

## **BAGGING REGRESI LOGISTIK ORDINAL PADA STATUS GIZI BALITA**

**Muhammad Sjahid Akbar<sup>1</sup>, Adatul Mukarromah<sup>2</sup>, Lalita Paramita<sup>3</sup>**

<sup>1,2</sup> Staf Pengajar Jurusan Statistika ITS Surabaya

<sup>3</sup> Alumni Jurusan Statistika ITS Surabaya

### **Abstract**

World Health Organization-National Centre for Health Statistic (WHO-NCHS) is standart nutritional status used in Indonesia, it based on Kartu Menuju Sehat (KMS). These Indices can be expressed in terms of Z-score based Weight-for-Age. This Indices need comparison considering the fact which cause nutritional status not only Weight-for-Age. The aim from this research to obtain bagging ordinal logistics regression for WHO-NCHS nutritional status and new nutritional status. A new nutritional status expressed in terms of cluster, while classification function expressed from logit model of ordinal logistics regression. The result for new nutritional status bagging obtained at 60 bootstrap replicated that is 76.345%, this model can decrease misclassification until 22.046%. While bagging for WHO-NCHS nutritional status can increase accurate classification from single data set 75.863% at 150 bootstrap replicated.

**Keywords:** Child nutritional status, Bagging, Ordinal logistics regression.

### **1. Pendahuluan**

Masalah gizi dapat terjadi pada seluruh kelompok usia, bahkan masalah gizi suatu kelompok umur tertentu akan mempengaruhi status gizi pada periode siklus kehidupan berikutnya (*intergenerational impact*). Kabupaten Nganjuk menjadi daerah dengan tingkat kurang gizi tinggi, yaitu sedikitnya 5022 bayi di bawah usia lima tahun mengalami kekurangan gizi, bahkan 381 di antaranya dinyatakan dalam kondisi gizi buruk<sup>[7]</sup>. Pada tahun 2005 LPPM Unair bekerjasama dengan Balitbangda Kabupaten Nganjuk melakukan suatu penelitian mengenai kondisi status gizi balita di Nganjuk untuk mengetahui jumlah penderita kekurangan energi protein (KEP). Penelitian tentang klasifikasi status gizi balita telah dilakukan dengan menggunakan metode Diskriminan. Variabel-variabel yang digunakan dalam penelitian tersebut adalah berat badan saat lahir, tinggi badan saat lahir, tinggi badan sekarang dan pendapatan rata-rata keluarga. Hasil penelitian tersebut menghasilkan ketepatan klasifikasi sebesar 42.8%<sup>[2]</sup>, sehingga diperlukan penelitian lebih lanjut untuk memperoleh ketepatan klasifikasi lebih tinggi atau meminimumkan kesalahan klasifikasi. Penelitian serupa dilakukan dengan tujuan untuk mengetahui faktor-faktor yang mempengaruhi status gizi balita dengan menggunakan metode regresi logistik ordinal<sup>[14]</sup>. Penelitian status gizi balita lainnya, dilakukan dengan mempertimbangkan faktor eksternal seperti pendapatan, pengeluaran keluarga, pengeluaran untuk pangan, jumlah anggota keluarga<sup>[13]</sup>. Metode yang digunakan adalah analisis *cluster*. Analisis selanjutnya bertujuan membandingkan antara model klasifikasi status gizi balita *cluster* dengan standart WHO-NCHS. Kedua model klasifikasi dibentuk dengan metode regresi logistik ordinal. Metode pengklasifikasian dengan menggunakan metode regresi logistik ordinal akan memberikan pendugaan parameter yang tidak stabil, artinya jika terdapat perubahan pada data menyebabkan perubahan yang signifikan pada model<sup>[4]</sup>. Sehingga untuk memperoleh parameter yang stabil pada model regresi logistik ordinal digunakan pendekatan *bootstrap*. Metode *bootstrap* yang digunakan adalah *bootstrap aggregating (bagging)*. Berdasarkan latar belakang tersebut timbul permasalahan yaitu bagaimana *bagging* regresi logistik ordinal pada klasifikasi status gizi balita *cluster* dan WHO-NCHS serta bagaimana perbandingan hasil ketepatan klasifikasi antara keduanya.

## 2. Bahan dan Metode

### 2.1 Analisis Kelompok

Analisis Kelompok atau *Cluster Analysis* adalah analisis statistika yang bertujuan mengelompokkan data sedemikian hingga data berada dalam kelompok yang mempunyai sifat yang relatif homogen <sup>[10]</sup>. Ada beberapa ukuran jarak yang digunakan dalam analisis kelompok. Jarak *euclidean* merupakan metode pengukuran jarak yang paling sering digunakan jika skala data adalah interval. Jika skala data adalah nominal maka metode pengukuran jarak yang digunakan adalah jarak *Chi-square*

$$d_{uv} = \text{Chisquare}(u, v) = \sqrt{\sum_{i=1}^p \frac{(u_i - E(u_i))^2}{E(u_i)} + \sum_{i=1}^p \frac{(v_i - E(v_i))^2}{E(v_i)}}, \quad (1)$$

### 2.2 Regresi logistik ordinal

Regresi logistik ordinal adalah suatu analisis regresi yang digunakan untuk menggambarkan hubungan antara variabel respon dengan sekumpulan variabel prediktor, dimana variabel respon bersifat ordinal, yaitu mempunyai lebih dari 2 kategori dan setiap kategori dapat diperingkat <sup>[8]</sup>.

