

Pendugaan Koefisien Regresi Logistik Biner Menggunakan Algoritma *Least Angle Regression*

Mamik Utami^{1*}, Anna Islamiyati², Sri Astuti Thamrin³

¹²³Departemen Statistika, Fakultas MIPA, Universitas Hasanuddin, Makassar,
90245, Indonesia

* Corresponding author, email: mamik.utami.mu@gmail.com

Abstract

Binary logistic regression is a statistical analysis method that aims to determine the relationship between variable which has two categories with the predictor variable that have categorical or continuous scale. The method that used to estimate logistic regression parameters is Maximum Likelihood Estimation (MLE) method. In estimating parameters, Least Angle Regression (LAR) algorithm is used to select the significant variables in order to get the best model from the estimation results of binary logistic regression coefficients. This LAR algorithm is applied to the risiko of stunting data in two-year-old-babies at Buntu Batu Health Center working area, Enrekang Regency, South Sulawesi in 2019. This results obtained in the estimation of binary logistic regression prediction model using LAR algorithm, the standard error value is 0.018 smaller than the standard error value of binary logistic regression, which is 0.025. This shows that the binary logistic regression model using LAR algorithm is better than the usual binary logistic regression model on the risk of stunting data. Based on the results obtained, the variables that significantly affect the risk of stunting in two-year-old-babies on 2019 are father's height, body length of birth, exclusive breastfeeding, history of infectious diseases, and history of immunization.

Keywords: Binary Logistic Regression, Least Angle Regression Algorithm, Stunting.

Abstrak

Regresi logistik biner adalah suatu metode analisis statistika yang bertujuan untuk mengetahui hubungan antara variabel respon dua kategori dengan variabel prediktor yang bersifat kategorik atau kontinu. Metode yang digunakan untuk mengestimasi parameter regresi logistik adalah metode *Maximum Likelihood Estimation* (MLE). Dalam pengestimasi parameter, pemilihan variabel untuk mendapatkan model terbaik sangatlah penting agar faktor-faktor yang diperoleh memberikan informasi yang akurat. Dalam penelitian ini, algoritma *Least Angle Regression* (LAR) digunakan dalam menyeleksi variabel yang signifikan agar mendapatkan model terbaik dari hasil pendugaan koefisien regresi logistik biner. Algoritma LAR ini diterapkan pada data risiko kejadian *stunting* pada bayi usia dua tahun atau baduta di wilayah kerja Puskesmas Buntu Batu, Kabupaten Enrekang, Sulawesi Selatan pada tahun 2019. Hasil yang diperoleh pada estimasi model dugaan regresi logistik biner menggunakan algoritma LAR yaitu nilai standar error sebesar 0.018 lebih kecil dibandingkan nilai standar error pada regresi logistik biner yaitu sebesar 0.025. Hal ini menunjukkan bahwa model regresi logistik biner menggunakan algoritma LAR lebih baik dibandingkan model regresi logistik biner biasa pada data risiko kejadian *stunting*. Berdasarkan hasil yang diperoleh maka variabel yang signifikan mempengaruhi risiko *stunting* pada baduta tahun 2019 di Kabupaten Enrekang adalah tinggi badan ayah, panjang badan lahir, ASI eksklusif, riwayat penyakit infeksi, dan riwayat imunisasi.

Kata Kunci: Regresi Logistik Biner, Algoritma *Least Angle Regression*, *Stunting*.

1. Pendahuluan

Regresi logistik biner merupakan suatu teknik analisis statistika yang digunakan untuk menganalisis hubungan antara variabel prediktor dengan variabel respon yang

bersifat biner atau dikotomus [1]. Dalam analisis regresi, estimasi parameter merupakan tujuan penting untuk mendapatkan estimator. Salah satu metode yang umumnya digunakan pada regresi logistik yaitu dengan metode *Maximum Likelihood Estimation* (MLE) [2]. Metode MLE merupakan dasar pendekatan dalam menduga parameter pada model regresi logistik yang pada dasarnya menggunakan pendekatan distribusi dengan memaksimalkan fungsi likelihoodnya. Dalam pengestimasi parameter pemilihan variabel sangatlah penting agar faktor-faktor yang diperoleh memberikan informasi dalam keakuratan prediksi.

