PENERAPAN *LEAST ABSOLUTE SHRINKAGE AND SELECTION OPERATOR* PADA *GEOGRAPHICALLY WEIGHTED LOGISTIC REGRESSION*

(Studi Kasus: *Unmet Need* KB di Sulawesi Selatan)

SKRIPSI

DIAN AYU PERMATA SARI RUSDY

H051191042

PROGRAM STUDI STATISTIKA DEPARTEMEN STATISTIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS HASANUDDIN

MAKASSAR

2024

PENERAPAN *LEAST ABSOLUTE SHRINKAGE AND SELECTION OPERATOR* PADA *GEOGRAPHICALLY WEIGHTED LOGISTIC REGRESSION*

(Studi Kasus: *Unmet Need* KB di Sulawesi Selatan)

SKRIPSI

Diajukan sebagai salah satu syarat memperoleh gelar Sarjana Statistika pada Program Studi Statistika Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin

DIAN AYU PERMATA SARI RUSDY

H051191042

PROGRAM STUDI STATISTIKA DEPARTEMEN STATISTIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS HASANUDDIN

MAKASSAR

MARET 2024

LEMBAR PERNYATAAN KEOTENTIKAN

Saya yang bertanda tangan di bawah ini menyatakan dengan sungguh-sungguh bahwa skripsi yang saya buat dengan judul:

Penerapan *Least Absolute Shrinkage And Selection Operator* Pada *Geographically Weighted Logistic Regression*

(Studi Kasus: *Unmet Need* KB di Sulawesi Selatan)

adalah benar hasil karya saya sendiri, bukan hasil plagiat dan belum pernah dipublikasikan dalam bentuk apapun.

Makassar, 27 Maret 2024

10000 METERAI TEMPEL 694DEALX081999033

Ayup

DIAN AYU PERMATA SARI RUSDY
NIM. H051191042

PENERAPAN *LEAST ABSOLUTE SHRINKAGE AND SELECTION OPERATOR* PADA *GEOGRAPHICALLY WEIGHTED LOGISTIC REGRESSION*

(Studi Kasus: *Unmet Need* KB di Sulawesi Selatan)

Disetujui Oleh:

Pembimbing Utama	Pembimbing Pertama
Sri Astuti Thamrin, S.Si., M.Stat., Ph.D	Dr. Anna Islamiyati, S.Si., M.Si.
NIP. 19740713 199903 2 001	NIP. 19770808 200501 2 002

Ketua Program Studi

Dr. Anna Islamiyati, S.Si., M.Si.
NIP. 19770808 200501 2 002

Pada 27 Maret 2024

HALAMAN PENGESAHAN

Skripsi ini diajukan oleh:

Nama	:	Dian Ayu Permata Sari Rusdy
NIM	:	H051191042
Program Studi	:	Statistika
Judul Skripsi	:	Penerapan *Least Absolute Shrinkage And Selection Operator* Pada *Geographically Weighted Logistic Regression* (Studi Kasus: *Unmet Need* KB di Sulawesi Selatan)

Telah berhasil dipertahankan dihadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Statistika pada Program Studi Statistika Fakultas Matematikan dan Ilmu Pengetahuan Alam Universitas Hasanuddin.

DEWAN PENGUJI

1. Ketua : Sri Astuti Thamrin, S.Si., M.Stat., Ph.D (.....................)
2. Sekretaris : Dr. Anna Islamiyati, S.Si., M.Si. (.....................)
3. Anggota : Drs Raupong, M.Si. (.....................)
4. Anggota : Dr. Nirwan, M.Si. (.....................)

Ditetapkan di : Makassar

Tanggal : 27 Maret 2024

KATA PENGANTAR

Assalamu'alaikum Warahmatullahi Wabarakatuh

Puji syukur penulis panjatkan kepada Allah *Subhanahu Wa Ta'ala* atas berkat dan rahmat-Nya yang telah dianugerahkan hingga saat ini, sehingga penulis dapat menyelesaikan tugas akhir ini. Shalawat serta salam senantiasa tercurahkan kepada baginda Rasulullah *Shallalahu 'Alaihi Wa Sallam* beserta keluarga dan para sahabatnya. *Alhamdulillahirobbil'alamin,* berkat nikmat kemudahan dan pertolongan yang diberikan oleh Allah *Subhanahu Wa Ta'ala*, penulis dapat menyelesaikan tugas akhir dengan judul **"Penerapan *Least Absolute Shrinkage And Selection Operator* Pada *Geographically Weighted Logistic Regression* (Studi Kasus: *Unmet Need* KB di Sulawesi Selatan)"** yang disusun sebagai salah satu syarat akademik untuk memperoleh gelar sarjana pada Program Studi Statistika Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin.

Dengan sepenuh hati, penulis menyadari bahwa dalam penyelesaian tugas akhir ini tidak akan terwujud tanpa bantuan dan dorongan dari berbagai pihak yang senantiasa memberikan dukungan dalam bentuk moril maupun materil. Walaupun penulis memiliki keterbatasan kemampuan dan pengetahuan, namun berkat bantuan serta dukungan yang diberikan, penulis dapat menyelesaikan tugas akhir ini dengan baik. Oleh karena itu, penulis mengungkapkan ucapan terima kasih yang setinggi-tingginya dan penghargaan yang tak terhingga kepada ibunda tercinta **Rahmatia Madjid** atas semua dukungan, doa, pengorbanan, cinta dan kasih sayang yang telah diberikan kepada penulis. Ucapan terima kasih juga penulis haturkan kepada kakak tersayang penulis **Surya Dedy Setyadi Rusdy** yang selalu menghibur dan memberi semangat kepada penulis dalam menyelesaikan skripsi ini.