Model yang dipakai untuk regresi logistik ordinal adalah model logit. Model logit tersebut adalah *cumulative logit models*. Peluang kumulatif,  $P(Y \leq j | x_i)$  didefinisikan sebagai berikut :

$$P(Y \leq j | x_i) = \frac{\exp\left(\alpha_j + \sum_{k=1}^p \beta_k x_{ik}\right)}{1 + \exp\left(\alpha_j + \sum_{k=1}^p \beta_k x_{ik}\right)} \quad (2)$$

dengan  $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ip})$  merupakan nilai pengamatan ke- $i$  ( $i = 1, 2, \dots, n$ ) dari setiap  $p$  variabel prediktor <sup>[6]</sup>.

Fungsi klasifikasi yang terbentuk bila terdapat  $j$  kategori respon adalah sejumlah  $j - 1$ . Fungsi pembeda dalam proses pengklasifikasian adalah *cumulative logit models*. Jika  $\phi_j(x_i) = P(Y = j | x_i)$  menyatakan peluang kategori respon ke- $j$  pada  $p$  variabel prediktor yang dinyatakan dalam vektor  $\mathbf{x}_i$  dan  $P(Y \leq j | x_i)$  menyatakan peluang kumulatif pada  $p$  variabel prediktor yang dinyatakan dalam vektor  $\mathbf{x}_i$  maka nilai  $\phi_j(x_i)$  diperoleh dengan persamaan berikut.

$$P(Y \leq j | x_i) = \phi_0(x_i) + \phi_1(x_i) + \dots + \phi_j(x_i).$$

Jika terdapat empat kategori respon dimana  $j = 1, 2, 3, 4$  maka nilai dari peluang kategori respon ke- $j$  diperoleh dengan

$$\phi_0(x_i) = P(Y \leq 0 | x_i) \quad (3)$$

$$\phi_1(x_i) = P(Y \leq 1 | x_i) - \phi_0(x_i) \quad (4)$$

$$\phi_2(x_i) = P(Y \leq 2 | x_i) - P(Y \leq 1 | x_i) \quad (5)$$

$$\phi_3(x_i) = 1 - P(Y \leq 2 | x_i) \quad (6)$$

Nilai  $\phi_j(x_i)$  pada persamaan (3), (4), (5), (6) akan dijadikan pedoman pengklasifikasian. Suatu pengamatan masuk dalam respon kategori- $j$  berdasarkan nilai  $\phi_j(x_i)$  yang terbesar. Penaksiran parameter model regresi logistik ordinal menggunakan metode *Maximum Likelihood Estimator*. Bentuk umum dari fungsi likelihood untuk sampel dengan  $n$  pengamatan bebas  $(y_i, x_i)$ ,  $i = 1, 2, \dots, n$  adalah

$$l(\beta) = \prod_{i=1}^n [\phi_0(x_i)^{y_{0i}} \phi_1(x_i)^{y_{1i}} \phi_2(x_i)^{y_{2i}} \phi_3(x_i)^{y_{3i}}], \quad (7)$$

Sehingga didapatkan fungsi  $\ln$ -likelihood sebagai berikut.

$$L(\beta) = \sum_{i=1}^n y_{0i} \ln[\phi_0(x_i)] + y_{1i} \ln[\phi_1(x_i)] + y_{2i} \ln[\phi_2(x_i)] + y_{3i} \ln[\phi_3(x_i)]. \quad (8)$$

Maksimum  $\ln$ -likelihood dapat diperoleh dengan cara mendifferensialkan  $L(\beta)$  terhadap  $\beta$  dan menyamakannya dengan nol<sup>[1]</sup>. Nilai  $\beta$  ditaksir dengan metode Newton Raphson karena persamaannya bersifat nonlinier.

Pengujian signifikansi koefisien  $\beta$  terhadap variabel respon dilakukan secara parsial dan serentak. Pemeriksaan signifikansi koefisien  $\beta$  secara parsial dengan menggunakan statistik uji Wald. Sedangkan pemeriksaan signifikansi koefisien  $\beta$  secara serentak dengan menggunakan statistik uji G.

### 2.3 Bagging (Bootstrap Aggregating)

*Bagging* prediktor adalah metode untuk membangkitkan prediktor dalam beberapa versi dan menggunakannya untuk *aggregate* prediktor. Himpunan data  $\mathcal{L}$  terdiri dari  $\{(y_n, x_n), n = 1, \dots, N\}$  dengan  $y$  dapat berupa kelas label atau numerik respon. Jika input adalah  $x$  maka  $y$  diprediksi dengan  $\varphi(x, \mathcal{L})$  dimana  $\varphi(x, \mathcal{L})$  adalah prediktor. Prediktor diperoleh dengan melakukan replikasi *bootstrap* yang kemudian disebut  $\{\varphi(x, \mathcal{L}_k)\}$ . Replikasi *bootstrap* dilakukan sebanyak  $B$  kali sehingga  $\{\mathcal{L}^{(B)}\}$  dari  $\mathcal{L}$  dan dibentuk prediktor  $\{\varphi(x, \mathcal{L}^{(B)})\}$ .  $\{\mathcal{L}^{(B)}\}$  adalah resampling dengan pengembalian<sup>[4]</sup>.

### 2.4 Status Gizi Balita

Status gizi adalah ekspresi dari keadaan keseimbangan dalam bentuk variabel tertentu untuk dapat dikatakan bahwa status gizi merupakan indikator baik buruknya penyediaan makanan sehari-hari. Status gizi yang baik diperlukan untuk mempertahankan derajat kebugaran dan kesehatan, membantu pertumbuhan bagi anak<sup>[9]</sup>. Faktor-faktor yang Mempengaruhi Status Gizi Balita terbagi menjadi dua<sup>[13]</sup>, yaitu.

1. Faktor gizi internal, merupakan faktor yang menjadi dasar pemeriksaan tingkat kebutuhan gizi seseorang, meliputi status kesehatan, umur dan jenis kelamin.
2. Faktor gizi eksternal, faktor yang berpengaruh di luar diri seseorang yang meliputi pendapatan, pendidikan, jumlah anggota keluarga, dan pengetahuan gizi.