Algoritma *Least Angle Regression* (LAR) pertama kali dikembangkan oleh Efron, dkk (2004) dimana algoritma LAR merupakan salah satu metode untuk mendapatkan model terbaik dari hasil pendugaan koefisien. Selain untuk mendapatkan hasil estimasi koefisien, algoritma LAR juga dapat melakukan seleksi variabel yang berkaitan dengan pemilihan model terbaik [3]. Dalam proses penyeleksian algoritma LAR yaitu memasukkan variabel prediktor yang memenuhi kriteria tertentu hingga seluruh variabel telah diproses dan didapatkan model terbaik. Beberapa penelitian yang telah dilakukan sebelumnya diantaranya Lee dan Jun (2017) dalam teknik *machine learning* yang menggunakan *Logistic Model Tree* (LMT) dengan membandingkan penggunaan metode *Simple* Logistik dan LAR-Logistik dalam pemilihan variabel yang menghasilkan LAR-Logistik memiliki model yang lebih akurat [4]. Selanjutnya Wahdaniah (2015) menjelaskan bahwa penerapan algoritma LAR untuk Model Cox yang diaplikasikan pada data penyakit kanker paru-paru memiliki nilai error yang jauh lebih kecil dibandingkan dengan Model Cox biasa [5]. Berdasarkan ulasan tersebut, dapat diketahui bahwa pengaplikasian algoritma LAR ke dalam suatu metode dapat menghasilkan model yang lebih optimal dan efisien.

Penelitian ini digunakan dalam mengestimasi parameter pada model regresi logistik biner menggunakan algoritma LAR. Tetapi, pada umumnya metode MLE sulit untuk mendapatkan solusi analitik dari persamaan *likelihood*-nya. Oleh karena itu, digunakan metode iterasi untuk memperoleh nilai penduga terlebih dahulu yaitu menggunakan *Iteratively Reweighted Least Square* (IRLS) agar diperoleh solusi numeriknya sehingga dapat dilanjutkan penyeleksian variabel menggunakan algoritma LAR.

Stunting atau kerdil merupakan kondisi balita yang memiliki panjang atau tinggi badan yang kurang dari panjang atau tinggi anak seusianya. *Stunting* disebabkan oleh berbagai faktor dari dalam maupun dari luar. Faktor-faktor tersebut dapat mempengaruhi perkembangan motorik anak sehingga mempengaruhi tingkat kecerdasan otak dan anak yang menderita *stunting* akan lebih rentan terhadap penyakit. Oleh sebab itu, akan dicari faktor-faktor yang paling berpengaruh terhadap risiko kejadian *stunting* dengan menggunakan algoritma LAR pada model regresi logistik.

2. Material dan Metode

2.1 Analisis Regresi Linear Berganda

Analisis regresi linear berganda merupakan salah satu analisis yang bertujuan untuk mengetahui pengaruh suatu variabel terhadap suatu atau beberapa variabel lain. Menurut Montgomery dan Runger (2011) bentuk umum persamaan regresi linear berganda dengan parameter β dan k variabel prediktor dapat dituliskan dengan model berikut :

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + \varepsilon_i \quad (1)$$

dengan ε adalah eror, $\beta_0, \beta_1, \beta_2, \dots, \beta_k$ merupakan parameter, y merupakan variabel respon x_1, x_2, \dots, x_{ik} adalah variabel bebas pada kejadian ke- i . Dalam notasi matriks dapat ditulis sebagai berikut [6] :

$$Y = X\beta + \varepsilon \quad (2)$$

dengan $\varepsilon_i \sim N(0, \sigma^2)$.

2.2 Metode *Ordinary Least Square*

Metode yang digunakan untuk mengestimasi parameter model regresi linier berganda adalah metode kuadrat terkecil atau sering juga disebut dengan metode *Ordinary Least Square* (OLS). Metode ini bertujuan meminimumkan jumlah kuadrat galat (JKG), yakni meminimumkan persamaan berikut :

$$\begin{aligned} JKG &= \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \\ &= \sum_{i=1}^n \varepsilon_i^2 \\ &= \varepsilon^T \varepsilon \end{aligned}$$

Sehingga didapatkan hasil nilai seminimal mungkin dengan S adalah :

$$\begin{aligned} S &= \varepsilon^T \varepsilon \\ &= Y'Y - 2\beta'X'Y + \beta'X'X\beta \end{aligned}$$

Untuk meminimumkannya dapat diperoleh dengan melakukan turunan pertama S terhadap β .

$$\begin{aligned} \frac{\partial(S)}{\partial\beta} \Big|_{\beta=\hat{\beta}} &= 0 \\ X'X\hat{\beta} &= X'Y \end{aligned}$$

