

Pemilihan Metode Terbaik *Support Vector Machine* (SVM) Dan Regresi Logistik Biner Untuk Klasifikasi Status Kemiskinan Rumah Tangga Di Provinsi Lampung Tahun 2019

Rengganis Woro Maharsi^{1*}, dan Sulistyo Hadi^{2*}

¹Program Studi Statistika Fakultas Sains dan Analitika Data Institut Teknologi Sepuluh Nopember (ITS), Surabaya, Indonesia

¹Badan Pusat Statistika (BPS) Kabupaten Tanggamus
Kota Agung, Tanggamus, Lampung

*Email korespondensi: ¹rengganis@bps.go.id, ²sulistyo.hadi@bps.go.id

Abstrak

Metode pengklasifikasian terdiri dari metode nonlinier, yaitu *Support Vector Machine* (SVM), dan metode linier, yaitu Regresi Logistik Biner. Data kemiskinan di Indonesia belum menjadi dasar kebijakan yang tepat sasaran untuk permasalahan pengentasan kemiskinan. Kontraksi Ekonomi ketika Pandemi Covid-19, Provinsi Lampung mengalami peningkatan pertumbuhan penduduk miskin. Pada Maret 2020, terjadi penurunan persentase penduduk miskin sekitar 0,28% dibandingkan Maret 2019. Jumlah penduduk miskin secara absolut ternyata naik, terdapat sekitar 45.464 penduduk miskin baru Tahun 2020 dan 34.553 penduduk miskin baru Tahun 2021. Data yang digunakan bersumber dari Survei Sosial Ekonomi Nasional (Susenas) Tahun 2019. Penelitian ini menggunakan data training (80%) dan testing (20%) dari Susenas 2019. Karakteristik rumah tangga miskin di Provinsi Lampung Tahun 2019 adalah KRT berjenis kelamin perempuan, tidak menamatkan SD, jumlah ART lebih banyak, tinggal di perdesaan, jenis dinding menggunakan Tanah/Bambu/Lainnya, sumber air minum bukan dari air kemasan dan sumur/bor/pompa/ledeng, bekerja di sektor pertanian atau tidak bekerja/penerima pendapatan, tidak memiliki jaminan kesehatan, dan tidak memiliki emas minimal 10 gr. Penelitian ini menunjukkan metode pengklasifikasian terbaik adalah SVM dengan nilai akurasi pada data *testing* sebesar 84,66%, *APER* 15,34%, *Sensitivity* 75%, *Specificity* 84,72%, *AUC* 79,86%, *FP rate* 2,98%, dan *FN rate* 0,18%.

Kata kunci: Klasifikasi, SVM, Logistik, Biner, Kemiskinan, *APER*.

Abstract

The classification method consists of a nonlinear method, namely Support Vector Machine (SVM), and a linear method, namely Binary Logistics Regression. Poverty data in Indonesia has not become the basis for targeted policies for poverty alleviation problems. Economic contraction during the Covid-19 pandemic, Lampung Province experienced an increase in the growth of the poor population. In March 2020, there was a decrease in the percentage of the poor by around 0.28% compared to March 2019. The number of poor people in absolute terms turned out to be increasing, there were around 45,464 new poor people in 2020 and 34,553 new poor people in 2021. The data used were sourced from the Socio-Economic Survey. National (Susenas) 2019. This study uses training (80%) and testing (20%) data from Susenas 2019. The characteristics of poor households in Lampung Province in 2019 are female household heads, not completing elementary school, the number of household members is more, living in rural areas, the type of wall uses soil/bamboo/other, drinking water sources are not from bottled water and wells/drills/pumps/plumbing, work in the agricultural sector or do not work/recipient of income, do not have health insurance, and do not have a minimum of 10 gr gold. This study shows that the best classification method is SVM with an accuracy value of 84.66% on testing data, *APER* 15.34%, *Sensitivity* 75%, *Specificity* 84.72%, *AUC* 79.86%, *FP rate* 2.98%, and *FN rate* 0.18%.

Keywords: Classification, SVM, Logistics, Binary, Poverty, *APER*

1. Pendahuluan

Metode statistik yang digunakan untuk melakukan pengklasifikasian diantaranya *Support Vector Machine* (SVM) dan Regresi Logistik Biner. Metode klasifikasi dapat digunakan untuk membentuk dan mengevaluasi model dengan variabel respon berupa data kategorik [7]. Metode klasifikasi terdiri dari tiga, yaitu metode klasifikasi linier (*Linier Classification Models*), nonlinier (*Nonlinier Classification Models*), dan metode

klasifikasi berbasis pohon dan aturan (Trees/Rule-Based Models). Penelitian ini melakukan perbandingan dua metode yaitu metode klasifikasi linier (Regresi Logistik Biner) dan metode Klasifikasi nonlinier (SVM).

Indonesia saat ini masih memiliki berbagai permasalahan terkait kemiskinan. Masalah utamanya adalah belum ada metode pengklasifikasian status kemiskinan rumah tangga yang memberikan tingkat akurasi yang tinggi, sehingga di masyarakat sering terjadi kesalahan target/sasaran kebijakan terkait pengentasan kemiskinan [18].

Tahun 2019, Indonesia mencatat sejarah baru yaitu pencapaian satu digit persentase kemiskinan yaitu 9,41 persen. Namun, persentase kemiskinan di Provinsi Lampung masih mencapai 12,62 persen pada Tahun 2019. Provinsi Lampung mengalami penurunan persentase kemiskinan Tahun 2019 dan kembali naik Tahun 2020 diakibatkan Pandemi Covid-19. Pandemi Covid-19 memberikan dampak langsung pada penduduk usia kerja. Sekitar 52 ribu pekerja menjadi penganggur dan 549,7 ribu pekerja mengalami pengurangan jam kerja. Tahun 2021, persentase kemiskinan kembali turun karena mulai tersedianya lapangan pekerjaan bagi penduduk, bantuan langsung tunai oleh pemerintah pusat dan daerah, serta membaiknya nilai tukar petani.