Sumber lain menyebutkan selain faktor-faktor di atas yang mempengaruhi status gizi balita, yaitu:

1. Status pemberian Air Susu Ibu (ASI).
2. Berat Badan Lahir Rendah (BBLR).
3. Konsumsi makanan.
4. Faktor perilaku.

### 2.5 Penentuan Status Gizi dengan Cara *Zscore* BB/U Standart Baku Antropometri WHO-NCHS.

Perhitungan nilai *Zscore* untuk status gizi dirumuskan sebagai berikut.

1. Bila “nilai riil” hasil pengukuran BB/U lebih besar atau sama dengan nilai median, maka.

$$Zscore = \frac{\text{nilai riil} - \text{nilai median}}{SD \text{ upper}}$$

2. Bila “nilai riil” hasil pengukuran BB/U lebih kecil dari nilai median, maka.

$$Zscore = \frac{\text{nilai riil} - \text{nilai median}}{SD \text{ lower}}$$

Nilai median disesuaikan dengan umur, diperoleh dari Tabel baku antropometri WHO NCHS. Penilaian status gizi diberikan dengan ketentuan berikut.

**Tabel 1.** Penilaian Status Gizi Balita

STATUS GIZI	<i>Zscore</i>
Status gizi lebih	$Zscore > 2$
Status gizi normal (baik)	$-2 < Zscore \leq 2$
Status gizi sedang (kurang)	$-3 < Zscore \leq -2$
Status gizi buruk	$Zscore \leq -3$

Sumber : Depkes RI dalam Soegianto, 2000

### 2.6 Variabel penelitian

Sumber data yang digunakan pada penelitian ini adalah data hasil survei Kurang Energi Protein (KEP) tahun 2005 di Kabupaten Nganjuk yang merupakan hasil kerjasama LPPM Unair dengan Balitbangda Kabupaten Nganjuk. Sedangkan variabel-variabel yang digunakan dalam penelitian ini meliputi.

1. Variabel respon yaitu Status gizi balita yang berskala data ordinal dengan deskripsi sebagai berikut:
 

Gizi buruk, dengan kode 0	Gizi lebih, dengan kode 2
Gizi sedang (kurang), dengan kode 1	Gizi normal (baik), dengan kode 3
2. Variabel prediktor yaitu faktor- faktor yang diduga mempengaruhi status gizi balita

**Tabel 2.** Faktor- faktor yang diduga mempengaruhi status gizi balita

Variabel penelitian	Keterangan	Kriteria obyektif & Satuan Variabel	Skala Data
Faktor internal	Umur (X1)	Bulan	Interval
	Jenis kelamin (X2)	1. Laki-laki 2. Perempuan	Nominal
	Berat Badan Lahir (BBL) (X3)	1. $\leq 2,5$ kg 2. $2,6 - 3,5$ kg 3. $\geq 3,6$ kg	Ordinal
	Berat Badan sekarang (X4)	Kg	Interval
	Tinggi Badan Lahir (X5)	Cm	Interval
	Tinggi Badan sekarang (X6)	Cm	Interval

**Lanjutan Tabel 3.** Faktor- faktor yang diduga mempengaruhi status gizi balita

Faktor eksternal	Pendapatan (X7)	1. $\leq$ Rp.362.000 2. Rp. 362001 - Rp. 529510 3. $\geq$ Rp.529511	Ordinal
	Pengeluaran untuk pangan (Presentase terhadap pendapatan) (X8)	1. $\leq$ 30% 2. 31%-60% 3. $\geq$ 61%	Ordinal
	Jumlah anggota keluarga (X9)	1. $<$ 4 orang 2. $\geq$ 4 orang.	Ordinal
	Pendidikan (X10)	1.SD 2. SMP 3. $\geq$ SMA	Ordinal
Status ASI	Frekuensi pemberian ASI (X11)	1. Tidak 2. 1-5 kali perhari 3. $\geq$ 6 kali perhari	Ordinal
Perilaku kebersihan	Peralatan makan dicuci (X12)	1. Tidak 2. Ya	Nominal
	Pemakaian sabun (X13)	1. Tidak 2. Ya	Nominal

Langkah-langkah analisis dalam penelitian ini dibagi menjadi 3 bagian.

1. Klasifikasi dengan penilaian status gizi balita hasil pengelompokan.
2. Klasifikasi dengan penilaian status gizi balita BB/U WHO-NCHS.
3. Membandingkan ketepatan klasifikasi antara penilaian status gizi balita BB/U WHO-NCHS dan hasil pengelompokan analisis *cluster*.

Sedangkan algoritma *bagging* untuk regresi logistik ordinal adalah sebagai berikut.

1. Mengambil sampel *bootstrap* sebanyak  $n$  dari *data set*  $\mathcal{L}$  dengan pengulangan sebanyak  $n$ . Pengambilan sampel sedemikian hingga setiap variabel *aggregate* dalam setiap observasi.
2. Memodelkan regresi logistik ordinal hasil sampel *bootstrap*  $\mathcal{L}_B$ .
3. Menghitung peluang kumulatif, peluang masing-masing kategori respon untuk setiap observasi dan menghitung ketepatan klasifikasi. Kesalahan klasifikasi pada langkah ini disebut  $e_B$
4. Mengulang langkah 1-4 sebanyak B kali (Replikasi *bootstrap*).
5. Memperoleh ketepatan klasifikasi *bagging* dari rata-rata ketepatan klasifikasi setiap pengulangan sampai B. Sehingga kesalahan klasifikasi *bagging* untuk replikasi B kali adalah  $\bar{e}_B$
6. Membentuk model *bagging* regresi logistik ordinal dari rata-rata setiap parameter pada peng-ulangan sampai B.
7. Untuk memperoleh hasil yang lebih baik maka replikasi *bootstrap* dilakukan sebanyak mungkin, replikasi *bootstrap* yang biasa digunakan adalah 50 sampai 200 <sup>[5]</sup>.