Sehingga solusi untuk estimasi parameter regresi adalah:

$$\hat{\beta}_{OLS} = (X'X)^{-1}X'Y \quad (3)$$

2.3 Regresi Logistik Biner

Model regresi logistik biner termasuk dalam sebaran keluarga Eksponensial. Sebaran keluarga eksponensial yang dimaksud adalah sebaran Bernoulli, yaitu sebaran dari peubah acak yang hanya mempunyai 2 kategori yaitu 0 dan 1 [7]. Variabel respon pada regresi logistik biner terdiri 2 kategori, misalkan $y = 1$ menyatakan hasil yang diperoleh ‘sukses’ dan $y = 0$ menyatakan hasil yang diperoleh ‘gagal’. Karena variabel y hanya memiliki dua kategori, maka variabel y tersebut mengikuti distribusi Bernoulli (Hosmer & Lemeshow, 2000) dengan fungsi probabilitas:

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

Menurut Hosmer dan Lemeshow (2000), model regresi logistik dengan melibatkan k variabel prediktor diformulasikan sebagai berikut :

$$\pi(x_i) = \frac{\exp(\beta_0 + \sum_{j=1}^k \beta_j x_{ij})}{1 + \exp(\beta_0 + \sum_{j=1}^k \beta_j x_{ij})} \quad (4)$$

dengan $i = 1, 2, \dots, n$ dan $j = 1, 2, \dots, k$.

Fungsi $\pi(x_i)$ pada regresi logistik merupakan fungsi non linear sehingga perlu dilakukan transformasi ke dalam bentuk logit untuk memperoleh fungsi linear agar dapat dilihat hubungan antara variabel respon dan variabel prediktornya.

Transformasi logit dari model (4), yang linear dalam parameter-parameternya adalah:

$$g(x_i) = \ln\left(\frac{\pi(x_i)}{1 - \pi(x_i)}\right) = \beta_0 + \sum_{j=1}^k \beta_j x_{ij} \quad (5)$$

2.4 Metode *Maximum Likelihood Estimation*

Salah satu metode yang umumnya digunakan untuk mendapatkan estimator pada regresi logistik yaitu dengan metode estimasi likelihood maksimum *Maximum Likelihood Estimation* (MLE).

Dengan mengacu pada model regresi logistik biner yaitu :

$$\pi(x_i) = \frac{\exp(\mathbf{x}'_i \boldsymbol{\beta})}{1 + \exp(\mathbf{x}'_i \boldsymbol{\beta})} \quad (6)$$

Berdasarkan model (6), parameter yang akan ditaksir adalah $\boldsymbol{\beta}$ dengan menggunakan metode estimasi maksimum likelihood. Diketahui y_i berdistribusi Bernoulli, sehingga fungsi kepadatan peluangnya dinyatakan sebagai berikut :

$$f(y_i) = \pi(x_i)^{y_i}(1 - \pi(x_i))^{(1-y_i)}$$

Untuk mendapatkan nilai $\boldsymbol{\beta}$ dengan memaksimalkan fungsi *likelihood* digunakan bentuk logaritma natural dari fungsi *likelihood*, yang kemudian disebut fungsi *log-likelihood*.

Fungsi *log-likelihood* dituliskan dalam bentuk :

$$\ell = \sum_{i=1}^n [-\ln(1 + \exp(\mathbf{x}'_i \boldsymbol{\beta})) + y_i(\mathbf{x}'_i \boldsymbol{\beta})] \quad (7)$$

2.5 Iteratively Reweighted Least Square

Metode iterasi IRLS adalah salah satu metode numerik untuk menyelesaikan persamaan nonlinear secara iteratif. Dalam metode ini dibutuhkan turunan pertama dan turunan kedua dari fungsi *log-likelihoodnya*. Metode ini merupakan pengembangan dari metode *fisher scoring* (Agresti, 2002). Penduga parameter dengan metode *fisher scoring* pada iterasi ke- $t + 1$ dalam proses iterasi $t = 0, 1, 2, \dots$ adalah sebagai berikut:

$$\widehat{\boldsymbol{\beta}}_{(t+1)} = \widehat{\boldsymbol{\beta}}_{(t)} - \mathbf{H}_{(t)}^{-1} \mathbf{v}_{(t)}, \quad (8)$$

dengan:

- $\widehat{\boldsymbol{\beta}}_{(t)}$: vektor untuk $\boldsymbol{\beta}$ pada iterasi ke- t
- $\mathbf{H}_{(t)}$: turunan kedua dari fungsi kriteria ℓ
- $\mathbf{v}_{(t)}$: turunan pertama dari fungsi kriteria ℓ

Proses iterasi berhenti sampai diperoleh penduga parameter yang konvergen yaitu $\sum_{i=1}^n |e^{g(x)}|$ konvergen dimana $\sum_{i=1}^n |\varepsilon_i^m - \varepsilon_i^{m-1}| < 0,001$.

