

## METODE BOOTSTRAP AGGREGATING REGRESI LOGISTIK BINER UNTUK KETEPATAN KLASIFIKASI KESEJAHTERAAN RUMAH TANGGA DI KOTA PATI

Ridha Ramandhani<sup>1</sup>, Sudarno<sup>2</sup>, Diah Safitri<sup>3</sup>

<sup>1</sup> Mahasiswa Jurusan Statistika FSM UNDIP

<sup>2,3</sup> Dosen Jurusan Statistika FSM UNDIP

### ABSTRAK

Kesejahteraan merupakan salah satu aspek yang cukup penting untuk menjaga dan membina terjadinya stabilitas sosial dan ekonomi. Berbagai penelitian yang telah dilakukan mengenai kesejahteraan mengindikasikan bahwa banyak faktor yang mempengaruhi kesejahteraan rumah tangga. Faktor-faktor yang mempengaruhi kesejahteraan rumah tangga antara lain jenis kelamin kepala rumah tangga, usia kepala rumah tangga, lapangan usaha kepala rumah tangga, jumlah anggota rumah tangga, bahan bakar utama untuk memasak, pengalaman membeli raskin dan ada atau tidaknya anggota keluarga yang menguasai penggunaan telepon seluler/HP. Dalam penelitian ini dilakukan kajian tentang klasifikasi kesejahteraan rumah tangga di Kota Pati dengan tujuan untuk mengidentifikasi faktor-faktor apa saja yang mempengaruhi kesejahteraan rumah tangga di Kota Pati. Dari hasil kajian dengan menggunakan metode *Bootstrap Aggregating (Bagging)* regresi logistik biner diperoleh tiga variabel prediktor yang berpengaruh signifikan terhadap kesejahteraan rumah tangga di Kota Pati, yaitu jenis kelamin kepala keluarga, jumlah anggota rumah tangga, dan penguasaan telepon seluler dengan tingkat akurasi sebesar 79,87%. Hasil analisis *bagging* regresi logistik biner dengan replikasi *bootstrap* sebesar 50, 60, 70, 80, 90, 100, 150, 200, 626, dan 1000 kali menunjukkan bahwa terdapat konsistensi pada setiap pengulangan.

Kata Kunci : Klasifikasi, Regresi Logistik Biner, *Bootstrap Aggregating*

## 1. PENDAHULUAN

### 1.1 Latar Belakang

Kesejahteraan merupakan salah satu aspek yang cukup penting untuk menjagakan membina terjadinya stabilitas sosial dan ekonomi. Kondisi tersebut juga diperlukan untuk meminimalkan terjadinya kecemburuan sosial dalam masyarakat (Sugiarto, 2007). Krisis ekonomi di Indonesia yang terjadi pada tahun 1998 menyebabkan kemunduran berbagai kegiatan ekonomi, terganggunya kegiatan produksi dan distribusi. Permasalahan tersebut membawa dampak yang serius pada peningkatan jumlah pengangguran dan penduduk miskin. Bahkan ketika pertumbuhan ekonomi yang sempat menurun akibat krisis dapat teratasi dan dapat dipulihkan, kemiskinan tetap saja sulit untuk ditanggulangi (Arianti dkk., 2008).

Dalam penulisan Tugas Akhir ini metode yang digunakan adalah regresi logistik dimana akan digunakan untuk mengetahui pola kesejahteraan rumah tangga dengan faktor-faktor yang mempengaruhinya dan nilai ketepatan klasifikasi kesejahteraan rumah tangga di Kota Pati. Nilai ketepatan klasifikasi tersebut dapat ditingkatkan dengan metode *Bootstrap Aggregating (Bagging)* Regresi Logistik Biner.

### 1.2 Tujuan

Tujuan dari penelitian dalam Tugas Akhir ini adalah mengetahui model dan ketepatan klasifikasi kesejahteraan rumah tangga di Kota Pati berdasarkan analisis Regresi Logistik Biner dan mengetahui ketepatan klasifikasi kesejahteraan rumah tangga di Kota Pati dengan metode *Bootstrap Aggregating (Bagging)* Regresi Logistik Biner.