Penghargaan yang tulus dan ucapan terima kasih dengan penuh keikhlasan juga penulis ucapkan kepada:

1. Bapak **Prof. Dr. Ir. Jamaluddin Jompa, M.Sc.**, selaku Rektor Universitas Hasanuddin beserta seluruh jajarannya.

2. **Bapak Dr. Eng. Amiruddin,** selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin beserta seluruh jajarannya.
3. **Ibu Dr. Anna Islamiyati, S.Si., M.Si.**, sebagai Ketua Departemen Statistika sekaligus Pembimbing Pertama penulis yang telah membekali ilmu kepada penulis selama menjadi mahasiswa di Departemen Statistika, serta dengan ikhlas meluangkan waktu dan pemikirannya untuk memberikan masukan dan arahan dalam penulisan tugas akhir ini.
4. **Ibu Sri Astuti Thamrin, S.Si., M.Stat., Ph.D**, selaku Penasehat Akademik sekaligus Pembimbing Utama yang dengan sabar telah meluangkan banyak waktunya dan pemikirannya untuk membimbing serta memberikan arahan dan dorongan semangat kepada penulis dari awal perkuliahan hingga selesainya penulisan tugas akhir ini.
5. **Bapak Drs Raupong, M.Si.** dan **Bapak Dr. Nirwan, M.Si.**, selaku Tim Penguji yang telah bersedia meluangkan waktunya untuk memberikan saran dan kritikan yang membangun dalam penulisan tugas akhir ini.
6. Segenap **Dosen Pengajar dan Staf Departemen Statistika** yang telah memberikan banyak ilmu dan bantuan kepada penulis selama menempuh pendidikan sarjana di Departemen Statistika.
7. Sahabat tercinta penulis **Calon Orang Sukses**, **A. Nurhikmah**, **Siti Nurhalisa Thamrin**, dan **Nurhilda Alqur'aini,** terima kasih telah menjadi sahabat dan saudara penulis. Terima kasih telah memberikan semangat dalam menyelesaikan perkuliahan, serta senantiasa mendengarkan segala keluh kesah penulis dalam hal apapun.
8. Sahabat terbaik penulis sejak awal perkuliahan **Refa Joyce Semida**, **Evlyn Pricilia Kondy**, **Nurul Magfirah**, **Rahma**, **Marhama**, **Wa Ode Sitti Amni**, dan **Nurmasyita Nasruddin**, terima kasih telah meluangkan waktunya untuk mendengar keluh kesah penulis, memberi semangat, dukungan dan menghibur penulis.
9. **Wahyu Dwi Rahmawati**, **Rahmah Ningsih Dwika Julia**, **Melinda Sari** **a Putri Laode**, **Inda Mutmainna**, **Arief Rahman Nur**, **Agus** **awan**, **Muhammad Ferdiansyah**, dan **Ummul Auliyah Syam** yang meluangkan waktunya untuk memberi bantuan kepada penulis

menghadapi kendala dalam perkuliahan serta dalam penyelesaikan tugas akhir ini.

10. Seluruh teman seperjuangan di **STATISTIKA 2019.** Terima kasih atas canda tawa, suka maupun duka yang telah kalian berikan selama perkuliahan.
11. Teman-teman **KKN 108 Pulau Sarappo.** Terima kasih untuk cerita suka dan duka perjuangannya.
12. **Keluarga Besar LDK MPM Unhas** dan **Keluarga Besar Himastat FMIPA Unhas** terima kasih atas ilmu dan pengalamannya yang telah menjadi tempat belajar bagi penulis.
13. Kepada seluruh pihak yang tidak sempat penulis sebutkan satu persatu. Terima kasih atas semua dukungan, partisipasi, dan apresiasinya yang diberikan kepada penulis.

Penulis juga menyadari bahwa skripsi ini masih jauh dari kata sempurna, namun ini hasil terbaik yang dapat diberikan oleh penulis pada penelitian ini. Oleh karena itu dengan segala kerendahan hati penulis mengucapkan permohonan maaf yang sebesar-besarnya. Akhir kata, semoga tulisan ini dapat memberikan manfaat untuk berbagai pihak.

Makassar, 27 Maret 2024

Dian Ayu Permata Sari Rusdy
NIM. H051191042

PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIK

Sebagai civitas akademik Universitas Hasanuddin, saya yang bertanda tangan di bawah ini:

Nama	: Dian Ayu Permata Sari Rusdy
NIM	: H051191042
Program Studi	: Statistika
Departemen	: Statistika
Fakultas	: Matematika dan Ilmu Pengetahuan Alam
Jenis Karya	: Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Hasanuddin **Hak Bebas Royalti Non-eksklusif** (***Non-exclusive Royalty- Free Right***) atas tugas akhir saya yang berjudul:

"Penerapan *Least Absolute Shrinkage And Selection Operator* Pada *Geographically Weighted Logistic Regression* (Studi Kasus: *Unmet Need* KB di Sulawesi Selatan)"

Beserta perangkat yang ada (jika diperlukan). Terkait dengan hal di atas, maka pihak universitas berhak menyimpan, mengalih-media/format-kan, mengelola dalam bentuk pangkalan data (*database*), merawat dan mempublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di Makassar, 27 Maret 2024

