

***PENALIZED MAXIMUM LIKELIHOOD ESTIMATION DENGAN
ALGORITMA GRADIENT DESCENT PADA MODEL REGRESI LOGISTIK
MULTINOMIAL***

SKRIPSI

**MUHAMMAD ALFAN IRSYADI HUTAGALUNG
190803102**



**PROGRAM STUDI SARJANA MATEMATIKA
DEPARTEMEN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS SUMATERA UTARA
MEDAN
2024**

DAFTAR ISI

DAFTAR ISI	i
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	3
BAB 2 TINJAUAN PUSTAKA	4
2.1 Regresi Logistik	4
2.2 Regresi Logistik Biner	4
2.3 Regresi Logistik Multinomial	5
2.4 <i>Log Odds</i> atau Logit	5
2.5 Maximum Likelihood Estimation	6
2.6 Koefisien Determinasi	8
2.7 Regularisasi	8
2.7.1 Regularisasi L_1 atau LASSO (<i>Least Absolute Shrinkage and Selection Operator</i>)	8
2.7.2 Regularisasi L_2 atau Ridge	9
2.8 Algoritma <i>Gradient Descent</i>	9
2.9 Tabel klasifikasi	10
BAB 3 METODE PENELITIAN	12
BAB 4 HASIL DAN PEMBAHASAN	13
4.1 Estimasi Parameter Model Regresi Logistik Multinomial menggunakan Metode MLE	13
4.2 Estimasi Parameter Model Regresi Logistik Multinomial menggunakan Metode LASSO dan Ridge	14
4.2.1 Menggunakan Metode LASSO	14
4.3 Pembahasan 2	16
4.4 Pembahasan 3	18
BAB 5 PENUTUP	19
5.1 Kesimpulan	19
5.2 Saran	19
DAFTAR PUSTAKA	20
LAMPIRAN	22

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Maximum Likelihood Estimation (MLE) merupakan metode umum yang digunakan untuk mengestimasi parameter dari regresi logistik. MLE bertujuan untuk mendapatkan nilai estimator parameter yang optimal dengan memaksimalkan fungsi likelihood. Metode ini merupakan metode yang populer dalam ilmu statistika untuk mengestimasi parameter distribusi probabilitas berdasarkan data yang diamati. Regresi logistik dengan variable dependen yang memiliki dua kategori nominal dinamakan regresi logistik biner, sedangkan regresi logistik dengan variabel dependen yang memiliki lebih dari dua kategori nominal dinamakan regresi logistik multinomial.

Namun, regresi logistik memiliki beberapa keterbatasan, diantaranya tidak boleh ada multikolinearitas yakni korelasi yang kuat antar variabel prediktor. Malau dkk. (Malau, Sawaluddin, Sutarman and Joseph, 2023) menunjukkan bahwa regresi logistik dapat mengalami overfitting ketika terdapat kombinasi linear antar variabel independen dan ketika model menunjukkan performa baik pada data latih namun buruk pada data uji. Ini menunjukkan adanya korelasi antara multikolinearitas dan overfitting dalam regresi logistik. Multikolinearitas dapat memengaruhi stabilitas model dengan membuat matriks analisis regresi menjadi singular, mengakibatkan ketidakstabilan nilai estimasi. Sementara itu, overfitting terjadi karena kompleksitas model yang berlebihan. Stabilitas model regresi logistik mengacu pada kemampuan model untuk mempertahankan kinerjanya pada data uji yang belum pernah dilihat sebelumnya. Sementara itu, Model yang terlalu kompleks cenderung tidak dapat digeneralisasi pada data baru yang tidak memiliki pola atau karakteristik yang sebanding atau konsisten. Oleh karena itu, penanganan multikolinearitas dan overfitting menjadi penting dalam pengembangan model regresi logistik yang stabil dan dapat diandalkan. Firth (Firth, 1993) mengusulkan Metode Penalized Maximum Likelihood Estimation dapat digunakan untuk mengatasi hal tersebut. Metode ini melibatkan penambahan penalti pada fungsi likelihood yang digunakan dalam estimasi parameter.

Penalized Maximum Likelihood Estimation (PMLE) adalah sebuah pendekatan statistik yang digunakan untuk mengatasi masalah kompleksitas model dalam analisis regresi. PMLE membantu mengontrol kompleksitas model dengan menerapkan penalti

pada koefisien regresi, yang pada gilirannya membantu mencapai keseimbangan yang lebih baik, sehingga menghasilkan model yang lebih stabil dan generalisasi yang lebih baik pada data yang tidak terlihat sebelumnya. (Lee, 2020)

Untuk mengestimasi parameter MLE dapat dilakukan dengan beberapa pendekatan, diantaranya metode Newton-Raphson. Akan tetapi, Metode Newton-Raphson memiliki beberapa permasalahan seperti ketergantungan pada titik awal yang dekat dengan solusi, serta sensitivitas terhadap pemilihan titik awal yang tepat. Untuk mengatasi ini, algoritma *gradient descent* menjadi solusi yang potensial. Algoritma *gradient descent* dapat menyelesaikan masalah ketergantungan pada titik awal dengan lebih baik karena tidak memerlukan titik awal yang presisi. Selain itu, sensitivitas terhadap pemilihan titik awal juga bisa diatasi dengan lebih baik menggunakan algoritma *gradient descent* karena metode ini lebih adaptif dan fleksibel dalam menemukan solusi tanpa bergantung secara krusial pada titik awal yang diberikan. Adaptif pada Algoritma *gradient descent* merujuk pada metode optimisasi yang menyesuaikan dengan laju pembelajaran (learning rate) berdasarkan karakteristik data dan proses optimisasi, memungkinkan penemuan solusi yang lebih stabil bahkan dengan variasi besar pada titik awal, yang merupakan kelemahan yang umum pada metode Newton-Raphson.