Kontraksi Ekonomi pada saat Pandemi Covid-19 di Provinsi Lampung berakibat meningkatnya laju pertumbuhan penduduk miskin di Provinsi Lampung. Pada Maret 2020, terjadi penurunan persentase penduduk miskin sekitar 0,28% dibandingkan Maret 2019. Namun, jumlah penduduk miskin ternyata naik, terdapat sekitar 45.464 penduduk miskin baru tahun 2020 dan 34.553 penduduk miskin baru tahun 2021 (Data diolah Badan Pusat Statistik/BPS, Provinsi Lampung). Pemerintah harus menangkap sinyal “warning” ini, sehingga dapat menentukan arah kebijakan program pengentasan kemiskinan yang tepat sasaran, efektif, dan efisien.

BPS mendefinisikan kemiskinan sebagai ketidakmampuan suatu penduduk dari sisi ekonomi untuk memenuhi kebutuhan dasar makanan dan non makanan yang diukur dengan pendekatan pengeluaran rumah tangga. Status kemiskinan penduduk ditentukan berdasarkan Garis Kemiskinan (GK) yang diperoleh dari pendekatan pengeluaran rumah tangga per kapita dari hasil Susenas. Penduduk Miskin adalah penduduk yang memiliki pengeluaran per kapita berada di bawah GK [2].

Kemiskinan merupakan masalah multidimensional sehingga pemerintah kesulitan dalam menentukan kebijakan yang tepat untuk mengatasinya. Kajian yang mendalam mengenai faktor-faktor yang mempengaruhi status kemiskinan rumah tangga dibutuhkan agar pemerintah dapat memiliki dasar dalam memutuskan suatu program perlu dilaksanakan atau tidak terkait masalah kemiskinan. Variabel yang tidak berhubungan langsung dengan uang (karakteristik perumahan, pendidikan kepala rumah tangga/KRT, kepemilikan aset, dan fasilitas sanitasi) mempengaruhi status kemiskinan penduduk [12]. BPS juga telah melakukan penelitian serupa pada Tahun 2000 dan menentukan terdapat delapan (8) variabel yang kuat mempengaruhi status kemiskinan rumah tangga yaitu, luas lantai per kapita, jenis lantai, ketersediaan air bersih, jenis jamban, kepemilikan aset, pendapatan per bulan, pengeluaran makanan dan konsumsi lauk pauk [3].

Kemiskinan dipengaruhi oleh lima (5) variabel non moneter yaitu pendidikan, jenis pekerjaan, gender, akses terhadap pelayanan kesehatan dasar dan infrastruktur, dan kondisi geografis. Rumah tangga akan beresiko menjadi rumah tangga miskin jika KRT tidak memiliki pendidikan yang memadai, pekerjaan di sektor pertanian, KRT berjenis kelamin perempuan, kurangnya akses terhadap pelayanan kesehatan dasar dan infrastruktur, serta kondisi geografis yang sulit.

Penyelesaian masalah kemiskinan dan pemerataan akan mendukung ekonomi berbasis inovasi dan teknologi di Indonesia. Kunci Indonesia mampu bersaing dengan negara maju seperti China, Korea Selatan, dan Jepang harus mampu mengentaskan kemiskinan terlebih dahulu. Provinsi Lampung harus dapat mengentaskan kemiskinan agar perekonomiannya dapat kembali bertumbuh dan kemiskinan kembali teratasi. Oleh karena itu, diperlukan penelitian untuk pemilihan metode terbaik dan menentukan faktor-faktor yang mempengaruhi status kemiskinan rumah tangga di Provinsi Lampung 2019..

2. Metodologi Penelitian

Beberapa tahapan dan teori dasar yang digunakan dalam penelitian ini adalah sebagai berikut.

2.1. Sumber Data

Penelitian ini menggunakan data sekunder dari Susenas tahun 2019. Penelitian ini tidak menggunakan data tahun 2020 dan 2021 karena menghindari bias pengukuran akibat Pandemi Covid-19. Unit penelitian terdiri dari rumah tangga (ruta) sampel Susenas 2019 di Provinsi Lampung. Data dibagi menjadi data *training* (80%) dan data *testing* (20%).

2.2 Variabel Penelitian

Variabel respon dan variabel prediktor dalam penelitian ini ditunjukkan oleh Tabel 1.

