (3)$$

Apabila antar pengamatan diasumsikan independen, maka fungsi likelihood dari pengamatan yang independen adalah sebagai berikut.

$$l(\boldsymbol{\beta}) = \prod_{i=1}^n f(y_i) = \prod_{i=1}^n \pi(\mathbf{x}_i)^{y_i} (1 - \pi(\mathbf{x}_i))^{1 - y_i} \quad (4)$$

Fungsi *likelihood*  $l(\boldsymbol{\beta})$  kemudian diubah ke persamaan ln.

$$L(\boldsymbol{\beta}) = \ln l(\boldsymbol{\beta}) = \sum_{j=0}^n \left[ \sum_{i=1}^n y_i x_{ij} \right] \beta_j - \sum_{i=1}^n \ln \left[ 1 + \exp \left( \sum_{j=0}^p \beta_j x_{ij} \right) \right] \quad (5)$$

Selanjutnya  $L(\boldsymbol{\beta})$  diturunkan terhadap  $\beta_j$  dan hasilnya sama dengan 0 (6).

$$\frac{\partial L(\boldsymbol{\beta})}{\partial \beta_j} = \sum_{i=1}^n y_i x_{ij} - \sum_{i=1}^n x_{ij} \left( \frac{\exp \left( \sum_{j=0}^p \beta_j x_{ij} \right)}{1 + \exp \left( \sum_{j=0}^p \beta_j x_{ij} \right)} \right) = 0 \quad (6)$$

Guna mengestimasi parameter  $\beta$ , digunakan metode numerik, yaitu Metode iterasi *Newton Raphson*, sedangkan untuk estimasi varians dan kovarians, diperoleh dari turunan kedua fungsi  $\ln likelihood L(\boldsymbol{\beta})$ . Berdasarkan turunan kedua fungsi  $\ln likelihood$ , dapat diperoleh matriks varians dan kovarians dari estimasi parameter melalui invers matriks [5].

$$\text{cov}(\hat{\boldsymbol{\beta}}) = \{ \mathbf{X}' \text{diag} [\hat{\pi}_i (1 - \hat{\pi}_i)] \mathbf{X} \}^{-1} \quad (7)$$

$\text{diag} [\hat{\pi}_i (1 - \hat{\pi}_i)]$  pada persamaan (7) merupakan  $n \times n$  matriks dia-gonal dengan elemen diagonal utama yaitu,  $\hat{\pi}_i (x_i) (1 - \hat{\pi}_i)$  di-mana akar kuadrat dari elemen-elemen diagonal utama adalah estimasi parameter model.

**C. Pengujian Signifikansi Parameter**

Pengujian signifikansi parameter secara parsial digunakan untuk mengetahui pengaruh masing-masing variabel prediktor terhadap variabel respon secara individu. Pengujian parameter

secara parsial menggunakan statistik uji *Wald*[4]. Hipotesis yang digunakan adalah sebagai berikut.

$$H_0 : \beta_j = 0$$

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

Statistik Uji :

$$W = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \quad (8)$$

Statistik Uji  $W$  pada persamaan (8) mengikuti distribusi normal dengantaraf signifikansi sebesar  $\alpha$ , sehingga akan diperoleh keputusan tolak  $H_0$  jika nilai  $|W| > Z_{\alpha/2}$  atau  $p\text{-value} < \alpha$ .

Uji signifikansi parameter secara serentak digunakan untuk mengetahui pengaruh variabel prediktor terhadap variabel respon secara bersamaan atau serentak. Pengujian secara serentak dilakukan dengan menggunakan statistik uji *Likelihood Ratio Test* atau uji  $G$  [4]. Hipotesis yang digunakan adalah sebagai berikut.

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

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

Statistik Uji:

$$G = 2 \left\{ \sum_{i=1}^n [\ln(\hat{\pi}_i) + (1 - y_i) \ln(1 - \hat{\pi}_i)] - [n_1 \ln(n_1) + n_0 \ln(n_0) - n \ln(n)] \right\} \quad (9)$$

Keterangan:

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

Statistik uji  $G$  pada persamaan (9) mengikuti distribusi *chi-square* dengan taraf signifikansi sebesar  $\alpha$  dan derajat bebas  $p$  yang merupakan banyaknya prediktor pada model. Keputusan tolak  $H_0$  jika nilai  $G > \chi_{(\alpha, p)}^2$  atau  $p\text{-value} < \alpha$ [4].

