

KLASIFIKASI TINGKAT KESEJAHTERAAN KELUARGA DI KECAMATAN SIDEMEN MENGGUNAKAN *BOOTSTRAP* *AGGREGATING (BAGGING)* REGRESI LOGISTIK ORDINAL

I Gusti Ngurah Sentana Putra^{1§}, Made Susilawati², I Putu Winada Gautama³

¹Program Studi Matematika– Universitas Udayana [Email: sentanaputra24@gmail.com]

²Program Studi Matematika – Universitas Udayana [Email: mdsusilawati@unud.ac.id]

³Program Studi Matematika – Universitas Udayana [Email: winadagautama@unud.ac.id]

[§]*Corresponding Author*

ABSTRACT

This research was conducted to determine the variables that have a significant impact on the stages of a well-off family in Sidemen Sub-district based on indicators obtained from the BKKBN and to classify the stages of a well-off family. This study used secondary data obtained from the stage of well-being data, Sidemen Sub-district, Karangasem Regency from BKKBN, totaling 1796 families. The method used is ordinal logistic regression and bagging ordinal logistic regression. Based on the logit regression model of ordinal logistic regression and ordinal logistic regression bagging, there are fourteen variables that have a significant effect on the dependent variable, namely marital status, type of insurance, age of head of household, occupation of head of household, having a source of income, eating a variety of food, having savings, accessing information from online media, families have ever recreated together, families have ever participated in social/community activities, the largest type of floor, main source of drinking water, ownership of a house/building, and children are still in school. The classification accuracy level in testing data using the ordinal logistic regression method was 79.4%, while the classification accuracy level using the bagging ordinal logistic regression method with 50,000 replications was 82.78%, so bagging showed an increase in classification by 3.38%.

Keywords: *Bootstrap Aggregating, Ordinal Logistic Regression, Modified Hosmer-Lemeshow Test, prosperous family stages, Sidemen Sub-district.*

1. PENDAHULUAN

Kesejahteraan merupakan standar yang menjadi perhatian pemerintah dalam menerangkan keadaan masyarakat. Hal ini disebabkan kesejahteraan menjadi salah satu tujuan dan bahan evaluasi pemerintah dalam menjamin taraf hidup masyarakat. Kesejahteraan dapat ditinjau dari kesejahteraan keluarga. Kemampuan keluarga untuk mencukupi kebutuhan sehingga mencapai kehidupan layak, nyaman dan sehat dapat diartikan menjadi kesejahteraan keluarga.

Menurut Badan Kependudukan dan Keluarga Berencana Nasional (BKKBN), keluarga yang sejahtera terdiri dari beberapa kategori, yaitu keluarga prasejahtera, keluarga sejahtera I, keluarga sejahtera II, keluarga sejahtera III, dan keluarga sejahtera III-plus. Status keluarga tersebut menentukan tingkat kesejahteraan mereka. Sebuah data yang

dikeluarkan oleh Ombudsman pada 13 Mei 2020 melalui siaran pers mengungkapkan bahwa sebanyak 72 persen dari keluhan terkait bantuan sosial berhubungan dengan penyaluran yang tidak tepat sasaran. Hal ini disebabkan oleh kurangnya integrasi data antara pemerintah pusat dan daerah sehingga penyaluran bantuan menjadi tidak akurat (Purwaningtyas, 2020).

Upaya peningkatan kesejahteraan keluarga dilakukan dengan pemetaan tingkat kesejahteraan keluarga sehingga dapat digunakan model klasifikasi untuk mengidentifikasi tingkat kesejahteraan keluarga dan menggambarkan faktor yang signifikan mempengaruhi tingkat kesejahteraan keluarga.

Metode analisis yang memperlihatkan korelasi antar variabel terikat dikotomi dan bertingkatan dengan variabel bebas yang berskala ordinal, nominal, interval atau kontinu

disebut analisis regresi logistik ordinal. Regresi logistik ordinal memberikan pendugaan parameter yang tidak stabil sehingga apabila data mengalami perubahan mengakibatkan perubahan signifikan pada model. Permasalahan ini kemudian diperlukan pendekatan bootstrap dalam menyelesaikannya.

Metode bootstrap merupakan metode yang bekerja dengan tidak memperhatikan asumsi distribusi dikarenakan menggunakan sampel data asli sebagai populasi. Cara kerja metode bootstrap dengan melakukan resampling pada data awal kemudian membentuk data baru. Pada metode regresi logistik yang memiliki penduga parameter tidak stabil dapat menggunakan metode bootstrap aggregating (bagging). Metode ini bertujuan untuk memperbaiki ketepatan dan meningkatkan tingkat akurasi prediksi dari penduga atau algoritma tertentu (Breiman, 1994).

Berdasarkan uraian di atas penelitian ini menggunakan metode Bootstrap Aggregating Regresi Logistik Ordinal untuk mengklasifikasikan tingkat kesejahteraan keluarga di Kecamatan Sidemen, Kabupaten Karangasem, dan untuk mengidentifikasi faktor-faktor signifikan yang memengaruhi tingkat kesejahteraan keluarga, serta untuk mendapatkan hasil klasifikasi tahapan kesejahteraan keluarga.

2. METODE PENELITIAN

2.1. Regresi Logistik Ordinal

Hosmer & Lemeshow (2013) menjelaskan bahwa regresi logistik ordinal adalah metode regresi yang memiliki variabel terikat dalam bentuk dikotomi yang memiliki tiga atau lebih kemungkinan dan memiliki sifat tingkatan. Menurut Agresti (2013) membentuk fungsi logit peluang kumulatif ke- j merupakan salah satu metode untuk membuat model regresi ordinal.