Tabel 1. Variabel Penelitian

Nama Variabel	Skala Pengukuran	Keterangan
(1)	(2)	(3)
Y = Klasifikasi Rumah Tangga Miskin	Nominal	1 : Miskin 0 : Tidak Miskin
X ₁ = Jenis Kelamin KRT	Nominal	1 : Perempuan 0 : Laki-laki
X ₂ = Pendidikan Terakhir KRT	Nominal	1 : PT/SMA/Sederajat 2 : SMP/Sederajat 3 : SD/Sederajat 4 : Tidak Tamat SD
X ₃ = Jumlah Anggota Rumah Tangga (ART)	Rasio jumlah orang
X ₄ = Status Wilayah Tempat Tinggal	Nominal	1 : Perdesaan 0 : Perkotaan
X ₅ = Jenis Atap Rumah Terluas	Nominal	1 : Seng/Asbes 2 : Sirap/Bambu/Ijuk/Rumbia/Lainnya 0 : Beton/Genteng
X ₆ = Jenis Dinding Rumah Terluas	Nominal	1 : Kayu/Bambu/ Lainnya 0 : Tembok
X ₇ = Jenis Lantai Rumah Terluas	Nominal	1 : Ubin/tegel/Teraso/Kayu/ Papan/BataMerah 2 : Tanah/Bambu/Lainnya 0 : Marmer/Granit/Keramik/Parket/ Vinil/Karpet
X ₈ = Status Penggunaan Jamban	Nominal	1 : Bersama 2 : Umum/Tidak ada 0 : Sendiri
X ₉ = Sumber Air Minum Utama	Nominal	1 : Sumur/Bor/Pompa/Ledeng 2 : Lainnya 0 : Air Kemasan/Isi Ulang
X ₁₀ = Sumber Air Mandi/Cuci/dll. Utama	Nominal	1 : Sumur/Bor/Pompa/Ledeng 2 : Lainnya 0 : Air Kemasan/Isi Ulang
X ₁₁ = Lapangan Usaha Pekerjaan Utama KRT	Nominal	1 : Sektor Pertanian 2 : Tidak bekerja/penerima pendapatan 0 : Sektor Non Pertanian
X ₁₂ = Status Kedudukan Pekerjaan Utama KRT	Nominal	1 : Karyawan/pegawai 2 : Pekerja Bebas/pekerja keluarga 0 : Berusaha
X ₁₃ = Status Gangguan Fungsional KRT	Nominal	1 : Punya 0 : Tidak Punya
X ₁₄ = Kepemilikan Jaminan Kesehatan	Nominal	1 : Tidak Punya

		0 : Punya
X ₁₅ = Mobil	Nominal	1 : Tidak Punya
		0 : Punya
X ₁₆ = Tanah/Lahan	Nominal	1 : Tidak Punya
		0 : Punya
X ₁₇ = Emas minimal 10 gr	Nominal	1 : Tidak Punya
		0 : Punya

2.3 Metode Analisis

Analisis deskriptif digunakan untuk menjelaskan karakteristik rumah tangga miskin di Provinsi Lampung Tahun 2019 [13]. Analisis inferensia adalah suatu teknik analisis guna menyimpulkan suatu parameter berdasarkan nilai dari suatu sampel. Pada penelitian ini, analisis inferensia yang akan digunakan adalah uji independensi, regresi logistik biner, *SVM*, uji kelaikan model klasifikasi, dan ukuran ketepatan klasifikasi [6].

Uji Independensi

Penelitian ini akan menggunakan uji *chi-square* untuk menguji independensi antara variabel prediktor dan respon. Hipotesis uji ini adalah $H_0: P_{ij} = P_i \cdot P_j$ dengan statistik uji $\chi^2_{hit} = \sum_{i,j} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$. O_{ij} = nilai observasi / pengamatan baris ke- i kolom ke- j , E_{ij} = nilai ekspektasi baris ke- i kolom ke- j , χ^2_{hit} = Nilai *Chi-Square* hitung (Statistik Uji). Tolak H_0 jika $\chi^2_{hit} > \chi^2_{\alpha(i-1)(j-1)}$ atau $p\text{-value} \leq \alpha$. Jika tolak H_0 , maka dengan tingkat kepercayaan $(1 - \alpha)100\%$ cukup bukti bahwa antara variabel prediktor dan variabel respon memiliki hubungan.

Support Vector Machine (SVM)

SVM adalah metode klasifikasi data dengan mendapatkan fungsi pemisah (*hyperplane* / $\vec{X}_i^T \vec{\theta} + b = 0$) antar kelas. Variabel respon (y_i) pada *SVM* dikategorikan menjadi dua, yaitu -1 untuk satu kelas dan +1 untuk kelas lainnya. \vec{X}_i^T tidak melibatkan elemen pertama, yaitu vektor bernilai 1 dan merupakan *intercept* dalam regresi logistik biner. $\vec{\theta}$ adalah vektor penimbang. Interpretasi $\vec{\theta}$ akan sama dengan $\vec{\beta}$ pada regresi. Semua data akan terpisah menjadi dua kelas jika memenuhi kondisi sebagai berikut [19]:

$$\vec{X}_i^T \vec{\theta} + b \leq -1, \text{ untuk } y_i = -1$$

$$\vec{X}_i^T \vec{\theta} + b \geq +1, \text{ untuk } y_i = +1$$

Kedua *hyperlane* tersebut sejajar dan memiliki jarak $d_+ = d_- = \frac{1}{\|\vec{\theta}\|}$, dengan $\|\vec{\theta}\| = \sqrt{(\theta_1^2 + \theta_2^2 + \dots + \theta_p^2)}$. Akibatnya, maksimal *margin* = $\frac{2}{\|\vec{\theta}\|}$ dan *hyperplane* terbaik adalah dengan optimasi fungsi permasalahan $\min \frac{1}{2} \|\vec{\theta}\|^2$. Batasan fungsi tersebut adalah $y_i(\vec{X}_i^T \vec{\theta} + b) - 1 \geq 0, i = 1, 2, \dots, n$ [19].

Regresi Logistik Biner

Variabel respon pada *binary logistic* dikategorikan menjadi dua, yaitu $y = 1$ “Ruta Miskin” dan $y = 0$ “Ruta Tidak Miskin”. Variabel respon akan mengikuti distribusi Bernoulli untuk setiap observasi tunggal, sehingga fungsi probabilitasnya adalah $f(y) = \pi^y(1 - \pi)^{1-y}, y = 0, 1$. Ketika $y = 0$ “Ruta Tidak Miskin”, maka $f(y) = 1 - \pi$ dan jika $y = 1$ “Ruta Miskin” maka $f(y) = \pi$. Selanjutnya, Hosmer dan Lemeshow (2000) menuliskan fungsi *binary logistic* dengan banyak variabel prediktor sebesar p adalah

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p}}{1 + e^{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p}}.$$