2.6 Algoritma Least Angle Regression

Algoritma *Least Angle Regression* (LAR) merupakan suatu metode algoritma untuk mendapatkan hasil estimasi koefisien serta dapat melakukan seleksi variabel yang berkaitan dengan pemilihan model terbaik yang dikembangkan oleh Efron, dkk (2004). Dalam pengestimasiannya, algoritma LAR menggunakan program komputasi dengan bantuan *software*.

Adapun langkah-langkah pada algoritma LAR (Hastie, dkk (2004)) :

- a. Membuat semua koefisien parameter sama dengan nol $\beta_1, \beta_2, \dots, \beta_k = 0$.
- b. Mencari variabel prediktor \mathbf{x} yang paling berkorelasi terhadap y .
- c. Menduga koefisien parameter $\boldsymbol{\beta}$ yang memiliki korelasi tertinggi dengan y .
- d. Menghitung korelasi antara variabel prediktor yang tersisa dengan sisaan terbaru.
- e. Mengulang langkah sampai seluruh p variabel prediktor masuk ke dalam model atau hingga proses iterasi berakhir.
- f. Mengeluarkan variabel dari gugus variabel aktif jika terdapat koefisien yang bernilai nol.

Proses iterasi berakhir ketika seluruh variabel telah masuk ke dalam model sehingga didapatkan hasil optimal dari metode algoritma LAR [8].

2.7 Uji Kesesuaian Model

Dalam menguji kesesuaian model dalam regresi logistik digunakan uji *Hosmer and Lemeshow Goodness of Fit Test*. Uji ini dilakukan untuk menguji hipotesis nol untuk

mendapatkan bukti bahwa data empiris yang digunakan sesuai dengan model. Hipotesis yang digunakan adalah sebagai berikut :

H_0 : Model sesuai (tidak terdapat perbedaan yang signifikan antara hasil pengamatan dengan kemungkinan hasil prediksi model).

H_1 : Model tidak sesuai (terdapat perbedaan yang signifikan antara hasil pengamatan dengan kemungkinan hasil prediksi model).

Statistik uji *Hosmer and Lemeshow* mengikuti distribusi *Chi-Square* dengan $df = g - 2$ dimana g adalah banyaknya kelompok (Hosmer and Lemeshow, 1980) dengan rumus sebagai berikut :

$$\chi_{HL}^2 = \sum_{s=1}^g \frac{(O_s - n'_s \bar{\pi}_s)^2}{n'_s \bar{\pi}_s (1 - \bar{\pi}_s)} \quad (9)$$

dengan :

O_s = jumlah nilai y pada grup ke- s

$\bar{\pi}_s$ = rata-rata dari $\hat{\pi}$ untuk grup ke- s

n'_s = jumlah sampel dalam grup ke- s

g = banyaknya grup

$s = 1, 2, \dots, 10$

Kriteria keputusan : Untuk menguji kesesuaian model, nilai *Chi-square* yang diperoleh dibandingkan dengan nilai *Chi-square* pada tabel *Chi-square* dengan $df = g - 2$. Jika $\chi_{HL}^2 > \chi_{(g-2)}^2$ maka H_0 diterima dan H_1 ditolak.

2.8 Odds Ratio

Interpretasi koefisien dalam model regresi logistik dilakukan dalam bentuk *odds ratio* (Hosmer dan Lemeshow, 1989). Rasio odds berhubungan dengan transformasi logit, seperti yang diketahui agar menjadi bentuk yang linear fungsi logistik perlu ditransformasi sedemikian rupa. Transformasi logit dapat dituliskan :

$$\ln \left(\frac{\pi(x_i)}{1 - \pi(x_i)} \right) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik}$$

Sehingga model *odds ratio* dituliskan pada persamaan (2.15) sebagai berikut:

$$OR = \frac{\pi(x_i)}{1 - \pi(x_i)} = \exp \left(\beta_0 + \sum_{j=1}^k \beta_j x_{ij} \right) \quad (10)$$