### 3. ANALISIS DAN PEMBAHASAN

#### 3.1 Klasifikasi status gizi balita *cluster*

Faktor-faktor eksternal yang diduga mempengaruhi status gizi balita antara lain pendapatan, pengeluaran keluarga, pengeluaran untuk pangan, jumlah anggota keluarga.

Status gizi balita *cluster* dibentuk dari hasil pengelompokan balita berdasarkan faktor-faktor eksternal. Penentuan status gizi dilihat dari karakteristik pada masing-masing kelompok sebagai berikut.

Kelompok 1 terdiri dari 40 balita yang mempunyai kedekatan yang erat. Pendapatan keluarga pada kelompok ini 47,5% berpenghasilan Rp.362.001-Rp.529.510, sedangkan dilihat dari proporsi pengeluaran untuk pangan kelompok 1 dominan pada kategori 3 yaitu mengeluarkan lebih besar 60% dari total pendapatannya untuk pangan, sedangkan pendidikan ibu pada kelompok ini 100% adalah berpendidikan SD. Berat badan saat lahir (BBL) juga menjadi penciri pada penentuan status gizi kelompok ini, rata-rata BBL pada kelompok ini adalah 3,2.

**Tabel 3.** Karakteristik Kelompok Hasil *Cluster*

F	Pendapatan		Proporsi Pengeluaran untuk pangan		Pendidikan Terakhir Ibu		Rata-rata BBL
			Kategori	( % )			
40	<= Rp.362000	(35%)	< 30%	( 0%)	SD	(100%)	3,2
	Rp.362001 - Rp.529510	(47,5%)	30%-60%	(20%)			
	> Rp.529510	(17,5%)	> 60%	(80%)			
62	<= Rp.362000	(16,1%)	< 30%	(4,8%)	SMP	(46,8%)	3,1
	Rp.362001- Rp.529510	(37,1%)	30%-60%	(40,3%)	SMA	(45,2%)	
	> Rp.529510	(46,8%)	> 60%	(54,8%)	Diploma	(3,2%)	
					S1/S2/S3	(4,8%)	
14	<= Rp.362000	(57,1%)	< 30%	(0%)	SD	(100%)	2,6
	Rp.362001- Rp.529510	(14,3%)	30%-60%	(21,4%)			
	> Rp.529510	(28,6%)	> 60%	(78,6%)			
29	<= Rp.362000	(69%)	< 30%	( 0%)	SD	(55,2%)	3,0
	Rp.362001- Rp.529510	(13,8%)	30%-60%	(13,8%)	SMP	(24,1%)	
	> Rp.529510	(17,2%)	> 60%	(86,2%)	SMA	(20,7%)	

Kelompok 2 terdiri dari 62 balita, 46,8% berpendapatan >Rp.529.510. Selain itu, 54,8% kelompok ini mengeluarkan pendapatan >60% untuk kebutuhan konsumsi. Jika Dilihat dari pendidikan terakhir ibu, kelompok 2 didominasi oleh SMA (45.2%) dan SMP (46,8%). Sedangkan rata-rata BBL balita pada kelompok dua adalah 3,1.

Kelompok 3 mempunyai anggota sebanyak 14 balita, sangat berbeda dengan kelompok 1 dan 2, pada kelompok 3 sebesar 57,15% anggotanya berpendapatan keluarga dibawah UMR, sedangkan jika dilihat dari proporsi pengeluaran untuk pangan sebanyak 78,6% kelompok ini pengeluaran untuk panganya >60%. Pada kelompok 3 didominasi

balita yang pendidikan ibunya adalah SD sebesar 57,1 dan rata-rata BBL balita pada kelompok ini adalah 2,6.

Pada kelompok 4 terdiri dari 29 balita. Deskripsi pada kelompok ini hampir sama dengan kelompok 3, pendapatan keluarga pada kelompok 4 sebesar 69% berada dibawah UMR, proporsi pengeluaran untuk pangan >60% sebanyak 86,2% dari anggota kelompok dan pendidikan terakhir ibu adalah SD sebesar 55.2%, Yang menjadi pembeda paling mencolok adalah rata-rata BBL pada kelompok ini yaitu 3,0 Kg.

Hasil deskripsi pada Tabel 3 menunjukkan kelompok 1 dan 2 cenderung berstatus gizi baik atau lebih. Jika dilihat dari rata-rata BBL nya yang besar dari kelompok 2, proporsi pengeluaran untuk pangan didominasi >60% dan pendidikan terakhir ibu yang 100% adalah SD, kelompok 1 lebih cenderung berstatus gizi lebih. Sehingga untuk selanjutnya, pada penelitian ini kelompok 1 ditentukan sebagai kelompok balita berstatus gizi lebih dan kelompok 2 kelompok balita berstatus gizi baik. Sedangkan untuk kelompok 3 adalah kelompok balita dengan status gizi buruk ditandai dengan pendapatan keluarga yang didominasi dibawah UMR dan didukung oleh BBL balita yaitu 2,64 kg, yang menurut ilmu kesehatan akan cenderung berstatus gizi buruk. Sehingga kelompok 3 pada penelitian ini adalah kelompok balita berstatus gizi buruk dan kelompok 4 adalah kelompok balita berstatus gizi kurang.