## 2. TINJAUAN PUSTAKA

### 2.1 Kesejahteraan Rumah Tangga

Menurut Cahyat dkk. (2007), kesejahteraan merupakan kemampuan untuk memenuhi kebutuhan dasar yang bersifat material (kebendaan) maupun bukan material, yang mencakup aspek gizi dan kesehatan, pengetahuan, dan kekayaan materi. Sedangkan kemiskinan merupakan suatu situasi dimana seseorang atau rumah tangga mengalami kesulitan untuk memenuhi kebutuhan dasar, sementara lingkungan pendukungnya kurang memberikan peluang untuk meningkatkan kesejahteraan secara berkesinambungan atau untuk keluar dari kerentanan.

Untuk mengukur kemiskinan, Badan Pusat Statistik (BPS) menggunakan konsep kemampuan memenuhi kebutuhan dasar (*basic needs approach*). Dengan pendekatan ini, kemiskinan dipandang sebagai ketidakmampuan dari sisi ekonomi untuk memenuhi kebutuhan dasar makanan dan bukan makanan yang diukur dari sisi pengeluaran. Penduduk miskin adalah penduduk yang memiliki rata-rata pengeluaran perkapita perbulan dibawah Garis Kemiskinan (GK). Garis Kemiskinan (GK) merupakan penjumlahan dari Garis Kemiskinan Makanan (GKM) dan Garis Kemiskinan Non Makanan (GKNM). Angka Garis Kemiskinan Kota Pati pada tahun 2014 adalah sebesar Rp 332.228,00 (BPS, 2014).

### 2.2. Faktor – Faktor Yang Mempengaruhi Kesejahteraan Rumah Tangga

Termuat dalam SKPD (Strategi Penanggulangan Kemiskinan Daerah) Provinsi Jawa Tengah Tahun 2015-2018, variabel–variabel yang menjadi ciri kesejahteraan suatu keluarga antara lain: lapangan usaha dari pekerjaan utama kepala rumah tangga, status kedudukan dalam pekerjaan utama, jenis kelamin, umur, status perkawinan, dan ijazah tertinggi yang dimiliki kepala rumah tangga.

Jumlah anggota rumah tangga diduga mempunyai keterkaitan erat dengan kesejahteraan rumah tangga karena kemiskinan dihitung berdasar pengeluaran dan jumlah anggota rumah tangga. Makin besar jumlah anggota rumah tangga akan makin besar pula resiko untuk menjadi miskin apabila pendapatannya tidak meningkat (Faturochman dan Molo, 1995). Umur kepala rumah tangga juga berkaitan dengan kesejahteraan rumah tangga walaupun hubungannya tidak begitu jelas, akan tetapi ada kecenderungan bahwa kepala rumah tangga yang lebih sejahtera lebih tua dibandingkan kepala rumah tangga yang kurang sejahtera. Jenis karakteristik lain adalah karakteristik jenis pekerjaan. Kemampuan mayoritas rumah tangga untuk keluar dari kemiskinan akan bergantung pada upah mereka dari pekerjaan yang dilakukan. Jadi penting untuk menguji hubungan antara kesejahteraan dengan jenis pekerjaan anggota rumah tangga yang berada dalam usia kerja (Pratama, 2011). Dillon dan Hermanto dalam Faturochman dan Molo (1995) mengungkapkan bahwa kenyataannya, sebagian penduduk atau rumah tangga miskin di desa masih mengandalkan pertanian sebagai pekerjaan utamanya akan tetapi usaha-usaha di luar pertanian tetap menjadi sumber pendapatan komplementer dan alternatif bagi keluarga. Sedangkan rumah tangga miskin di kota lebih banyak mengandalkan penghasilan dari sektor-sektor jasa atau lebih dikenal dengan sektor informal. Sedangkan karakteristik umum penduduk miskin menurut Rusastra dan Togar (2007) adalah sebagian besar tinggal di desa, bekerja di sektor pertanian, sifat pekerjaan adalah informal serta status pekerjaan sebagai pekerja keluarga yang tidak di bayar. Salah satu karakteristik umum penduduk miskin lainnya menurut *The World Bank* (2006) adalah sifat pekerjaan yang bersifat informal serta status pekerjaan sebagai pekerja keluarga yang tidak dibayar.