Model logit kumulatif memiliki variabel terikat berupa data tingkatan atau tahapan yang dinotasikan dengan $0,1,2,3 \dots J$, di mana J adalah banyaknya tingkatan variabel terikat ordinal dan variabel bebas berupa data kontinu, nominal atau kategorik yang dinotasikan $X = (X_1, X_2, \dots, X_p)$ dengan p adalah banyaknya variabel bebas.

Asumsikan bahwa variabel terikat Y berskala ordinal terdiri dari $K+1$ dan dinyatakan dengan $0,1,2, \dots, J$. Peluang besyarat dirumuskan secara umum dengan $Y = k$ pada

vektor x dari p variabel bebas adalah $P[Y = k|x] = \phi_k(x)$. Bentuk persamaan tersebut dapat dituliskan sebagai berikut

$$P(Y = k|x) = \frac{e^{g_k(x)}}{\sum_{i=1}^K e^{g_i(x)}}$$

dengan $g_0(x) = 0$, dan $k = 0,1,2,3$.

Model regresi logistik ordinal dapat diperoleh menggunakan model odds proportional sebagai salah satu interpretasinya. Model logit diperoleh dengan membandingkan peluang kumulatif dari variabel terikat yang kurang dari atau sama dengan kategori ke- k , diberikan p variabel bebas yang dinyatakan dalam vektor x , yaitu $P(Y \leq k|x)$, dengan peluang dari kategori terikat yang lebih besar dari kategori ke- k , (x) yang didefinisikan sebagai berikut.

$$\begin{aligned} \text{logit}[P(Y \leq k|x)] &= c_k(x) \\ &= \ln \left[\frac{\Pr(Y \leq k|x)}{\Pr(Y > k|x)} \right] \\ &= \ln \left[\frac{\phi_0(x) + \phi_1(x) + \dots + \phi_k(x)}{\phi_{k+1}(x) + \phi_{k+2}(x) + \dots + \phi_K(x)} \right] \\ &= \tau_k - x'\beta \end{aligned}$$

Untuk $k = 0,1,2, \dots, K-1$ (Hosmer & Lemeshow, 2013)

Menurut Hosmer & Lemeshow (2013) metode penduga parameter regresi logistik ordinal menggunakan metode kemungkinan maksimum (*maximum likelihood*). Bentuk umum dari fungsi *likelihood* untuk nilai variabel terikat y_i yang diasumsikan saling bebas dengan sampel sebanyak n amatan, $i = 1, 2, \dots, n$; adalah (Hosmer & Lemeshow, 2013):

$$l(\beta) = \prod_{i=1}^n [\phi_0(x_i)^{z_{0i}} \times \phi_1(x_i)^{z_{1i}} \times \dots \times \phi_K(x_i)^{z_{Ki}}]$$

di mana banyak fungsi parameter yang tidak diketahui ditunjukkan dengan K , fungsi parameter yang tidak diketahui ditunjukkan dengan $\phi_h(x_i)$, dan $z' = (z_0, z_1, \dots, z_k)$ dibentuk dari variabel terikat ordinal. Estimasi kemungkinan maksimum diperoleh dari vektor $\beta'_k = (\beta_{0k}, \beta_{1k}, \dots, \beta_{pk})$ dengan memaksimalkan fungsi *likelihood* bersama dari persamaan. Logaritma dari fungsi *likelihood* dapat dituliskan sebagai berikut (Hosmer & Lemeshow, 2013):

$$l(\beta) = \sum_{i=0}^n z_{0i} \ln[\phi_0(x_i)] + z_{1i} \ln[\phi_1(x_i)] + \dots + z_{ki} \ln[\phi_k(x_i)]$$

2.2. Pendugaan Parameter Simultan dan Parsial

Uji signifikansi parameter dalam regresi logistik ordinal mencakup uji signifikansi parsial dan simultan. Uji signifikansi simultan bertujuan untuk mengevaluasi pengaruh variabel secara keseluruhan terhadap pembentukan model. Uji signifikansi simultan menggunakan uji G yang dirumuskan sebagai berikut:

$$G = -2 \ln \frac{L_0}{L_p} \quad (1)$$

dengan

L_0 = likelihood tanpa variabel bebas

L_p = likelihood dengan variabel bebas

Pengujian signifikansi secara parsial bertujuan untuk mengetahui pengaruh variabel dalam model secara individu. Pengujian signifikansi parameter secara parsial menggunakan uji wald yang dirumuskan sebagai berikut

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

Statistik uji yang digunakan adalah statistik uji wald dengan $\hat{\beta}_k$ adalah nilai estimasi parameter variabel bebas ke-k, $SE(\hat{\beta}_k)$ adalah nilai galat baku dari estimasi variabel bebas ke-k.

2.3. Stepwise Regresi Logistik Ordinal

Stepwise Regression adalah metode memasukkan dan mengeluarkan variabel dari model (Hosmer & Lemeshow, 2013). Metode ini merupakan kombinasi dari maju dan mundur. Kriteria regresi logistik menggunakan *P-value*. Nilai *P-value* dapat ditentukan dengan statistik z, t, wald atau *likelihood ratio* (Hilbe, 2009).

2.4. Uji Kesesuaian Model

Uji goodness of fit pada model regresi logistik ordinal dapat dilakukan dengan menggunakan uji Hosmer & Lemeshow. Uji statistik yang digunakan dalam uji Hosmer & Lemeshow adalah berdasarkan pengelompokan nilai perkiraan probabilitas yang diamati pada setiap variabel bebas (Hosmer & Lemeshow,

2013). Statistik uji Hosmer & Lemeshow dapat dirumuskan sebagai berikut::

$$\hat{C} = \sum_{r=1}^g \frac{(C_r - n'r\bar{p}_{1r})^2}{n'r\bar{p}_{1r}(1 - \bar{p}_{1r})} \quad (3)$$

dengan g menyatakan banyaknya kategori, C_r menyatakan jumlah nilai Y pada grup ke-r, $n'r$ menyatakan jumlah observasi dalam grup ke-r dan \bar{p}_{1r} menyatakan rata rata taksiran probabilitas sukses grup ke-r.

