

Penerapan Metode *Rare Event Weighted Logistic Regression* untuk Kasus Data *Imbalanced* (Studi Kasus: Jenis Perceraian di Kabupaten Bojonegoro)

Cinde Pristi Kurnia Merdiko dan Vita Ratnasari
Departemen Statistika, Institut Teknologi Sepuluh Nopember (ITS)
e-mail: vita_ratna@statistika.its.ac.id

Abstrak—Jawa Timur menjadi provinsi penyumbang angka perceraian tertinggi di Indonesia beberapa tahun terakhir. Pada tahun 2019, jumlah kasus perceraian di Jawa Timur meningkat sebesar 6.052 kasus dari tahun sebelumnya dan menyumbang 21,6% perceraian di Indonesia. Kabupaten Bojonegoro menjadi salah satu penyumbang tingginya angka perceraian di Jawa Timur. Perceraian menurut kasusnya dibagi menjadi dua yaitu cerai gugat dan cerai talak sehingga salah satu metode yang bisa digunakan untuk mengidentifikasi faktor signifikan penyebab perceraian adalah metode regresi logistik. Akan tetapi, kasus perceraian di Kabupaten Bojonegoro tahun 2020 merupakan data *imbalanced* dengan perbandingan cerai gugat dan cerai talak adalah 71,4% dan 28,6% sehingga pada penelitian ini digunakan metode pengembangan regresi logistik untuk data *imbalanced* yaitu *Rare Event Weighted Logistic Regression (RE-WLR)*. Hasil penelitian ini menunjukkan bahwa usia penggugat atau pemohon dan tergugat atau termohon ketika menikah, pekerjaan penggugat atau permohon dan tergugat atau termohon, serta sebab perceraian berpengaruh signifikan terhadap jenis perceraian di Kabupaten Bojonegoro dengan nilai ketepatan klasifikasi AUC yang dihasilkan sebesar 85,2% dan termasuk dalam *Good Classification*.

Kata Kunci—Jenis Perceraian, *Rare Event Weighted Logistic Regression (RE-WLR)*.

I. PENDAHULUAN

PERCERAIAN merupakan pengakhiran suatu perkawinan karena sesuatu sebab dengan keputusan hakim atas tuntutan dari salah satu pihak atau kedua belah pihak dalam perkawinan [1]. Putusnya perkawinan yang disebabkan karena perceraian dapat terjadi karena dua hal, yaitu talak (suami menjatuhkan cerai kepada istri) ataupun berdasarkan gugatan perceraian dari istri kepada suami. Terdapat beberapa alasan yang mengakibatkan perceraian di antaranya adalah zina, meninggalkan tempat tinggal bersama dengan itikad jahat selama 5 tahun, mendapat hukuman penjara 5 tahun atau lebih karena dipersalahkan melakukan kejahatan, dan penganiayaan berat yang dilakukan suami terhadap istri atau sebaliknya.

Kasus perceraian di Indonesia tahun 2014 hingga 2019 terus mengalami peningkatan. Salah satu provinsi yang menjadi penyumbang besar kasus perceraian di Indonesia adalah Provinsi Jawa Timur. Jawa Timur menjadi provinsi pertama dengan jumlah kasus perceraian tertinggi di Indonesia mulai tahun 2014 hingga 2018 [2]. Pada tahun 2019, Jawa timur menyumbang 21,6% dari total kasus perceraian di Indonesia. Tingginya kasus perceraian di Jawa Timur diikuti dengan meningkatnya kasus perceraian pada beberapa kabupaten/kota, salah satunya adalah Kabupaten

Bojonegoro. Kasus perceraian di Kabupaten Bojonegoro mengalami peningkatan dari tahun 2018 dan pada tahun 2020 terdapat 2.577 kasus perceraian dimana 56,3% penyebabnya adalah masalah ekonomi.

Dalam upaya menurunkan jumlah kasus perceraian yang terjadi di Kabupaten Bojonegoro, salah satu cara yang dapat dilakukan adalah dengan mengidentifikasi faktor-faktor yang menyebabkan terjadinya perceraian. Beberapa penelitian tentang kasus perceraian telah dilakukan sebelumnya, penelitian terakhir adalah penelitian Nisva tahun 2020 yang memodelkan faktor yang mempengaruhi jenis perceraian di Kabupaten Lumajang menggunakan Regresi Logistik Biner [3]. Penelitian tersebut menghasilkan 7 variabel penyebab jenis perceraian yang signifikan dari 10 variabel yang digunakan. Namun pada penelitian tersebut dihasilkan nilai *specificity* yang tinggi sementara nilai *sensitivity* yang relatif rendah yakni secara berurutan sebesar 0,9 dan 0,7. Hal ini diduga disebabkan oleh penggunaan metode regresi logistik biner tanpa memperhatikan data *imbalanced* sehingga dapat mengurangi ketepatan klasifikasi pada kelas minoritas.

Penelitian ini menggunakan data kasus perceraian dan faktor-faktor yang diduga mempengaruhi jenis perceraian di Kabupaten Bojonegoro tahun 2020 dimana kasus perceraianya terdiri dari 71,4% cerai gugat dan 28,6% cerai talak [4]. Hal tersebut mengindikasikan adanya ketidakseimbangan data atau termasuk dalam kategori data *imbalanced*. Salah satu penanganan kasus data *imbalanced* pada regresi logistik adalah menggunakan metode *Rare Event Weighted Logistic Regression ((RE-WLR)*. RE-WLR merupakan pengembangan metode regresi logistik dengan melakukan regularisasi, pembobotan, dan koreksi bias sehingga dapat dihasilkan model yang baik untuk data yang *imbalanced* [5]. Penelitian ini akan menggunakan metode RE-WLR dan hasil dari penelitian ini diharapkan dapat menjadi dasar terkait upaya menurunkan jumlah kasus perceraian di Kabupaten Bojonegoro berdasarkan karakteristik dan faktor signifikan penyebab jenis perceraian di Kabupaten Bojonegoro.

II. TINJAUAN PUSTAKA

A. Statistika Deskriptif

Statistika deskriptif merupakan metode hanya memberikan informasi karakteristik suatu keadaan secara deskriptif tanpa mengambil suatu keputusan atau menarik kesimpulan (inferensia) dari data atau keadaan yang sedang diamati [6]. Statistika deskriptif juga dapat berupa penyajian data seperti

diagram dan grafik.