### 2.3. Regresi Logistik

Menurut Agresti (1990), regresi logistik merupakan metode yang menghubungkan antara variabel respon yang bersifat kategorik dengan variabel prediktor.

### 2.4. Regresi Logistik Biner

Hosmer dan Lemeshow (2000) menyatakan bahwa regresi logistik biner digunakan untuk menjelaskan hubungan antara beberapa variabel prediktor X terhadap variabel respon Y yang bersifat dikotomi atau biner. Dikotomi artinya variabel respon Y hanya bernilai 1 untuk keberadaan suatu karakteristik dan bernilai 0 untuk ketidakberadaan suatu karakteristik.

Persamaan model regresi logistik menurut Hosmer dan Lemeshow (2000) adalah:

$$\pi(x_i) = \frac{e^{\beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi}}}{1 + e^{\beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi}}} \quad (1)$$

dengan fungsi logit  $g(x_i)$  yaitu:

$$g(x_i) = \ln \left[ \frac{\pi(x_i)}{1 - \pi(x_i)} \right] = \beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi} \quad (2)$$

Persamaan (1) dan (2) dapat disederhanakan menjadi:

$$\pi(x_i) = \frac{e^{g(x_i)}}{1 + e^{g(x_i)}} \quad (3)$$

#### 2.4.1. Penaksiran Parameter Model Regresi Logistik Biner

Menurut Agresti (1990), estimasi parameter regresi logistik yang dapat menggunakan metode *Maximum Likelihood Estimator* (MLE) yang membutuhkan turunan pertama dan kedua dari fungsi log likelihood. Hosmer dan Lemeshow (2000) mengatakan, pada dasarnya metode *Maximum Likelihood Estimator* menghasilkan nilai parameter yang memaksimalkan fungsi log likelihood dengan uji signifikansi parameter yang digunakan adalah uji Rasio Likelihood dan uji Wald.

#### 2.5.1. Uji Rasio Likelihood (Uji Keseluruhan)

Hosmer dan Lemeshow (2000) mengatakan bahwa, uji Rasio Likelihood merupakan uji signifikansi parameter secara keseluruhan atau bersama-sama. Berikut langkah-langkah uji Rasio Likelihood:

Hipotesis :

$H_0$  :  $\beta_1 = \beta_2 = \dots = \beta_p = 0$ , artinya tidak ada pengaruh antara variabel prediktor dengan variabel respon.

$H_1$  : paling sedikit ada satu  $\beta_j \neq 0$ ,  $j = 1, 2, \dots, p$ , artinya paling tidak ada satu variabel prediktor yang berpengaruh terhadap variabel respon.

Statistik Uji:

$$G = -2 \ln \left( \frac{\log \text{likelihood tanpa variabel bebas}}{\log \text{likelihood dengan variabel bebas}} \right)$$

Pada taraf signifikansi  $\alpha$ ,  $H_0$  akan ditolak apabila nilai  $G > \chi^2_{(\alpha;p)}$  atau p-value  $< \alpha$ , dengan kesimpulan bahwa secara keseluruhan atau secara bersama-sama variabel prediktor mempengaruhi variabel respon.

### 2.5.2. Uji Wald (Uji Parameter Secara Individu)

Bewick *et al.*, (2005) mengatakan, uji Wald digunakan untuk menguji signifikansi koefisien parameter dalam model. Berikut langkah-langkah uji Wald:

Hipotesis :

$H_0$  :  $\beta_j = 0$ , artinya tidak ada pengaruh variabel prediktor ke-j terhadap variabel respon,  $j = 1, 2, \dots, p$

$H_1$  :  $\beta_j \neq 0$ , artinya ada pengaruh variabel prediktor ke-j terhadap variabel respon,  $j = 1, 2, \dots, p$

Statistik Uji yang digunakan adalah:

$$W_j = \left\{ \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \right\}^2$$

$H_0$  akan ditolak apabila nilai  $W_j > \chi^2_{(\alpha;1)}$  atau p-value  $< \alpha$ , dengan kesimpulan bahwa  $\beta_j$  signifikan atau dengan kata lain secara parsial variabel X berpengaruh terhadap variabel respon.