**D. Uji Kesesuaian Model**

Uji kesesuaian model regresi logistik disebut juga dengan *Goodness of Fit* [4]. Pengujian ini dilakukan untuk melihat apa-kah model telah sesuai atau tidak. Hipotesis uji kesesuaian model adalah sebagai berikut.

$$H_0 : \text{Model Sesuai}$$

$$H_1 : \text{Model tidak sesuai}$$

Statistik Uji:

$$\hat{\chi} = \sum_{k=1}^g \frac{(o_k - n'_k \bar{\pi}_k)^2}{n'_k \bar{\pi}_k (1 - \bar{\pi}_k)} \quad (10)$$

Statistik uji  $\hat{\chi}$  pada persamaan (10) mengikuti distribusi *chi-square* dengan taraf signifikansi  $\alpha$  dan derajat bebas  $g-2$ . Keputusan tolak  $H_0$  apabila nilai  $\hat{\chi} > \chi_{(g-2)}^2$  atau  $p\text{-value} < \alpha$ . Model yang didapat kemudian digunakan dalam pengklasifikasian. Untuk mendapatkan hasil pengklasifikasian terlebih dahulu perlu ditetapkan nilai *cut off*. Apabila  $\pi(\mathbf{x}) \geq$  nilai *cut off*, maka dikategorikan kedalam kategori 1, sebaliknya dikategorikan kedalam kategori 0. Nilai yang paling umum digunakan untuk *cut off* adalah sebesar 0,5[4].

**E. Support Vector Machine**

Konsep SVM secara sederhana dapat dijelaskan yakni seba-gai usaha mencari *hyperplane* terbaik yang berfungsi sebagai pemisah dua buah kelas atau kategori pada suatu

inputspace. Pemisah kedua kelas, yang disebut *hyperplane*, ditentukan dari parameter  $w$  dan  $b$  (11).

$$\langle w, x \rangle + b = 0 \tag{11}$$

Himpunan vektor dikatakan terpisah optimal oleh *hyperplane* jika dapat dipisahkan tanpa *error* dan jarak antara vektor terdekat dengan *hyperplane* maksimal [6]. Pemisahan *hyper-plane* dengan bentuk *canonical* harus memenuhi batasan (12).

$$y[\langle w, x \rangle + b] \geq 1, \quad i = 1, 2, \dots, l \tag{12}$$

*Hyperplane* yang optimal didapatkan dari memaksimalkan

margin  $\rho(w, b) = \frac{2}{w}$  dan meminimalkan  $\phi(w) = \frac{1}{2} w^2$  [6]. Masalah ini dapat dipecahkan dengan berbagai teknik komputasi, di antaranya adalah dengan *Lagrange Multiplier* (13).

$$L(w, b, \alpha) = \frac{1}{2} w^2 - \sum_{i=1}^l \alpha_i [y_i (\langle w, x_i \rangle + b) - 1] \tag{13}$$

$\alpha_i$  pada persamaan (13) merupakan pengganda fungsi *lagrange*. Nilai optimal (13) dapat dicari dengan mentransformasi (13) kedalam *dual space* (14)

$$\hat{\alpha} = \arg \min_{\alpha} \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j x_i \cdot x_j - \sum_{k=1}^l \alpha_k \tag{14}$$

Pada umumnya, masalah dalam dunia nyata jarang yang ber-sifat *linear separable*, akan tetapi bersifat *non-linear separable*. Untuk menyelesaikan problem non-linear, SVM dimodifikasi dengan memasukkan fungsi Kernel. Pada kasus non-linear ini, *hyperplane* yang memisahkan data secara optimal ditambahkan variabel *slack(t<sub>i</sub>)* yang harus meminimalkan:

$$\phi(w) = \frac{1}{2} w^2 + C \sum_{i=1}^l t_i \tag{15}$$

Penggunaan variabel *slack(t<sub>i</sub>)* digunakan untuk mengatasi pembatas  $y_i[\langle w, x_i \rangle + b] + t_i \geq 1$  yang tidak terpenuhi dengan cara memberikan penalti untuk data yang tidak memenuhi pembatas tersebut. Pemberian penalti ini dilakukan dengan menerapkan konstanta ongkos  $C$ .

Optimasi (15) dapat dipecahkan dengan teknik komputasi, di antaranya adalah dengan *Lagrange Multiplier*.

$$L(w, b, \alpha, t_i, \beta) = \frac{1}{2} w^2 + C \sum_{i=1}^l t_i - \sum_{i=1}^l \alpha_i [y_i (\langle w, x_i \rangle + b) - 1 + t_i] - \sum_{j=1}^l \beta_j t_j \tag{16}$$

$\alpha_i$  dan  $\beta_i$  pada persamaan (16) adalah pengganda fungsi *lag-range*. Nilai optimal dari persamaan tersebut dapat dicari dengan mentransformasi (16) kedalam *dual space* (17).

$$\hat{\alpha} = \arg \min_{\alpha} \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j K(x_i, x_j) - \sum_{k=1}^l \alpha_k \tag{17}$$

dimana  $K(x_i, x_j)$  merupakan fungsi kernel. Adapun fungsi kernel yang biasa digunakan dalam SVM adalah sebagai berikut.