B. Regresi Logistik Biner

Regresi logistik biner merupakan suatu metode analisis data yang digunakan untuk mencari hubungan antara variabel respon (y) yang bersifat biner atau dikotomis dengan variabel prediktor (x) yang bersifat kualitatif, kuantitatif, atau kombinasi dari keduanya [7]. *Outcome* dari variabel respon y mengikuti distribusi binomial dengan p adalah banyaknya variabel prediktor untuk n pengamatan. Fungsi probabilitas untuk setiap observasi dituliskan pada persamaan 1.

$$f(y_i) = \pi_i^{y_i}(1 - \pi_i)^{1-y_i}; y = 0,1 \quad (1)$$

Dimana ketika $y = 0$ maka $f(y) = 1 - \pi$ dan ketika $y = 1$ maka $f(y) = \pi$. Model regresi logistik dan bentuk logitnya secara berurutan dapat dituliskan pada persamaan 2 dan 3.

$$\pi(x_i) = \pi_i = \frac{e^{(\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip})}}{1 + e^{(\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip})}} = \frac{e^{x_i^T \beta}}{1 + e^{x_i^T \beta}} = \frac{1}{1 + e^{-x_i^T \beta}} \quad (2)$$

$$g(x_i) = \ln\left(\frac{\pi(x_i)}{1 - \pi(x_i)}\right) = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} \quad (3)$$

Metode yang digunakan dalam mengestimasi parameter model regresi logistik adalah metode *Maximum Likelihood Estimator* (MLE). Estimasi parameter menggunakan metode MLE dilakukan dengan memaksimalkan fungsi *likelihood* yang dituliskan pada persamaan 4 dan 5.

$$L(\beta) = \prod_{i=1}^n f(x_i) = \prod_{i=1}^n \pi(x_i)^{y_i} (1 - \pi(x_i))^{1-y_i} \quad (4)$$

$$\ln L(\beta) = \sum_{i=1}^n \left\{ y_i \ln \left[\frac{e^{x_i^T \beta}}{1 + e^{x_i^T \beta}} \right] + (1 - y_i) \ln \left[\frac{1}{1 + e^{x_i^T \beta}} \right] \right\} \quad (5)$$

Estimasi parameter dengan metode MLE tidak menghasilkan solusi yang berbentuk *closed form*, sehingga estimasi parameter pada regresi logistik biner dengan metode MLE ini diperoleh menggunakan metode optimasi numerik yaitu iterasi *Newton-Raphson* [8].

C. Regularization

Permasalahan yang sering muncul pada proses klasifikasi adalah terjadinya *over fitting* pada data *training* dimana model yang dihasilkan bagus untuk digunakan pada data *training* tetapi menghasilkan performa yang jelek untuk data baru (data *testing*) [7]. Salah satu pendekatan yang dapat mengurangi *over fitting* adalah menggunakan *regularization* dengan fungsi *regularized log likelihood* didefinisikan sebagai berikut.

$$\ln L(\beta) = \sum_{i=1}^n \ln \left[\frac{e^{y_i x_i^T \beta}}{1 + e^{x_i^T \beta}} \right] - \frac{\lambda}{2} \|\beta\|^2 \quad (6)$$

Dimana nilai $\|\beta\|^2 = \beta_1^2 + \beta_2^2 + \dots + \beta_p^2$ adalah *penalty* dan $\lambda > 0$ adalah parameter regularisasi. Penambahan unsur $\frac{\lambda}{2} \|\beta\|^2$ berfungsi sebagai regularisasi atau penalti yang digunakan untuk meningkatkan kinerja generalisasi dan mengatasi masalah *over fitting*.

D. Rare Event Weighted Logistic Regression (RE-WLR)

Metode RE-WLR merupakan pengembangan dari *Rare Event Weighted Kernel Logistic Regression* (RE-WKLR) yang dikembangkan oleh Maalouf dan Siddiqi tahun 2014 untuk mengatasi permasalahan data *imbalanced* berukuran besar. RE-WLR merupakan regularisasi, *weighting*, dan *bias correction* pada regresi logistik [9]. Data *imbalanced* atau

rare events merupakan suatu kejadian dimana frekuensi kejadian substansial lebih kecil daripada frekuensi kejadian pada umumnya.

Metode RE-WLR dimulai dengan melakukan regularisasi pada model regresi logistik seperti pada persamaan 7, kemudian dilanjutkan dengan memboboti fungsi *log likelihood* tersebut yang dituliskan pada persamaan 7 (menjadi fungsi *regularized weighted log likelihood*).

$$\ln L_W(\beta) = \sum_{i=1}^n w_i \ln \left[\frac{e^{y_i x_i^T \beta}}{1 + e^{x_i^T \beta}} \right] - \frac{\lambda}{2} \|\beta\|^2 \quad (7)$$

Dimana $w_i = \left(\frac{\tau}{y}\right) y_i + \left(\frac{1-\tau}{1-y}\right) (1 - y_i)$, \bar{y} adalah proporsi kejadian dalam sampel/data *training*, τ adalah proporsi kejadian dalam *populasi*/keseluruhan data, dan w_i adalah bobot pada RE-WLR. Selanjutnya vektor gradien dan matriks Hessian dapat diperoleh dari turunan pertama dan kedua fungsi *regularized weighted log likelihood* yang secara berurutan dituliskan pada persamaan 8 dan 9.

$$g(\beta) = X^T W (y - \pi) - \lambda \beta \quad (8)$$

$$H(\beta) = -X^T D X - \lambda I \quad (9)$$

Keterangan:

$$W = \text{diag}(w_1, w_2, \dots, w_n) \quad (10)$$

$$D = \text{diag}(v_i w_i); v_i = \pi(x_i)(1 - \pi(x_i)) \quad (11)$$

Estimasi parameter *regularized weighted log likelihood* menggunakan metode iterasi Newton-Raphson dengan rumus *update* Newton-Raphson ditunjukkan pada persamaan 12.

$$\hat{\beta}^{(t+1)} = (X^T D X + \lambda I)^{-1} X^T D z^{(t)} \quad (12)$$

Dengan $z^{(t)} = X \hat{\beta}^{(t)} + D^{-1}(y - p)$ merupakan variabel dependen yang disesuaikan. Berdasarkan persamaan 12 diperoleh subproblem untuk *weighted least square* yaitu invers matriks Hessian bernilai sangat besar sehingga proses komputasinya sangat lama. Untuk mengatasi hal tersebut, maka digunakan metode *Conjugate Gradient* (CG) *Linear* yang dilakukan dengan memberikan nilai limit tertentu pada jumlah iterasi untuk menghindari perhitungan komputer yang terlalu lama. Setelah diperoleh estimasi parameternya, maka dilanjutkan dengan koreksi bias yang bertujuan untuk mengatasi bias pada proses estimasi parameter dengan metode MLE pada data yang *imbalanced*. Vektor *bias correction* dituliskan pada persamaan 13.