Dari estimasi parameter model regresi logistik yang telah diperoleh, kemudian dilakukan uji *Goodness of Fit*. Menurut Bewick *et al.*, (2005), uji *Goodness of Fit* digunakan untuk mengukur seberapa baik model dapat menggambarkan variabel respon. Hosmer dan Lemeshow (2000), mengatakan bahwa uji *Goodness of Fit* merupakan uji yang dilakukan untuk mengetahui apakah ada perbedaan antara prediksi dan hasil observasi (model sesuai atau tidak).

### 2.5.3 Uji Goodness of Fit

Dalam Hosmer dan Lemeshow (2000), berikut langkah-langkah uji *Goodness of Fit*:

Hipotesis :

$H_0$  : Model sesuai (tidak ada perbedaan antara hasil observasi dengan hasil prediksi)

$H_1$  : Model tidak sesuai (terdapat perbedaan antara hasil observasi dengan hasil prediksi)

Statistik Uji :

$$\hat{C}(\text{Hosmer-Lemeshow}) = \sum_{k=1}^g \frac{(o_k - n'_k \bar{\pi}_k)^2}{n'_k \bar{\pi}_k (1 - \bar{\pi}_k)}$$

dengan:

$g$  = jumlah grup

$n'_k$  = jumlah subjek pada grup ke-k

$O_k = \sum_{j=1}^{C_k} y_j$ , jumlah nilai variabel respon pada grup ke-k

$\bar{\pi}_k = \sum_{j=1}^{C_k} \frac{m_j \hat{\pi}(x_j)}{n'_k}$ , rata-rata taksiran probabilitas dimana  $m_j$  adalah banyaknya subjek pada  $C_k$  kombinasi variabel prediktor.

Dengan daerah penolakan  $H_0$  adalah  $\hat{C} > \chi^2_{(\alpha;g-2)}$ . Jika  $H_0$  benar, maka distribusi statistik uji  $\hat{C}$  mengikuti distribusi *chi-square* dengan derajat bebas  $g-2$ .

## 2.6. Prosedur Klasifikasi

Menurut Rencher (2002), tingkat kesalahan klasifikasi atau biasa disebut *Apparent Error Rate (APER)* digunakan untuk menilai kemampuan prosedur klasifikasi dalam menentukan keanggotaan kelompok. Bisa juga menggunakan sebaliknya, yaitu *Correct Error Rate* atau tingkat ketepatan klasifikasi.

**Tabel 1.** Matriks Konfusi

Hasil Observasi	Taksiran	
	y <sub>1</sub>	y <sub>2</sub>
y <sub>1</sub>	n <sub>11</sub>	n <sub>12</sub>
y <sub>2</sub>	n <sub>21</sub>	n <sub>22</sub>

$$APER(\%) = \frac{n_{12} + n_{21}}{n_{11} + n_{12} + n_{21} + n_{22}}$$

Dari perhitungan nilai *Apparent Error Rate (APER)* yang telah diuraikan tersebut, maka dapat dilihat nilai errornya, sehingga untuk mencari nilai ketepatannya dapat menggunakan *1-Apparent Error Rate*.