Yang Menyatakan,

DIAN AYU PERMATA SARI RUSDY

NIM. H051191042

ABSTRAK

Analisis regresi merupakan analisis yang digunakan untuk menaksir pola hubungan antara variabel respon dengan variabel prediktor. Salah satu analisis regresi yang umum digunakan adalah regresi logistik biner. Regresi logistik biner digunakan apabila variabel responnya bertipe kategorik dengan dua kemungkinan yaitu sukses atau gagal. Regresi logistik biner bersifat global, sehingga kurang tepat diterapkan pada data spasial. Oleh karena itu, regresi logistik biner dikembangkan menjadi *geographically weighted logistic regression* (GWLR). GWLR mempertimbangkan faktor lokasi yang dimasukkan ke dalam model melalui fungsi pembobot. GWLR tidak dapat mengatasi multikolinearitas. Multikolinearitas dapat menyebabkan hasil estimasi parameter model menjadi tidak signifikan, sehingga perlu diatasi. Salah satu metode untuk mengatasi multikolinearitas adalah *least absolute shrinkage and selection operator* (LASSO). Penelitian ini bertujuan untuk mendapatkan model GWLR dengan LASSO dan faktor yang berpengaruh, serta memperoleh performa model GWLR dengan LASSO terhadap *unmet need* KB di Sulawesi Selatan. Data yang digunakan adalah data *unmet need* KB, tingkat pendidikan, dan pengeluaran masyarakat di Sulawesi Selatan pada Tahun 2021. Nilai AIC model GWLR dengan LASSO, yaitu 31.918 lebih kecil dari nilai AIC model GWLR tanpa LASSO, yaitu 38.879; yang berarti GWLR dengan metode LASSO dapat memodelkan *unmet need* KB lebih baik dibandingkan model GWLR. Dari hasil penelitian, diperoleh bahwa status *unmet need* KB di 22 Kabupaten/Kota dipengaruhi oleh persentase wanita dengan pendidikan SMP/setara atau lebih rendah, banyak wanita usia subur, persentase suami/keluarga yang menolak terhadap KB, dan banyak petugas lapangan KB; serta 2 Kabupaten/Kota dipengaruhi oleh banyak wanita usia subur, persentase suami/keluarga yang menolak terhadap KB, dan banyak petugas lapangan KB.

Kata Kunci: *Unmet Need*, Keluarga Berencana, Regresi Logistik Biner, GWLR, Multikolinearitas, LASSO.

ABSTRACT

Regression analysis is used to estimate the pattern of association between response and predictor variables. A commonly known regression analysis is binary logistic regression. It is used in cases of categorized response variables with two categories: success or failure. Binary logistic regression is a global model, resulting in inappropriate implementation on spatial data. Therefore, it is developed into geographically weighted logistic regression (GWLR). GWLR considers location factors incorporated into the model through a weight function. GWLR is unable to overcome multicollinearity. Multicollinearity may cause the estimated parameters to be insignificant, thus it needs to be handled. One of the methods to deal with multicollinearity is least absolute shrinkage and selection operator (LASSO). This study focuses on estimating the GWLR model with LASSO and the influential factors, as well as obtaining the performance of the GWLR model with LASSO on unmet need for family planning (FP) in South Sulawesi. The data used are the unmet need for FP, education level, and expenditure data in South Sulawesi by 2021. The AIC score of GWLR model with LASSO, which is 31.918, is less than the AIC score of the GWLR model without LASSO, which is 38.879; indicating that GWLR with the LASSO is able to represent the unmet need for FP better than GWLR model. The results indicate that the status of unmet need for FP in 22 districts/cities are determined by the percentage of women with secondary school education or below, the number of reproductive-age women, the percentage of husbands/families who refuse FP, and the number of FP staff; while 2 districts/cities are determined by the number of reproductive-age women, the percentage of husbands/families who refuse family planning, and the number of FP staff.

Keywords: *Unmet Need, Family Planning, Binary Logistic Regression, GWLR, Multicollinearity, LASSO.*

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

DAFTAR LAMPIRAN

BAB I
PENDAHULUAN

1.1 Latar Belakang

Analisis regresi merupakan salah satu analisis dalam statistika yang digunakan untuk menaksir pola hubungan antara variabel respon dan variabel prediktor. Variabel respon merupakan variabel yang nilainya dipengaruhi oleh variabel prediktor. Sedangkan variabel prediktor merupakan variabel yang dianggap memiliki pengaruh terhadap variabel respon. Salah satu analisis regresi yang umum digunakan adalah regresi logistik biner (Alwi dkk., 2018).

Regresi logistik biner merupakan metode yang digunakan untuk menggambarkan hubungan antara variabel respon dengan satu atau lebih variabel prediktor. Variabel respon dari regresi logistik biner merupakan variabel bertipe kategorik dengan dua kemungkinan, yaitu sukses atau gagal (Kumar dkk., 2022). Analisis regresi logistik biner merupakan model yang yang bersifat global atau tidak melibatkan unsur lokasi geografis, sehingga kurang tepat apabila diterapkan pada data yang dipengaruhi oleh lokasi secara geografis atau yang biasa disebut dengan data spasial (Hastuti, 2022). Untuk mengatasi hal tersebut, metode regresi logistik biner kemudian dikembangkan menjadi *geographically weighted logistic regression* (GWLR).