$$B(\hat{\beta}) = (X^T D X)^{-1} X^T D \xi \quad (13)$$

Dimana elemen ke- i vektor ξ untuk model WLR adalah $\xi_i = 0,5 Q_{ii} (1 + w_1 \hat{\pi}_i - w_1)$ dan Q_{ii} merupakan elemen diagonal matriks Q yang setelah diboboti menjadi $Q = X(X^T D X)^{-1} X^T$ serta D telah didefinisikan pada persamaan 11. *Bias* juga diperoleh melalui iterasi dengan algoritma CG *Linear* seperti pada saat memperoleh estimasi parameter *regularized weighted log likelihood*. Selanjutnya, dapat dituliskan estimasi parameter dengan koreksi bias pada RE-WLR pada persamaan 14.

$$\tilde{\beta} = \hat{\beta} - B(\hat{\beta}) \quad (14)$$

Setelah diperoleh estimasi parameter terkoreksi menggunakan algoritma linier CG, maka dilanjutkan dengan memasukkan estimasi parameter tersebut ke dalam

Tabel 5. Ketepatan Klasifikasi

		Predicted Membership		Total
		π_1	π_0	
Actual Membership	π_1	n_{11}	n_{10}	n_1
	π_0	n_{01}	n_{00}	n_0

Tabel 6. Kriteria Nilai AUC

Nilai AUC	Kriteria
0,9 – 1,0	Excellent Classification
0,8 – 0,9	Good Classification
0,7 – 0,8	Fair Classification
0,6 – 0,7	Poor Classification
0,5 – 0,6	Failed

Tabel 7. Variabel Penelitian

Variabel	Nama Variabel	Kategori	Skala
Y	Jenis Perceraian	0: Cerai Gugat	Nominal
		1: Cerai Talak	
X ₁	Usia penggugat/pemohon ketika menikah	-	Rasio
		Usia tergugat/termohon ketika menikah	
X ₂	Usia tergugat/termohon ketika menikah	-	Rasio
		Usia tergugat/termohon ketika menikah	
X ₃	Pendidikan penggugat/pemohon	0: SD/ sederajat	Ordinal
		1: SMP/ sederajat	
X ₄	Pendidikan tergugat/termohon	2: SMA/ sederajat	Ordinal
		3: Perguruan Tinggi	
X ₅	Pekerjaan penggugat/pemohon	Sama dengan kategori X ₃	Nominal
		0: Tidak bekerja	
X ₆	Pekerjaan tergugat/termohon	1: Petani/Peternak/ Nelayan	Nominal
		2: ART/TKI/ Buruh/Kuli/ Tukang/Sopir/ Ojek/Kernet/ Satpam/ Security/Baby sitter	
X ₇	Lama pernikahan	3: Karyawan/Sales	Nominal
		4: Jasa/Pedagang/ Wirusaha/ Swasta/ Kontraktor	
X ₈	Keadaan perceraian	5: Guru/ Pegawai/ PNS/TNI/ POLRI/ Tenaga Kesehatan/ Tenaga Professional Lain	Nominal
		0: Ba'da Duhul	
X ₉	Sebab perceraian	1: Qobla Duhul	Nominal
		0: Meninggalkan Kewajiban	
		1: Moral	
		2: Penganiayaan	

persamaan logit untuk memperoleh prediksi peluang. Formula persamaan RE-WLR dituliskan pada persamaan 15.

$$\tilde{\pi}_i = \frac{e^{x_i^T \beta}}{1 + e^{x_i^T \beta}} \quad (15)$$

Pengujian signifikansi parameter dari model RE-WLR

Tabel 1. Karakteristik Jenis Perceraian Menurut Usia Ketika Menikah dan Lama Pernikahan

Variabel	Y	Mean	Varians	Min	Maks
Usia Penggugat	Cerai Gugat	31,642	75,004	17,000	70,000
	Cerai Talak	37,001	104,806	20,000	78,000
Usia Tergugat	Cerai Gugat	36,492	82,497	18,000	76,000
	Cerai Talak	33,300	105,697	17,000	73,000
Lama Pernikahan	Cerai Gugat	8,915	54,031	0,000	44,000
	Cerai Talak	8,476	63,973	0,000	44,000

Tabel 2. Signifikansi Parameter Model RE-WLR Terbaik

Koefisien	Estimasi	St. Error	t-hitung	P-Value
Konstanta	-1,768	0,373	22,443	0,000
X ₁	0,282	0,016	299,527	0,000
X ₂	-0,266	0,016	272,879	0,000
X ₅₍₁₎	2,595	0,314	68,368	0,000
X ₅₍₂₎	2,906	0,290	100,352	0,000
X ₅₍₃₎	1,537	0,273	31,729	0,000
X ₅₍₄₎	1,527	0,274	31,051	0,000
X ₅₍₅₎	1,009	0,403	6,275	0,012
X ₆₍₁₎	-1,898	0,253	56,425	0,000
X ₆₍₂₎	-2,891	0,252	132,029	0,000
X ₆₍₃₎	-1,505	0,217	47,802	0,000
X ₆₍₄₎	-1,709	0,217	62,099	0,000
X ₆₍₅₎	-1,367	0,371	13,539	0,001
X ₉₍₁₎	-0,726	0,819	0,786	0,375
X ₉₍₂₎	0,919	0,150	37,405	0,000

Tabel 3. Hasil Uji Kesesuaian Model

Deviance Chi-Square	$\chi^2_{(2052;0,05)}$	Derajat Bebas	P-Value
1157,313	2162,601	2056	1,000

Tabel 4. Hasil Ketepatan Klasifikasi Model RE-WLR Terbaik

Observasi	Prediksi	
	Cerai Talak	Cerai Gugat
Cerai Talak	446	144
Cerai Gugat	76	1396

secara parsial, dapat menggunakan statistik uji *Wald* [7]. Hipotesis pengujian dari uji *Wald* dituliskan sebagai berikut. $H_0: \beta_j = 0$ (variabel prediktor ke- j tidak berpengaruh signifikan terhadap variabel respon)

$H_0: \beta_j \neq 0; j = 1, 2, \dots, p$ (variabel prediktor ke- j berpengaruh signifikan terhadap variabel respon)

Statistik Uji:

$$W_j^2 = \frac{\tilde{\beta}_j^2}{(SE(\tilde{\beta}_j))^2}, \text{ dengan } j = 1, 2, \dots, p \quad (16)$$