1. Linear :  $K(x, x') = x' \cdot x$
2. Polynomial :  $K(x, x') = (\langle x, x' \rangle + 1)^p$
3. Gaussian Radial Basis Function:  $K(x, x') = \exp\left(-\frac{\|x - x'\|^2}{2\sigma^2}\right)$

Hasil klasifikasi data  $x$  dapat diperoleh dari persamaan (18)

$$f(x) = \text{sgn}\left(\sum_{i \in SV} \alpha_i y_i K(x_i, x) + b\right) \tag{18}$$

#### F. Evaluasi Klasifikasi

Hasil klasifikasi dievaluasi untuk melihat tingkat kebaikannya. Ukuran yang dipakai yakni nilai Apparent Error Rate (APER). Nilai APER (19) menyatakan nilai proporsi sampel yang salah diklasifikasikan oleh fungsi klasifikasi [7].

Tabel 1. *Confussion Matrix*

	Data Prediksi		Total
	$\pi_1$	$\pi_2$	
Data Aktual $\pi_1$	$n_{1C}$	$n_{1M} = n_1 - n_{1C}$	$n_1$
$\pi_2$	$n_{2M} = n_2 - n_{2C}$	$n_{2C}$	$n_2$

dimana,

$n_{1C}$  = Banyak data kelas  $\pi_1$  tepat diklasifikasikan ke kelas  $\pi_1$

$n_{2C}$  = Banyak data kelas  $\pi_2$  tepat diklasifikasikan ke kelas  $\pi_2$

$n_{1M}$  = Banyak data kelas  $\pi_1$  salah diklasifikasikan ke kelas  $\pi_2$

$n_{2M}$  = Banyak data kelas  $\pi_2$  salah diklasifikasikan ke kelas  $\pi_1$

Berdasarkan *confussionmatrix* pada Tabel 1, dapat dihitung nilai APER dan ketepatan klasifikasi menggunakan rumus:

$$APER = \frac{n_{1M} + n_{2M}}{n_1 + n_2} \tag{19}$$

$$\text{Ketepatan Klasifikasi} = 1 - APER$$

#### G. Kesejahteraan Sosial

Kesejahteraan sosial dalam arti luas mencakup berbagai tindakan yang dilakukan manusia untuk mencapai tingkat kehidupan masyarakat yang lebih baik. Kemiskinan sendiri merupakan salah satu masalah pembangunan kesejahteraan sosial. Pada tahun 2000 BPS melakukan Studi Penentuan Kriteria Penduduk Miskin untuk mengetahui karakteristik-karakteristik rumah tangga yang mampu mencirikan kemiskinan. Diperoleh 8 variabel yang dianggap layak untuk penentuan rumah tangga miskin di lapangan yaitu luas lantai per kapita, jenis lantai, ke-tersediaan air bersih/air minum, jenis jamban/WC, kepemilikan asset, total pendapatan per bulan, pengeluaran untuk makanan, konsumsi lauk pauk. Selain itu, terdapat pendataan Sosial Eko-nomi Penduduk 2005 (PSE05) yang dimaksudkan untuk mendapatkan data mikro berupa direktori rumah tangga miskin yang menerima Bantuan Langsung Tunai (BLT). Penentuan rumah tangga penerima BLT pada PSE05 didasarkan 14 indikator yaitu luas lantai rumah, jenis lantai, jenis dinding rumah, fasilitas tempat buang air besar, sumber air minum, penerangan yang digunakan, bahan bakar yang digunakan frekuensi makan dalam sehari, kebiasaan membeli daging/ayam/susu, kemampuan membeli pakaian, kemampuan berobat ke puskesmas, pekerjaan kepala rumah tangga, pendidikan kepala rumah tangga, dan kepemilikan asset [8].

### III. METODOLOGI PENELITIAN

#### A. Sumber Data

Data yang digunakan adalah data sekunder dari *Indonesian Family Life Surveys-East (IFLS-East)*, hanya data survei pada Provinsi Papua saja.

**B. Variabel Penelitian**

Variabel penelitian yang terdiri dari variabel respon (Y) dan variabel prediktor (X). Variabel respon terdiri dari kategori rumah tangga miskin (0) atau tidak miskin (1). Variabel prediktor (X) terdiri dari 17 variabel yang mencakup 5 aspek.