Diasumsikan bahwa statistika uji Hosmer & Lemeshow mengikuti sebaran khi-kuadrat (χ^2) dengan derajat bebas $g - 2$. Apabila nilai $\hat{C} > \chi^2_{(\alpha, g-2)}$ maka keputusan yang diperoleh adalah menolak H_0 yang berarti model tidak menunjukkan data secara layak.

2.5. Uji Hosmer-Lemeshow Modifikasi

Uji *chi-square* meningkat dengan ukuran sampel yang menyebabkan uji Hosmer & Lemeshow juga mengalami kenaikan. Uji kesesuaian idealnya menghasilkan penolakan model regresi yang baik yang independen dengan ukuran sampel (Paul, 2013). Menurut (Yu *et.al*, 2017), uji kesesuaian dilakukan dengan cara membandingkan kuasa uji dengan parameter non sentral (λ) untuk mencapai model yang ideal. Langkahnya dengan mengalikan λ dengan konstanta akan menstabilkan kuasa uji. Konstanta khusus ini adalah fungsi dari ukuran sampel. Jika H_0 diterima, $\hat{C}_g \sim \chi_{g-2}$ diketahui, tetapi jika H_0 ditolak, statistik Hosmer-Lemeshow yang digunakan adalah distribusi chi-kuadrat nonsentral sebagai berikut:

$$\hat{C}_g \sim \chi_{g-2}(\lambda) \quad (4)$$

dimana λ merupakan parameter non-sentralis dengan estimasi sebagai berikut :

$$\hat{\lambda} = (\hat{C}_g) - (g - 2) \quad (5)$$

dengan \hat{C}_g merupakan nilai uji Hosmer & Lemeshow.

Menurut Paul (2013) ukuran sampel $1000 < n \leq 25000$ maka penentuan g dapat menggunakan rumus sebagai berikut :

$$g = \max \left(10, \min \left\{ \frac{k}{2}, \frac{n-k}{2}, 2 + 8 \left(\frac{n}{1000} \right)^2 \right\} \right)$$

dimana k merupakan jumlah n sukses. Prosedur pengujian hosmer & lemeshow modifikasi dapat digunakan tahap sebagai berikut (Yu *et.al*, 2017):

Tahap 1 : Melakukan uji Hosmer & Lemeshow test dengan $g = 10$ dari ukuran data

n , kemudian hitung λ . Tahap 2 : Definisikan modifikasi λ_c dengan $\lambda_c = c\lambda = c(\widehat{C}_g) - (g - 2)$. Tahap 3 : jika $\lambda_c < 0$ maka H_0 diterima namun jika $\lambda_c \geq 0$, maka bangkitkan secara acak r dari distribusi X_{g-2, λ_c}^2 . Z_c merupakan nilai kritis dari modifikasi hosmer & lemeshow dimana Z_c merupakan penyelesaian dari $\frac{1}{c} \int_0^\infty (1 - G_{m-2, x(z_c)}) f_{m-2}(\frac{x}{c} + (m - 2)) dx = \alpha$. Jika $r > z_c$ maka tolak H_0 . Nilai c dalam tahap 2 ditentukan dengan rumus sebagai berikut :

$$c = \frac{n_0}{n} \sqrt{\frac{g_0}{g}} \quad (6)$$

dengan n_0 merupakan ukuran sampel grup yang didefinisikan oleh peneliti, n merupakan ukuran sampel data, g_0 merupakan jumlah grup dimana jumlah yang ditentukan 10 dan g merupakan grup.

2.6. Bootstrap Aggregating

Metode *Bagging* digunakan untuk memperbaiki dan meningkatkan ketepatan akurasi dengan reduksi variansi dari prediktor pada metode regresi dan klasifikasi. Dataset direplikasi dengan *bootstrap* kemudian dibentuk versi bergandanya. Dasar pemikiran dari *bagging* adalah membangkitkan prediktor dengan versi berganda dengan menggunakan *bootstrap resampling*,

2.7. Akaike Information Criterion

AIC adalah metode yang dikembangkan oleh Akaike (1973) untuk membandingkan model statistik dalam memprediksi data. AIC adalah salah satu metode dalam memilih model terbaik, dan didasarkan pada metode maximum likelihood. AIC digunakan untuk memilih model yang paling sesuai dengan data yang diperoleh, dan dapat dihitung menggunakan persamaan:

$$AIC = 2k - 2\ln(L) \quad (7)$$

dengan k merupakan jumlah variabel bebas yang digunakan dan L merupakan log-likelihood estimator. Menurut Fathurahman (2009), dalam model regresi, model terbaik dapat diidentifikasi dengan nilai AIC yang paling kecil.