Uji Serentak dan Parsial Koefisien Regresi Logistik Biner

Uji serentak menggunakan statistik uji $G = -2 \ln \frac{\left(\frac{n_1}{n}\right)^{n_1} \left(\frac{n_0}{n}\right)^{n_0}}{\sum_{i=1}^n \pi^{y_i} (1 - \pi)^{1-y_i}}$, dengan $n_1 = \sum_{i=1}^n y_i, n_0 = \sum_{i=1}^n (1 - y_i), n = n_1 + n_0$. Hipotesis yang diuji adalah $H_0: \beta_1 = \beta_2 = \dots = \beta_p = 0$. Tolak H_0 jika nilai $G > \chi^2_{\alpha(p)}$. Jika tolak H_0 , maka dengan tingkat kepercayaan $(1 - \alpha)100\%$ terbukti bahwa secara bersama-sama variabel prediktor signifikan mempengaruhi variabel respon. Uji parsial menggunakan statistik uji Wald yaitu $W =$

$\frac{\hat{\beta}_j}{se(\hat{\beta}_j)}$, dengan $\hat{\beta}_j$ adalah koefisien regresi tiap variabel prediktor, dan $se(\hat{\beta}_j)$ adalah standar error dari koefisien regresi. Hipotesis yang diuji adalah $H_0 : \beta_j = 0$. Tolak H_0 jika nilai $W^2 > \chi^2_{\alpha(1)}$. Jika tolak H_0 , maka dengan tingkat kepercayaan $(1 - \alpha)100\%$ terbukti bahwa variabel prediktor ke-j signifikan mempengaruhi variabel respon.

Pengujian Kesesuaian Model

Dalam regresi logistic dikenal *Hosmer-Lemeshow Goodness-of-fit test* untuk menguji kesesuaian model yang terbentuk. Statistik uji yang digunakan adalah $\hat{C} = \sum_{k=1}^g \frac{(O_k - n'_k \bar{\pi}_k)^2}{n'_k \bar{\pi}_k (1 - \bar{\pi}_k)}$, dengan O_k = Observasi pada grup ke-k ($\sum_{j=1}^{C_k} y_i$ dengan C_k : respon (0,1)), $\bar{\pi}_k$ = Rata-rata taksiran peluang ($\sum_{j=1}^{C_k} \frac{m_j \hat{\pi}_j}{n'_k}$), n'_k = Banyak observasi pada grup ke-k, g = Jumlah grup (kombinasi kategori dalam model serentak). Tolak H_0 jika nilai $\hat{C} > \chi^2_{\alpha(db)}$. Jika gagal tolak H_0 , dengan tingkat kepercayaan $(1 - \alpha)100\%$ terbukti bahwa model yang terbentuk telah sesuai dengan hasil pengamatan [6].

Pemilihan kriteria kebaikan model ada beberapa indikator yang dapat digunakan diantaranya *sensitivity*, *spesifitas*, keakuratan, *False Positif Rate (FP Rate)*, *False Negative Rate (FN Rate)*, dan *APER (Apparent Error Rate)*. Jika dilihat dengan grafik ada dua indikator yaitu *Receiver Operator Characteristic (ROC)* dan *Area Under the Receiver Operating Characteristics Curve (AUC)*. ROC adalah plot dari *true positive rate* dengan *true negative rate* sebagai fungsi dari model batas untuk pengklasifikasian kelas positif. AUC adalah metrik untuk menghitung keseluruhan kinerja suatu model klasifikasi berdasarkan luas daerah di bawah kurva ROC sehingga nilai AUC tersebut berada dalam interval 0 sampai 1. Semakin tinggi AUC, semakin baik model dalam memprediksi 0 kelas sebagai 0 dan 1 kelas sebagai 1.

3. Hasil dan Pembahasan

3.1 Analisis Deskriptif

Lampung Utara merupakan kabupaten dengan angka kemiskinan tertinggi hingga 2021 [16]. Kemiskinan di Lampung Utara mencapai 19,63%, dengan kedalaman kemiskinan mencapai 3,3 poin, dan tingkat keparahan kemiskinan mencapai 0,82 poin. Kemiskinan di Lampung Utara merupakan masalah yang kompleks. Perekonomian Lampung Utara 36,90% masih ditopang oleh Lapangan Usaha Sektor Pertanian, namun mayoritas petani berstatus pekerja bebas tanpa memiliki lahan pertanian sendiri. Akibatnya, daya beli dan pendapatan rumah tangga sektor pertanian akan rendah. Harga-harga untuk produk pertanian seperti singkong, sawit, dan karet di Lampung Utara juga tergolong rendah, sehingga tidak mampu mendongkrak pendapatan yang diterima oleh petani.

Karakteristik rumah tangga miskin di Provinsi Lampung 14,1% tinggal di wilayah perkotaan, 14,4% KRT berjenis kelamin laki-laki, 10,1% KRT berpendidikan maksimal SD, 11,2% bekerja di sektor pertanian, 5,9% berstatus buruh atau pekerja bebas, 3% KRT memiliki gangguan fungsional, 6,9 persen tidak memiliki jaminan kesehatan. Berdasarkan kelayakan perumahan, rumah tangga miskin di Provinsi Lampung memiliki karakteristik 12,6% beratap beton/genteng, 5,9% berdinding kayu/bambu/lainnya, 13,6 berlantai bukan marmer/granit/keramik/ parket/vinil/karpet, 3% masih menggunakan jamban bersama, 0,3% sumber air minumnya tidak dari air kemasan/isi ulang/sumur/bor/pompa/ledeng, 0,6% masih mandi/cuci/dll di sungai dan sejenisnya. Berdasarkan kepemilikan aset, rumah tangga miskin di Provinsi Lampung memiliki karakteristik 0,3% memiliki tabungan emas minimal 10 gr, 0,1% memiliki mobil, dan 14,1% memiliki tanah/lahan.