Tabel 2. Variabel Penelitian

Aspek	Variabel Prediktor
Kependudukan	X <sub>1</sub> Banyaknya Anggota Rumah Tangga
	X <sub>2</sub> Umur Kepala Rumah Tangga
	X <sub>3</sub> Status Perkawinan Kepala Rumah Tangga
	X <sub>4</sub> Jenis Kelamin Kepala Rumah Tangga

Tabel 2. Variabel Penelitian (Lanjutan)

Aspek	Variabel Prediktor
Pendidikan	X <sub>5</sub> Pendidikan Terakhir yang pernah diikuti Kepala Rumah Tangga
	X <sub>6</sub> Status Kepemilikan Rumah
	X <sub>7</sub> Sumber Air Minum
	X <sub>8</sub> Sumber Pene-rangan/ Listrik
Perumahan	X <sub>9</sub> Bahan Bakar Memasak
	X <sub>10</sub> Luas Lantai Rumah
	X <sub>11</sub> Jenis Lantai Rumah
	X <sub>12</sub> Jenis Dinding Rumah
	X <sub>13</sub> Tempat Buang Air Besar
Sosial Ekonomi	X <sub>14</sub> Status Pekerjaan Kepala Rumah Tangga
	X <sub>15</sub> Pelayanan Kesehatan Gratis
	X <sub>16</sub> Membeli Raskin
Teknologi Informasi	X <sub>17</sub> Ada/Tidak Anggota RT Menggunakan Telepon Seluler

**C. Langkah Analisis**

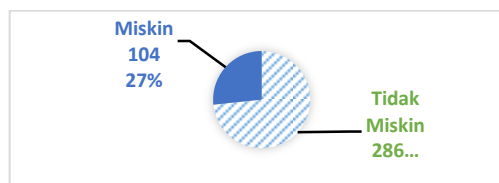
Langkah analisis yang dilakukan pada penelitian ini adalah sebagai berikut:

1. Mendeskripsikan karakteristik kesejahteraan rumah tangga di Provinsi Papua
2. Membagi data menjadi proporsi data *training* dan *testing* sebesar 70:30 dan 80:20
3. Mendapatkan hasil klasifikasi dari analisis regresi logistik biner.
4. Mendapatkan hasil klasifikasi dengan metode SVM
5. Membandingkan ketepatan klasifikasi yang diperoleh dari metode regresi logistik dan metode SVM
6. Menarik Kesimpulan

**IV. ANALISIS DAN PEMBAHASAN**

**A. Karakteristik Kesejahteraan Rumah Tangga di Papua**

Pengkategorian miskin dan tidak miskin dapat dilihat dari total pengeluaran rumah tangga per kapita per bulan yang dibandingkan dengan garis kemiskinan di Provinsi Papua yang telah ditetapkan oleh BPS, yakni pada tahun 2012, garis kemiskinan Provinsi Papua adalah Rp297,502,00. Pengeluaran kurang dari garis kemiskinan maka dikategorikan rumah tangga miskin, sebaliknya apabila pengeluaran lebih dari garis kemiskinan maka dikategorikan rumah tangga tidak miskin.



Gambar 1. Presentase Rumah Tangga Kategori Miskin dan Tidak Miskin

Gambar 1 menunjukkan bahwa kategori rumah tangga miskin di Provinsi Papua sebanyak 27%, yakni sebesar 104 rumah tangga dari 390 rumah tangga yang di survei, sedangkan kategori rumah tangga tidak miskin di Provinsi Papua sebesar 73%, yaitu dari 390 rumah tangga yang di survei terdapat 286 rumah tangga yang dikategorikan sebagai rumah tangga tidak miskin.

**B. Klasifikasi Kesejahteraan Rumah Tangga Provinsi Papua dengan Metode Regresi Logistik Biner**

Data yang digunakan merupakan data *training* pada partisi data 70:30 dan 80:20. Pada analisis regresi logistik, dilakukan uji signifikansi parameter terlebih dahulu baik secara univariabel maupun multivariabel.

Hasil perhitungan taksiran parameter secara univariabel pada partisi data *training* 70% didapatkan 5 variabel yang signifikan mempengaruhi variabel respon secara parsial dimana variabel tersebut memiliki nilai *p-value* <  $\alpha=0,05$ . Kelima variabel tersebut adalah variabel banyaknya anggota rumah tangga ( $x_1$ ), status kepemilikan rumah ( $x_6$ ), adanya sumber penerangan atau sumber listrik ( $x_8$ ), luas lantai rumah ( $x_{10}$ ), dan tempat buang air besar ( $x_{13}$ ). Pada partisi data *training* 80%, didapatkan hasil dari analisa regresi logistik biner secara univariabel menggunakan data *training* 80% yaitu didapatkan empat variabel signifikan mempengaruhi variabel respon, dimana variabel-variabel tersebut memiliki nilai *p-value* kurang dari  $\alpha=0,05$ . Keempat variabel tersebut adalah variabel banyaknya anggota rumah tangga ( $x_1$ ), status kepemilikan rumah ( $x_6$ ), adanya sumber penerangan atau sumber listrik ( $x_8$ ), dan luas lantai rumah ( $x_{10}$ ).