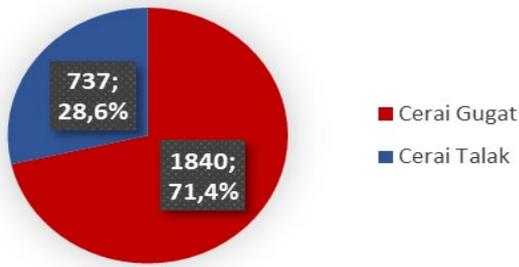
Statistik uji W^2 mengikuti distribusi *Chi-Squared* derajat bebas satu. Formula $SE(\tilde{\beta})$ untuk model RE-WLR adalah sebagai berikut [9].

$$SE(\tilde{\beta}) = \sqrt{\left(\frac{n}{n+p}\right)^2 (X^T D X)^{-1}} \quad (17)$$

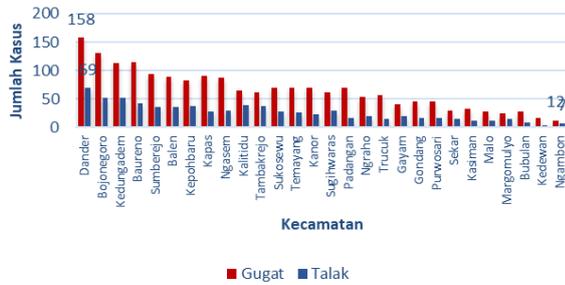
Daerah penolakan uji *Wald* adalah jika $W^2 > \chi^2_{(1,\alpha)}$ atau $p\text{-value} < \alpha$.

E. Uji Kesesuaian Model

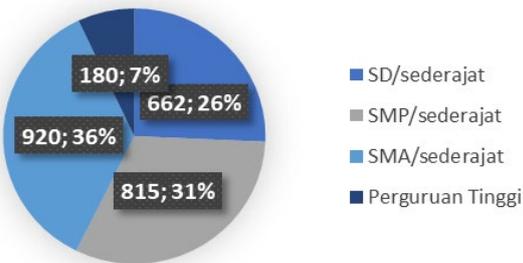
Uji kesesuaian model atau *goodness of fit test* dilakukan untuk mengetahui apakah tidak terdapat perbedaan antara hasil observasi dengan kemungkinan hasil prediksi [7]. Hipotesis pengujian pada uji kesesuaian model dituliskan



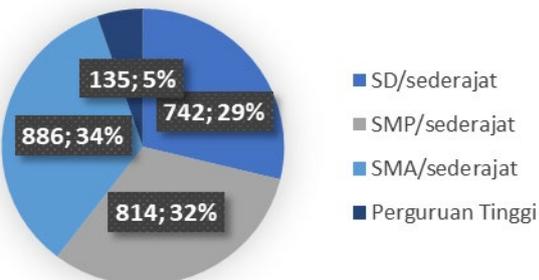
Gambar 5. Jumlah kasus perceraian berdasarkan jenisnya.



Gambar 6. Persebaran jenis kasus perceraian pada tiap kecamatan di Kabupaten Bojonegoro.



Gambar 7. Karakteristik pendidikan terakhir pengugat/pemohon.



Gambar 8. Karakteristik pendidikan terakhir tergugat/termohon.

sebagai berikut.

H_0 : Model sesuai (tidak terdapat perbedaan antara hasil observasi dengan kemungkinan prediksi model)

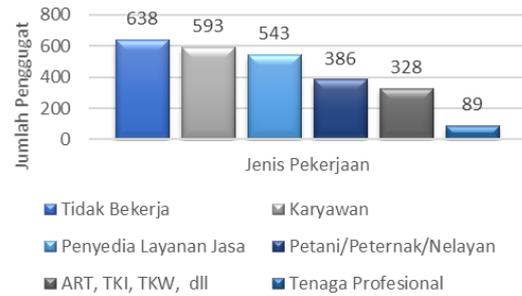
H_1 : Model tidak sesuai (terdapat perbedaan antara hasil observasi dengan kemungkinan prediksi model)

Statistik Uji:

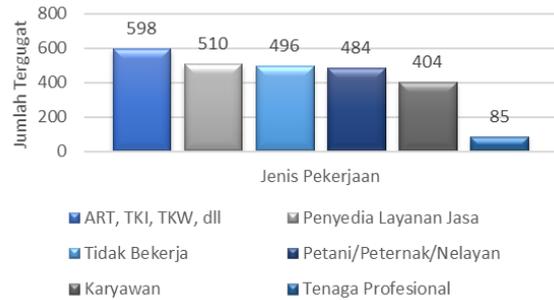
$$DEV(\hat{\beta}) = -2 \ln L(\hat{\beta}) \tag{18}$$

$$= -2 \times \sum_{i=1}^n \ln \left[\frac{e^{y_i x_i^T \hat{\beta}}}{1 + e^{x_i^T \hat{\beta}}} \right] - \frac{\lambda}{2} \|\hat{\beta}\|^2$$

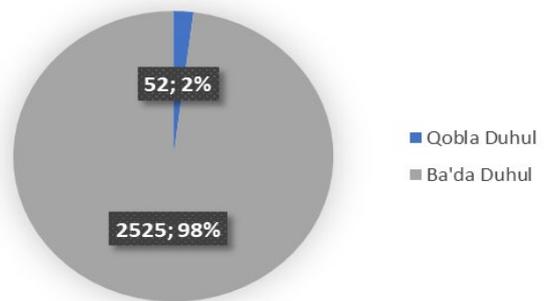
Daerah penolakan uji kesesuaian model adalah jika $DEV(\hat{\beta}) > \chi^2_{(n-p-1, \alpha)}$.



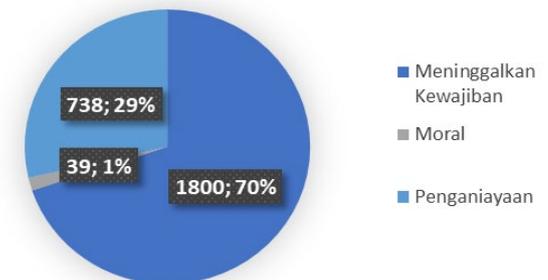
Gambar 1. Karakteristik jenis pekerjaan pengugat/pemohon.



Gambar 2. Karakteristik jenis pekerjaan tergugat/termohon.



Gambar 3. Karakteristik keadaan perceraian.