## 2.8. Bootstrap Aggregating (Bagging)

*Bagging* merupakan suatu metode yang pertama kali dikenalkan oleh Breiman (1994). Menurut Breiman (1994), metode ini digunakan sebagai alat untuk memperbaiki stabilitas dan kekuatan prediksi dengan cara mereduksi variansi dari suatu prediktor. Sebuah himpunan data  $\mathcal{F}$  terdiri dari  $\{(y_n, x_n), n = 1, \dots, N\}$  dengan  $y$  dapat berupa kelas label atau numerik respon dan  $x$  merupakan variabel prediktor. Jika input adalah  $x$ , maka  $y$  diprediksi dengan  $\varphi(x, \mathcal{F})$  dimana  $\varphi(x, \mathcal{F})$  adalah parameter. Selanjutnya dilakukan pembootstrapan data asli sehingga replikasi bootstrap ke- $k$  diperoleh parameter  $\{\varphi(x, \mathcal{F}_k)\}$ . Replikasi *bootstrap* dilakukan sebanyak  $B$  kali sehingga  $\{\mathcal{F}^{(B)}\}$  dari  $\mathcal{F}$  dan dibentuk parameter  $\{\varphi(x, \mathcal{F}^{(B)})\}$ .  $\{\mathcal{F}^{(B)}\}$  adalah resampling dengan pengembalian. Jika  $y$  merupakan data numerik, prosedur nyata untuk menggantikan  $\varphi(x, \mathcal{F})$  dengan mengambil rata-rata dari  $\{\varphi(x, \mathcal{F}_k)\}$ . Jika  $\varphi(x, \mathcal{F})$  memprediksi sebuah kelas  $j \in \{1, \dots, r\}$ , maka salah satu metode dengan menggabungkan  $\{\varphi(x, \mathcal{F}_k)\}$  untuk memprediksi kelas. Ambil *bootstrap* sampel dengan pengulangan  $\{\mathcal{F}^{(B)}\}$  dari  $\mathcal{F}$  dan membentuk  $\{\varphi(x, \mathcal{F}^{(B)})\}$ . Jika  $y$  merupakan data kategorik, maka untuk menentukan kategorik dengan menggunakan  $\{\varphi(x, \mathcal{F}^{(B)})\}$  dari  $\varphi_B(x)$ , dimana  $\varphi_B(x) = \text{average}_B \varphi(x, \mathcal{F}^{(B)})$ . *Bagging* bekerja dengan baik pada metode klasifikasi yang menghasilkan parameter yang tidak stabil, dimana perubahan kecil pada data setakan menghasilkan perubahan besar pada model yang diperoleh. Beberapa metode yang tidak stabil adalah *neural network*, regresi, klasifikasi, dan regresi pohon. Sedangkan contoh metode yang stabil adalah *k-nearest neighbor* (Breiman, 1994).

## 3. METODOLOGI PENELITIAN

### 3.1 Sumber Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang merupakan hasil dari Survei Sosial Ekonomi Nasional (SUSENAS) tahun 2014.

### 3.2 Variabel Penelitian

Penelitian ini melibatkan beberapa variabel yang terdiri dari variabel prediktor ( $x$ ) beserta variabel respon ( $y$ ). Variabel respon yang digunakan dalam penelitian ini dibedakan menjadi 2 kategori, yaitu (0) miskin dan (1) tidak miskin. Penentuan kategori

untuk miskin dan tidak miskin ditentukan dari besarnya pengeluaran perkapita rumah tangga per bulan tahun 2014. Untuk variabel prediktor berikut adalah penjelasannya:

**Tabel 2.** Variabel Prediktor

Variabel Prediktor	Arti
X1	Jenis kelamin kepala rumah tangga
X2	Usia kepala rumah tangga
X3	Status perkawinan kepala rumah tangga
X4	Ijazah tertinggi kepala rumah tangga
X5	Lapangan usaha kepala rumah tangga
X6	Status pekerjaan utama kepala rumah tangga
X7	Jumlah anggota rumah tangga
X8	Bahan bakar energi utama untuk memasak
X9	Pengalaman menerima layanan kesehatan gratis
X10	Pengalaman membeli raskin
X11	Ada atau tidaknya anggota keluarga yang menguasai penggunaan HP
X12	Ada atau tidaknya anggota keluarga yang menguasai penggunaan PC/Laptop

#### 4. HASIL DAN PEMBAHASAN

##### 4.1 Model Awal Regresi Logistik Biner

Berdasarkan pengolahan data didapat model awal regresi logistik biner. Berikut adalah model yang dihasilkan.