Hasil regresi logistik ordinal diperoleh variabel yang secara individu signifikan terhadap status gizi balita *cluster*. Pengujian dilakukan dengan membandingkan nilai  $W^2$  dengan distribusi normal standart pada tingkat kesalahan  $\alpha$ ,  $H_0$  ditolak bila  $|W| > Z_{\alpha/2}$ . Diperoleh variabel yang signifikan berpengaruh antara lain Usia ( $X_1$ ), BBL( $X_3$ ), berat badan sekarang( $X_4$ ), pendapatan( $X_7$ ), pendidikan terkhir ibu ( $X_{10}$ ), Frekuensi pemberian ASI( $X_{11}$ ), keterlibatan ayah( $X_{14}$ ). Variabel ini kemudian masuk ke dalam model *multiple* regresi logistik ordinal.

**Tabel 4.** Multiple Regresi Logistik Ordinal (Status gizi *cluster*)

Variabel		Coef	Wald	Odds	P-value
Constant (1)		3,45671	3,19		0,001*
Constant (2)		6,53178	5,34		0,000*
Constant (3)		9,45014	6,75		0,000*
Umur	$X_1$	0,199458	6,72	1,22	0,000*
BBL	$X_3$				
2,6 – 3,5 kg		-1,75397	-2,70	0,17	0,007*
≥ 3,6 kg		-2,29054	-2,59	0,10	0,010*
BB sekarang (BB_S)	$X_4$	-0,935991	-6,11	0,39	0,000*
Pendapatan					
Rp.362001-Rp.529510	$X_7$	-1,54162	-3,04	0,21	0,002*
≥ Rp.529511		-0,818740	-1,53	0,44	0,125
Pendidikan					
SMP	$X_{10}$	-2,50658	-4,66	0,08	0,000*
≥ SMA		-2,93448	-4,94	0,05	0,000*
Frekuensi ASI	$X_{11}$				
1-5 kali		-3,4893	-2,10	0,03	0,036*
≥ 6 kali		0,495809	0,88	1,64	0,376
Keterlibatan Ayah	$X_{14}$				
Cukup		2,45754	2,29	11,69	0,022*
Banyak		-0,0100782	-0,02	0,99	0,982

\*signifikan pada  $\alpha = 0.05$

Pada pengujian serentak diperoleh kesimpulan bahwa bahwa terdapat satu atau lebih variabel bebas yang berpengaruh secara signifikan terhadap status gizi balita, yaitu ditunjukkan dari nilai  $G$  sebesar 168.095 dan signifikansi sebesar 0,000. Pada pengujian parsial model *multiple* regresi logistik ordinal diperoleh ketujuh variabel signifikan berpengaruh. Hal ini ditunjukkan dari nilai uji  $|W|$  yang lebih besar dari nilai  $z_{\alpha/2}$  (1,96) atau dari nilai signifikansi yang kurang dari  $\alpha = 0,05$ . Sehingga diperoleh model logit sebagai berikut.

$$\text{Logit 1: } P(Y \leq 0 | x_i) = 3,45671 + 0,199458(\text{umur}) - 1,75397(\text{BBL\_2500-3500gr}) - 2,29054(\text{BBL\_} > 3600\text{gr}) - 0,935991(\text{BB\_S}) - 1,54162(\text{pendapatan\_Rp,362001-529510}) - 0,81874(\text{pendapatan\_} > \text{Rp.529510}) - 2,50658(\text{pendidikan\_SMP}) - 2,93448(\text{pendidikan\_SMA}) - 3,4893(\text{ASI\_1-5kali}) + 2,45754(\text{ketelibatn ayah\_cukup}) \quad (9)$$

$$\text{Logit 2: } P(Y \leq 1 | x_i) = 6,53178 + 0,199458(\text{umur}) - 1,75397(\text{BBL\_2500-3500gr}) - 2,29054(\text{BBL\_} > 3600\text{gr}) - 0,935991(\text{BB\_S}) - 1,54162(\text{pendapatan\_Rp.362001-529510}) - 0,81874(\text{pendapatan\_} > \text{Rp.529510}) - 2,50658(\text{pendidikan\_SMP}) - 2,93448(\text{pendidikan\_SMA}) - 3,4893(\text{asi\_1-5kali}) + 2,45754(\text{ketelibatn ayah\_cukup}) \quad (10)$$

$$\text{Logit 3: } P(Y \leq 1 | x_i) = 9,45014 + 0,199458(\text{umur}) - 1,75397(\text{BBL\_2500-3500gr}) - 2,29054(\text{BBL\_} > 3600\text{gr}) - 0,935991(\text{BB\_S}) - 1,54162(\text{pendapatan\_Rp.362001-529510}) - 0,81874(\text{pendapatan\_} > \text{Rp.529510}) - 2,50658(\text{pendidikan\_SMP}) - 2,93448(\text{pendidikan\_SMA}) - 3,4893(\text{ASI\_1-5kali}) + 2,45754(\text{ketelibatn ayah\_cukup}) \quad (1)$$

Pengklasifikasian balita dilakukan berdasarkan fungsi klasifikasi yang diperoleh pada persamaan (9), (10), dan (11). Sehingga diperoleh ketepatan klasifikasi seperti pada Tabel 5.

**Tabel 5.** Ketepatan klasifikasi *data set* tunggal status gizi balita *cluster*

	Aktual				Total
	0	1	2	3	
Prediksi					
0	5	2	0	0	7
1	8	15	8	0	31
2	1	10	25	6	42
3	0	2	7	56	65
Total	14	29	40	62	145

Sehingga dapat dihitung besarnya ketepatan klasifikasi

$$\text{APER} = \left( \frac{8}{145} + \frac{1}{145} + \frac{2}{145} + \frac{10}{145} + \frac{2}{145} + \frac{8}{145} + \frac{7}{145} + \frac{6}{145} \right) \times 100\% = 30,3\%$$

Sehingga ketepatan klasifikasinya adalah 69.7%. Fungsi klasifikasi (9), (10), (11) merupakan model *data set* tunggal. Parameter yang dihasilkan dari model regresi cenderung tidak stabil<sup>[4]</sup>. Untuk melihat kestabilan parameter dan memperoleh keakuratan model yang lebih baik maka dilakukan resampling pada data. Resampling pada pengamatan terkelompok yang melibatkan beberapa variabel prediktor dilakukan dengan metode *bagging* yang merupakan pengambilan sampel dengan pengembalian untuk *data set* yang terdiri dari respon ( $y$ ) dan variabel prediktor ( $x$ ).