Dalam konteks optimisasi numerik, tujuan utama algoritma *gradient descent* adalah untuk menemukan nilai minimum (atau maksimum) dari suatu fungsi matematis. Prosesnya dimulai dari titik awal yang dipilih, dan kemudian nilai titik tersebut diperbarui secara iteratif dengan langkah-langkah yang proposional terhadap negatif gradien dari fungsi tersebut. Dengan kata lain, jika gradien menunjukkan arah di mana fungsi meningkat, maka langkah-langkah yang diambil oleh algoritma *gradient descent* akan berlawanan arah untuk mengurangi nilai fungsi. Beberapa varian dari algoritma *gradient descent* yang menggunakan teknik adaptif antara lain adalah AdaGrad, Adadelta, RMSprop, Adam, dan Nadam.

Penelitian ini juga memiliki kontribusi terhadap pengembangan metode statistik. Penerapan PMLE dengan algoritma *gradient descent* dalam analisis regresi logistik multinomial adalah bagian dari perkembangan metode statistik yang terus berlanjut untuk meningkatkan kualitas analisis data. Hal ini dapat memotivasi penelitian lebih lanjut dalam pengembangan teknik statistik yang lebih canggih.

1.2 Rumusan Masalah

Berdasarkan uraian dari latar belakang penelitian ini, MLE merupakan metode umum yang digunakan untuk mengestimasi parameter dari regresi logistik. Namun, regresi logistik memiliki beberapa batasan, diantaranya tidak adanya gejala multikolinearitas.

Firth (Firth, 1993) mengusulkan sebuah estimator likelihood yang dikenai penalti dan telah terbukti mengurangi masalah tersebut yang dikenal sebagai PMLE. Untuk mengestimasi parameter MLE ataupun PMLE dapat dilakukan dengan beberapa pendekatan salah satunya menggunakan metode Newton-Raphson. Akan tetapi, Metode Newton-Raphson memiliki beberapa permasalahan seperti ketergantungan pada titik awal yang dekat dengan solusi, serta sensitivitas terhadap pemilihan titik awal yang tepat.

1.3 Batasan Masalah

Adapun batasan masalah dari penelitian ini adalah sebagai berikut:

1. Penelitian ini akan menggunakan *penalized gradient descent* dengan regularisasi L_1 (LASSO) dan L_2 (Ridge)
2. Penelitian ini akan mengimplementasikan metode MLE dan PMLE *gradient descent* pada model regresi logistik multinomial dengan menggunakan bahasa pemrograman Python

1.4 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah mengestimasi parameter dan melakukan perbandingan performa pada model regresi logistik multinomial menggunakan metode MLE dan PMLE dengan algoritma *gradient descent*

1.5 Manfaat Penelitian

Adapun manfaat dari penelitian ini adalah mendorong perkembangan metode statistik dengan menggabungkan MLE dan *penalized gradient descent* berpotensi menginspirasi penelitian lebih lanjut dalam pengembangan teknik statistik yang lebih canggih

BAB 2

TINJAUAN PUSTAKA

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

2.1 Regresi Logistik

Regresi Logistik adalah regresi dengan variabel respon berskala kategorik (nominal atau ordinal). Regresi logistik terbagi atas regresi logistik biner, regresi logistik multinomial, serta regresi logistik ordinal. Regresi logistik biner adalah regresi logistik dengan variabel respon berskala nominal dua kategori. Regresi logistik multinomial adalah regresi logistik dengan variabel respon berskala nominal lebih dari dua kategori. Regresi logistik ordinal adalah regresi logistik dengan variabel respon berskala ordinal.

2.2 Regresi Logistik Biner

Regresi logistik biner adalah teknik statistik yang umum digunakan dalam berbagai bidang, termasuk penelitian medis, epidemiologi, dan ilmu sosial. Metode ini sangat cocok untuk menganalisis variabel respons biner, di mana hasilnya hanya bisa "ya" atau "tidak", seperti kehadiran atau ketiadaan penyakit atau keberhasilan atau kegagalan suatu peristiwa. Ini memungkinkan peneliti untuk memodelkan hubungan antara satu atau lebih variabel prediktor dan probabilitas hasil biner.

Regresi logistik biner melibatkan pemodelan logaritma natural dari peluang hasil biner. Fungsi logistik, juga dikenal sebagai fungsi sigmoid, mengubah kombinasi linear variabel prediktor menjadi nilai probabilitas antara 0 dan 1. Persamaan yang biasanya digunakan dalam regresi logistik biner telah didokumentasikan dengan baik dalam berbagai buku teks statistik dan artikel jurnal, memberikan dasar yang kuat bagi peneliti untuk mengaplikasikannya (Shedriko, 2021).