GWLR merupakan metode statistik yang dapat digunakan untuk memodelkan hubungan antara variabel respon dengan variabel prediktor, serta mempertimbangkan faktor lokasi dengan asumsi bahwa data berdistribusi bernoulli (Salsabila dkk., 2023). Pada model GWLR, faktor lokasi dimasukkan kedalam model melalui fungsi pembobot. Fungsi pembobot ini akan mewakili letak data observasi satu terhadap data observasi lainnya. Model GWLR bersifat lokal sehingga memiliki koefisien regresi yang masing-masing bergantung pada lokasi dimana data tersebut diamati (Lestari dkk., 2020). Namun, model GWLR memiliki kekurangan, yaitu tidak dapat mengatasi masalah multikolinearitas.

M...ltikolinearitas terjadi karena adanya korelasi yang cukup tinggi di antara ...rediktor. Adanya multikolinearitas dapat diketahui dengan menggunakan ...*ance inflation factor* (VIF). Nilai VIF akan menjadi semakin besar jika ...korelasi yang semakin besar di antara variabel prediktor (Khariyani dkk., 2022). Jika VIF memiliki nilai lebih dari 10, maka multikolinearitas akan memberikan pengaruh yang serius pada estimasi model (Nakyeyune dkk., 2022). Adanya multikolinearitas akan menyebabkan hasil taksiran parameter model menjadi tidak signifikan, sehingga multikolinearitas ini perlu diatasi (Sriningsih dkk., 2018). Metode yang dapat digunakan untuk mengatasi multikolinearitas adalah *least absolute shrinkage and selection operator* (LASSO).

LASSO merupakan metode yang dapat digunakan untuk mengatasi multikolinearitas yaitu dengan melakukan penyeleksian variabel pada model. Penyeleksian ini dilakukan dengan cara menyusutkan nilai koefisien regresi. Penyusutkan nilai koefisien regresi oleh LASSO dapat dilakukan hingga mendekati nol atau tepat nol (Chen dkk., 2020). Penyusutan nilai ini menjadi kelebihan dari LASSO karena dapat menyeleksi variabel prediktor pada model sehingga hanya variabel yang dianggap berpengaruh saja yang dimasukkan ke dalam model sehingga dapat memudahkan dalam menginterpretasikan model (Pardede dkk., 2022). LASSO dapat diterapkan di berbagai bidang, salah satunya dalam bidang kesehatan, yaitu pada kasus *unmet need* Keluarga Berencana (KB).

Unmet need KB didefinisikan sebagai wanita yang memiliki usia produktif dan tidak ingin memiliki anak lagi ataupun ingin menunda anak yang berikutnya tetapi tidak menggunakan metode kontrasepsi (Yolanda & Destri, 2019). Tingginya angka *unmet need* KB di Indonesia menjadi salah satu permasalahan yang sampai saat ini belum terselesaikan dengan baik. Data statistik rutin Badan Kependudukan dan Keluarga Berencana Nasional (BKKBN) menunjukkan bahwa pada tahun 2020 terdapat 12,13 persen dari total Pasangan Usia Subur (PUS) merupakan *unmet need* KB. Kemudian angka ini naik menjadi 12,21 persen pada tahun 2021 (BKKBN, 2021). Hal ini menunjukkan bahwa angka *unmet need* KB mengalami peningkatan secara nasional. Peningkatan ini disebabkan tingginya *unmet need* KB di beberapa Provinsi, salah satunya Sulawesi Selatan.

Berdasarkan data statistik rutin BKKBN, angka *unmet need* KB di Sulawesi Selatan mencapai 13,12 persen pada tahun 2020. Angka ini kemudian meningkat 13,70 persen pada tahun berikutnya (BKKBN, 2021). Peningkatan angka dikan Sulawesi Selatan sebagai salah satu provinsi dengan angka *unmet* yang cukup tinggi di Indonesia. Angka *unmet need* KB yang tinggi akan meningkatkan risiko kehamilan yang tidak diinginkan. Hal ini juga akan berpengaruh terhadap peningkatan risiko kematian ibu dan bayi (Purba dkk., 2020).

Penelitian sebelumnya telah dilakukan oleh Meyfield dkk (2018) dan diperoleh kesimpulan bahwa metode GWLR lebih baik dalam memodelkan data spasial dibandingkan regresi logistik. Penelitian juga dilakukan oleh Lestari dkk (2020) dengan kesimpulan pemodelan data dengan metode LASSO pada *geographically weighted regression* (GWR) memberikan hasil yang lebih baik dibandingkan metode GWR. Penelitian lain juga dilakukan oleh Amalah (2020) yang melakukan estimasi parameter model GWLR pada data yang mengandung multikolinearitas dengan metode *ridge*.

Berdasarkan latar belakang dan penelitian-penelitian sebelumnya, peneliti tertarik untuk melakukan penelitian mengenai penerapan LASSO pada *geographically weighted logistic regression*, dengan data yang digunakan adalah data *unmet need* KB di Provinsi Sulawesi Selatan tahun 2021.

1.2 Rumusan Masalah

Rumusan masalah yang dibahas pada penelitian ini adalah:

1. Bagaimana model GWLR dengan LASSO dan faktor apa saja yang berpengaruh secara signifikan pada data *unmet need* KB di Sulawesi Selatan tahun 2021?
2. Bagaimana performa model GWLR dengan LASSO pada data *unmet need* KB di Sulawesi Selatan tahun 2021?

1.3 Batasan Masalah

Batasan masalah pada penelitian ini adalah:

1. Fungsi pembobot yang digunakan adalah fungsi pembobot *fixed gaussian*.
2. Data yang digunakan adalah data *unmet need* KB di Provinsi Sulawesi Selatan tahun 2021.