Ketujuh variabel yang masuk dalam model *multiple* regresi logistik akan diperlakukan resampling *bagging*. Sampel bootstrap diambil sebanyak  $n$  data yaitu 145 data, kemudian direplikasi bootstrap sebanyak 50, 60, 70, 80, 90, 100, 150 dan 200. Pada setiap pengambilan sampel akan dibentuk model *multiple* regresi logistik ordinal sehingga akan diperoleh nilai ketepatan klasifikasi sebanyak  $B$  dalam setiap  $B$  replikasi bootstrap.

#### Estimasi *Bagging Class Probability*

Perhitungan ketepatan klasifikasi dilakukan pada setiap iterasi Hasil perhitungan ketepatan pada kemudian dirata-rata sehingga menghasilkan ketepatan klasifikasi *bagging* regresi logistik ordinal. Kesalahan klasifikasi  $e_B$  pada tiap iterasi dihitung dengan perhitungan  $(1 - \text{ketepatan klasifikasi})$ . Keberhasilan *bagging* diukur dari seberapa besar *bagging* dapat menurunkan kesalahan klasifikasi dari model *data set* tunggal. Tabel 6 merupakan hasil dari *bagging* dengan 50, 60, 70, 80, 90, 100, 150 dan 200 replikasi bootstrap.

**Tabel 6.** Hasil *bagging* regresi logistik ordinal status gizi *cluster*

Replikasi Bootstrap	Rata-rata ketepatan klasifikasi	$\bar{e}_B$	$e_s$	Penurunan kesalahan klasifikasi	Varians ketepatan klasifikasi
50 kali	75,7%	24,3%	30,3%	19,8%	0,0016
60 kali	<b>76,3%</b>	<b>23,7%</b>	30,3%	<b>22,0%</b>	0,0018
70 kali	75,9%	24,1%	30,3%	20,6%	0,0019
80 kali	75,7%	24,3%	30,3%	19,8%	0,0022
90 kali	75,8%	24,2%	30,3%	20,2%	0,0022
100 kali	75,5%	24,5%	30,3%	19,4%	0,0028
150 kali	76,0%	24,0%	30,3%	20,8%	0,0018
200 kali	75,3%	24,7%	30,3%	18,4%	0,0023

Tabel 6 memberikan informasi bahwa dengan 60 replikasi *bootstrap* diperoleh rata-rata ketepatan klasifikasi terbesar yaitu sebesar 76.3%, sehingga berdasarkan hasil diatas maka dapat disimpulkan bahwa diperoleh *bagging* prediktor terbaik adalah pada replikasi bootstrap sebanyak 60 kali. Model *bagging* ini dapat meningkatkan ketepatan klasifikasi dari model *data set* tunggal yaitu sebesar 69.7% menjadi 76.3% atau dengan kata lain *bagging* dapat menurunkan kesalahan klasifikasi sebesar 22% dari model *data set* tunggal. Varians ketepatan klasifikasi yang kecil yaitu mendekati 0 menunjukkan bahwa ketepatan klasifikasi pada setiap pengambilan sampel untuk  $B$  replikasi bootstrap stabil.

### 3.2 Klasifikasi status gizi balita WHO-NCHS

Standar penilaian status gizi di Indonesia adalah standart WHO-NCHS, sehingga pada penelitian ini akan dilakukan klasifikasi status gizi balita dengan respon yang digunakan adalah status gizi balita standart WHO-NCHS dengan indikator pengukuran berat badan berdasarkan umur.

Pada model *multiple* regresi logistik ordinal diperoleh variabel yang secara signifikan terhadap status gizi balita WHO-NCHS adalah Usia ( $X_1$ ), BBLR( $X_3$ ), berat badan sekarang( $X_4$ ), pendapatan( $X_7$ ), dan keterlibatan ayah ( $X_{14}$ ).

**Tabel 7. Multiple Regresi Logistik Ordinal (WHO-NCHS)**

Variabel		Coef	Wald	Odds	P-value
Constant (1)		2,77611	2,95		0,003*
Constant (2)		5,31818	5,14		0,000*
Constant (3)		6,17282	5,76		0,000*
Umur	X <sub>1</sub>	0,126171	5,62	1,14	0,000*
BBL	X <sub>3</sub>				
2,6 – 3,5 kg		-1,6783	-2,78	0,18	0,005*
≥ 3,6 kg		-2,2673	-2,61	0,15	0,009*
BB sekarang	X <sub>4</sub>	-0,746637	-5,43	0,47	0,000*
Pendapatan	X <sub>7</sub>				
Rp. 362001 - Rp529510		-1,710210	-3,31	0,17	0,001*
≥ Rp.529511		-0,330332	-0,71	0,21	0,479
Keterlibatan	X <sub>14</sub>				
Cukup		2,565310	2,71	25,69	0,007*
Banyak		0,402915	0,98	1,56	0,325

Pada uji serentak diperoleh  $G$  sebesar 92.330 dan signifikansi sebesar 0,000. Hal ini mengindikasikan bahwa dengan  $\alpha = 0,05$  dapat menolak  $H_0$ , artinya ada satu atau lebih variabel bebas yang berpengaruh secara signifikan terhadap status gizi balita. Model logitnya adalah sebagai berikut.