3.2 Analisis Inferensia

3.2.1 Uji Independensi

Uji independensi digunakan untuk melihat apakah ada hubungan antara variabel prediktor dengan variabel respon. Statistik uji yang digunakan adalah statistik uji Khi Kuadrat [6]. Tabel 2 menunjukkan Status Gangguan Fungsional KRT dan Kepemilikan Tanah/Lahan tidak cukup bukti untuk menyatakan bahwa memiliki hubungan dengan status kemiskinan rumah tangga. Sehingga, semua variabel prediktor dapat digunakan untuk dilakukan analisis klasifikasi dengan regresi logistik biner, kecuali kedua variabel tersebut.

Tabel 2. Uji Independensi variabel prediktor terhadap variabel respon

Variabel	χ^2	<i>p-value</i>	Kesimpulan
(1)	(2)	(3)	(4)
Jenis Kelamin KRT	10,885	0,001	Ada hubungan
Pendidikan Terakhir KRT	116,345	0,000	Ada hubungan
Status Wilayah Tempat Tinggal	166,209	0,000	Ada hubungan
Jenis Atap Rumah Terluas	6,150	0,046	Ada Hubungan
Jenis Dinding Rumah Terluas	167,795	0,000	Ada Hubungan
Jenis Lantai Rumah Terluas	252,694	0,000	Ada Hubungan
Status Penggunaan Jamban	68,086	0,000	Ada Hubungan
Sumber Air Minum Utama	146,205	0,000	Ada Hubungan
Sumber Air Mandi/Cuci/dll. Utama	8,913	0,012	Ada Hubungan
Lapangan Usaha Pekerjaan Utama KRT	79,985	0,000	Ada hubungan
Status Kedudukan Pekerjaan Utama KRT	69,707	0,000	Ada hubungan
Status Gangguan Fungsional KRT	0,258	0,612	Tidak ada hubungan
Kepemilikan Jaminan Kesehatan	3,229	0,072	Ada Hubungan
Kepemilikan Mobil	128,304	0,000	Ada Hubungan
Kepemilikan Tanah/Lahan	0,0248	0,619	Tidak ada hubungan
Kepemilikan Emas minimal 10 gr	97,270	0,000	Ada Hubungan

3.2.2 Analisis Regresi Logistik Biner

3.2.2.1 Uji Serentak dan Parsial

Hasil uji serentak dengan uji *Omnibus Test of Model Coefficients* (Khai Kuadrat hitung) adalah variabel-variabel prediktor mempengaruhi secara signifikan pada klasifikasi status kemiskinan rumah tangga [6]. Kemudian, dilanjutkan uji parsial dengan statistik Uji *Wald* menggunakan Metode *Enter* (SPSS).

Tabel 3. Uji Parsial Variabel Prediktor Terhadap Variabel Respon

Variabel	Wald Hitung	p-value	<i>Odds Ratio</i>
(1)	(2)	(3)	(4)
Jenis Kelamin KRT	4,061	0,044	1,300
Pendidikan Terakhir KRT (Tidak Tamat SD)	7,974	0,005	1,388
(SD/Sederajat)	3,449	0,063	1,221
(SMP/Sederajat)	10,665	0,001	1,471
Jumlah Anggota Rumah Tangga (ART)	543,017	0,000	1,804
Status Wilayah Tempat Tinggal	57,089	0,000	2,407
Jenis Atap Rumah Terluas (Seng/Asbes)	1,971	0,160	0,873
(Sirap/Bambu/Ijuk/ Rumbia/Lainnya)	0,960	0,327	1,6282
Jenis Dinding Rumah Terluas (Ubin/tegel/Teraso/ Kayu/Papan/Bata Merah)	52,575	0,000	2,091
(Tanah/Bambu/Lainnya)	27,695	0,000	2,370
Status Penggunaan Jamban (Bersama)	25,005	0,000	1,623
Sumber Air Minum Utama	51,401	0,000	2,351
(Sumur/Bor/Pompa/Ledeng)			
(Lainnya)	13,173	0,000	3,294
Sumber Air Mandi/Cuci/dll. Utama	0,000	1,000	
(Sumur/Bor/Pompa/Ledeng)	0,000	1,000	28834341,48
(Lainnya)			22458892,70
Lapangan Usaha Pekerjaan Utama KRT	9,107	0,002	1,298
Status Kedudukan Pekerjaan Utama KRT	0,483	0,487	1,069
(Karyawan/Pegawai)			
(Tidak Bekerja/Penerima Pendapatan)	24,336	0,000	1,636
Kepemilikan Jaminan Kesehatan	4,726	0,030	1,167
Kepemilikan Mobil	2,212	0,137	1,197
Kepemilikan Emas minimal 10 gr	31,515	0,000	7,949

Tabel 3 mengidentifikasi dengan tingkat kesalahan 5%, determinan penentu klasifikasi kemiskinan rumah tangga adalah Jenis Kelamin KRT, Pendidikan Terakhir KRT, Jumlah ART, Status Wilayah Tempat Tinggal, Jenis Dinding Terluas, Sumber Air Minum Utama, Lapangan Usaha Pekerjaan Utama KRT, Status Kedudukan Pekerjaan Utama KRT, Kepemilikan Jaminan Kesehatan, dan Kepemilikan Emas minimal 10 gr. Karakteristik yang beresiko menjadi rumah tangga miskin di Provinsi Lampung Tahun 2019 adalah KRT dengan Jenis kelamin perempuan, tidak menamatkan SD, memiliki jumlah ART lebih banyak, tinggal di perdesaan, jenis dinding menggunakan Tanah/Bambu/Lainnya, sumber air minumannya bukan dari air kemasan dan sumur/bor/pompa/ledeng, bekerja di sektor pertanian atau tidak bekerja/penerima pendapatan, tidak memiliki jaminan kesehatan, dan tidak memiliki emas minimal 10 gr.