2.8. Confusion Matrix

Ketepatan klasifikasi pada penelitian ini menggunakan *confusion matrix*. *Confusion matrix* merupakan perhitungan ketepatan klasifikasi dengan membandingkan hasil prediksi data dari proses klasifikasi dengan data

sebenarnya dari keseluruhan data. *Confusion matrix* menghitung *accuracy*, *precision* dan *recall*. *Precision* merupakan ketetapan dari klasifikasi yang diprediksi benar. *Recall* merupakan ketetapan dari klasifikasi yang diidentifikasi benar. *Accuracy* merupakan ketetapan jumlah data yang diprediksi benar

2.9. Jenis dan Sumber Data

Pada penelitian ini, data yang digunakan bersumber dari data sekunder yang diperoleh dari BKKBN, yang merupakan data tahapan keluarga sejahtera di Kecamatan Sidemen Kabupaten Karangasem. Populasi yang diteliti meliputi semua variabel yang memiliki karakteristik tersendiri. Populasi dalam penelitian ini ditentukan melalui data jumlah keluarga Kecamatan Sidemen, Karangasem berdasarkan tingkatan kesejahteraan keluarga dengan jumlah populasi sebanyak 11.206 keluarga. Dalam penelitian ini digunakan data dua desa yaitu desa lokasari dan desa telagatawang sejumlah 1796 keluarga dengan pertimbangan bahwa desa lokasari dan desa telagatawang memiliki seluruh tahapan keluarga sejahtera secara merata. Variabel terikat yang digunakan dalam penelitian ini diambil dari tahapan keluarga sejahtera yaitu keluarga pra sejahtera, keluarga sejahtera I, keluarga sejahtera II dan keluarga sejahtera III sedangkan variabel bebas yang digunakan merupakan beberapa indikator keluarga sejahtera dari BKKBN yaitu banyak anggota keluarga (X_1), jenis kelamin kepala keluarga (X_2), status perkawinan (X_3), jenis asuransi (X_4), umur kepala keluarga (X_5), pendidikan kepala keluarga (X_6), pekerjaan kepala keluarga (X_7), ibadah secara rutin (X_8), memiliki sumber penghasilan (X_9), makan makanan beragam (X_{10}), memiliki tabungan/simpanan (X_{11}), mengakses informasi dari media online (X_{12}), keluarga pernah berekreasi bersama di luar rumah (X_{13}), keluarga ikut serta kegiatan sosial/gotong (X_{14}), jenis atap rumah terluas (X_{15}), jenis dinding rumah terluas (X_{16}), jenis lantai rumah terluas (X_{17}), sumber air minum utama (X_{18}), luas rumah/bangunan keseluruhan (X_{19}), kepemilikan rumah/bangunan tempat tinggal (X_{20}), anak masih sekolah (X_{21}).

2.10. Tahapan Analisis Data

Penelitian ini menggunakan metode analisis regresi logistik ordinal dan *bagging* regresi logistik ordinal dengan tahapan analisis

data sebagai berikut:

1. Melakukan data *splitting* menjadi 80% data training dan 20% data testing.
2. Melakukan Uji Multikolinearitas
3. Melakukan estimasi parameter model
4. Uji signifikansi parameter simultan dengan uji G dan uji signifikansi parsial dengan uji wald
5. Menguji kesesuaian model dengan uji Hosmer & Lemeshow dan kesesuaian model dengan uji Hosmer & Lemeshow modifikasi
6. Seleksi variabel bebas dengan *stepwise logistic regression*
7. Interpretasi nilai AIC model awal dan model akhir regresi logistik ordinal.
8. Melakukan model akhir menggunakan uji model odds proporsional (proportional odds model) dan interpretasi terhadap model akhir yang sudah diperoleh
9. Menghitung ketepatan pengklasifikasian model Regresi Logistik Ordinal
10. Mengambil n sampel L_B secara random pada data training L (variabel terikat dan variabel bebas berdasarkan model regresi logistik berganda) untuk masing-masing replikasi.
11. Membentuk model menggunakan hasil *bootstrap*
12. Menggunakan data testing untuk melakukan prediksi nilai variabel bebas berdasarkan model Bagging regresi logistik ordinal yang telah dibentuk pada tahap sebelumnya.
13. Langkah 10 hingga 12 diulang sebanyak B kali, di mana B merupakan jumlah replikasi yang ditentukan sebelumnya.
14. Model Bagging regresi logistik ordinal diperoleh pada replikasi B kali.
15. Menentukan kelas maksimum diprediksi menggunakan *majority voting* dan menghitung kesalahan klasifikasi bagging.
16. Replikasi Bootstrap yang dilakukan pada penelitian ini adalah 50.000 kali

3. HASIL DAN PEMBAHASAN

Regresi logistik dengan prosedur bootstrap dalam kumpulan data besar direkomendasikan menggunakan proporsi pemisahan 70/30, 80/20, dan 50/50 untuk mempertahankan akurasi dan mengurangi waktu komputasi model klasifikasi. Tahap awal penelitian ini dilakukan dengan membagi 80% dari 1796 data yang diperoleh dari tahapan keluarga sejahtera di kecamatan Sidemen sebagai data *training*.

3.1. Uji Multikolinearitas

Pada model regresi logistik ordinal langkah selanjutnya sebelum melakukan pemodelan adalah uji multikolinearitas untuk mengetahui variabel-variabel yang saling berkorelasi.

Menurut Hosmer & Lemeshow (2013) Analisis regresi ordinal harus uji multikolinearitas atau tidak terdapat variabel yang multikolinearitas dengan variabel lainnya.

Berdasarkan uji multikolinearitas tidak terdapat nilai GVIF dibawah 10 sehingga variabel bebas dalam penelitian tidak memiliki korelasi dengan satu atau lebih variabel bebas lainnya. Hal ini diperkuat dengan tidak adanya nilai $GVIF^{\frac{1}{2*df}} \leq 10^{\frac{1}{2*df}}$.

3.2. Uji Parameter Simultan

Pengujian parameter secara simultan dapat digunakan uji G dengan hipotesis sebagai berikut

$H_0: \beta_i = 0$, tidak terdapat pengaruh variabel bebas secara simultan terhadap tahapan keluarga sejahtera

$H_1: \beta_i \neq 0$, terdapat minimal satu variabel bebas berpengaruh secara simultan terhadap tahapan keluarga sejahtera.