$$\text{Logit1: } P(Y \leq 0 | x_i) = 2,77611 + 0,126171 (\text{umur}) - 1,6783 (\text{BBL2}) - 2,2673 (\text{BBL3}) - 0,746637 (\text{BB\_S}) - 1,7102 (\text{pendaptn2}) + 2,56531 (\text{keterlibatn ayah2}) \quad (2)$$

$$\text{Logit2: } P(Y \leq 1 | x_i) = 5,31818 + 0,126171 (\text{umur}) - 1,6783 (\text{BBL2}) - 2,2673 (\text{BBL3}) - 0,746637 (\text{BB\_S}) - 1,7102 (\text{pendaptn2}) + 2,56531 (\text{keterlibatn ayah2}) \quad (3)$$

$$\text{Logit3: } P(Y \leq 2 | x_i) = 6,17282 + 0,126171 (\text{umur}) - 1,6783 (\text{BBL2}) - 2,2673 (\text{BBL3}) - 0,746637 (\text{BB\_S}) - 1,7102 (\text{pendaptn2}) + 2,56531 (\text{keterlibatn ayah2}) \quad (4)$$

Hasil klasifikasi status gizi balita WHO-NCHS seperti pada Tabel 8.

**Tabel 8. Hasil klasifikasi Status gizi WHO-NCHS**

Prediksi	Aktual				Total
	0	1	2	3	
0	7	1	0	0	8
1	7	20	0	5	32
2	0	0	0	0	0
3	0	7	15	83	105
Total	14	28	15	88	145

Diperoleh ketepatan klasifikasi 75.862% dan kesalahan klasifikasi 24.138%.

Selanjutnya analisis dilanjutkan dengan *bagging*, prosedur *bagging* sama dengan pada status gizi *cluster* resampling pada *data set* dengan status gizi WHO\_NCHS dilakukan sesuai dengan banyaknya data asli yaitu 145 dan dilakukan replikasi bootstrap 50, 60, 70, 80, 90, 100, 150 dan 200. Dari setiap pengulangan akan dibentuk model *multiple* regresi logistik ordinal sehingga akan diperoleh nilai ketepatan klasifikasi pada setiap pengulangan. Sehingga diperoleh rata-rata ketepatan klasifikasi dari hasil *bagging* sebagai berikut.

**Tabel 9.** Hasil *bagging* regresi logistik ordinal status gizi WHO-NCHS

Replikasi Bootstrap	Rata-rata ketepatan klasifikasi	$\bar{e}_B$	$e_s$	Penurunan kesalahan klasifikasi	Varians ketepatan klasifikasi
50 kali	76,1%	23,9%	24,1%	1,1%	0,0011
60 kali	76,0%	24,0%	24,1%	0,4%	0,0009
70 kali	76,3%	23,7%	24,1%	1,8%	0,0008
80 kali	76,0%	24,0%	24,1%	0,7%	0,0011
90 kali	76,2%	23,8%	24,1%	1,4%	0,0012
100 kali	76,5%	23,5%	24,1%	2,6%	0,0012
150 kali	<b>76,6%</b>	<b>23,4%</b>	24,1%	<b>3,0%</b>	0,0009
200 kali	76,1%	23,9%	24,1%	1,1%	0,0011

Tabel 9 memberikan informasi bahwa *bagging* dengan replikasi 150 kali memberikan ketepatan klasifikasi terbesar yaitu 76.6%, sedangkan *bagging* dengan replikasi bootstrap sebanyak 60 dan 80 kali memberikan ketepatan klasifikasi yang terkecil yaitu 76%. Sehingga model *bagging* yang memberikan prediktor paling baik adalah dengan replikasi bootstrap 150 kali. Kesalahan klasifikasi pada model ini adalah sebesar 23.4% atau dengan kata lain model mampu menurunkan kesalahan klasifikasi pada model *data set* tunggal sebesar 3%. Varians ketepatan klasifikasi yang kecil yaitu mendekati 0 menunjukkan bahwa ketepatan klasifikasi pada setiap pengambilan sampel untuk B replikasi bootstrap sudah stabil.

### 3.3 Perbandingan Hasil Klasifikasi status gizi balita *cluster* dan Status gizi balita WHO-NCHS.

Pada model regresi logistik ordinal pada *data set* tunggal diperoleh 7 faktor yang mempengaruhi status gizi balita *cluster* hasil *cluster* yaitu umur, BBLR, berat badan sekarang, pendapatan, pendidikan terakhir ibu, frekuensi pemberian ASI, dan Keterlibatan ayah dalam mengasuh. Sedangkan faktor-faktor yang mempengaruhi status gizi balita standart WHO-NCHS adalah umur, BBLR, pendapatan, berat badan sekarang, keterlibatan ayah dalam mengasuh.. Model yang diperoleh dengan respon status gizi balita *cluster* pada *data set* tunggal mampu mengklasifikasikan balita dengan ketepatan sebesar 69.655%, setelah dilakukan *bagging* diperoleh parameter stabil dan ketepatan klasifikasi terbesar dengan replikasi bootstrap 70 kali, yaitu 76.345%, Sedangkan dengan respon status gizi WHO-NCHS model menghasilkan ketepatan klasifikasi sebesar 75.862%, setelah dilakukan *bagging* diperoleh ketepatan klasifikasi terbaik pada replikasi bootstrap 150 kali yaitu sebesar 76.584%. Dilihat dari kenaikan kesalahan klasifikasinya *bagging* status gizi balita WHO-NCHS mengalami penurunan yang cukup kecil yaitu sebesar 2.991% ,sedangkan pada *bagging* status *cluster* penurunannya mencapai 22.046%. Hal ini mengindikasikan bahwa parameter model regresi logistik ordinal WHO-NCHS pada *data set* tunggal lebih stabil daripada parameter regresi logistik ordinal pada status gizi balita

*cluster*. Ini disebabkan karena pada saat pembentukan model *multiple* logistik ordinal status gizi balita *cluster* terdapat kategori yang pengaruhnya kecil terhadap model dan tidak signifikan pada tingkat signifikansi 5%. Jika dilihat dari hasil ketepatan klasifikasi, baik model status gizi balita *cluster* maupun status gizi balita standart WHO-NCHS memberikan hasil yang hampir sama, namun jika dilihat dari kestabilan parameternya, model klasifikasi status gizi balita WHO-NCHS lebih stabil, sehingga model ini lebih tepat untuk pengklasifikasian balita Nganjuk.