3.2.2.2 Uji Kelayakan Model Dan Penentuan Ketepatan Klasifikasi Dengan Regresi Logistik Biner

Uji kelayakan model regresi logistik biner diperoleh dengan menggunakan Hosmer and Lemeshow Test dengan pendekatan statistik uji Khai Kuadrat [6]. Berdasarkan *output* SPSS, menunjukkan bahwa $\chi^2 = 5,844$ dan $p\text{-value}$ (0,665) > 0,05. Sehingga, model regresi logistik biner yang terbentuk sesuai dengan data pengamatan.

Tabel 4. Hasil Klasifikasi Status Kemiskinan Data Training

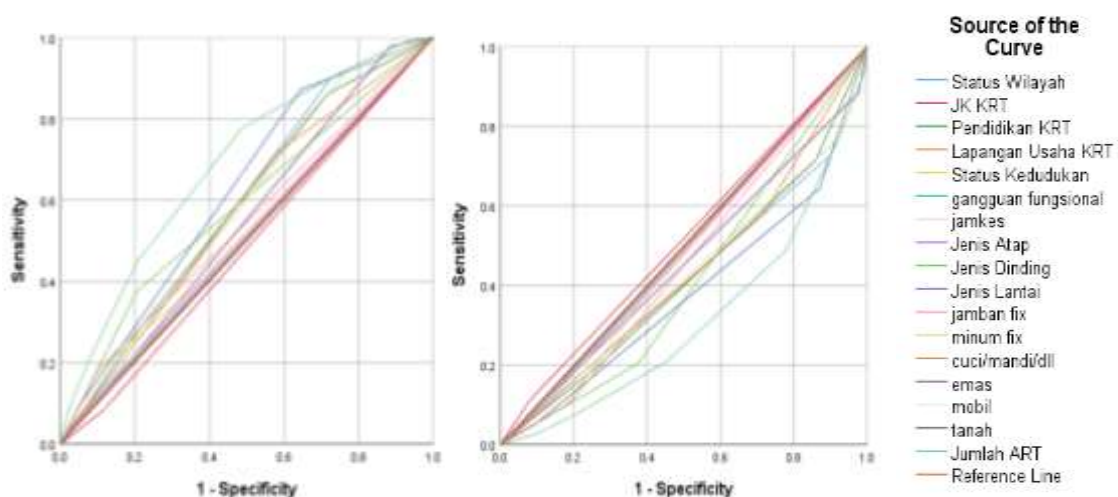
Observed		Predicted		
		Status Kemiskinan Ruta		Persentase Benar
		Tidak Miskin	Miskin	
Status Kemiskinan Ruta	Tidak Miskin	6391	123	98,1
	Miskin	1023	186	15,4
Overall Percentage				85,2

Tabel 5. Hasil Klasifikasi Status Kemiskinan Data Testing

Observed		Predicted		
		Status Kemiskinan Ruta		Persentase Benar
		Tidak Miskin	Miskin	
Status Kemiskinan Ruta	Tidak Miskin	1598	254	86,28
	Miskin	30	48	61,54
Overall Percentage				85,29

Berdasarkan Tabel 4 dan Tabel 5, tabel klasifikasi, model regresi logistik biner mampu memprediksi status kemiskinan rumah tangga dengan benar sebesar 85,2% pada data training dan 85,29% pada data testing. Artinya, model klasifikasi status kemiskinan rumah tangga di Provinsi Lampung Tahun 2019 mampu mengklasifikasikan dengan benar dengan tingkat akurasi sebesar 85,29%.

Nilai *Apparent Error Rate* (APER) data testing pada model sebesar 14,71%. Nilai ini mengindikasikan terdapat 14 – 15 rumah tangga yang salah pengklasifikasian status kemiskinnannya per 100 rumah tangga. *Specificity* model ini mencapai 98,11% (sangat baik), tapi *sensitivity* sangat rendah hanya 15,38%. *FP Rate* pada data *training* sebesar 1,89% dan data *testing* sebesar 15,90%. *FN Rate* pada data *training* sebesar 84,62% dan data *testing* sebesar 38,46%.



Gambar 3. Kurva ROC Prediksi Ruta Miskin dan Tidak Miskin pada Data Testing

Tabel 6. Nilai AUC tiap Variabel Prediktor berdasarkan Variabel Respon

Variabel	Nilai AUC		Variabel	Nilai AUC	
	Ruta Miskin	Ruta Tidak Miskin		Ruta Miskin	Ruta Tidak Miskin
(1)	(2)	(3)	(1)	(2)	(3)
Jenis Kelamin KRT	0,484	0,516	Mandi/Cuci/dll. Utama		
Pendidikan Terakhir KRT	0,584	0,416	Lapangan Usaha		
Status Wilayah Tempat Tinggal	0,587	0,413	Pekerjaan Utama KRT	0,569	0,431
Jenis Atap Rumah			Status Kedudukan		
Terluas	0,512	0,488	Pekerjaan Utama KRT	0,526	0,474
Jenis Dinding Rumah			Status Gangguan Fungsional KRT	0,503	0,497
Terluas	0,586	0,414	Kepemilikan Jaminan		
Jenis Lantai Rumah			Kesehatan	0,514	0,486
Terluas	0,622	0,378	Kepemilikan Mobil	0,552	0,448
Status Penggunaan Jamban	0,542	0,458	Kepemilikan Tanah/Lahan	0,502	0,498
Sumber Air Minum Utama	0,579	0,421	Kepemilikan Emas minimal 10 gr	0,546	0,454
Sumber Air	0,507	0,493	Jumlah ART	0,687	0,313

Berdasarkan Gambar 3 dan Tabel 6, variabel prediktor tidak ada yang dominan mampu mengklasifikasikan rumah tangga di Provinsi Lampung sebagai ruta tidak miskin karena semua AUC bernilai kurang dari 0,5. Selain jenis kelamin KRT, semua variabel prediktor memiliki nilai AUC di atas 0,5 lebih dalam mengklasifikasikan status rumah tangga di Provinsi Lampung sebagai rumah tangga miskin.