Menurut Hosmer & Lemeshow (Hosmer and Lemeshow, 2000), persamaan

regresi logistik adalah sebagai berikut:

$$Pr(Y = 1|X = x) = \pi(x) = \frac{1}{1 + e^{-g(x)}} \quad (2.1)$$

Dengan:

$$g(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p \quad (2.2)$$

$$= \beta_0 + \sum_{i=1}^p \beta_i x_i \quad (2.3)$$

Dimana

$Pr(Y = 1|X = x)$ = kelas positif yang diestimasi

β_0 = konstanta

β_i = koefisien parameter variabel prediktor ke-i (i=1,2,3,...,p)

x_i = variabel prediktor ke-i

2.3 Regresi Logistik Multinomial

Menurut Agresti Agresti (2013). Model regresi logistik multinomial dinyatakan dengan:

$$\pi_k(X) = Pr(Y = k|X) = \frac{\exp(\alpha_k + \beta_k^T X)}{1 + \sum_{h=1}^{K-1} \exp(\alpha_h + \beta_h^T X)} \quad (2.4)$$

Dimana

$Pr(Y = k|X_m)$ = peluang kelas-k yang diestimasi dengan observasi ke-m

α = konstanta

β_i = koefisien parameter variabel prediktor ke-i (i=1,2,3,...,n)

X_{mi} = variabel prediktor ke-i pada observasi ke-m

2.4 Log Odds atau Logit

Logit merupakan log natural (ln) dari *odds*. Dimana *odds* merupakan peluang kejadian sukses dibandingkan dengan peluang gagal. Secara umum fungsi logit variabel respon ke-j pada variabel prediktor ke-i dinyatakan sebagai berikut:

$$\begin{aligned} \ln \left(\frac{\pi_j(x)}{\pi_J(x)} \right) &= \alpha_j + \beta_{j1} x_1 + \beta_{j2} x_2 + \cdots + \beta_{jp} x_p \\ &= \alpha_j + \beta_j^T x, \end{aligned} \quad j = 1, 2, \dots, J - 1 \quad (2.5)$$

Persamaan 2.5 dapat ditulis ulang sebagai:

$$\begin{aligned}\frac{\pi_j(x)}{\pi_J(x)} &= \exp \alpha_j + \beta_j^T x \\ \pi_j(x) &= \pi_J(x) \exp \alpha_j + \beta_j^T x\end{aligned}\quad (2.6)$$

Keterangan:

x_i = vektor yang memuat nilai-nilai pengamatan ke- i dari variabel prediktor ke- j

$\pi_j(x_i)$ = peluang kategori respon ke- j pada variabel prediktor ke- i

$g_j(x_i)$ = fungsi logit variabel respon untuk kategori ke- j pada variabel prediktor ke- i

β_{jk} = koefisien model kategori ke- j variabel prediktor ke- k

x_{ki} = nilai variabel prediktor ke- k pengamatan ke- i

Karena $\pi_j(x) = P(Y = j|x)$ dengan x merupakan variabel prediktor dan $\sum_j \pi_j(x) = 1$, maka ,

$$\begin{aligned}\pi_J(x) &= 1 - \sum_{j=1}^{J-1} \pi_j(x) \\ \pi_J(x) &= 1 - \sum_{j=1}^{J-1} \pi_J(x) \exp (\alpha_j + \beta_j^T x) \\ \pi_J(x) &= 1 - \pi_J(x) \sum_{j=1}^{J-1} \exp (\alpha_j + \beta_j^T x) \\ \pi_J(x) + \pi_J(x) \sum_{j=1}^{J-1} \exp (\alpha_j + \beta_j^T x) &= 1 \\ \pi_J(x) \left(1 + \sum_{j=1}^{J-1} \exp (\alpha_j + \beta_j^T x) \right) &= 1 \\ \pi_J(x) &= \frac{1}{1 + \sum_{j=1}^{J-1} \exp (\alpha_j + \beta_j^T x)}\end{aligned}\quad (2.7)$$

Definisi 2.1. Asumsikan $\pi_j(x) = P(Y = j|x)$

2.5 Maximum Likelihood Estimation

Definisi 2.2. Misalkan y_1, y_2, \dots, y_m merupakan sebuah sampel acak berukuran m dari fungsi densitas peluang kontinu $f_m(y; \beta)$ dimana β adalah parameter yang tidak diketahui. Fungsi *likelihood*,

$$L(y; \beta) = \prod_{m=1}^M f_m(y; \beta) \quad (2.8)$$

Menurut Agresti(Agresti, 2013), fungsi *likelihood* pada regresi logistik multinomial dengan $f_m(y; \beta) = \prod_{k=1}^K \pi_k (X_m)^{y_{mk}}$ dan $\pi_k (X_m)^{y_{mk}}$ merupakan peluang bahwa observasi ke- k memiliki kelas y_k berdasarkan variabel X_m dan parameter model (β) seperti berikut:

$$L(y; \beta) = \prod_{m=1}^M \prod_{k=1}^K \pi_k (X_m)^{y_{mk}} \quad (2.9)$$

Dan fungsi *log likelihood*-nya:

$$\begin{aligned} \ln L(y; \beta) &= \sum_{m=1}^M \sum_{k=1}^K y_{mk} \ln (\pi_k (X_m)) \\ &= \sum_{m=1}^M \left\{ \sum_{k=1}^{K-1} y_{mk} (\alpha_k + \beta_k^T X_m) - \ln \left[1 + \sum_{k=1}^{K-1} \exp (\alpha_k + \beta_k^T X_m) \right] \right\} \end{aligned} \quad (2.10)$$