$$\pi(x_i) = \frac{e^{g(x_i)}}{1 + e^{g(x_i)}}$$

dengan:

$$g(x) = 2,762 - 1,437 X_{1_1} + 0,013 X_2 + 0,569 X_{3_1} - 0,082 X_{4_1} - 0,160 X_{4_2} + 0,109 X_{4_3} - 0,523 X_{4_4} + 1,091 X_{5_1} + 0,828 X_{5_2} + 0,015 X_{6_1} - 0,463 X_{6_2} - 0,899 X_7 + 37,995 X_{8_1} + 0,841 X_{8_2} - 1,547 X_{8_3} + 0,159 X_{8_4} - 0,272 X_{9_1} + 0,105 X_{10_1} + 0,916 X_{11_1} + 0,283 X_{12_1}$$

##### 4.2 Uji Rasio Likelihood (Uji Parameter Secara Keseluruhan)

Hipotesis

$H_0$  :  $\beta_1 = \beta_2 = \dots = \beta_{20} = 0$  (secara bersama – sama variabel prediktor tidak mempengaruhi model)

$H_1$  : paling sedikit ada satu  $\beta_j \neq 0$ , dengan  $j = 1, 2, \dots, 20$  (secara bersama-sama variabel prediktor mempengaruhi model)

Taraf Signifikansi

$$\alpha = 5\%$$

Statistik Uji

$$G = -2 \ln \left( \frac{\text{likelihood tanpa variabel bebas}}{\text{likelihood dengan variabel bebas}} \right) = 141,739$$

Kriteria Uji

$H_0$  ditolak jika  $G > \chi^2_{(0,05;20)}$

Keputusan

Karena  $(G = 141,739) > (\chi^2_{(0,05;20)} = 31,41)$  maka  $H_0$  ditolak

Kesimpulan

Jadi pada taraf signifikansi sebesar 5%, dapat disimpulkan bahwa secara bersama-sama variabel prediktor mempengaruhi model.

#### 4.3. Uji Wald (Uji Parameter Secara Individu)

Hipotesis

$H_0$  :  $\beta_j = 0, j = 1, \dots, 3$  (tidak ada pengaruh variabel prediktor ke-j terhadap variabel respon)

$H_1$  :  $\beta_j \neq 0, j = 1, \dots, 3$  (ada pengaruh variabel prediktor ke-j terhadap variabel respon)

Taraf Signifikansi

$\alpha = 5\%$

Statistik Uji

$$W = \left( \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \right)^2$$

Kriteria Uji

$H_0$  ditolak jika  $W > \chi^2_{(0,05;1)}$

**Tabel 3.** Nilai Wald untuk Setiap Parameter

Variabel Prediktor (X)	Wald (W)	$\chi^2_{(0,05;1)}$	Sign.	Keputusan
Jenis Kelamin ( $X_1$ )				
Laki-laki ( $X_{1\_1}$ )	4,891	3,841	0,027	$H_0$ ditolak
Jumlah Anggota Rumah Tangga ( $X_7$ )	70,544	3,841	0,000	$H_0$ ditolak
HP ( $X_{11}$ )				
Ya ( $X_{11\_1}$ )	16,581	3,841	0,000	$H_0$ ditolak

Kesimpulan

Jadi pada taraf signifikansi sebesar 5% dapat disimpulkan bahwa variabel  $X_1$ ,  $X_7$ ,  $X_{11}$  mempengaruhi model.

#### 4.4. Uji Goodness of Fit (Uji Kecocokan Model)

Setelah uji signifikansi parameter, dilakukan uji kecocokan model sebagai berikut:

Hipotesis

$H_0$  : Tidak ada perbedaan antara prediksi dan hasil observasi (model sesuai)

$H_1$  : Ada perbedaan antara prediksi dan hasil observasi (model tidak sesuai)

Taraf Signifikansi

$\alpha = 5\%$

Statistik Uji

$$\hat{C}(\text{Hosmer-Lemeshow}) = \sum_{k=1}^g \frac{(o_k - n'_k \bar{\pi}_k)^2}{n'_k \bar{\pi}_k (1 - \bar{\pi}_k)}$$

Kriteria Uji

$$H_0 \text{ ditolak jika } \widehat{C} > \chi^2_{(0,05;g-2)}$$

Keputusan

$$H_0 \text{ diterima karena } (\widehat{C} = 8,628) < \chi^2_{(0,05;7)} = 14,067)$$

Kesimpulan

Jadi pada taraf signifikansi sebesar 5%, dapat disimpulkan bahwa tidak ada perbedaan antara prediksi dan hasil observasi (model yang terbentuk sesuai).