1.4 Tujuan Penelitian

Tujuan penelitian ini adalah:

1. Mendapatkan model GWLR dengan LASSO dan informasi mengenai faktor yang berpengaruh secara signifikan terhadap *unmet need* KB di Sulawesi Selatan tahun 2021.
2. Memperoleh performa model GWLR dengan LASSO pada data *unmet need* KB di Sulawesi Selatan tahun 2021.

1.5 Manfaat Penelitian

Manfaat yang diharapkan dari penelitian ini adalah:

1. Diharapkan dapat menambah pemahaman bagi pembaca terkait pemodelan GWLR dengan LASSO.
2. Hasil analisis data diharapkan dapat memberikan gambaran mengenai faktor-faktor yang berpengaruh terhadap *unmet need* KB di Sulawesi Selatan.

BAB II
TINJAUAN PUSTAKA

2.1 Standarisasi Data

Standarisasi data perlu dilakukan apabila pada variabel prediktor terdapat perbedaan ukuran skala yang cukup besar. Perbedaan satuan yang cukup besar akan menyebabkan hasil analisis menjadi kurang valid sehingga memungkinkan terjadinya kesalahan dalam menginterpretasikan hasil analisis (Ulinnuh & Veriani, 2020). Standarisasi data dilakukan dengan menggunakan Persamaan (2.1) (Bau dkk., 2023).

$$Z_{ik} = \frac{x_{ik} - \bar{x}_k}{s_k} \tag{2.1}$$

Keterangan:
$i = 1,2,\dots,n$ adalah banyaknya observasi/data.
$k = 1,2,\dots,p$ adalah banyak variabel prediktor
Z_{ik} adalah data standar untuk data ke-i, variabel ke-k
x_{ik} adalah data ke-i, variabel ke-k
$\bar{x}_k$ adalah rata-rata variabel ke-k
s_k adalah standar deviasi variabel ke-k

2.2 Regresi Logistik Biner

Regresi logistik merupakan suatu metode analisis statistika untuk menggambarkan hubungan antara variabel respon yang bersifat dikotomus (berskala nominal atau ordinal dengan dua kategori) atau polikotomus (berskala nominal atau ordinal dengan lebih dari dua kategori) dengan sekumpulan variabel prediktor yang bersifat kontinu atau kategorik (Nisva & Ratnasari, 2020). Regresi logistik biner merupakan metode analisis data yang digunakan untuk mendeskripsikan hubungan antara variabel respon yang bersifat biner atau dikotomus dengan satu atau lebih variabel prediktor. Variabel respon regresi logistik biner terdiri dari dua kategori, yaitu sukses yang dinotasikan dengan 1 atau gagal yang dinotasikan dengan 0 (Lanfranchi dkk., 2020). Variabel respon pada regresi [illegible]iner mengikuti distribusi Bernoulli dengan fungsi probabilitas seperti yang [illegible]an pada Persamaan (2.2) (Sofiyat dkk., 2023).

$$f(y_i) = \pi(x_i)^{y_i}\left(1 - \pi(x_i)\right)^{1-y_i} \tag{2.2}$$

Keterangan:

$\pi(x_i)$ adalah peluang kejadian ke-i

y_i adalah peubah acak ke-i yang terdiri dari 0 dan 1

Dalam regresi logistik, variabel respon dituliskan sebagai $y = \pi(x_i) + \varepsilon$, dengan ε mempunyai salah satu dari kemungkinan dua nilai, yaitu apabila $y = 1$ maka $\varepsilon = 1 - \pi(x_i)$ dengan peluang $\pi(x_i)$ dan jika $y = 0$ maka $\varepsilon = -\pi(x_i)$ dengan peluang $1 - \pi(x_i)$ (Amalah dkk., 2023). Adapun model probabilitas regresi logistik biner dapat dituliskan sebagaimana pada Persamaan (2.3).

$$\begin{aligned}\pi(x) &= \frac{\exp\left(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p\right)}{1 + \exp\left(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p\right)} \\ &= \frac{\exp\left(\beta_0 + \sum_{k=1}^{p} \beta_k x_k\right)}{1 + \exp\left(\beta_0 + \sum_{k=1}^{p} \beta_k x_k\right)} \end{aligned} \tag{2.3}$$

Apabila Persamaan (2.3) dituliskan menjadi model logit, maka diperoleh Persamaan (2.4).

$$\begin{aligned} logit[\pi(x)] &= \ln\left(\frac{\pi(x)}{1 - \pi(x)}\right) \\ &= \ln\left(\frac{\dfrac{\exp\left(\beta_0 + \sum_{k=1}^{p} \beta_k x_k\right)}{1 + \exp\left(\beta_0 + \sum_{k=1}^{p} \beta_k x_k\right)}}{1 - \dfrac{\exp\left(\beta_0 + \sum_{k=1}^{p} \beta_k x_k\right)}{1 + \exp\left(\beta_0 + \sum_{k=1}^{p} \beta_k x_k\right)}}\right) \\ &= \ln\left(\frac{\dfrac{\exp\left(\beta_0 + \sum_{k=1}^{p} \beta_k x_k\right)}{1 + \exp\left(\beta_0 + \sum_{k=1}^{p} \beta_k x_k\right)}}{\dfrac{1 + \exp\left(\beta_0 + \sum_{k=1}^{p} \beta_k x_k\right)}{1 + \exp\left(\beta_0 + \sum_{k=1}^{p} \beta_k x_k\right)} - \dfrac{\exp\left(\beta_0 + \sum_{k=1}^{p} \beta_k x_k\right)}{1 + \exp\left(\beta_0 + \sum_{k=1}^{p} \beta_k x_k\right)}}\right) \\ &= \ln\left(\frac{\dfrac{\exp\left(\beta_0 + \sum_{k=1}^{p} \beta_k x_k\right)}{1 + \exp\left(\beta_0 + \sum_{k=1}^{p} \beta_k x_k\right)}}{\dfrac{1}{1 + \exp\left(\beta_0 + \sum_{k=1}^{p} \beta_k x_k\right)}}\right) \\ &= \ln\left(\exp\left(\beta_0 + \sum_{k=1}^{p} \beta_k x_k\right)\right) \end{aligned}$$