Pada regresi logistik, estimasi parameter (β) didapatkan dengan memaksimalkan fungsi *log-likelihood* nya. Karena fungsi logaritma natural adalah fungsi naik, sehingga untuk mencari nilai maksimum dari fungsi *log-likelihood* dapat ditentukan dengan turunan pertama dari fungsi *log-likelihood* terhadap β sama dengan nol, atau ditulis sebagai:

$$\frac{\partial (\ln L(y; \beta))}{\partial \beta} = 0 \quad (2.11)$$

Multikolinieritas diperkenalkan Allen & Frisch(Allen and Frisch, 1935). Multikolinieritas memperlihatkan adanya hubungan yang linier dengan tingkat sempurna di antara variabel-variabel prediktor dalam model regresinya. Multikolinieritas yaitu kondisi terjadi korelasi yang sangat kuat antara variabel prediktor (X) yang dilibatkan dalam pembangunan model regresi linier. Kolinieritas artinya hubungan linier tunggal, adapun multikolinieritas yaitu adanya lebih dari satu hubungan linier yang hampir sempurna. Pada penerapannya sering tidak dibedakan baik satu hubungan ataupun lebih di pergunakan penamaan multikolinieritas.

Multikolinieritas memiliki banyak dampak, yaitu:

1. Estimasi koefisien regresi yang tidak akurat.

2. Standar kesalahan yang terlalu tinggi pada estimasi koefisien regresi.
3. Pengujian t yang terdegradasi dan interval kepercayaan yang lebih lebar untuk pengujian signifikansi koefisien regresi.
4. Ketidaksignifikan palsu yang ditentukan oleh nilai-nilai p.
5. Penurunan dalam prediktabilitas model.

Menurut Young (Young, 2017), Salah satu ukuran untuk menguji efek multikolinieritas dengan *Variance Inflation Factors* (VIF). Nilai VIF dihitung dengan persamaan yaitu:

$$VIF_j = \frac{1}{1 - R_j^2} \quad (2.12)$$

Dimana R_j^2 merupakan koefisien determinasi yang diperoleh dengan melakukan regresi x_j terhadap variabel yang tersisa. Aturan umumnya adalah bahwa jika $VIF_j = 1$, maka tidak ada multikolinieritas. jika $1 < VIF_j < 5$, maka kemungkinan terdapat multikolinieritas yang sedang. Dan jika $VIF_j \geq 5$, maka terdapat multikolinieritas yang kuat.

2.6 Koefisien Determinasi

Dalam regresi logistik multinomial, koefisien determinasi atau *pseudo-R-squared* yang digunakan adalah koefisien determinasi Cox dan Snell, Nagelkerke, dan McFadden.

Nilai-nilai ini menunjukkan bahwa variasi data variabel prediktor dapat menjelaskan variasi data variabel respon, sisa dari 100% menunjukkan bahwa variasi variabel respon dijelaskan oleh variabel lain di luar variabel prediktor yang berada pada model

2.7 Regularisasi

2.7.1 Regularisasi L_1 atau LASSO (*Least Absolute Shrinkage and Selection Operator*)

Regularisasi L_1 , juga dikenal sebagai Regularisasi LASSO (Least Absolute Shrinkage and Selection Operator), menambahkan istilah hukuman ke dalam fungsi kerugian yang sebanding dengan nilai absolut koefisien model. Metode ini mengecilkan beberapa koefisien menjadi tepat nol, yang efektif melakukan seleksi pada variabel prediktor. Istilah regularisasi L_1 direpresentasikan sebagai:

$$Penalty(\beta_i, \lambda) = \lambda \sum_{i=1}^n |\beta_i| \quad (2.13)$$

Di sini, λ mengontrol kekuatan regularisasi, dan β_i adalah koefisien model yang terkait dengan variabel prediktor (Hastie, Tibshirani and Friedman, 2009).

2.7.2 Regularisasi L_2 atau Ridge

Regularisasi L_2 , juga dikenal sebagai Regularisasi Ridge, menambahkan istilah hukuman ke dalam fungsi kerugian yang sebanding dengan kuadrat koefisien model. Metode ini tidak memaksakan koefisien model untuk menjadi tepat nol, tetapi mengecilkan mereka menuju nol. Istilah regularisasi L_2 direpresentasikan sebagai:

$$Penalty(\beta_i, \lambda) = \lambda \sum_{i=1}^n \beta_i^2 \quad (2.14)$$

Di sini, λ mengontrol kekuatan regularisasi, dan β_i adalah koefisien model yang terkait dengan variabel prediktor (Bishop, 2006).

2.8 Algoritma *Gradient Descent*

Dalam matematika, algoritma *gradient descent* digunakan untuk menemukan minimum atau maksimum suatu fungsi. Fungsi yang dihadapi dapat merepresentasikan berbagai hal, seperti fungsi biaya dalam ekonomi, permukaan potensial dalam fisika, atau fungsi matematika kompleks lainnya. Prinsip dasar dari algoritma *gradient descent* adalah mengikuti arah di mana gradien (atau turunan) dari fungsi tersebut menunjukkan penurunan atau kenaikan paling cepat.