Statistik uji G mengikuti sebaran khi-kuadrat dengan derajat bebasnya yaitu $((K + 1) - 2)p = ((21 + 1) - 2)4 = 80$ dan uji *likelihood ratio* dengan nilai $G = 7624.422 > \chi_{0,05;80}^2 = 101.8795$ maka tolak H_0 , sehingga dapat disimpulkan terdapat pengaruh sedikitnya satu variabel bebas terhadap tahapan keluarga sejahtera.

3.3. Uji Parameter Parsial

Selanjutnya dilakukan pengujian parsial. Statistik uji yang digunakan adalah uji wald dengan hipotesis sebagai berikut:

$H_0: \beta_i = 0$, tidak terdapat pengaruh signifikan variabel bebas terhadap tahapan keluarga sejahtera.

$H_1: \beta_i \neq 0$, terdapat pengaruh signifikan variabel bebas terhadap tahapan keluarga sejahtera.

dengan $i = 1, \dots, 21$.

Pengujian parameter secara parsial terdapat lima variabel yang tidak signifikan dari dua puluh satu variabel yaitu Banyak Anggota Keluarga (X_1), Pendidikan Kepala Keluarga (X_6), Ibadah secara rutin (X_8), Jenis atap rumah terluas (X_{15}), dan Luas rumah/bangunan keseluruhan (X_{19}). Oleh karena itu, dapat

disimpulkan bahwa variabel-variabel lain memiliki pengaruh yang signifikan terhadap Tahapan Keluarga Sejahtera di Kecamatan Sidemen.

3.4. Uji Kesesuaian Model

Pengujian hosmer & lemeshow dilakukan untuk menguji kesesuaian model dengan hasil amatan dengan menggunakan hipotesis sebagai berikut:

H_0 : model regresi logistik ordinal sesuai (tidak terdapat perbedaan yang signifikan antara hasil pengamatan dengan kemungkinan hasil prediksi model)

H_1 : model regresi logistik ordinal tidak sesuai (terdapat perbedaan yang signifikan antara hasil pengamatan dengan kemungkinan hasil prediksi model)

Uji Hosmer & Lemeshow mengikuti sebaran khi-kuadrat (χ^2) dengan kriteria uji kesesuaian model Hosmer & Lemeshow adalah tolak H_0 jika $\hat{C} > \chi^2_{(\alpha, g-2)}$ atau $p\text{-value} \leq \alpha$. Dari tabel chi-square diperoleh $\chi^2_{(\alpha, 8)} = 8.7458$, dengan $g=10$. Nilai $\hat{C} > \chi^2_{(\alpha, g-2)} = 144.3$ yang diperoleh dari hasil uji hosmer & lemeshow sehingga kesimpulannya H_0 ditolak yang berarti adanya perbedaan signifikan antar hasil prediksi model dengan hasil amatan.

3.5. Uji Hosmer-Lemeshow Modifikasi

Uji *chi-square* meningkat dengan ukuran sampel yang menyebabkan uji Hosmer & Lemeshow juga mengalami kenaikan. Langkah selanjutnya dilakukan pengujian uji hosmer lemeshow modifikasi. Diketahui nilai dari $\hat{C} = 144.3$ maka diperoleh nilai $\lambda = 136.3$. Kemudian ditentukan $n_0 = 100$, $n = 1436$, $g_0 = 10$ dan $g = 10$, sehingga diperoleh nilai $\lambda_c = 9.491$.

Nilai $\lambda_c \geq 0$ maka keputusan dibangkitkan r dari distribusi X^2_{g-2, λ_c} dimana diperoleh 12.544, Nilai z_c pada tingkat signifikansi 5% dengan derajat kebebasan 32, tidak ada nilai kritis untuk ukuran sampel 1436, jadi nilai z_c diinterpolasi. Hasil interpolasi nilai kritis untuk ukuran sampel 1436 yaitu 15.284. $r < z_c = 12.544 < 15.284$ maka terima H_0 yang berarti tidak adanya perbedaan signifikan antara hasil prediksi model dengan data amatan.

3.6. Stepwise Regresi Logistik Ordinal

Penentuan model terbaik regresi logistik menggunakan metode *stepwise*, dengan tujuan

untuk memilih variabel terbaik yang mendominasi pembentukan fungsi regresi logistik ordinal. Kriteria pemilihan variabel menggunakan $p\text{-value} < 0.05$ dan nilai Wald tertinggi.

Proses *stepwise* berhenti pada *step* dua belas setelah variabel bebas yang dimasukkan tidak mempunyai pengaruh signifikan terhadap model. Step dua belas merupakan step terakhir dalam pemilihan variabel dengan variabel yang mempunyai pengaruh signifikan yaitu variabel status perkawinan (X_3), jenis asuransi (X_4), umur kepala keluarga (X_5), pekerjaan kepala keluarga (X_7), memiliki sumber penghasilan (X_9), makan makanan beragam (X_{10}), memiliki tabungan/simpanan (X_{11}), mengakses informasi dari media online (X_{12}), keluarga pernah berekreasi bersama di luar rumah (X_{13}), keluarga ikut serta kegiatan sosial/gotong (X_{14}), jenis lantai rumah terluas (X_{17}), sumber air minum utama (X_{18}), Kepemilikan rumah/bangunan tempat tinggal (X_{20}), dan anak masih sekolah (X_{21})

3.7. Metode Akaike Information Criterion

Variabel data yang diperoleh ditentukan model paling baik menggunakan model AIC. Pada penelitian ini, untuk menentukan model AIC terbaik dibandingkan nilai AIC dari model awal regresi logistik ordinal dengan menggunakan seluruh variabel bebas dengan model regresi ordinal dengan menggunakan variabel yang dipilih menggunakan metode *stepwise regression*.