$$= \beta_0 + \sum_{k=1}^{p} \beta_k x_k \tag{2.4}$$

Nilai β adalah parameter model dan $x_k, k = 1,2, \dots, p$ merupakan variabel prediktor ke-k (Alwi dkk, 2018).

2.3 Estimasi Parameter Regresi Logistik Biner

Estimasi parameter pada regresi logistik biner dilakukan dengan menggunakan metode *maximum likelihood estimation* (MLE). Metode MLE menghasilkan nilai parameter dengan memaksimalkan fungsi peluang dari data yang diamati dengan menggunakan fungsi *likelihood* yang ditampilkan pada Persamaan (2.5) (Wardana & Sari, 2020).

$$l(\boldsymbol{\beta}) = \prod_{i=1}^{n} [\pi(x_i)]^{y_i} [1 - \pi(x_i)]^{1-y_i}$$

$$= [\pi(x_i)]^{\sum_{i=1}^{n} y_i} [1 - \pi(x_i)]^{\sum_{i=1}^{n} (1-y_i)} \tag{2.5}$$

Adapun nilai $\pi(x_i)$ pada Persamaan (2.5) dituliskan pada Persamaan (2.6).

$$\pi(x_i) = \frac{\exp\left(\beta_0 + \sum_{k=1}^{p} \beta_k x_{ik}\right)}{1 + \exp\left(\beta_0 + \sum_{k=1}^{p} \beta_k x_{ik}\right)} = \frac{\exp(\boldsymbol{\beta}^T \boldsymbol{x}_i)}{1 + \exp(\boldsymbol{\beta}^T \boldsymbol{x}_i)} \tag{2.6}$$

$$\boldsymbol{\beta}^T = (\beta_0 \quad \beta_1 \quad \beta_2 \quad \beta_3 \quad \cdots \quad \beta_p)$$

$$\boldsymbol{x}_i^T = (1 \quad x_{i1} \quad x_{i2} \quad \cdots \quad x_{ip})$$

Untuk memudahkan perhitungan, fungsi $l(\boldsymbol{\beta})$ pada Persamaan (2.5) dapat dituliskan dengan menggunakan logaritma natural sehingga diperoleh Persamaan (2.7).

$$L(\boldsymbol{\beta}) = \ln l(\boldsymbol{\beta})$$

$$= \ln[\pi(x_i)]^{\sum_{i=1}^{n} y_i} [1 - \pi(x_i)]^{\sum_{i=1}^{n} (1-y_i)}$$

$$= \ln[\pi(x_i)]^{\sum_{i=1}^{n} y_i} + \ln[1 - \pi(x_i)]^{\sum_{i=1}^{n} (1-y_i)}$$

$$= \sum_{i=1}^{n} \left[y_i \ln \pi(x_i) + (1 - y_i) \ln\left(1 - \pi(x_i)\right)\right]$$

$$= \sum_{i=1}^{n} \left[y_i \ln \pi(x_i) + \ln\big(1 - \pi(x_i)\big) - y_i \ln\big(1 - \pi(x_i)\big) \right]$$

$$= \sum_{i=1}^{n} \left[y_i \big(\ln \pi(x_i) - \ln\big(1 - \pi(x_i)\big) \big) + \ln\big(1 - \pi(x_i)\big) \right]$$

$$= \sum_{i=1}^{n} \left[y_i \ln \left(\frac{\pi(x_i)}{1 - \pi(x_i)} \right) + \ln\big(1 - \pi(x_i)\big) \right] \tag{2.7}$$

Selanjutnya dengan mensubtitusi Persamaan (2.6) pada Persamaan (2.7) maka diperoleh Persamaan (2.8).

$$L(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left[y_i \ln \left(\frac{\frac{\exp(\boldsymbol{\beta}^T \boldsymbol{x}_i)}{1 + \exp(\boldsymbol{\beta}^T \boldsymbol{x}_i)}}{1 - \frac{\exp(\boldsymbol{\beta}^T \boldsymbol{x}_i)}{1 + \exp(\boldsymbol{\beta}^T \boldsymbol{x}_i)}} \right) + \ln \left(1 - \frac{\exp(\boldsymbol{\beta}^T \boldsymbol{x}_i)}{1 + \exp(\boldsymbol{\beta}^T \boldsymbol{x}_i)} \right) \right]$$

$$= \sum_{i=1}^{n} \left[y_i \ln(\exp(\boldsymbol{\beta}^T \boldsymbol{x}_i)) + \ln \left(\frac{1}{1 + \exp(\boldsymbol{\beta}^T \boldsymbol{x}_i)} \right) \right]$$

$$= \sum_{i=1}^{n} [y_i (\boldsymbol{\beta}^T \boldsymbol{x}_i) + \ln(1 + \exp(\boldsymbol{\beta}^T \boldsymbol{x}_i))^{-1}]$$

$$= \sum_{i=1}^{n} [y_i (\boldsymbol{\beta}^T \boldsymbol{x}_i) - \ln(1 + \exp(\boldsymbol{\beta}^T \boldsymbol{x}_i))] \tag{2.8}$$