#### 4.5. Model Akhir

Berdasarkan pengujian yang dilakukan, maka diperoleh model akhir sebagai berikut :

$$\pi(x_i) = \frac{e^{4,024-0,832X_{1_1} - 0,831 X_7+1,168 X_{11_1}}}{1 + e^{4,024-0,832X_{1_1} - 0,831 X_7+1,168 X_{11_1}}}$$

#### 4.6. Ketepatan Klasifikasi

**Tabel 4.** Tabel Hasil Klasifikasi Metode Regresi Logistik Biner

Hasil Observasi	Prediksi	
	Miskin	Tidak Miskin
Kelas 0 (Miskin)	29	107
Kelas 1(Tidak Miskin)	21	469

$$APER = \frac{21+107}{29+107+21+469} = 0,2044728435$$

Sehingga untuk presentase ketepatan klasifikasi adalah  $1 - APER = 1 - 0,2044728435 = 0,7955271565 = 79,55271565\%$

#### 4.7. Bagging Regresi Logistik Biner

Hasil pada model regresi logistik biner menyimpulkan bahwa variabel yang signifikan berpengaruh pada kesejahteraan rumah tangga adalah jenis kelamin ( $X_1$ ), jumlah anggota rumah tangga ( $X_7$ ), dan ada atau tidaknya anggota keluarga yang dapat menggunakan telepon seluler ( $X_{11}$ ). Ketiga variabel tersebut yang kemudian akan diperlakukan resampling *bootstrap*.

Pada Tugas Akhir ini, perhitungan ketepatan klasifikasi dilakukan pada setiap iterasi hasil perhitungan ketepatan pada setiap replikasi *bootstrap*, kemudian dirata-rata sehingga menghasilkan nilai ketepatan klasifikasi *bagging* regresi logistik biner. Tabel 5 merupakan hasil dari *bagging* dengan 50, 60, 70, 80, 90, 100, 150, 200, 626, dan 1000 replikasi *bootstrap*. Diperoleh hasil bahwa pada replikasi *bootstrap* sebanyak 90 diperoleh hasil penurunan ketepatan klasifikasi terbesar.

**Tabel 5.** Hasil *Bagging* Regresi Logistik Biner

Replikasi <i>Bootstrap</i>	Rata rata ketepatan klasifikasi	$e_B$	$\bar{e}_B$	Penurunan ketepatan klasifikasi
50	0,796034%	20,44728435 %	0,203966%	0,05068435 %
60	0,79815333%	20,44728435%	0,20184677 %	0,26260735 %
70	0,79757714%	20,44728435%	0,20242286 %	0,20499835 %

80	0,79851625%	20,44728435%	0,20148375 %	0,29890935 %
<b>90</b>	<b>0,79873556%</b>	<b>20,44728435%</b>	<b>0,20126444%</b>	<b>0,32084035 %</b>
100	0,795394%	20,44728435%	0,204606 %	-0,01331565 %
150	0,79745933%	20,44728435%	0,20254067%	0,19321735 %
200	0,795999%	20,44728435%	0,204001 %	0,04718435 %
626	0,79647284 %	20,44728435%	0,20352716%	0,09456835 %
1000	0,7956049 %	20,44728435%	0,2043951%	0,0077435 %

$\bar{e}_B$  = rata-rata kesalahan klasifikasi model bagging regresi logistik biner

$e_B$  = kesalahan klasifikasi model regresi logistik biner

## 5. PENUTUP

### 5.1 Kesimpulan

Berdasarkan hasil analisis dan pembahasan, diperoleh kesimpulan sebagai berikut:

1. Hasil analisis regresi logistik menunjukkan bahwa pola hubungan kesejahteraan rumah tangga di Kabupaten Pati dengan faktor-faktor yang mempengaruhinya digambarkan dalam model regresi logistik berikut:

$$\hat{\pi}(x_i) = \frac{e^{4,024 - 0,832 x_{1,1} - 0,831 x_7 + 1,168 x_{11,1}}}{1 + e^{4,024 - 0,832 x_{1,1} - 0,831 x_7 + 1,168 x_{11,1}}}$$