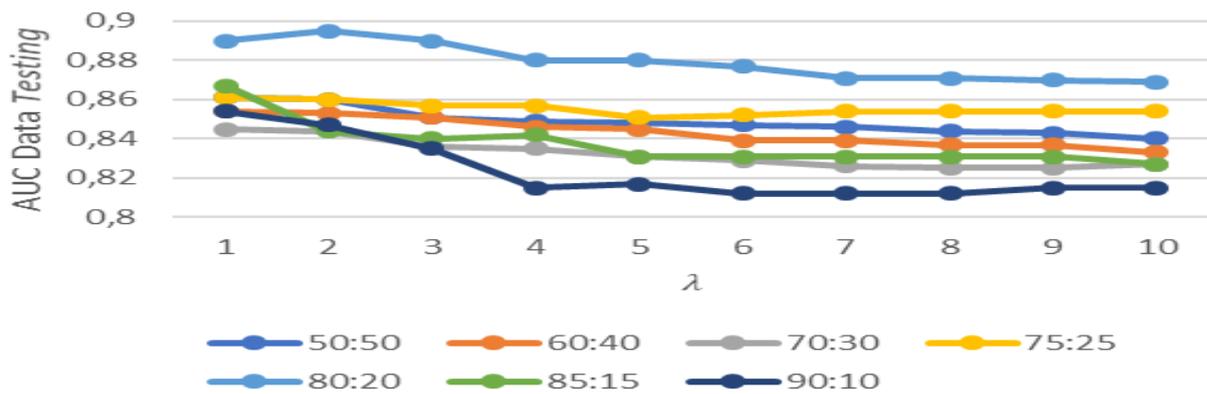


Gambar 4. Karakteristik sebab perceraian.

F. Evaluasi Ketepatan Klasifikasi

Regresi logistik biner merupakan salah satu metode klasifikasi, sehingga perlu diketahui ketepatan model regresi logistik dalam mengklasifikasikan suatu kategori yang sedang diamati. Perhitungan ketepatan klasifikasi dapat menggunakan bantuan tabel ketepatan klasifikasi yang dituliskan pada Tabel 1 [9]. Berdasarkan Tabel 1, beberapa simbol memiliki keterangan sebagai berikut:

- n_{11} = Jumlah prediksi π_1 yang tepat terklasifikasi pada π_1 (true positive)
- n_{10} = Jumlah prediksi π_1 yang salah terklasifikasi pada π_0 (false negative)



Gambar 9. AUC data *testing* setiap skema partisi data.



Gambar 10. AUC data *testing* skema 80:20 untuk setiap λ.

n_{00} = Jumlah prediksi π_0 yang tepat terklasifikasi pada π_0 (*true negative*)

n_{01} = Jumlah prediksi π_0 yang salah terklasifikasi pada π_1 (*false positive*)

Beberapa ukuran ketepatan klasifikasi yang digunakan pada penelitian ini adalah akurasi, sensitivitas, spesifisitas, dan *Area Under Receive Operating Characteristics Curve* (AUC). AUC menunjukkan luasan atau area di bawah kurva ROC, yaitu kurva yang digunakan untuk menunjukkan hubungan antara sensitivitas dan spesifisitas untuk setiap *cut off* yang mungkin. Sementara itu, sensitivitas disebut juga dengan *true positive rate*, yaitu proporsi kelas positif (kelas π_1) yang tepat terklasifikasi, sedangkan spesifisitas disebut juga *true negative rate*, yaitu proporsi kelas negatif (kelas π_0) yang tepat terklasifikasi. Masing-masing formula ukuran ketepatan klasifikasi dituliskan pada persamaan berikut.

$$Akurasi = 1 - \frac{n_{10} + n_{01}}{n_1 + n_2} \quad (19)$$

$$Sensitivitas = \frac{n_{11}}{n_{11} + n_{10}} \quad (20)$$

$$Spesifisitas = \frac{n_{00}}{n_{00} + n_{01}} \quad (21)$$

$$AUC = \frac{1}{2} (sensitivitas + spesifisitas) \quad (22)$$

Nilai AUC yang semakin mendekati 1 menunjukkan bahwa model klasifikasi yang terbentuk semakin akurat. Terdapat beberapa kriteria keakuratan hasil klasifikasi berdasarkan AUC yang ditampilkan pada Tabel 2 [10].

G. Perceraian

Menurut P.N.H. Simanjuntak dalam bukunya, Hukum Perdata Indonesia perceraian merupakan pengakhiran suatu

perkawinan karena sesuatu sebab dengan keputusan hakim atas tuntutan dari salah satu pihak atau kedua belah pihak dalam perkawinan [1]. Berdasarkan Peraturan Pemerintah No. 9 Tahun 1975, perceraian dibagi menjadi dua bagian yaitu cerai talak dan cerai gugat. Cerai talak merupakan perceraian yang dijatuhkan oleh suami terhadap istrinya, sementara cerai gugat merupakan perceraian yang didasarkan atas adanya suatu gugatan yang diajukan oleh pihak istri.

Alasan terjadinya suatu perceraian disebabkan oleh banyak hal. Berdasarkan penelitian Nisva tahun 2020, diketahui terdapat enam faktor yang signifikan mempengaruhi jenis perceraian di Kabupaten Lumajang yaitu usia penggugat/pemohon dan tergugat/termohon saat menikah, pendidikan tergugat/termohon, pekerjaan penggugat/pemohon dan tergugat/termohon, keadaan istri saat jatuh talak, dan penyebab perceraian [3].

III. METODOLOGI PENELITIAN

A. Sumber Data

Data yang digunakan pada penelitian ini adalah data sekunder yang diperoleh dari laporan perkara perceraian di Pengadilan Agama Bojonegoro Kelas 1A. Data tersebut merupakan data laporan perkara perceraian yang terjadi di Kabupaten Bojonegoro pada tahun 2020 dengan total jumlah perkara adalah 2.577 perkara. Unit *sampling* pada penelitian ini adalah perkara perceraian.

B. Variabel Penelitian

Variabel penelitian pada penelitian ini disajikan pada Tabel 3.

C. Langkah Analisis

Langkah-langkah analisis data pada penelitian ini adalah sebagai berikut.