Penelitian ini belum menghasilkan nilai AUC yang cukup untuk membentuk suatu model klasifikasi status kemiskinan rumah tangga. Jumlah Art dan jenis lantai adalah variabel yang memiliki nilai AUC di atas 0,6 yang menunjukkan performa model yang terbentuk masuk kategori *Poor* atau lemah. Variabel prediktor lainnya masuk kategori *Fail* atau gagal [9].

Nilai AUC data *training* dari model Regresi Logistik Biner mencapai 56,75 artinya model ini berkategori *Poor* atau lemah dalam mengklasifikasikan status kemiskinan rumah tangga di Provinsi Lampung tahun 2019. Sedangkan, Nilai AUC data *testing* dari model Regresi Logistik Biner mencapai 73,91 artinya model ini berkategori *Fair* atau cukup dalam mengklasifikasikan status kemiskinan rumah tangga di Provinsi Lampung tahun 2019 [9].

3.2.3 Analisis Support Vector Machine (SVM)

Berdasarkan Tabel 7 dan Tabel 8, tabel klasifikasi, model SVM mampu memprediksi status kemiskinan rumah tangga dengan benar sebesar 85,14% pada data *training* dan 84,66% pada data *testing*. Artinya, model SVM klasifikasi status kemiskinan ruta di Provinsi Lampung tahun 2019 mampu mengklasifikasikan dengan benar tingkat akurasi sebesar 84,66%.

Tabel 7. Hasil Klasifikasi Status Kemiskinan Data Training Metode SVM

Observed		Predicted		
		Status Kemiskinan Ruta		Persentase Benar
		Tidak Miskin	Miskin	
Status Kemiskinan Ruta	Tidak Miskin	6505	1139	85,10
	Miskin	9	70	88,61
Overall Percentage				85,14

Tabel 8. Hasil Klasifikasi Status Kemiskinan Data Testing Metode SVM

Observed		Predicted		
		Status Kemiskinan Ruta		Persentase Benar
		Tidak Miskin	Miskin	
Status Kemiskinan Ruta	Tidak Miskin	1625	293	84,72
	Miskin	3	9	0,75
Overall Percentage				84,66

Nilai *Apparent Error Rate* (APER) data *testing* pada model sebesar 15,34%. Nilai ini mengindikasikan terdapat 15 – 16 rumah tangga yang salah pengklasifikasian status kemiskinannya per 100 rumah tangga. *Specificity* model ini mencapai 85,10% dan *sensitivity* juga baik sebesar 88,61%. *FP Rate* pada data *training* sebesar 5,79% dan data *testing* sebesar 15,28%. *FN Rate* pada data *training* sebesar 0,14% dan data *testing* sebesar 25,00%.

Nilai AUC data *training* dari model SVM mencapai 86,85% artinya model ini berkategori *Good* atau baik dalam mengklasifikasikan status kemiskinan rumah tangga di Provinsi Lampung tahun 2019. Sebaliknya, Nilai AUC dari model SVM mencapai 79,86, artinya model SVM berkategori *Fair* atau cukup dalam mengklasifikasikan status kemiskinan ruta di Provinsi Lampung tahun 2019 [9].

3.2.4 Pemilihan Model Terbaik

Tabel 9 menunjukkan bahwa akurasi dan APER regresi logistik biner lebih baik dibandingkan SVM, namun perbedaannya hanya 0,63%. SVM memiliki *sensitivity* lebih baik, namun regresi logistik biner memiliki *specificity* lebih baik. Regresi logistik biner lebih baik dalam mengklasifikasikan dengan benar rumah tangga tidak miskin. SVM lebih baik dalam mengklasifikasikan dengan benar rumah tangga miskin. Pada penelitian ini terjadi *imbalanced* proporsi ruta miskin dan tidak miskin sehingga regresi logistik biner menghasilkan kesimpulan yang bias. Oleh karena itu, metode SVM memiliki peluang lebih kecil dalam kesalahan mengklasifikasikan status kemiskinan ruta di Provinsi Lampung.

Tabel 9. Perbandingan Ketepatan Klasifikasi SVM dan Regresi Logistik Biner Data Testing

Indikator	Nilai SVM	Nilai Regresi Logistik Biner	Keputusan
(1)	(2)	(3)	(4)
Akurasi	84,66	85,29	RegLog
APER	15,34	14,71	RegLog
Sensitivity	75,00	61,84	SVM
specificity	84,72	86,28	RegLog
FP Rate	15,28	13,71	RegLog
FN Rate	25,00	38,46	SVM
AUC	79,86	73,91	SVM

4. Kesimpulan

Metode terbaik dalam mengklasifikasikan status kemiskinan rumah tangga di Provinsi Lampung Tahun 2019 adalah Metode Klasifikasi Non-Linier yaitu SVM. Namun, metode ini masih memberikan nilai APER yang cukup tinggi yaitu 15,34%. Klasifikasi rumah tangga yang lebih beresiko untuk menjadi rumah tangga miskin di Provinsi Lampung adalah KRT berjenis kelamin perempuan, pendidikan terakhir KRT maksimal adalah SMP, jumlah ART lebih banyak, wilayah tempat tinggalnya adalah perdesaan, jenis dinding selain tembok, fasilitas jamban bersama, sumber air minum selain air mineral, sektor pekerjaan utama adalah pertanian, status kedudukan pekerjaan adalah tidak bekerja/penerima pendapatan, tidak memiliki jaminan kesehatan, dan tidak memiliki tabungan emas minimal 10gram.