Konsep ini mirip dengan mencari lembah (minimum) atau puncak (maksimum) dalam topografi fungsi matematika. Dalam algoritma *gradient descent*, titik awal dipilih, dan kemudian perlahan-lahan bergerak menuju minimum atau maksimum dengan mengikuti arah penurunan atau kenaikan gradien. Proses ini dilakukan secara iteratif hingga ditemukan minimum atau maksimum yang diinginkan.

Algoritma *gradient descent* memiliki berbagai variasi, termasuk metode *steepest descent*, metode *conjugate gradient*, dan lainnya, yang digunakan tergantung pada konteks dan sifat fungsi yang dianalisis. Ini adalah alat matematis penting dalam optimisasi dan digunakan dalam berbagai bidang matematika dan ilmu terapan untuk menyelesaikan masalah optimisasi.

Definisi 2.3. Berdasarkan Bishop (Bishop, 2006), Persamaan utama dari algoritma *gradient descent* dinyatakan sebagai berikut:

$$\theta_{t+1} = \theta_t - \alpha \nabla J(\theta_t) \quad (2.15)$$

Di sini, θ mewakili parameter yang sedang dioptimalkan, α (alpha) mengindikasikan tingkat pembelajaran, yang mengendalikan ukuran langkah, dan $\nabla J(\theta_t)$ mewakili gradien dari fungsi biaya atau kerugian yang berkaitan dengan parameter pada iterasi saat ini t . Persamaan ini menangkap aturan pembaruan fundamental dalam algoritma *gradient descent*, dimana parameter disesuaikan untuk meminimalkan fungsi biaya.

Pada PMLE, estimasi parameter (β) didapatkan dengan minimalkan fungsi biayanya ($\min l(y; \beta)$). Karena fungsi logaritma natural adalah fungsi naik, sehingga fungsi biayanya tersebut menjadi penjumlahan fungsi negatif *log-likelihood* dan fungsi penaltinya, atau ditulis sebagai:

$$\min_{\beta} l(y; \beta) = -\ln L(y; \beta) + \text{penalty}(\beta; \lambda) \quad (2.16)$$

Untuk memperoleh nilai minimum dari fungsi biaya tersebut, dapat diperoleh melalui turunan pertama jika,

$$\frac{\partial (l(y; \beta))}{\partial \beta} = 0 \quad (2.17)$$

Dengan,

$$\frac{\partial (l(y; \beta))}{\partial \beta} = -\frac{\partial (\ln L(y; \beta))}{\partial \beta} + \frac{\partial (\text{penalty}(\beta; \lambda))}{\partial \beta} \quad (2.18)$$

Sehingga untuk mengestimasi parameter β dengan algoritma *gradient descent* dapat dihitung sebagai berikut:

$$\beta_i^{(t+1)} = \beta_i^{(t)} - \alpha \cdot \frac{\partial (l(y; \beta^{(t)}))}{\partial \beta^{(t)}} \quad (2.19)$$

Keterangan:

$\beta^{(t)}$ = estimasi parameter PMLE pada iterasi ke- t

α = besar langkah pergerakan penurunan *gradient*

$l(y; \beta)$ = fungsi biaya

2.9 Tabel klasifikasi

Sebuah tabel klasifikasi atau yang lebih dikenal sebagai matriks konfusi adalah matriks yang merangkum kinerja klasifikasi dari sebuah model dengan mengacu pada beberapa data uji. Matriks tersebut menunjukkan berapa banyak prediksi yang benar dan salah per kelas, yang membantu dalam memahami kinerja model. Kasus khusus dari tabel klasifikasi sering digunakan dengan dua kelas, salah satunya ditetapkan sebagai kelas positif dan yang lainnya sebagai kelas negatif.

Dalam konteks ini, empat sel matriks tersebut ditandai sebagai *true positives* (TP), *false positives* (FP), *true negatives* (TN), dan *false negatives* (FN). Sejumlah ukuran kinerja klasifikasi didefinisikan dalam hal empat sel ini, seperti akurasi, presisi, recall, dan skor F1. Tabel klasifikasi sangat berguna ketika dataset tidak seimbang, dan akurasi saja tidak cukup untuk mengevaluasi kinerja model.

Tabel klasifikasi adalah cara yang ringkas dan terstruktur untuk mendapatkan informasi lebih mendalam tentang suatu pengklasifikasi, yang dihitung dengan memetakan hasil yang diharapkan (atau sebenarnya) ke hasil yang diprediksi oleh model. Ini memberikan banyak informasi tentang kinerja model, seperti jumlah TP, FP, FN, dan TN. Paket scikit-learn dalam bahasa pemrograman Python mengandung semua alat untuk menghitung tabel klasifikasi, dan dapat diperoleh dengan menggunakan fungsi "confusion_matrix" dan memasukkan distribusi label sebenarnya dan distribusi label yang diprediksi sebagai argumen. Tabel klasifikasi tidak terbatas pada klasifikasi biner dan dapat digunakan dalam pengklasifikasi multikelas juga.

Kelas Prediksi / Kelas Aktual	positif	negatif
positif	TP	FP
negatif	FN	TN

Tabel 2.1 Tabel klasifikasi biner atau dua kelas

BAB 3

METODE PENELITIAN

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

BAB 4

HASIL DAN PEMBAHASAN

Pada bab ini, akan diuraikan tentang hasil dan pembahasan penelitian ini, yang meliputi: estimasi parameter model regresi logistik multinomial menggunakan metode PMLE, pengolahan data set dengan regresi logistik multinomial dengan regularisasi L_1 dan L_2 menggunakan algoritma *gradient descent*.