Kemudian untuk memperoleh nilai estimasi $\boldsymbol{\beta}$, maka Persamaan (2.8) diturunkan terhadap $\boldsymbol{\beta}$ dan hasil turunan tersebut disamakan dengan nol, sehingga diperoleh Persamaan (2.9) (Lestari, 2020).

$$\frac{\partial L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \mathbf{0}$$

$$\frac{\partial}{\partial \boldsymbol{\beta}} \left(\sum_{i=1}^{n} [y_i (\boldsymbol{\beta}^T \boldsymbol{x}_i) - \ln(1 + \exp(\boldsymbol{\beta}^T \boldsymbol{x}_i))] \right) = \mathbf{0}$$

$$\sum_{i=1}^{n} y_i \boldsymbol{x}_i - \sum_{i=1}^{n} \frac{\boldsymbol{x}_i \exp(\boldsymbol{\beta}^T \boldsymbol{x}_i)}{1 + \exp(\boldsymbol{\beta}^T \boldsymbol{x}_i)} = \mathbf{0}$$

$$\sum_{i=1}^{n} \boldsymbol{x}_i \left[y_i - \frac{\exp(\boldsymbol{\beta}^T \boldsymbol{x}_i)}{1 + \exp(\boldsymbol{\beta}^T \boldsymbol{x}_i)} \right] = \mathbf{0}$$

$$\sum_{i=1}^{n} \boldsymbol{x}_i [y_i - \pi(x_i)] = \mathbf{0} \tag{2.9}$$

Persamaan (2.9) dapat dinyatakan dalam bentuk perkalian matriks sehingga diperoleh Persamaan (2.10).

$$\frac{\partial L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = X^T \big(\boldsymbol{y} - \boldsymbol{\pi}(\boldsymbol{x})\big) = \mathbf{0} \tag{2.10}$$

Keterangan:

X adalah matriks berukuran $n \times (p + 1)$ dengan elemen-elemennya yaitu

$$X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}$$

$\boldsymbol{y}^T = (y_1 \quad y_2 \quad \cdots \quad y_n)$

$\boldsymbol{\pi}(\boldsymbol{x})^T = (\pi(x_1) \quad \pi(x_2) \quad \cdots \quad \pi(x_n))$

Berdasarkan Persamaan (2.9), hasil turunan pertama fungsi $L(\boldsymbol{\beta})$ merupakan persamaan non linier, sehingga solusi eksak untuk mendapatkan penaksir $\boldsymbol{\beta}$ tidak dapat diperoleh secara analitik. Metode alternatif untuk memperoleh penaksir $\boldsymbol{\beta}$ adalah metode iteratif Newton-Raphson (Lestari, 2020).

$$\widehat{\boldsymbol{\beta}}^{(t+1)} = \widehat{\boldsymbol{\beta}}^{(t)} - \left(H\big(\widehat{\boldsymbol{\beta}}^{(t)}\big)\right)^{-1} \boldsymbol{g}\big(\widehat{\boldsymbol{\beta}}^{(t)}\big) \tag{2.11}$$

$t = 1,2, \ldots$ sampai konvergen

Keterangan:

$\boldsymbol{g}\big(\boldsymbol{\beta}^{(t)}\big) = \frac{\partial L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}$ merupakan vektor hasil turunan pertama fungsi log-*likelihood* sesuai Persamaan (2.10)

$H\big(\widehat{\boldsymbol{\beta}}^{(t)}\big)$ adalah matriks Hessian yang merupakan matriks hasil turunan kedua dari $\boldsymbol{\beta})$.

Matriks $H(\widehat{\boldsymbol{\beta}}^{(t)})$ yang digunakan pada Persamaaan (2.11) diperoleh pada Persamaan (2.12).