Nilai APER, FP rate, FN rate yang tinggi kemungkinan disebabkan karena adanya “*imbalanced*” data pada variabel status kemiskinan rumah tangga. Baik regresi logistik biner maupun SVM belum mampu mengatasi kasus *imbalanced* data. Oleh karena itu, sebelum dilakukan pembentukan model maka lebih baik dilakukan *oversampling* ataupun *undersampling* yang berguna untuk menyeimbangkan data menurut variabel respon (status kemiskinan). Nilai AUC masih tergolong lemah dan gagal, maka lebih baik digunakan metode *Entropy Based Fuzzy Support Vector Machine (EFSVM)* baik untuk linier dan nonlinier data

Daftar Pustaka:

- [1] Badan Pusat Statistik Provinsi Lampung. 2019. *Lampung Dalam Angka 2019*. Bandar Lampung: Badan Pusat Statistik Provinsi Lampung.
- [2] BPS. 2016. *Penghitungan dan Analisis Kemiskinan Makro Indonesia 2016*. Jakarta: Badan Pusat Statistik.
- [3] BPS. 2020. *Penghitungan dan Analisis Kemiskinan Makro Indonesia Tahun 2020*. Jakarta: Badan Pusat Statistik.

- Statistik.
- [4] Fan, Q., Wang, Z., Li, D., Gao, D., Zha, H. 2017. *Entropy-based fuzzy support vector machine for imbalanced datasets*. *Knowledge-Based Systems*, 115, hal. 87–99.
 - [5] Harumeka, Ajiwasesa. 2021. *Analisis Entropy Based Fuzzy Membership Sebagai Penimbang Pada Weighted Logistic Regression Untuk Klasifikasi Data Tidak Seimbang (Studi Kasus: Klasifikasi Rumah Tangga Miskin Di Jawa Timur)*. Departemen Statistika, ITS Surabaya.
 - [6] Hoshmer, D. W. dan Lemeshow, S. 2000. *Applied Logistic Regression*, Second Edition. John Wiley & Sons, Inc.
 - [7] Kuhn, M. dan Johnson, K. 2013. *Applied Predictive Modeling*. New York: Springer Science+Business Media.
 - [8] Maalouf, M. dan Siddiqi, M. 2014. *Weighted logistic regression for large-scale imbalanced and rare events data*. *Knowledge-Based Systems*, 59, hal.142–148.
 - [9] Mohanty, A. K., Beberta, S. dan Lenka, S. K. 2011. *Classifying Benign and Malignant Mass using GLCM and GLRLM based Texture Features from Mammogram*. *International Journal of Engineering Research and Applications (IJERA)*, 1(3), hal. 687–693.
 - [10] Octaviani, Pusphita Anna, dkk. 2014. *Penerapan Metode Klasifikasi Support Vector Machine (SVM) Pada Data Akreditasi Sekolah Dasar (SD) Di Kabupaten Magelang*, *Jurnal Gaussian*, Volume 3, Nomor 4, Tahun 2014, Halaman 811 – 820.
 - [11] Purwa, T. 2019. *Perbandingan Metode Regresi Logistik dan Random Forest untuk Klasifikasi Data Imbalanced (Studi Kasus: Klasifikasi Rumah Tangga Miskin di Kabupaten Karangasem, Bali Tahun 2017)*. *Jurnal Matematika, Statistika dan Komputasi*, 16(1), hal. 58.
 - [12] Setyobudi, B., Herwiyanti, E., Azhar, S., Rafinda, A. (2017) “Pengaruh Insentif Moneter Dan Non-Moneter Terhadap Kinerja Individual Peran Orientasi Individu Terhadap Uang,” *Sustainable Competitive Advantages*, 7(September), hal. 592–602.
 - [13] Sugiyono. 2013. *Metode Penelitian Manajemen: Pendekatan, Kuantitatif, Kualitatif, Kombinasi (Mixed Methods), Penelitian Tindakan (Action Research), Penelitian Evaluasi*. Bandung. hal.81-86.
 - [14] Kusnandar, VB. 2021. *Kemiskinan Lampung Meningkat Selama Pandemi Covid-19*. [Online]. Available at <https://databoks.katadata.co.id/datapublish/2021/07/29/kemiskinan-lampung-meningkat-selama-pandemi-covid-19> [Accessed 1 April 2022].
 - [15] Ulya, FN. *Kunci Mengentaskan Kemiskinan, Indonesia Harus Contoh 3 Negara Ini*. [Online]. Available at <https://money.kompas.com/read/2021/03/16/204500526/kunci-mengentaskan-kemiskinan-indonesia-harus-contoh-3-negara-ini-> [Accessed 1 April 2022].
 - [16] Napitupulu, J. 2022. *Kabupaten Termiskin di Lampung Menurut BPS Tahun 2021*. [Online]. Available at <https://kirka.co/kabupaten-termiskin-di-lampung-menurut-bps-tahun-2021/> [Accessed 1 April 2022].
 - [17] Dinas Kominfotik Provinsi Lampung. 2017. *Pengentasan Kemiskinan Melalui KUBE di Provinsi Lampung*. [Online]. Available at <https://lampungprov.go.id/detail-post/pengentasan-kemiskinan-melalui-kube-di-provinsi-lampung> [Accessed 1 April 2022].
 - [18] Gloria. 2021. *Pakar UGM: Mentalitas Miskin Buat Bansos Sering Salah Sasaran*. [Online]. Available at <https://ugm.ac.id/id/berita/21990-pakar-ugm-mentalitas-miskin-buat-bansos-sering-salah-sasaran> [Accessed 28 Maret 2022]
 - [19] Vapnik, N.N. 2000. *The Nature of Statistical Learning Theory*. Second Edition. New York. Springer-Verlag New York Inc.