4.1 Estimasi Parameter Model Regresi Logistik Multinomial menggunakan Metode MLE

Model regresi logistik multinomial dinyatakan dengan:

$$\pi_k(X) = Pr(Y = k|X) = \frac{\exp(\alpha_k + \beta_k^T X)}{1 + \sum_{h=1}^{K-1} \exp(\alpha_h + \beta_h^T X)} \quad (4.1)$$

Keterangan:

$Pr(Y = k|X_m)$ = peluang kelas-k yang diestimasi dengan observasi ke-m (m=1,2,3,...,M)

dengan menggunakan persamaan 2.11 pada 2.10:

$$\begin{aligned} \frac{\partial (\ln L(y; \beta))}{\partial \beta} &= \frac{\partial \left(\sum_{i=1}^n \left\{ \sum_{j=1}^{J-1} y_{ij} (\alpha_j + \beta_j^T x_i) - \ln \left[1 + \sum_{j=1}^{J-1} \exp (\alpha_j + \beta_j^T x_i) \right] \right\} \right)}{\partial \beta_{jk}} \\ &= \frac{\partial \left(\sum_{i=1}^n \sum_{j=1}^{J-1} y_{ij} (\alpha_j + \beta_j^T x_i) \right)}{\partial \beta_{jk}} - \frac{\partial \left(\sum_{i=1}^n \ln \left[1 + \sum_{j=1}^{J-1} \exp (\alpha_j + \beta_j^T x_i) \right] \right)}{\partial \beta_{jk}} \end{aligned} \quad (4.2)$$

karena

$$\sum_{j=1}^{J-1} y_{ij} (\alpha_j + \beta_j^T x_i) = \sum_{j=1}^{J-1} y_{ij} \left(\alpha_j + \sum_{k=1}^p \beta_{jk} x_{ik} \right) \quad (4.3)$$

$$= \sum_{j=1}^{J-1} \left(y_{ij} \alpha_j + \sum_{k=1}^p y_{ij} \beta_{jk} x_{ik} \right) \quad (4.4)$$

Maka dengan menggunakan turunan parsial terhadap β_{jk} didapat,

$$\begin{aligned} \frac{\partial \left(\sum_{j=1}^{J-1} y_{ij} (\alpha_j + \beta_j^T x_i) \right)}{\partial \beta_{jk}} &= \frac{\partial \left(\sum_{j=1}^{J-1} (y_{ij} \alpha_j + \sum_{k=1}^p y_{ij} \beta_{jk} x_{ik}) \right)}{\partial \beta_{jk}} \\ &= y_{ik} x_{ik} \end{aligned} \quad (4.5)$$

sehingga persamaan 4.2 menjadi

$$\begin{aligned} \frac{\partial (\ln L(y; \beta_{jk}))}{\partial \beta_{jk}} &= \sum_{i=1}^n y_{ij} x_{ik} - \frac{\exp(\alpha_j + \beta_j^T x_i)}{1 + \sum_{j=1}^{J-1} \exp(\alpha_j + \beta_j^T x_i)} x_{ik} \\ &= \sum_{i=1}^n y_{ij} x_{ik} - [\pi_j(x_i)] x_{ik} \\ &= \sum_{i=1}^n [y_{ij} - \pi_j(x_i)] x_{ik} \end{aligned} \quad (4.6)$$

$$\frac{\partial (\ln L(y; \beta))}{\partial \beta} = x^T (y - \pi) \quad (4.7)$$

4.2 Estimasi Parameter Model Regresi Logistik Multinomial menggunakan Metode LASSO dan Ridge

4.2.1 Menggunakan Metode LASSO

Pada PMLE dengan LASSO, mencari nilai minimum dari fungsi negatif *Log-Likelihood* dijumlahkan dengan hasil kali norma β terhadap ruang L_1 dan kekuatan regularisasi (λ) yang dapat ditulis seperti persamaan 2.16.

Untuk mencari nilai minimum tersebut, akan dilakukan differensiasi terhadap parameter model Regresi Logistik (β) seperti pada persamaan 2.18. Dengan mensubstitusikan 2.11

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy

pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetur adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacus congue quam, in hendrerit risus eros eget felis. Maecenas eget erat in sapien mattis porttitor. Vestibulum porttitor. Nulla facilisi. Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacus. Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus vel est. Curabitur consectetur.

Suspendisse vel felis. Ut lorem lorem, interdum eu, tincidunt sit amet, laoreet vitae, arcu. Aenean faucibus pede eu ante. Praesent enim elit, rutrum at, molestie non, nonummy vel, nisl. Ut lectus eros, malesuada sit amet, fermentum eu, sodales cursus, magna. Donec eu purus. Quisque vehicula, urna sed ultricies auctor, pede lorem egestas dui, et convallis elit erat sed nulla. Donec luctus. Curabitur et nunc. Aliquam dolor odio, commodo pretium, ultricies non, pharetra in, velit. Integer arcu est, nonummy in, fermentum faucibus, egestas vel, odio.