#### 4. Kesimpulan

Hasil pembahasan di atas maka dapat diambil kesimpulan sebagai berikut.

1. *Bagging* regresi logistik ordinal status gizi *cluster* mampu menurunkan kesalahan klasifikasi sebesar 22.046% dari *data set* tunggal. Prediktor terbaik diperoleh pada replikasi bootstrap sebanyak 60 kali. Ketepatan klasifikasi *bagging* sebesar 76.345% sedangkan pada *data set* tunggal sebesar 69.655%. Variabel prediktor yang dilibatkan dalam model adalah faktor-faktor yang mempengaruhi kelompok status gizi balita *cluster* yaitu umur, BBL, berat badan sekarang pendapatan, pendidikan terakhir ibu, frekuensi pemberian ASI dan intensitas keterlibatan ayah dalam mengasuh. Sedangkan status gizi balita *cluster* dibentuk dari analisis *cluster*.
2. Fungsi klasifikasi pada *data set* tunggal mampu mengklasifikasikan balita dengan ketepatan 75.862%, setelah dilakukan *bagging* diperoleh ketepatan klasifikasi sebesar 76.584% sehingga penurunan kesalahan *data set* tunggal dari  $e_m$  29.138% turun sebesar 2.991% menjadi  $\bar{e}_b$  23.416%. Variabel prediktor yang dilibatkan dalam model adalah faktor-faktor yang mempengaruhi status gizi balita menurut standart WHO-NCHS adalah umur, BBL, berat badan sekarang, pendapatan, keterlibatan ayah.
3. Hasil ketepatan klasifikasi, baik model status gizi balita *cluster* maupun status gizi balita standart WHO-NCHS memberikan hasil yang hampir sama, yaitu 76.55% untuk status gizi *cluster* dan 76.54% untuk status gizi standart WHO-NCHS, namun jika dilihat dari faktor-faktor yang dilibatkan pada pembentukan model, status gizi balita *cluster* melibatkan faktor eksternal yang secara tidak langsung mempengaruhi status gizi balita. Sehingga model *bagging* status gizi *cluster* dapat menjadi alternatif pilihan dalam pengklasifikasian balita

#### DAFTAR PUSTAKA

1. Agresti, A. *Categorical Data Analysis*. John Wiley and Sons. New York, 1990.
2. Akbar, Otok, Marice., Klasifikasi Status Gizi Balita dengan Pendekatan Diskriminan Bootstrap. *Prosiding Seminar Nasional Basic Science*, Universitas Brawijaya Malang, 2005.
3. Benufinit, S.H.N. (2005). Hubungan Antara Status gizi, Kebiasaan Ibu dalam Memberikan Makanan dan Sanitasi Lingkungan dengan Kejadian Diare Pada Balita. *Skripsi Jurusan Epidemiologi dan Biostatistika Fakultas Kesehatan Masyarakat Universitas Nusa Cendana*, Tidak dipublikasikan.
4. Breiman, L. *Bagging Predictor*. Technical report No.421. Department of statistics University of California., 1994.
5. Efron, B. dan Tibshirani, R.J. *An Introduction to the Bootstrap*, Chapman & Hall, Inc, New York., 1993.
6. Fahrmeir and Tutz. *Multivariate Statistical Modelling Based on Generalized Linear Models*. Springer Verlag. New-York. Inc., 1994.

7. Haryono. (2005). 'Maksimalkan TP PKK untuk Kelola Posyandu', Available: [http://Maksimalkan+ TP+ PKK+ untuk+Kelola+Posyandu%22++nganjuk](http://Maksimalkan+TP+PKK+untuk+Kelola+Posyandu%22++nganjuk).(tanggal akses: 2 Februari 2008).
8. Hosmer, D. W. and Lemeshow, S. *Applied Logistic Regression*. John Wiley and Sons, Inc. USA., 2000.
9. Irianto, D.P. *Panduan Gizi Lengkap Keluarga dan Olah-ragawan*. Andi, Yogyakarta, 2006.
10. Johnson, R. A. dan Wichern, D. W. *Applied Multivariate Statistical Analysis*. Prentice\_Hall.Inc. USA., 1998.
11. Mesah, S. (2005). Pengaruh Status Gizi dan Sanitasi Rumah Terhadap Kejadian Pnemonia pada Balita Penderita ISPA. *Skripsi Jurusan Epidemiologi dan Biostatistika Fakultas Kesehatan Masyarakat Universitas Nusa Cendana*, Tidak dipublikasikan.
12. Soegianto, B. *Pengenalan Dini Penyimpangan Pertumbuhan dan Tindak Lanjutnya Sebagai Salah Satu Cara Mencegah Terjadinya Manultrisi Pada Anak Balita*. Modul Ajar Akademi Gizi Surabaya., 2000.
13. Wahyuni, E. (2003). Hubungan Tingkat Pendidikan, Pengetahuan Gizi Ibu dan Tingkat Pendapatan Keluarga dengan status gizi balita, *Tugas Akhir Fakultas Kesehatan Masyarakat Universitas Airlangga*, Tidak dipublikasikan.
14. Wibawati., Faktor-Faktor yang Mempengaruhi Status Gizi Balita di Kabupaten "X". *Proseding Seminar Nasional UNESA*, 2007.