3. Hasil dan Diskusi

Berdasarkan model regresi logistik biner pada data kejadian *stunting* di Kabupaten Enrekang untuk parameter sebanyak 6 dan jumlah sampel sebanyak 180 adalah sebagai berikut:

$$\pi(x_i) = \frac{\exp(\beta_0 + \beta_1 x_{i1} + \dots + \beta_6 x_{i6})}{1 + \exp(\beta_0 + \beta_1 x_{i1} + \dots + \beta_6 x_{i6})} = \frac{\exp(\mathbf{x}'_i \boldsymbol{\beta})}{1 + \exp(\mathbf{x}'_i \boldsymbol{\beta})} \quad (11)$$

Berdasarkan persamaan (11), parameter β akan diestimasi menggunakan metode MLE dengan iterasi IRLS. Selanjutnya, menyeleksi variabel-variabel yang berpengaruh signifikan untuk mendapatkan model terbaik dengan menggunakan algoritma LAR.

3.1 Model Regresi Logistik Biner Menggunakan Algoritma LAR

Dalam pengestimasian regresi logistik biner menggunakan algoritma LAR, akan dilihat hasil estimasi yang dilakukan dengan perhitungan iterasi berbasis matriks melalui proses komputasi. Dalam proses analisis menggunakan algoritma LAR ini dilakukan dengan cara memasukkan seluruh variabel prediktornya dimana akan selalu dimulai dengan nilai koefisien variabel bernilai nol seluruhnya, lalu akan dicari variabel yang paling berkorelasi lalu diestimasi untuk mendapatkan penduga koefisien dan dimasukkan ke dalam gugus variabel aktif. Proses tersebut diulang hingga proses iterasi berakhir. Proses iterasi berakhir apabila seluruh variabel telah diestimasi dan masuk ke dalam model. Jika terdapat koefisien yang bernilai nol maka variabel tersebut akan dikeluarkan dari gugus variabel aktif karena dalam algoritma LAR dianggap tidak signifikan berpengaruh terhadap variabel respon.

$$\begin{aligned} \ln\left(\frac{\pi(x_i)}{1 - \pi(x_i)}\right) &= \beta_0 + \sum_{j=1}^k \beta_j x_{ij} \\ &= (13.592 - 0.032 X_1 - 0.188 X_3 + 1.972 X_4 - 1.308 X_5 + \\ &\quad 0.797 X_6) \end{aligned}$$

3.2 Pemilihan Model Terbaik

Perbandingan nilai koefisien regresi yang didapat melalui metode regresi logistik biner dengan MLE dan algoritma LAR dapat dilihat pada **Tabel 1**:

Tabel 1. Nilai Standar Error pada Model

Model Regresi Logistik	Standar Error
Regresi Logistik Biner	0.025
Regresi Logistik Biner dengan Algoritma LAR	0.018

Berdasarkan Tabel 1 diperoleh bahwa nilai standar error pada regresi logistik biner menggunakan algoritma LAR sebesar 0.025 lebih kecil dibandingkan dengan nilai standar error pada regresi logistik biner biasa yaitu 0.018. Berdasarkan nilai standar error kedua model tersebut menunjukkan bahwa model regresi logistik menggunakan algoritma LAR lebih baik dibandingkan regresi logistik biner biasa. Meskipun selisih dari nilai standar eror kecil akan tetapi hasil yang diperoleh menunjukkan bahwa pendekatan dengan regresi logistik menggunakan algoritma LAR memiliki persentase keakuratan yang lebih tinggi. Selisih kecil dapat saja disebabkan oleh kondisi data riil.

Tabel 2. Nilai Odds Ratio Regresi Logistik dengan Algoritma LAR

Variabel	$\hat{\beta}$	$Exp(\hat{\beta})$
----------	---------------	--------------------

Tinggi badan ayah (X_1)	-0.032	0.968
Panjang badan lahir (X_3)	-0.188	0.828
ASI eksklusif (X_4)	1.972	7.185
Riwayat penyakit infeksi (X_5)	-1.308	0.270
Riwayat imunisasi (X_6)	0.797	2.219

Berdasarkan Tabel 2 interpretasi nilai *odds ratio* dari masing-masing variabel prediktor sebagai berikut :