Sed commodo posuere pede. Mauris ut est. Ut quis purus. Sed ac odio. Sed vehicula hendrerit sem. Duis non odio. Morbi ut dui. Sed accumsan risus eget odio. In hac habitasse platea dictumst. Pellentesque non elit. Fusce sed justo eu urna porta tincidunt. Mauris felis odio, sollicitudin sed, volutpat a, ornare ac, erat. Morbi quis dolor. Donec pellentesque, erat ac sagittis semper, nunc dui lobortis purus, quis congue purus metus ultricies tellus. Proin et quam. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent sapien turpis, fermentum vel, eleifend faucibus, vehicula eu, lacus.

Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Donec odio elit, dictum in, hendrerit sit amet, egestas sed, leo. Praesent feugiat sapien aliquet odio. Integer vitae justo. Aliquam vestibulum fringilla lorem. Sed neque lectus, consectetur at, consectetur sed, eleifend ac, lectus. Nulla facilisi. Pellentesque eget lectus. Proin eu metus. Sed porttitor. In hac habitasse platea dictumst. Suspendisse eu lectus. Ut mi mi, lacinia sit amet, placerat et, mollis vitae, dui. Sed ante tellus, tristique ut, iaculis eu, malesuada ac, dui. Mauris nibh leo, facilisis non, adipiscing quis, ultrices a, dui.

Morbi luctus, wisi viverra faucibus pretium, nibh est placerat odio, nec commodo wisi enim eget quam. Quisque libero justo, consectetur a, feugiat vitae, porttitor eu, libero. Suspendisse sed mauris vitae elit sollicitudin malesuada. Maecenas ultricies eros sit amet ante. Ut venenatis velit. Maecenas sed mi eget dui varius euismod. Phasellus aliquet volutpat odio. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Pellentesque sit amet pede ac sem eleifend consectetur. Nullam elementum, urna vel imperdiet sodales, elit ipsum pharetra ligula, ac pretium ante justo a nulla. Curabitur tristique arcu eu metus. Vestibulum lectus. Proin mauris. Proin eu nunc eu urna hendrerit faucibus. Aliquam auctor, pede consequat laoreet varius, eros tellus scelerisque quam, pellentesque hendrerit ipsum dolor sed augue. Nulla nec lacus.

4.3 Pembahasan 2

Suspendisse vitae elit. Aliquam arcu neque, ornare in, ullamcorper quis, commodo eu, libero. Fusce sagittis erat at erat tristique mollis. Maecenas sapien libero, molestie et, lobortis in, sodales eget, dui. Morbi ultrices rutrum lorem. Nam elementum ullamcorper leo. Morbi dui. Aliquam sagittis. Nunc placerat. Pellentesque tristique sodales est. Maecenas imperdiet lacinia velit. Cras non urna. Morbi eros pede, suscipit ac, varius vel, egestas non, eros. Praesent malesuada, diam id pretium elementum, eros sem dictum tortor, vel consectetur odio sem sed wisi.

Sed feugiat. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Ut pellentesque augue sed urna. Vestibulum diam eros, fringilla et, consectetur eu, nonummy id, sapien. Nullam at lectus. In sagittis ultrices mauris. Curabitur malesuada erat sit amet massa. Fusce blandit. Aliquam erat volutpat. Aliquam euismod. Aenean vel lectus. Nunc imperdiet justo nec dolor.

Etiam euismod. Fusce facilisis lacinia dui. Suspendisse potenti. In mi erat, cursus id, nonummy sed, ullamcorper eget, sapien. Praesent pretium, magna in eleifend egestas, pede pede pretium lorem, quis consectetur tortor sapien facilisis magna. Mauris quis magna varius nulla scelerisque imperdiet. Aliquam non quam.

Aliquam porttitor quam a lacus. Praesent vel arcu ut tortor cursus volutpat. In vitae pede quis diam bibendum placerat. Fusce elementum convallis neque. Sed dolor orci, scelerisque ac, dapibus nec, ultricies ut, mi. Duis nec dui quis leo sagittis commodo.

Aliquam lectus. Vivamus leo. Quisque ornare tellus ullamcorper nulla. Mauris porttitor pharetra tortor. Sed fringilla justo sed mauris. Mauris tellus. Sed non leo. Nullam elementum, magna in cursus sodales, augue est scelerisque sapien, venenatis congue nulla arcu et pede. Ut suscipit enim vel sapien. Donec congue. Maecenas urna mi, suscipit in, placerat ut, vestibulum ut, massa. Fusce ultrices nulla et nisl.

Etiam ac leo a risus tristique nonummy. Donec dignissim tincidunt nulla. Vestibulum rhoncus molestie odio. Sed lobortis, justo et pretium lobortis, mauris turpis condimentum augue, nec ultricies nibh arcu pretium enim. Nunc purus neque, placerat id, imperdiet sed, pellentesque nec, nisl. Vestibulum imperdiet neque non sem accumsan laoreet. In hac habitasse platea dictumst. Etiam condimentum facilisis libero. Suspendisse in elit quis nisl aliquam dapibus. Pellentesque auctor sapien. Sed egestas sapien nec lectus. Pellentesque vel dui vel neque bibendum viverra. Aliquam porttitor nisl nec pede. Proin mattis libero vel turpis. Donec rutrum mauris et libero. Proin euismod porta felis. Nam lobortis, metus quis elementum commodo, nunc lectus elementum mauris, eget vulputate ligula tellus eu neque. Vivamus eu dolor.