Selanjutnya, dilakukan analisis regresi logistik biner secara multivariabel untuk melihat pengaruh setiap variabel prediktor terhadap variabel respon secara bersamaan pada proporsi data *training* 70% dan 80%. Berikut adalah hasil analisis regresi logistik biner secara multivariabel pada dua proporsi data.

Tabel 3. Uji Serentak Proporsi data Training 70% dan 80%.

Model serentak	Chi-square	df	P-value
Proporsi data <i>training</i> 70%	84.302	13	0.000
Proporsi data <i>training</i> 80%	63.795	5	0.000

Tabel 3 menunjukkan bahwa model secara multivariabel memiliki nilai *p-value* kurang dari  $\alpha=0,05$ , sehingga pada uji hipotesis menolak hipotesis awal, maka dapat dikatakan model memiliki paling tidak satu nilai koefisien yang tidak sama dengan nol. Dapat disimpulkan bahwa model multivariat telah sesuai, dan memiliki setidaknya satu variabel prediktor yang memiliki pengaruh signifikan terhadap variabel respon.

Analisis selanjutnya yaitu membentuk model berdasarkan variabel-variabel yang dinyatakan signifikan pada analisis se-

belumnya. Berikut adalah pembentukan model menggunakan data *training* 70% dan 80%

Tabel 4. Model Regresi Logistik Biner pada Data *Training* 70% dan 80%

Paertisi Data	Variabel	B	Wald	P-value
Data Training 70%	x1	-0.267	8.369	0.004*
	x6(1)	-1.108	1.824	0.177
	x6(2)	0.646	0.385	0.535
	x8(1)	0.670	2.950	0.086
	x10	0.018	5.314	0.021*
	x13(1)	0.820	0.605	0.437
	x13(2)	0.349	0.118	0.732
	x13(3)	0.118	0.012	0.911
	x13(4)	-0.264	0.056	0.813
	x13(5)	0.147	0.019	0.891
	x13(6)	-0.907	0.837	0.360
	x13(7)	-23.594	0.000	1.000
	x13(8)	-3.165	3.985	0.046*
	Constant	1.706	1.918	0.166
Data Training 80%	x1	-.216	7.354	.007*
	x6(1)	-1.728	5.087	.024*
	x6(2)	-.314	.124	.725
	x8(1)	.813	6.537	.011*
	x10	.025	12.254	.000*
	Constant	1.775	4.731	.030*

\*Parameter signifikan pada alpha=5%

Variabel prediktor yang signifikan mempengaruhi kesejahteraan rumah tangga di Provinsi Papua pada proporsi data *training* 70% adalah banyaknya anggota rumah tangga ( $x_1$ ), luas lantai rumah ( $x_{10}$ ), dan tempat buang air besar yaitu pada danau ( $x_{13(8)}$ ), dimana nilai *p-value* ketiga variabel tersebut kurang dari  $\alpha = 0,05$ . Model klasifikasi yang terbentuk adalah sebagai berikut.

$$g(x) = -0,267X_1 + 0,018X_{10} - 3,165X_{13(8)}$$

Variabel prediktor yang signifikan mempengaruhi kesejahteraan rumah tangga di Provinsi Papua pada proporsi data *training* 80% yaitu banyaknya anggota rumah tangga ( $x_1$ ), status kepemilikan rumah milik sendiri ( $x_{6(1)}$ ), ada/tidak sumber penerangan atau sumber listrik ( $x_{8(1)}$ ), dan luas lantai rumah ( $x_{10}$ ). Model klasifikasi yang terbentuk adalah sebagai berikut.

$$g(x) = 1,775 - 0,216X_1 - 1,728X_{6(1)} + 0,813X_{8(1)} + 0,025X_{10}$$

C. Uji Kesesuaian Model Regresi Logistik Biner

Pengujian kesesuaian model menggunakan Uji Hosmer dan Lemeshow. Hasil uji Hosmer dan Lemeshow pada proporsi data *training* 70% dan 80% disajikan pada Tabel 5.

Tabel 5. Uji Hosmer dan Lemeshow

Partisi Data	Chi-square	df	Sig.
70:30	9.228	8	0.323
80:20	6.124	8	0.633

Hasil uji kesesuaian model menunjukkan bahwa untuk pada data *training* 70% dan 80%, pada hipotesis keduanya gagal me-nolak  $H_0$  dikarenakan *p-value* lebih dari  $\alpha = 0,05$ . Oleh karena itu, dapat disimpulkan bahwa model yang didapat dari proporsi data *training* 70% dan 80% telah sesuai.