- a. Tinggi Badan Ayah (X_1)
Seorang ayah dengan tinggi badan pendek merupakan faktor risiko kejadian *stunting* pada baduta 12-24 bulan di wilayah kerja puskesmas buntu batu kabupaten Enrekang. Peningkatan 1 cm tinggi badan ayah dapat menurunkan risiko baduta terkena *stunting* di Kabupaten Enrekang sebesar 0.968 dengan menganggap variabel lain konstan.
- b. Panjang Badan Lahir (X_3)
Riwayat panjang badan lahir pendek baduta merupakan faktor risiko kejadian *stunting* pada baduta 12-24 bulan di wilayah kerja puskesmas buntu batu kabupaten Enrekang. Peningkatan 1 cm panjang badan lahir bayi dapat menurunkan risiko bayi terkena *stunting* di Kabupaten Enrekang sebesar 0.828 dengan menganggap variabel lain konstan.
- c. ASI Eksklusif (X_4)
Tidak memiliki riwayat ASI eksklusif merupakan faktor risiko kejadian *stunting* pada baduta 12-24 bulan di wilayah kerja puskesmas Buntu Batu, Kabupaten Enrekang. Peluang seorang baduta yang tidak diberikan ASI eksklusif dapat meningkatkan risiko *stunting* sebesar 7.185 lebih besar dibandingkan peluang seorang baduta yang diberikan ASI eksklusif dan menganggap variabel lain konstan.
- d. Riwayat Penyakit Infeksi (X_5)
Memiliki riwayat penyakit infeksi saluran pernapasan (ISPA) dalam tiga bulan terakhir merupakan faktor risiko kejadian *stunting* pada baduta 12-24 bulan di wilayah kerja puskesmas Buntu Batu, Kabupaten Enrekang. Peluang seorang baduta yang tidak memiliki riwayat penyakit ISPA dapat menurunkan risiko *stunting* sebesar 0.270 lebih besar dibandingkan peluang seorang baduta yang memiliki riwayat penyakit ISPA dan menganggap variabel lain konstan.
- e. Riwayat Imunisasi (X_6)
Peluang seorang baduta yang tidak diberikan imunisasi secara lengkap dapat meningkatkan risiko *stunting* sebesar 2.219 lebih besar dibandingkan peluang seorang baduta yang diberikan imunisasi secara lengkap dan menganggap variabel lain konstan.

4. Kesimpulan

Hasil estimasi model dugaan regresi logistik biner melalui MLE dengan algoritma LAR yang bersesuaian dengan data risiko kejadian *stunting* di Kabupaten Enrekang adalah sebagai berikut:

$$\hat{g}(x) = 13.592 - 0.032X_1 - 0.188X_3 + 1.972X_4 - 1.308X_5 + 0.797X_6$$

Faktor-faktor yang paling berpengaruh terhadap risiko kejadian *stunting* pada baduta tahun 2019 di wilayah kerja Puskesmas Buntu Batu, Kabupaten Enrekang dengan menggunakan algoritma LAR adalah pengaruh ASI Eksklusif, riwayat penyakit infeksi, tinggi badan ayah, panjang badan lahir bayi, dan riwayat imunisasi berdasarkan nilai korelasi setiap koefisien yang ada.

Daftar Pustaka

- [1] Hosmer, D., & Lemeshow, S. *Applied Logistic Regression*. USA: John Wiley & Sons, Inc. 2000.
- [2] Aldrich, J. R. A. Fisher and the Making of Maximum. *Statistical Science*, 12(3), 162-176, 1997.
- [3] Efron, B., Hastie, T., Johnstone, I., & Tibshirani, R. Least Angle Regression. *The Annals of Statistics*, 32(2), 407-499, 2004.
- [4] Lee, S., & Jun, C. H. Fast Incremental Learning of Logistic Model Tree. *Expert Systems with Applications*. 2017.
- [5] Wahdaniah. *Penerapan Algoritma Least Angle Regression Pada Least Absolute Shrinkage And Selection Operator Untuk Model Cox*. Universitas Hasanuddin. 2015.
- [6] Montgomery, D. C., & Runger, G. C. *Applied Statistics and Probability for Engineers Fifth Edition*. New York: John Wiley & Sons. 2011.
- [7] Islamiyati, A. Estimasi Parameter Model Regresi Logistik Biner. *Jurnal Matematika, Statistika, dan Komputasi*, 11(2), 122-128, 2015.
- [8] Hastie, T., Tibshirani, R., & Friedman, J. *The Elements of Statistical Learning Data Mining, Inference, and Prediction (2nd Edition)*. New York: Springer. 2009.