Nulla in ipsum. Praesent eros nulla, congue vitae, euismod ut, commodo a, wisi. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Aenean nonummy magna non leo. Sed felis erat, ullamcorper in, dictum non, ultricies ut, lectus. Proin vel arcu a odio lobortis euismod. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Proin ut est. Aliquam odio. Pellentesque massa turpis, cursus eu, euismod nec, tempor congue, nulla. Duis viverra gravida mauris. Cras tincidunt. Curabitur eros ligula, varius ut, pulvinar in, cursus faucibus, augue.

Nulla mattis luctus nulla. Duis commodo velit at leo. Aliquam vulputate magna et leo. Nam vestibulum ullamcorper leo. Vestibulum condimentum rutrum mauris. Donec id mauris. Morbi molestie justo et pede. Vivamus eget turpis sed nisl cursus tempor. Curabitur mollis sapien condimentum nunc. In wisi nisl, malesuada at, dignissim sit amet, lobortis in, odio. Aenean consequat arcu a ante. Pellentesque porta elit sit amet orci. Etiam at turpis nec elit ultricies imperdiet. Nulla facilisi. In hac habitasse platea dictumst. Suspendisse viverra aliquam risus. Nullam pede justo, molestie nonummy, scelerisque eu, facilisis vel, arcu.

4.4 Pembahasan 3

Curabitur tellus magna, porttitor a, commodo a, commodo in, tortor. Donec interdum. Praesent scelerisque. Maecenas posuere sodales odio. Vivamus metus lacus, varius quis, imperdiet quis, rhoncus a, turpis. Etiam ligula arcu, elementum a, venenatis quis, sollicitudin sed, metus. Donec nunc pede, tincidunt in, venenatis vitae, faucibus vel, nibh. Pellentesque wisi. Nullam malesuada. Morbi ut tellus ut pede tincidunt porta. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam congue neque id dolor.

Donec et nisl at wisi luctus bibendum. Nam interdum tellus ac libero. Sed sem justo, laoreet vitae, fringilla at, adipiscing ut, nibh. Maecenas non sem quis tortor eleifend fermentum. Etiam id tortor ac mauris porta vulputate. Integer porta neque vitae massa. Maecenas tempus libero a libero posuere dictum. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Aenean quis mauris sed elit commodo placerat. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Vivamus rhoncus tincidunt libero. Etiam elementum pretium justo. Vivamus est. Morbi a tellus eget pede tristique commodo. Nulla nisl. Vestibulum sed nisl eu sapien cursus rutrum.

Nulla non mauris vitae wisi posuere convallis. Sed eu nulla nec eros scelerisque pharetra. Nullam varius. Etiam dignissim elementum metus. Vestibulum faucibus, metus sit amet mattis rhoncus, sapien dui laoreet odio, nec ultricies nibh augue a enim. Fusce in ligula. Quisque at magna et nulla commodo consequat. Proin accumsan imperdiet sem. Nunc porta. Donec feugiat mi at justo. Phasellus facilisis ipsum quis ante. In ac elit eget ipsum pharetra faucibus. Maecenas viverra nulla in massa.

Nulla ac nisl. Nullam urna nulla, ullamcorper in, interdum sit amet, gravida ut, risus. Aenean ac enim. In luctus. Phasellus eu quam vitae turpis viverra pellentesque. Duis feugiat felis ut enim. Phasellus pharetra, sem id porttitor sodales, magna nunc aliquet nibh, nec blandit nisl mauris at pede. Suspendisse risus risus, lobortis eget, semper at, imperdiet sit amet, quam. Quisque scelerisque dapibus nibh. Nam enim. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nunc ut metus. Ut metus justo, auctor at, ultrices eu, sagittis ut, purus. Aliquam aliquam.

BAB 5

PENUTUP

5.1 Kesimpulan

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

5.2 Saran

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

DAFTAR PUSTAKA

Bibliografi

- Agresti, A. (2013), *Categorical Data Analysis*, 3 edn, John Wiley and Sons.
- Allen, R. G. D. and Frisch, R. (1935), 'Statistical confluence analysis by means of complete regression systems', *Econ. J. (London)* **45**, 741.
- Bishop, C. M. (2006), *Pattern Recognition and Machine Learning (Information Science and Statistics)*, Springer-Verlag.
- Firth, D. (1993), 'Bias reduction of maximum likelihood estimates', *Biometrika* **80**, 27–38.
- Hastie, T., Tibshirani, R. and Friedman, J. (2009), *The elements of statistical learning*, Springer New York.
- Hosmer, D. W. and Lemeshow, S. (2000), *Applied Logistic Regression*, Wiley.
- Lee, S. (2020), 'Logistic regression procedure using penalized maximum likelihood estimation for differential item functioning', *J. Educ. Meas.* **57**, 443–457.
- Malau, T. L., Sawaluddin, Sutarman and Joseph, T. (2023), 'Analisis metode logistik regresi ensemble untuk klasifikasi dengan pra-pemrosesan menggunakan principal component analysis', *IJM: Indonesian Journal of Multidisciplinary* **1**, 707–722.
URL: <https://journal.csspublishing.com/index.php/ijm/article/view/190>
- Shedriko, S. (2021), 'Binary logistic regression in determining affecting factors student graduation in a subject', *JURNAL TEKNOLOGI DAN OPEN SOURCE* **4**, 114–120.

LAMPIRAN

Lampiran 1. Nama Lampiran Pertama

Lampiran 2. Nama Lampiran Kedua