1. Mengumpulkan data kasus perceraian di Kabupaten Bojonegoro.
2. Mendeskripsikan karakteristik data kasus perceraian di Kabupaten Bojonegoro.
3. Menentukan skema pembagian data *training* dan data *testing* menggunakan *sampling* stratifikasi dengan langkah-langkah dan skema partisi data 50:50, 60:40, 70:30, 75:25, 80:20, 85:15, dan 90:10 serta parameter regularisasi 1-10 dengan *increment* 1 serta dicobakan parameter regularisasi tambahan yaitu 0,5; 1,5; dan 2,5 untuk skema partisi data terpilih.
4. Memperoleh model RE-WLR terbaik untuk klasifikasi jenis perceraian di Kabupaten Bojonegoro dengan tahapan sebagai berikut:
 - a. Mendapatkan estimasi parameter model RE-WLR.
 - b. Menghitung probabilitas optimum ($\tilde{\pi}_i$) dari masing-masing observasi menggunakan persamaan 14.
 - c. Mendapatkan *confusion matrix* data *training* dan data *testing* sehingga dapat dihitung nilai *sensitivity* (persamaan 20), *specificity* (persamaan 21), dan AUC (persamaan 22) dari model RE-WLR.
 - d. Memilih skema partisi data dan parameter regularisasi (λ) terbaik berdasarkan nilai AUC tertinggi data *testing*.
 - e. Menetapkan model RE-WLR terbaik.
5. Melakukan uji kesesuaian model RE-WLR terbaik untuk data *training*.
6. Menguji signifikansi parameter parsial model RE-WLR terbaik untuk data *training*. Jika terdapat variabel prediktor yang tidak signifikan, maka variabel prediktor tersebut dikeluarkan dari model dan dimodelkan ulang (eliminasi *backward*).
7. Menginterpretasikan model RE-WLR terbaik untuk data *training*.
8. Mengidentifikasi ketepatan klasifikasi model RE-WLR terbaik untuk data *training*.
9. Menarik kesimpulan dan saran.

IV. HASIL DAN PEMBAHASAN

A. Karakteristik Kasus Perceraian di Kabupaten Bojonegoro

Perceraian menurut jenisnya terbagi menjadi 2, yaitu cerai gugat dan cerai talak. Perbandingan jumlah cerai gugat dan cerai talak di Kabupaten Bojonegoro tahun 2020 disajikan pada Gambar 1.

Gambar 1 menunjukkan bahwa terdapat 1.840 kasus cerai gugat atau sebesar 71,4% dari 2.577 kasus perceraian yang ada di Kabupaten Bojonegoro, sementara kasus cerai talak sebesar 28,6% atau sebanyak 737 kasus. Kondisi ini merupakan kondisi data *imbalanced* dimana kelas minoritas yaitu cerai talak memiliki persentase kurang dari 35% [11]. Selanjutnya persebaran jenis perceraian pada masing-masing kecamatan di Kabupaten Bojonegoro disajikan pada Gambar 2.

Gambar 2 menunjukkan bahwa pada seluruh kecamatan di Kabupaten Bojonegoro didominasi oleh kasus cerai gugat, dimana kecamatan dengan jumlah kasus perceraian tertinggi

adalah Kecamatan Dander dengan 227 kasus. Sementara kecamatan dengan kasus perceraian terendah adalah Kecamatan Ngambon dengan total 19 kasus.

Pada Tabel 4 disajikan karakteristik jenis perceraian menurut variabel usia ketika menikah dan lama pernikahan. Tabel 4 memberikan informasi bahwa rata-rata pasangan yang bercerai di Kabupaten Bojonegoro tahun 2020 adalah pasangan yang berusia sekitar 30 tahun dengan variasi yang tinggi. Selain itu diketahui pula bahwa rata-rata lama pernikahan pasangan cerai di Kabupaten Bojonegoro tahun 2020 pada kedua jenis perceraian adalah pasangan yang telah menikah selama 8 hingga 9 tahun dan juga terjadi variasi yang tinggi. Selanjutnya, Gambar 3 menyajikan karakteristik pendidikan terakhir penggugat/pemohon.

Penggugat/pemohon merupakan pihak yang mengajukan perceraian. Pada tahun 2020, penggugat/pemohon perceraian di Kabupaten Bojonegoro mayoritas adalah memiliki pendidikan terakhir SMA/ sederajat yaitu sebanyak 920 orang atau 36% dari total penggugat/pemohon. Sementara hanya 180 orang atau 7% penggugat/pemohon dengan pendidikan terakhir Perguruan Tinggi. Sementara itu Gambar 4 menyajikan karakteristik pendidikan terakhir pihak tergugat/termohon (Tabel 5).

Gambar 4 menginformasikan bahwa mayoritas pendidikan terakhir tergugat/termohon sama dengan pendidikan terakhir penggugat/pemohon yaitu SMA/ sederajat yaitu sebanyak 886 orang atau 34% dari total tergugat/termohon. Sementara hanya 135 orang atau 5% tergugat/termohon yang berpendidikan terakhir Perguruan Tinggi. Penjelasan berikutnya adalah mengenai pekerjaan penggugat/pemohon dan tergugat/termohon yang secara berurutan disajikan pada Gambar 5 dan Gambar 6.

Berdasarkan Gambar 5, diketahui bahwa pekerjaan sebanyak 638 penggugat/pemohon adalah tidak bekerja seperti ibu rumah tangga, pensiunan, mahasiswa, serta yang menjawab tidak tahu. Sementara itu, jenis pekerjaan dengan jumlah paling rendah adalah jenis pekerjaan tenaga profesional, yaitu hanya terdiri dari 89 orang. Pada Gambar 5 diketahui bahwa tergugat/termohon mayoritas memiliki pekerjaan sebagai asisten rumah tangga, TKI/TKW, buruh, dan lain-lain yang sejenis sejumlah 598 orang. Selain itu, juga diinformasikan bahwa tergugat/termohon yang bekerja sebagai tenaga profesional hanya 85 orang, lebih sedikit dibandingkan penggugat/pemohon.

Selanjutnya dijelaskan karakteristik keadaan perceraian pada Gambar 7. Pada Gambar 7 diketahui bahwa 98% kasus perceraian di Kabupaten Bojonegoro tahun 2020 adalah keadaan *ba'da duhul* (perceraian yang dilakukan setelah terjadi hubungan suami istri) dan hanya 2% yang termasuk dalam *qobla duhul*. Berikutnya adalah karakteristik sebab perceraian yang disajikan pada Gambar 8.

Gambar 8 menunjukkan bahwa sebanyak 1800 kasus perceraian atau sebesar 70% dari 2.577 kasus perceraian di Kabupaten Bojonegoro tahun 2020 disebabkan karena masalah meninggalkan kewajiban yang meliputi meninggalkan salah satu pihak, dihukum penjara, cacat badan, kawin paksa, murtad, ekonomi, dan poligami. Sementara itu, alasan sebab terjadinya perceraian yang paling sedikit yaitu hanya 39 kasus atau 1% dari total kasus perceraian adalah terkait dengan moral seperti zina, mabuk, mandat, dan judi.