2. Pada model regresi logistik biner terdapat tiga variabel prediktor yang berpengaruh signifikan terhadap variabel respon, ketiga variabel tersebut adalah jenis kelamin kepala rumah tangga ( $X_1$ ), jumlah anggota rumah tangga ( $X_7$ ), dan ada atau tidaknya anggota keluarga yang mengusai telepon seluler/HP ( $X_{11}$ ).
3. Dengan menggunakan metode regresi logistik biner diperoleh hasil ketepatan klasifikasi sebesar 79,55271565% sedangkan hasil analisis *bagging* regresi logistik biner menunjukkan bahwa pada 90 kali replikasi bootstrap diperoleh nilai ketepatan klasifikasi terbesar yaitu 79,873556%.
4. Penurunan persentase tingkat kesalahan model *bagging* dengan 90 kali replikasi sebesar 0,32084035%, merupakan penurunan persentase tingkat kesalahan terbesar.
5. Model yang digunakan adalah model *bagging* regresi logistik biner dengan persamaan logit  $g(x) = -4,80929 - 0,86966 X_{1,1} + 0,854362 X_7 + 1,209059 X_{12,1}$

## DAFTAR PUSTAKA

- Agresti, A. 1990. *Categorical Data Analysis*. John Wiley and Sons. New York.
- Arianti, N. N., Nabiu, M., dan Yuliarso, M. Z. 2008. *Analisis Faktor-Faktor Yang Mempengaruhi Tingkat Kemiskinan Regional di Kota Bengkulu (Suatu Tinjauan Terhadap Perubahan Struktur Perekonomian Dari Sektor Pertanian ke Sektor Industri dan Jasa) Vol. No.1*. Bengkulu : Fakultas Pertanian Universitas Bengkulu.
- Badan Pusat Statistik (BPS). 2014. *Jawa Tengah dalam Angka 2014*. Semarang : BPS.
- Bewick, V., Cheek, L., and Ball, J. 2005. *Statistics review 14: Logistic regression. Critical Care (London, England), Vol. 9*.
- Breiman, L. 1994. *Bagging Predictor*. Technical report No. 421. Departement of Statistics University of California.

- Cahyat, A., Gonner, C., dan Haug, M. 2007. *Mengkaji Kemiskinan dan Kesejahteraan Rumah Tangga : Sebuah Panduan dengan Contoh dari Kutai Barat, Indonesia*. Bogor: CIFOR.
- Faturochman, dan Molo, M. 1995. *Kemiskinan dan Kependudukan di Pedesaan Jawa : Analisis data SUSENAS 1992*. Yogyakarta : Pusat Penelitian Kependudukan Universitas Gajah Mada.
- Hosmer, D. W. and Lemeshow. 2000. *Applied Logistic Regression*. USA : John Wiley and Sons.
- Pemerintah Provinsi Jawa Tengah. 2015. *Dokumen Strategi Penanggulangan Kemiskinan Daerah Tahun 2015-2018*. Semarang : Pemerintah Provinsi Jawa Tengah.
- Pratama. 2011. *Klasifikasi Kesejahteraan Rumah Tangga di Jawa Timur dengan Pendekatan Multivariate Adaptive Regression Spline-Bootstrap Aggregating (MARS BAGGING)*. Surabaya : Jurusan Statistika FMIPA-ITS.
- Rencher, A.C. 2002. *Methods of Multivariate Analysis*. USA : John Wiley and Sons.
- Rusastra, I.W., dan Togar, A.N. 2007. *Karakteristik Wilayah dan Keluarga Miskin di Perdesaan : Basis Perumusan dan Intervensi Kebijakan*. Bogor : Pusat Analisis Sosial Ekonomi Pertanian.
- Sugiarto, E. 2007. *Teori Kesejahteraan Sosial Ekonomi dan Pengukurannya*. Jurnal Eksekutif, Vol. 4 No. 2, Agustus 2007, hal. 263-269
- The World Bank*. 2006. *Making the New Indonesia Work for the Poor*. USA