Tabel 3.1 Nilai AIC model regresi logistik ordinal

	Nilai AIC
Model Regresi Logistik Ordinal	1564.126
Model Regresi logistik Ordinal setelah <i>stepwise regression</i>	1560.967

Model terbaik untuk regresi berdasarkan nilai AIC terkecil. Berdasarkan Tabel 3.1, model regresi ordinal setelah *stepwise regression* memiliki nilai AIC sebesar 1560.967 yang artinya lebih kecil dibandingkan dengan model regresi logistik ordinal sehingga model regresi ordinal setelah *stepwise regression* adalah model terbaik.

3.8. Interpretasi Regresi Logistik Ordinal

Selanjutnya dilakukan analisis setiap koefisien dari tahapan model keluarga sejahtera sehingga diperoleh nilai odds ratio dari masing-masing koefisien dengan nilai berikut ini:

Tabel 3.2. Nilai Odds Ratio Masing-Masing Koefisien Regresi Logistik Ordinal

	Variabel	Coef	exp (β_i)
Status Perkawinan (X_3)	Belum Kawin		
	Kawin	0.9332	2.5425
	Cerai Hidup	-0.7793	0.4587
	Cerai Mati	0.3561	1.4277
Jenis Asuransi (X_4)	BPJS-PBI/ Jamkesmas/Jamkesda		
	BPJS-non BPI	0.7910	2.2057
	Swasta	-2.3852	0.09207
	Tidak memiliki	-3.4355	0.0322
Status Perkawinan (X_3)	Belum Kawin	0.0156	1.0157
Pekerjaan Kepala Keluarga (X_7)	Tidak		
	Iya	-0.7312	0.4813
Memiliki sumber penghasilan (X_9)	Iya		
	Tidak	-3.6206	0.0267
Makan makanan beragam (X_{10})	Iya		
	Tidak	-4.7823	0.0084
Memiliki tabungan/simpanan (X_{11})	Iya		
	Tidak	-1.4918	0.2249
Mengakses informasi dari media online (X_{12})	Iya		
	Tidak	-0.4946	0.6098
Keluarga pernah berekreasi bersama di luar rumah (X_{13})	Iya		
	Tidak	-2.3583	0.0946
Jenis lantai rumah terluas (X_{17})	Keramik		
	Semen	-0.0782	0.9248
	Tanah	-2.9688	0.05136
	Lainnya	-0.4838	0.6164
Sumber air minum utama (X_{18})	Air Kemasan/ isu ulang		
	PDAM	1.0586	2.8824
	Sumur Bor	5.1299	169.0106
	Sumur Terlindung	1.4960	4.46390
	Air permukaan	1.1391	3.1240
	Lainnya	0.8609	2.3653
Kepemilikan rumah/bangunan tempat tinggal (X_{20})	Milik Sendiri		
	Kontrak/Sewa	-2.0499	1.28744
	Menumpang	0.0626	1.0646
	Lainnya	0.9441	2.5708
Anak masih sekolah (X_{21})	Tidak		
	Iya	0.6369	1.8906

Metode regresi logistik ordinal dapat diinterpretasikan bahwa *odds ratio* $OR = 1$ berarti kelompok mempunyai peluang yang sama. Apabila $OR > 1$ maka kelompok pertama variabel mempunyai peluang yang lebih besar dibanding dengan kelompok kedua. Sebaliknya apabila $OR < 1$, kelompok pertama mempunyai peluang yang lebih kecil dibandingkan dengan kelompok kedua. Berdasarkan Tabel 3.2 sebagai contoh Odds Ratio $X_{3(2)}$ (Status Perkawinan) = 2,5425, artinya status perkawinan kawin dapat meningkatkan tahapan keluarga sejahtera dibandingkan status perkawinan tidak kawin sebesar 2,5425 kali. Odds Ratio $X_{3(3)}$ (Status

Perkawinan) = 0,4587, artinya status perkawinan cerai hidup dapat menurunkan tahapan keluarga sejahtera dibandingkan status perkawinan tidak kawin sebesar 0,4587 kali. Odds Ratio $X_{3(4)}$ (Status Perkawinan) = 1,4277, artinya status perkawinan cerai mati dapat meningkatkan tahapan keluarga sejahtera dibandingkan status perkawinan tidak kawin sebesar 1,4277 kali.

3.9. Klasifikasi Regresi Logistik Ordinal

Persamaan model logit tahapan keluarga sejahtera dibentuk fungsi klasifikasi sehingga diperoleh hasil pengklasifikasian tahapan

keluarga sejahtera. Pada penelitian ini proses klasifikasi menggunakan data *testing* sebanyak 360 data dari data tahapan keluarga sejahtera

kecamatan sidemen. Dari persamaan model logit diperoleh hasil klasifikasi tahapan keluarga sejahtera sebagai berikut:

Tabel 3.3. *Confusion Matrix* Regresi Logistik Ordinal

Keanggotaan sebenarnya	Keanggotaan prediksi				Total
	Keluarga Prasejahtera	Keluarga Sejahtera I	Keluarga Sejahtera II	Keluarga Sejahtera III	
Keluarga Prasejahtera	51	0	15	0	66
Keluarga Sejahtera I	13	0	19	0	32
Keluarga Sejahtera II	2	0	220	1	223
Keluarga Sejahtera III	0	0	24	15	39
	66	0	278	16	360

Berdasarkan Tabel, dengan *confusion matrix* tingkat ketepatan klasifikasi dapat dihitung sebagai berikut:

$$\text{Accuracy} = \frac{51 + 0 + 220 + 15}{360} = 0.794 \quad (4.4)$$

diperoleh persentase ketepatan klasifikasi untuk tingkat keluarga kaya di Kecamatan Sidemen menggunakan data uji adalah 79,4% dengan kesalahan klasifikasi 20,6%.