D. Evaluasi Ketepatan Klasifikasi Regresi Logistik Biner

Evaluasi klasifikasi menggunakan regresi logistik biner dengan partisi data 70:30 dan 80:20 disajikan pada Tabel 6.

Tabel 6. Evaluasi Klasifikasi Regresi Logistik Pada Partisi 70:30 dan 80:20

Kriteria	70:30		80:20	
	Training 70%	Testing 30%	Training 80%	Testing 20%
APER	21.92%	18.80%	22.76%	23.08%
Ketepatan Klasifikasi	78.02%	81.20%	77.24%	76.92%

Ketepatan klasifikasi data *testing* pada proporsi data 80:20 lebih baik dibandingkan pada proporsi data 70:30. Didapatkan ketepatan klasifikasi data *testing* pada partisi data 80:20 yakni sebesar 81.20%.

E. Klasifikasi dengan Metode Support Vector Machine

Klasifikasi menggunakan metode SVM dibagi menjadi dua bagian, yakni menggunakan seluruh variabel prediktor dan variabel prediktor yang signifikan pada analisis regresi logistik.

Tabel 7. Ketepatan Klasifikasi SVM dengan Seluruh Variabel Prediktor

Fungsi Kernel	Parameter	C	70:30		80:20	
			70%	30%	80%	20%
Linear	0	10	46.15%	50.43%	44.55%	53.85%
		100	46.15%	50.43%	44.55%	53.85%
		1000	46.15%	50.43%	44.55%	53.85%
Polynomial	p = 1	10	78.75%	76.07%	79.49%	73.08%
		100	79.12%	76.07%	79.49%	76.92%
		1000	79.12%	76.07%	79.49%	76.92%
	p = 2	10	100%	74.36%	99.04%	<b>80.77%</b>

Tabel 7. Ketepatan Klasifikasi SVM dengan Seluruh Variabel Prediktor (Lanjutan)

Fungsi Kernel	Parameter	C	70:30		80:20	
			70%	30%	80%	20%
Polynomial	p = 2	100	100%	74.36%	100.00%	75.64%
		1000	100%	74.36%	100.00%	75.64%
		10	100%	74.36%	100.00%	74.36%
	p = 3	100	100%	74.36%	100.00%	75.64%
		1000	100%	74.36%	100.00%	74.36%
		10	100%	79.49%	100.00%	79.49%
Radial Basis Function (RBF)	$\sigma = 2$	100	100%	79.49%	100.00%	79.49%
		1000	100%	79.49%	100.00%	79.49%
		10	94.87%	75.21%	92.95%	75.21%
	$\sigma = 5$	100	99.63%	75.21%	99.36%	75.21%
		1000	100.00%	72.65%	100.00%	72.65%
		10	86.45%	<b>80.34%</b>	100.00%	80.34%
$\sigma = 10$	100	93.41%	73.50%	86.22%	73.50%	
	1000	98.17%	73.50%	91.67%	73.50%	

Ketepatan klasifikasi data *testing* tertinggi pada partisi data 70:30 yakni sebesar 80,34% dengan menggunakan kernel RBF ( $\sigma = 10$  dan  $C = 10$ ). Pada partisi data 80:20, didapatkan ketepatan klasifikasi data *testing* tertinggi menggunakan kernel Polynomial ( $p = 2$ ,  $C = 10$ ) sebesar 80.77%.

Ketepatan klasifikasi yang didapat menggunakan variabel prediktor yang signifikan pada analisis regresi logistik biner pada partisi data 70:30 dan 80:20 disajikan pada Tabel 8.

Tabel 8. Ketepatan Klasifikasi SVM dengan Variabel Prediktor Terpilih

Fungsi Kernel	Parameter	C	70:30		80:20	
			70%	30%	80%	20%
Linear	0	10	53.11%	55.56%	63.46%	78.21%
		100	50.55%	54.70%	60.26%	76.92%
		1000	50.55%	54.70%	59.94%	76.92%
Polynomial	p = 1	10	73.63%	72.65%	73.72%	71.79%
		100	73.63%	72.65%	73.72%	71.79%
		1000	73.63%	72.65%	73.72%	71.79%
	p = 2	10	79.49%	81.20%	76.60%	78.21%
		100	79.49%	81.20%	76.60%	78.21%
		1000	79.49%	81.20%	76.60%	78.21%
p = 3	10	81.68%	76.92%	79.81%	78.21%	