B. Pemodelan Rare Event Weighted Logistic Regression (RE-WLR)

1) Penentuan Skema Partisi Data Terbaik dan Parameter Regularisasi Optimum

Pembentukan model RE-WLR berbeda dengan pemodelan pada Regresi Logistik pada umumnya. Perbedaannya terletak pada asumsi bebas multikolinieritas. Pada pemodelan RE-WLR, tidak memerlukan identifikasi multikolinieritas seperti pada pemodelan Regresi Logistik dikarenakan pada RE-WLR sudah terdapat tahapan regularisasi yang bertujuan untuk menangani kasus multikolinieritas seperti pada Regresi Ridge. Oleh karena itu, pada penelitian ini pemodelan RE-WLR dimulai dengan membentuk model dengan 70 skema yang meliputi 7 skema partisi data dan 10 parameter regularisasi. Selanjutnya, ditentukan model dengan skema partisi dan parameter regularisasi yang menghasilkan nilai AUC data *testing* tertinggi.

Pada Gambar 9 diketahui bahwa skema partisi data yang menghasilkan nilai AUC data *testing* tertinggi adalah skema 80:20. Oleh karena itu, pada tahapan selanjutnya, digunakan skema partisi data 80:20 kemudian dibandingkan untuk setiap nilai λ . Pada penelitian ini, digunakan nilai λ tambahan yaitu 0,5; 1,5; dan 2,5 dengan perbandingan nilai AUC data *testing* disajikan pada Gambar 10.

Gambar 10 menginformasikan bahwa terdapat dua λ yang menghasilkan nilai AUC data *testing* tertinggi yaitu $\lambda = 1,5$ dan $\lambda = 2$ dengan nilai AUC data *testing* sebesar 0,895. Untuk menentukan nilai λ optimum, digunakan pembandingan lain yaitu nilai AUC data *training* dimana model dengan $\lambda = 1,5$ memiliki nilai AUC data *training* yang lebih tinggi yaitu 0,859 sementara pada model dengan $\lambda = 2$ memiliki nilai AUC data *training* 0,858. Berdasarkan hal tersebut, terpilih skema pemodelan terbaik adalah skema dengan partisi data 80:20 dan $\lambda = 1,5$ (Tabel 6).

2) Pembentukan Model RE-WLR dengan Seluruh Variabel

Model RE-WLR dengan skema terpilih yang melibatkan seluruh variabel prediktor dituliskan pada persamaan 21 berikut.

$$\tilde{\pi}_i(x_i) = \frac{e^{x_i^T \tilde{\beta}}}{1 + e^{x_i^T \tilde{\beta}}} \tag{21}$$

Dimana:

$$\begin{aligned} x_i^T \tilde{\beta} = & -1,286 + 0,282X_1 - 0,262X_2 - 0,153X_{3(1)} + 0,359X_{3(2)} + 0,100X_{3(3)} + \\ & 0,007X_{4(1)} - 0,470X_{4(2)} - 0,138X_{4(3)} + 2,618X_{5(1)} + 2,950X_{5(2)} + \\ & 1,527X_{5(3)} + 1,540X_{5(4)} + 1,113X_{5(5)} - 1,917X_{6(1)} - 2,869X_{6(2)} + \\ & -1,462X_{6(3)} - 1,655X_{6(4)} - 1,386X_{6(5)} - 0,011X_7 - 0,488X_{8(1)} + \\ & -0,724X_{9(1)} + 0,920X_{9(2)} \end{aligned}$$

Berdasarkan hasil pengujian signifikansi untuk model lengkap, diketahui bahwa dengan taraf signifikansi (α) sebesar 5% maka diperoleh enam variabel yang signifikan berpengaruh dalam model. Variabel tersebut yaitu X_1 (usia penggugat/ pemohon ketika menikah), X_2 (usia tergugat/termohon ketika menikah), X_4 (pendidikan tergugat/termohon), X_5 (pekerjaan penggugat/pemohon), X_6 (pekerjaan tergugat/termohon), dan X_9 (sebab perceraian).

3) Pembentukan Model RE-WLR Terbaik dengan Eliminasi Backward

Setelah diperoleh model dengan skema partisi data dan parameter regularisasi yang akan digunakan, selanjutnya perlu ditentukan model terbaik yang hanya melibatkan variabel prediktor yang berpengaruh signifikan terhadap respon yaitu menggunakan eliminasi *backward*. Melalui eliminasi *backward*, diperoleh model RE-WLR terbaik dengan lima variabel prediktor signifikan pada taraf signifikansi 5% yaitu variabel X_1 (usia penggugat/ pemohon ketika menikah), X_2 (usia tergugat/termohon ketika menikah), X_5 (pekerjaan penggugat/pemohon), X_6 (pekerjaan tergugat/termohon), dan X_9 (sebab perceraian). Hasil pengujian signifikansi parameter parsial model terbaik selengkapnya disajikan pada Tabel 7.

Setelah diketahui signifikansi parameter model secara parsial yang disajikan pada Tabel 7 maka perlu diketahui apakah model terbaik yang terbentuk menghasilkan model yang sudah sesuai untuk mengklasifikasikan jenis perceraian di Kabupaten Bojonegoro.

Tabel 7 menunjukkan nilai *Deviance Chi-Square* < *Chi-Square* tabel serta *p-value* > 0,05 sehingga dapat disimpulkan bahwa pada taraf signifikansi 5% model terbaik yang diperoleh sudah sesuai, artinya tidak terdapat perbedaan yang signifikan antara hasil pengamatan dengan kemungkinan hasil prediksi model. Berikutnya, model RE-WLR terbaik yang terbentuk dituliskan pada persamaan 22.

$$\tilde{\pi}_i(x_i) = \frac{e^{x_i^T \tilde{\beta}}}{1 + e^{x_i^T \tilde{\beta}}} \tag{22}$$

Dimana:

$$\begin{aligned} x_i^T \tilde{\beta} = & -1,768 + 0,282X_1 - 0,266X_2 + 2,595X_{5(1)} + 2,906X_{5(2)} + 1,537X_{5(3)} + \\ & 1,527X_{5(4)} + 1,009X_{5(5)} - 1,898X_{6(1)} - 2,891X_{6(2)} - 1,505X_{6(3)} - \\ & -1,709X_{6(4)} - 1,367X_{6(5)} - 0,726X_{9(1)} + 0,919X_{9(2)} \end{aligned}$$

Berikut merupakan interpretasi hasil model RE-WLR terbaik menggunakan nilai *odds ratio* untuk masing-masing variabel prediktor signifikan.

a. *Odds Ratio* X_1 : **OR = exp (10*0,282) = 16,820 ≈ 17**

Artinya setiap kenaikan 10 tahun usia penggugat/pemohon, maka kecenderungan terklasifikasi ke jenis perceraian cerai talak sebesar 17 kali dibanding cerai gugat.

b. *Odds Ratio* X_2 : **OR = exp (10*-0,266) = 0,070**

Artinya setiap kenaikan 10 tahun usia tergugat/termohon, maka kecenderungan untuk terklasifikasi ke dalam cerai gugat 14,261 ≈ 15 kalinya dibandingkan untuk terklasifikasi ke dalam cerai talak.