Tingkat akurasi klasifikasi sebesar 79,4 persen menunjukkan bahwa model klasifikasi cukup efektif dan memiliki performa yang baik dalam melakukan klasifikasi data. Namun, terdapat kemungkinan terjadinya kesalahan dalam prediksi, yang biasanya disebut sebagai "*error rate*" atau tingkat kesalahan. Oleh karena itu, perlu dilakukan evaluasi dan peningkatan pada model klasifikasi tersebut untuk meningkatkan tingkat akurasi dan mengurangi tingkat kesalahan dalam prediksi.

3.10. Bagging Regresi Logistik Ordinal

Fungsi klasifikasi yang dibentuk berdasarkan model logit adalah model *data set* tunggal. Menurut Breiman (1996), model regresi logistik ordinal menghasilkan parameter yang

tidak stabil. Parameter yang tidak stabil dalam regresi logistik ordinal dapat diperbaiki dengan melakukan *resampling* data.

Proses *resampling* menggunakan sampel yang dikembalikan dengan mengikutsertakan beberapa variabel terikat dan variabel bebas. Pada penelitian ini, variabel yang diikutsertakan merupakan variabel pada model logit regresi logistik ordinal dari data *training* secara acak. Metode *bootstrap aggregating* regresi logistik ordinal dilakukan replikasi sebanyak 50.000 kali yang kemudian ditentukan prediksi replikasi menggunakan data *testing*. Prediksi variabel terikat maksimum pada replikasi dapat menggunakan *Majority voting*, kemudian prediksi maksimum klasifikasi pada replikasi dilakukan klasifikasi untuk memperoleh kesalahan klasifikasi *bagging* regresi logistik ordinal.

3.11. Interpretasi Bagging Regresi Logistik Ordinal

Selanjutnya dilakukan analisis setiap koefisien dari tahapan model keluarga sejahtera sehingga diperoleh nilai odds ratio dari masing-masing koefisien dengan nilai disajikan pada Tabel 3.4.

Tabel 3.4. Nilai Odds Ratio Masing-Masing Koefisien Bagging Regresi Logistik Ordinal

	Variabel	Koefisien	$\exp(\beta_i)$
Status Perkawinan (X_3)	Belum Kawin		
	Kawin	0.9555	2.6001
	Cerai Hidup	-0.3950	0.6737
	Cerai Mati	0.1854	1.2038
Jenis Asuransi (X_4)	BPJS-PBI/ Jamkesmas/Jamkesda		
	BPJS-non BPI	1.0367	2.8199
	Swasta	-1.3326	0.2638
	Tidak memiliki	-3.3743	0.0342
Status Perkawinan (X_3)	Belum Kawin	0.0169	1.0171
Pekerjaan Kepala Keluarga (X_7)	Tidak		
	Iya	-0.3141	0.7305
Memiliki sumber penghasilan (X_9)	Iya		
	Tidak	-4.0547	0.0173
Makan makanan beragam (X_{10})	Iya		
	Tidak	-5.8406	0.0029
Memiliki tabungan/simpanan (X_{11})	Iya		
	Tidak	-1.7181	0.1794
Mengakses informasi dari media online (X_{12})	Iya		
	Tidak	-0.4091	0.6642
Keluarga pernah berekreasi bersama di luar rumah (X_{13})	Iya		
	Tidak	-2.4698	0.0846
Jenis lantai rumah terluas (X_{17})	Keramik		
	Semen	-0.0063	0.9937
	Tanah	-3.8872	0.0205
	Lainnya	1.0043	2.7300
Sumber air minum utama (X_{18})	Air Kemasan/ isu ulang		
	PDAM	0.6824	1.9787
	Sumur Bor	-0.7684	0.4637
	Sumur Terlindung	1.2299	3.42097
	Air permukaan	0.5951	1.8132
Kepemilikan rumah/bangunan tempat tinggal (X_{20})	Lainnya	0.5854	1.7957
	Milik Sendiri		
	Kontrak/Sewa	-0.4122	0.6622
	Menumpang	-0.13921	0.8700
Anak masih sekolah (X_{21})	Lainnya	1.0206	2.7749
	Tidak		
	Iya	0.6386	1.8939

Metode regresi logistik ordinal dapat diinterpretasikan bahwa *odds ratio* $OR = 1$ berarti kelompok mempunyai peluang yang sama. Apabila $OR > 1$ maka kelompok pertama variabel mempunyai peluang yang lebih besar dibanding dengan kelompok kedua. Sebaliknya apabila $OR < 1$, kelompok pertama mempunyai peluang yang lebih kecil dibandingkan dengan kelompok kedua. Berdasarkan Tabel 3.4 sebagai contoh Odds Ratio $X_{3(2)}$ (Status Perkawinan) = 2,6001, artinya status perkawinan kawin dapat meningkatkan tahapan keluarga sejahtera dibandingkan belum kawin sebesar 2,6001 kali. Odds Ratio $X_{3(3)}$ (Status Perkawinan) = 0,6737, artinya status perkawinan cerai hidup dapat

menurunkan tahapan keluarga sejahtera dibandingkan belum kawin sebesar 0,6737 kali. Odds Ratio $X_{3(4)}$ (Status Perkawinan) = 1,2038, artinya status perkawinan cerai mati dapat meningkatkan tahapan keluarga sejahtera dibandingkan belum kawin sebesar 1,2038 kali.