Radial Basis Function (RBF)	σ = 2	100	79.49%	81.20%	76.60%	78.21%
		1000	81.68%	76.92%	80.13%	79.49%
		10	80.59%	<b>82.05%</b>	77.88%	75.64%
	σ = 4	100	84.62%	79.49%	81.41%	<b>80.77%</b>
		1000	86.45%	73.50%	83.01%	74.36%
		10	79.12%	81.20%	77.56%	76.92%
	σ = 6	100	80.95%	81.20%	76.92%	76.92%
		1000	79.49%	81.20%	78.85%	79.49%
		10	78.75%	80.34%	75.00%	74.36%
	σ = 6	100	80.22%	<b>82.05%</b>	77.88%	74.36%
		1000	80.95%	81.20%	77.24%	79.49%

Pada Tabel 8, ditunjukkan bahwa terdapat dua model yang menghasilkan ketepatan klasifikasi data *testing* tertinggi pada partisi data 70:30 yakni sebesar 82,05% pada kernel RBF (σ=2, C=10) dan RBF (σ=6, C=100). Diantara kedua model tersebut, dipilih model dengan kernel RBF (σ=2, C=10) sebagai model yang menghasilkan ketepatan data *testing* tertinggi dikarenakan selain ketepatan klasifikasi data *testing*, ketepatan klasifikasi data *training* model tersebut juga lebih tinggi. Pada partisi data 80:20, ketepatan klasifikasi data *testing* tertinggi menggunakan fungsi kernel RBF (σ=2 dan C=10), yakni sebesar 80,77%.

F. Perbandingan Hasil Klasifikasi

Perbandingan hasil klasifikasi dengan metode regresi logistik biner dan SVM pada partisi data 70:30 dan 80:20 berdasarkan nilai ketepatan klasifikasinya disajikan pada Tabel 10. Pada Tabel 9, ditunjukkan bahwa klasifikasi dengan SVM lebih baik dibandingkan dengan menggunakan regresi logistik biner, baik pada data *training* maupun data *testing*. Hal ini dapat dili-hat dari nilai ketepatan klasifikasi dengan metode SVM yang lebih tinggi dibandingkan dengan metode regresi logistik biner.

Tabel 9. Perbandingan Hasil Klasifikasi Metode Regresi Logistik dan SVM

Partisi data	Regresi Logistik	SVM	
		Seluruh Variabel	Variabel Signifikan
70:30	70%	78.02%	80.59%
	30%	81.20%	<b>82.05%</b>
80:20	80%	77.24%	81.41%
	20%	76.92%	80.77%

Model klasifikasi yang menghasilkan nilai ketepatan klasifikasi tersebut menggunakan fungsi kernel *radial basis function* (RBF) dengan parameter σ = 2 dan C=10. Pembentukan model tersebut disajikan pada tabel sebagai berikut.

Tabel 10. Model Klasifikasi SVM kernel RBF (σ = 2, C=10)

Alpha (Lagrange Multipliers)									
i	α <sub>i</sub>	i	α <sub>i</sub>	i	α <sub>i</sub>	i	α <sub>i</sub>	i	α <sub>i</sub>
1	0	56	8.5364	111	0	166	10	221	10
2	0	57	10	112	0	167	10	222	0
3	10	58	10	113	10	168	0	223	10
4	0	59	10	114	0	169	0	224	10
5	0	60	10	115	0	170	10	225	10
6	0	61	8.5364	116	0	171	10	226	10
7	0	62	0	117	10	172	10	227	10
8	0	63	10	118	0	173	0	228	3.0865
9	0	64	10	119	10	174	10	229	10
10	0	65	0	120	0.9607	175	0	230	10
11	0	66	10	121	10	176	2.5645	231	0
12	0.8031	67	10	122	5.6687	177	0	232	10
13	0	68	10	123	0	178	9.1567	233	0
14	0	69	0	124	0	179	2.5645	234	0