c. *Odds Ratio* $X_{5(1)}$: **OR = exp (2,595) = 13,401 ≈ 14**

Artinya penggugat/pemohon yang memiliki pekerjaan sebagai petani/peternak/nelayan memiliki kecenderungan 14 kali untuk terklasifikasi dalam jenis perceraian cerai talak dibandingkan dengan penggugat/pemohon yang tidak bekerja.

d. *Odds Ratio* $X_{6(1)}$: **OR = exp (-1,898) = 0,150**

Artinya tergugat/termohon dengan pekerjaan sebagai petani/peternak/nelayan memiliki kecenderungan sebesar 6,667 ≈ 7 kali untuk terklasifikasi pada jenis perceraian cerai

gugat dibandingkan dengan tergugat/termohon yang tidak bekerja.

e. *Odds Ratio* $X_{9(2)}$: $OR = \exp(0,919) = 2,507 \approx 3$

Artinya kasus perceraian yang disebabkan karena penganiayaan memiliki kecenderungan untuk terklasifikasi dalam jenis cerai talak sebesar 3 kali dibandingkan kasus perceraian yang disebabkan karena meninggalkan kewajiban.

Langkah berikutnya adalah mengukur ketepatan klasifikasi dari model terbaik yang terbentuk untuk mengetahui kebaikan model RE-WLR dalam mengklasifikasikan jenis perceraian di Kabupaten Bojonegoro.

Berdasarkan nilai-nilai yang diperoleh pada Tabel 7, maka dapat ditentukan ukuran-ukuran ketepatan klasifikasi sebagai berikut.

$$\text{Sensitivitas} = \frac{446}{446 + 144} = \frac{446}{590} = 0,756$$

$$\text{Spesifisitas} = \frac{1396}{1396 + 76} = \frac{1396}{1472} = 0,948$$

$$\begin{aligned} AUC &= \frac{1}{2}(\text{sensitivitas} + \text{spesifisitas}) \\ &= \frac{1}{2}(0,756 + 0,948) = 0,852 \end{aligned}$$

Berdasarkan nilai ukuran ketepatan klasifikasi yang diperoleh, dapat diketahui bahwa nilai sensitivitas yang diperoleh adalah 0,756 yang menunjukkan bahwa model RE-WLR yang terbentuk mampu mengklasifikasikan cerai talak dengan tepat sebesar 75,6%. Sementara nilai spesifisitas sebesar 0,948 berarti bahwa model RE-WLR mampu mengklasifikasikan cerai gugat dengan tepat sebesar 94,8%. Selanjutnya, untuk menentukan apakah model klasifikasi yang dihasilkan tergolong baik maka dapat diketahui melalui nilai AUC yaitu sebesar 0,852 atau 85,2% yang mana termasuk dalam kategori baik atau *Good Classification*.

V. KESIMPULAN

Berdasarkan hasil analisis yang dilakukan, maka dapat diperoleh kesimpulan sebagai berikut: (1) Jenis perceraian di Kabupaten Bojonegoro tahun 2020 merupakan data *imbalanced* dengan perbandingan cerai gugat dan cerai talak adalah 71,4% dan 28,6% dari total 2.577 kasus. Kasus perceraian tertinggi terjadi di Kecamatan Dander dan

terendah di Kecamatan Ngambon. Penggugat wanita memiliki rata-rata usia yang lebih muda dibandingkan dengan penggugat laki-laki. Mayoritas pendidikan pasangan cerai adalah SMA/ sederajat. Pasangan cerai gugat memiliki lama pernikahan yang relatif lebih lama dibandingkan pasangan cerai talak. 98% kasus perceraian merupakan keadaan perceraian ba'da duhul dengan sebab perceraian 70% karena masalah meninggalkan kewajiban. (2) Faktor-faktor yang berpengaruh signifikan terhadap jenis perceraian di Kabupaten Bojonegoro menggunakan model RE-WLR dengan skema partisi data 80:20 dan $\lambda = 1,5$ meliputi usia penggugat/pemohon ketika menikah (X_1), usia tergugat/termohon ketika menikah (X_2), pekerjaan penggugat/pemohon (X_5), pekerjaan tergugat/termohon (X_6), dan sebab perceraian (X_9). Model menghasilkan nilai ketepatan klasifikasi AUC sebesar 85,2% dan termasuk dalam kategori *Good Classification*.

DAFTAR PUSTAKA

- [1] P. Simanjuntak, *Hukum Perdata Indonesia*, 1st ed. Jakarta: PT. Fajar Interpratama Mandiri, 2017.
- [2] BPS, *Statistik Indonesia 2020*, 1st ed. Jakarta: Badan Pusat Statistik, 2020.
- [3] T. M. T. Nisva and V. Ratnasari, "Analisis regresi logistik biner pada faktor-faktor yang mempengaruhi jenis perceraian di Kabupaten Lumajang," *Inferensi*, vol. 3, no. 1, pp. 29–36, 2020.
- [4] K. I. Pengadilan Agama Bojonegoro, *Laporan Faktor-Faktor Penyebab Terjadinya Perceraian pada Pengadilan Agama Bojonegoro Tahun 2020*, 1st ed. Kabupaten Bojonegoro: Pengadilan Agama Bojonegoro Kelas 1A, 2020.
- [5] I. Menarianti, "Klasifikasi data mining dalam menentukan pemberian kredit bagi nasabah koperasi," *J. Ilm. Teknosains*, vol. 1, no. 1, 2015.
- [6] R. E. Walpole, *Pengantar Metode Statistika, Diterjemahkan: Bambang Sumantri*, 3rd ed. Jakarta: PT. Gramedia Pustaka Utama, 1955.
- [7] L. D. W. H. S., *Applied Logistic Regression*, 2nd ed. Canada: John Wiley & Sons, Inc., 2000.
- [8] X. Yan and X. Su, *Linear Regression Analysis: Theory and Computing*, 1st ed. Singapura: World Scientific, 2009.
- [9] M. Maalouf and M. Siddiqi, "Weighted logistic regression for large-scale imbalanced and rare events data," *Knowledge-Based Syst.*, vol. 59, pp. 142–148, 2014.
- [10] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis*, 6th ed. New Jersey: Person Prentice Hall, 2007.
- [11] H. Li and J. Sun, "Forecasting business failure: the use of nearest-neighbour support vectors and correcting imbalanced samples--Evidence from the Chinese hotel industry," *Tour. Manag.*, vol. 33, no. 3, pp. 622–634, 2012.