3.12. Klasifikasi Bagging Regresi Logistik Ordinal

Persamaan model logit tahapan keluarga sejahtera dibentuk fungsi klasifikasi sehingga diperoleh hasil pengklasifikasian tahapan keluarga sejahtera Pada penelitian ini proses klasifikasi menggunakan data *testing* sebanyak 30 data dari data tahapan kelyarga sejahtera

kecamatan sidemen. Dari persamaan model logit diperoleh hasil klasifikasi tahapan keluarga sejahtera sebagai berikut :

Tabel 4. 5. *Confusion Matrix* Bagging Regresi Logistik Ordinal

Keanggotaan sebenarnya	Keanggotaan prediksi				Total
	Keluarga Prasejahtera	Keluarga Sejahtera I	Keluarga Sejahtera II	Keluarga Sejahtera III	
Keluarga Prasejahtera	53	14	2	0	69
Keluarga Sejahtera I	0	0	0	0	0
Keluarga Sejahtera II	13	18	221	15	267
Keluarga Sejahtera III	0	0	0	24	24
	66	32	223	39	360

Berdasarkan Tabel, dengan *confusion matrix* tingkat ketepatan klasifikasi dapat dihitung sebagai berikut:

$$Accuracy = \frac{53 + 0 + 221 + 24}{360} = 0.8278 \quad (12)$$

diperoleh persentase ketepatan klasifikasi untuk tingkat keluarga kaya di Kecamatan Sidemen menggunakan data uji adalah 82,78% dengan kesalahan klasifikasi 17,22%.

Tingkat akurasi klasifikasi sebesar 82,78 persen dalam kasus ini menunjukkan bahwa model klasifikasi tersebut cukup efektif dan memiliki performa yang baik dalam melakukan klasifikasi data. Tingkat akurasi yang tinggi seperti ini dapat memberikan keyakinan pada pengguna model dalam pengambilan keputusan.

4. KESIMPULAN DAN SARAN

Berdasarkan model logit Regresi Logistik Ordinal dan bagging Regresi Logistik Ordinal, terdapat empat belas variabel yang berpengaruh signifikan terhadap variabel bebas yaitu status perkawinan, jenis asuransi, umur kepala keluarga, pekerjaan kepala keluarga, memiliki sumber penghasilan, makan makanan beragam, memiliki tabungan/simpanan, mengakses informasi dari media online, keluarga pernah berekreasi bersama di luar rumah, keluarga ikut serta kegiatan sosial/gotong, jenis lantai rumah terluas, sumber air minum utama, kepemilikan rumah/bangunan tempat tinggal, dan anak masih sekolah.

Metode regresi logistik ordinal dengan data training memiliki akurasi klasifikasi sebesar 79,40%, sedangkan metode bagging regresi logistik ordinal dengan 50.000 iterasi memiliki akurasi sebesar 82,78%. Berdasarkan hal

tersebut, proses bootstrap aggregating meningkatkan ketepatan klasifikasi sebesar 3,21% sehingga bagging regresi logistik ordinal dapat mengklasifikasikan tahapan tahapan keluarga sejahtera lebih tinggi dibanding dengan regresi logistik ordinal.

Pada analisis dan pembahasan klasifikasi tingkat kesejahteraan keluarga di kecamatan sidemen menggunakan bootstrap aggregating (Bagging) regresi logistik ordinal adapun beberapa saran yang dapat diberikan penulis pada penelitian selanjutnya untuk dapat menggunakan proporsi pemisahan data yang lebih beragam. Pada penelitian selanjutnya untuk menggunakan nilai n_0 dalam uji hosmer lemeshow modifikasi yang lebih besar. Pada penelitian tahapan keluarga sejahtera dapat menggunakan metode regresi logistik ordinal untuk menaikkan tingkat ketepatan klasifikasi. Berdasarkan penelitian ini dianjurkan untuk perangkat kecamatan dan desa setempat untuk dapat memperhatikan proses input dan penentuan klasifikasi data serta mempertimbangkan variabel-variabel yang signifikan berpengaruh dalam menyusun program kerja kecamatan atau desa setempat.

DAFTAR PUSTAKA

- Agresti, A. (2013). *Categorical Data Analysis* (3rd Ed). Wiley.
- Akaike, H. (1998). *Information Theory And An Extension Of The Maximum Likelihood Principle*. In E. Parzen, K. Tanabe, & G. Kitagawa (Eds.), *Selected Papers Of Hirotugu Akaike* (Pp. 199–213).
- Breiman, L. (1996). *Bagging Predictors*. *Machine Learning*, 24(2), 123–140.

- Hilbe, J. M. (2009). *Logistic Regression Models*. Crc Press.
- Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied Logistic Regression (Third Edition)*. Wiley.
- Nattino, G., Pennell, M. L., & Lemeshow, S. (2020). Assessing The Goodness Of Fit Of Logistic Regression Models In Large Samples: A Modification Of The Hosmer-Lemeshow Test. *Biometrics*, 76(2), 549–560.
- Paul, P., Pennell, M. L., & Lemeshow, S. (2013). Standardizing The Power Of The Hosmer-Lemeshow Goodness Of Fit Test In Large Data Sets. *Statistics In Medicine*, 32(1), 67–80.
- Yu, W., Xu, W., & Zhu, L. (2017). A modified Hosmer–Lemeshow test for large data sets. *Communications in Statistics - Theory and Methods*, 46(23), 11813–11825.
- Vrigazova, B. (2021). The Proportion for Splitting Data into Training and Test Set for the Bootstrap in Classification Problems. *Business Systems Research Journal*, 12(1), 228–242.
- Yu, W., Xu, W., & Zhu, L. (2017). A modified Hosmer–Lemeshow test for large data sets. *Communications in Statistics - Theory and Methods*, 46(23), 11813–11825..