15	10	70	0	125	10	180	10	235	4.3184
16	0	71	6.7782	126	0	181	10	236	1.2577
17	0	72	0	127	10	182	0	237	0
18	0	73	1.015	128	8.6856	183	0	238	0
19	0	74	9.8039	129	1.4951	184	10	239	0.281
20	0	75	10	130	0	185	10	240	0
21	0	76	10	131	0	186	10	241	0
22	2.4999	77	10	132	0	187	7.8139	242	0
23	0	78	10	133	10	188	0	243	0
24	0	79	10	134	0	189	0	244	0
25	0	80	0	135	0	190	0	245	10
26	3.3656	81	0	136	0	191	0	246	0
27	0	82	0	137	0	192	0	247	0
28	0	83	0	138	0	193	0	248	0
29	0	84	10	139	0	194	0	249	0
30	0	85	10	140	0	195	0	250	0
31	0	86	0	141	10	196	0.281	251	0
32	10	87	10	142	0	197	0	252	0
33	0	88	10	143	0	198	10	253	0
34	10	89	10	144	0	199	0	254	0
35	10	90	10	145	10	200	0	255	10
36	10	91	1.5806	146	0	201	0	256	1.2577
37	0	92	10	147	10	202	0	257	0
38	9.1567	93	10	148	0	203	0	258	3.0368
39	10	94	10	149	0	204	0	259	3.3954
40	10	95	4.0114	150	0	205	0	260	0
41	10	96	0	151	10	206	0	261	0
42	10	97	0	152	10	207	0	262	0
43	10	98	0	153	0	208	0	263	10
44	10	99	2.7092	154	10	209	10	264	0
45	10	100	10	155	0	210	10	265	9.6037
46	2.3889	101	10	156	0	211	10	266	0

Tabel 10. Model Klasifikasi SVM kernel RBF (σ = 2, C=10) (Lanjutan)

Alpha (Lagrange Multipliers)									
i	α <sub>i</sub>	i	α <sub>i</sub>	i	α <sub>i</sub>	i	α <sub>i</sub>	i	α <sub>i</sub>
47	0	102	0.5329	157	10	212	10	267	0.6467
48	0	103	0	158	10	213	0	268	0
49	10	104	0	159	0	214	10	269	1.6463
50	9.6765	105	10	160	10	215	10	270	10
51	10	106	10	161	10	216	10	271	5.4773
52	10	107	10	162	0	217	10	272	0
53	10	108	0	163	10	218	10	273	10
54	10	109	0	164	0	219	10		
55	10	110	0	165	0	220	10		

Beta = 0

Nilai alpha dan beta (Tabel 10) digunakan dalam pembentukan persamaan model klasifikasi pada model SVM dengan kernel RBF (σ=2, C=10). Persamaan model klasifikasi yaitu:

$$f(x) = \text{sgn} \left( \sum_{i \in SV} \alpha_i y_i K(x_i, x) + b \right)$$

Hasil klasifikasi yaitu apabila  $f(x) \geq 0$  maka dikategorikan ke dalam kategori rumah tangga tidak miskin, sedangkan apabila  $f(x) < 0$  maka dikategorikan rumah tangga miskin.

V. KESIMPULAN DAN SARAN

Klasifikasi dengan metode SVM lebih baik apabila dibandingkan menggunakan regresi logistik biner, baik menggunakan data *training* maupun data *testing* pada partisi data 70:30 dan 80:20. Model klasifikasi yang menghasilkan ketepatan klasifikasi data *testing* tertinggi, yakni sebesar 82,05%, menggunakan kernel RBF (σ=2 dan C=10).

Saran yang dapat diberikan peneliti yaitu untuk penelitian selanjutnya, sebaiknya lebih banyak fungsi kernel yang

diguna-kan pada metode SVM dalam membandingkan hasil klasifikasi. Selain itu, pada penelitian selanjutnya, dalam membandingkan metode SVM dapat dilakukan dengan metode lainnya yang menonjol pada kasus data *non linear separable* untuk mendapatkan perbandingan yang lebih sensitif.

#### DAFTAR PUSTAKA

- [1] Badan Pusat Statistik, 2014. Diakses 10 Januari 2014, dari: <http://www.bps.go.id/webbeta/frontend/linkTabelStatis/view/id/1488>.
- [2] E. Q. Aeni, "Pendekatan Cart Arcing untuk Klasifikasi Kesejahteraan Rumah Tangga di Propinsi Jawa Tengah," Institut Teknologi Sepuluh Nopember, Surabaya, 2010.
- [3] E. S. Ningrum, "Klasifikasi Kesejahteraan Rumah Tangga di Kota Malang dengan Pendekatan Bagging Regresi Logistik," Institut Teknologi Sepuluh Nopember, Surabaya, 2012.
- [4] D. W. Hosmer, S. Lemeshow and X. R. Sturdivant, Applied Logistic Regression, 3rd ed., New York: John Wiley & Sons, 2013.
- [5] A. Agresti, An Introduction to Categorical Data Analysis, New York: John Wiley & Sons, Inc, 2007.
- [6] S. R. Gunn, "Support Vector Machines for Classification and Regression," University of Southampton, Southampton, 1998.
- [7] R. A. Johnson and D. W. Wichern, "Applied Multivariate Statistical Analysis, Sixth Edition," United States of America, Pearson Prantice Hall, 2007.
- [8] Badan Pusat Statistik, Perhitungan dan Analisis Kemiskinan Makro Indonesia tahun 2014, Jakarta: Badan Pusat Statistik, 2014.