$$
\begin{aligned}
H(\widehat{\boldsymbol{\beta}}^{(t)}) &= \frac{\partial^2 L(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}} = \frac{\partial}{\partial\boldsymbol{\beta}}\left(\frac{\partial L(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\right) \\
&= \frac{\partial}{\partial\boldsymbol{\beta}}\left(\sum_{i=1}^{n}\boldsymbol{x}_i[y_i - \pi(x_i)]\right) \\
&= \frac{\partial}{\partial\boldsymbol{\beta}}\left(\sum_{i=1}^{n}\boldsymbol{x}_i\left[y_i - \frac{\exp(\boldsymbol{\beta}^T\boldsymbol{x}_i)}{1+\exp(\boldsymbol{\beta}^T\boldsymbol{x}_i)}\right]\right) \\
&= \sum_{i=1}^{n}\boldsymbol{x}_i\left[\frac{\partial}{\partial\boldsymbol{\beta}}\left(y_i - \frac{\exp(\boldsymbol{\beta}^T\boldsymbol{x}_i)}{1+\exp(\boldsymbol{\beta}^T\boldsymbol{x}_i)}\right)\right] \\
&= \sum_{i=1}^{n}\boldsymbol{x}_i\left(0 - \left[\frac{\boldsymbol{x}_i\exp(\boldsymbol{\beta}^T\boldsymbol{x}_i)\,(1+\exp(\boldsymbol{\beta}^T\boldsymbol{x}_i)) - \boldsymbol{x}_i\exp(\boldsymbol{\beta}^T\boldsymbol{x}_i)\exp(\boldsymbol{\beta}^T\boldsymbol{x}_i)}{(1+\exp(\boldsymbol{\beta}^T\boldsymbol{x}_i))^2}\right]\right) \\
&= -\sum_{i=1}^{n}\boldsymbol{x}_i\left(\frac{\boldsymbol{x}_i\exp(\boldsymbol{\beta}^T\boldsymbol{x}_i) + \boldsymbol{x}_i\exp(\boldsymbol{\beta}^T\boldsymbol{x}_i)\exp(\boldsymbol{\beta}^T\boldsymbol{x}_i) - \boldsymbol{x}_i\exp(\boldsymbol{\beta}^T\boldsymbol{x}_i)\exp(\boldsymbol{\beta}^T\boldsymbol{x}_i)}{(1+\exp(\boldsymbol{\beta}^T\boldsymbol{x}_i))^2}\right) \\
&= -\sum_{i=1}^{n}\boldsymbol{x}_i\left(\frac{\boldsymbol{x}_i\exp(\boldsymbol{\beta}^T\boldsymbol{x}_i)}{(1+\exp(\boldsymbol{\beta}^T\boldsymbol{x}_i))^2}\right) \\
&= -\sum_{i=1}^{n}\boldsymbol{x}_i\left(\boldsymbol{x}_i\left[\frac{\exp(\boldsymbol{\beta}^T\boldsymbol{x}_i)}{1+\exp(\boldsymbol{\beta}^T\boldsymbol{x}_i)}\right]\left[\frac{1}{1+\exp(\boldsymbol{\beta}^T\boldsymbol{x}_i)}\right]\right) \\
&= -\sum_{i=1}^{n}\boldsymbol{x}_i\,(\boldsymbol{x}_i[\pi(x_i)][1-\pi(x_i)]) \\
&= -\sum_{i=1}^{n}\boldsymbol{x}_i\,\boldsymbol{x}_i^T[\pi(x_i)][1-\pi(x_i)] \qquad (2.12)
\end{aligned}
$$

Persamaan (2.12) dapat dinyatakan dalam bentuk perkalian matriks sehingga diperoleh Persamaan (2.13).

$$H(\widehat{\boldsymbol{\beta}}^{(t)}) = \frac{\partial^2 L(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}} = -X^T V X \qquad (2.13)$$

V merupakan matriks diagonal berukuran $n \times n$ dengan elemen diagonal:

$$
\begin{bmatrix}
\pi(x_1)(1-\pi(x_1)) & 0 & \cdots & 0 \\
0 & \pi(x_2)(1-\pi(x_2)) & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \pi(x_n)(1-\pi(x_n))
\end{bmatrix}
$$

Iterasi dimulai dengan penentuan nilai awal $\boldsymbol{\beta}^{(0)} = \left(\beta_0^{(0)}, \beta_1^{(0)}, \dots, \beta_p^{(0)}\right)$ dan akan berhenti apabila telah konvergen, yaitu jika $\|\boldsymbol{\beta}^{(t+1)} - \boldsymbol{\beta}^{(t)}\| \leq \varepsilon$ dengan ε merupakan bilangan yang sangat kecil. Penaksir parameter model regresi logistik biner merupakan nilai $\widehat{\boldsymbol{\beta}}^{(t+1)}$ pada iterasi terakhir (Lestari, 2020).

2.4 Multikolinearitas

Multikolinearitas merupakan suatu kondisi adanya hubungan linier atau korelasi yang tinggi antara variabel prediktor pada model regresi. Adanya multikolinearitas dapat mempengaruhi interpretasi koefisien regresi sehingga memungkinkan terjadinya kesalahan dalam pengambilan keputusan serta pengujian signifikansi parsial menjadi tidak signifikan (Kasim, 2021). Pendeteksian multikolinearitas dapat dilakukan dengan menggunakan nilai *variance inflation factor* (VIF). Nilai VIF yang lebih besar dari 10 mengindikasikan bahwa terdapat multikolinearitas pada model. Nilai VIF dapat diidentifikasi menggunakan Persamaan (2.14).

$$VIF_k = \frac{1}{1 - R_k^2}, k = 1,2, \dots, p \tag{2.14}$$

Nilai R_k^2 adalah koefisien determinasi antara x_k dengan variabel prediktor lainnya pada model (Sriningsih dkk, 2018).

2.5 *Least Absolute Shrinkage and Selection Operator*

Least absolute shrinkage and selection operator (LASSO) merupakan metode untuk mengatasi multikolinearitas dengan cara mereduksi koefisien parameter β mendekati nol atau menjadi nol sehingga menghasilkan parameter β yang lebih sedikit dan model yang dihasilkan mudah diinterpretasikan dibandingkan dengan model awal (Hidayah & Adnan, 2021). Penduga koefisien LASSO dituliskan pada Persamaan (2.15).

$$\widehat{\boldsymbol{\beta}}^{LASSO} = \arg\min_{\beta} \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{k=1}^{p} x_{ki}\beta_k \right)^2, \sum_{k=1}^{p} |\beta_k| \leq t \tag{2.15}$$

Nilai t merupakan parameter yang mengontrol penyusutan pada estimasi parameter model, dengan nilai $t \geq 0$ (He dkk., 2019). Misalkan $\hat{\beta}_k$ merupakan nilai estimasi parameter dengan metode *Ordinary Least Squares* (OLS), jika $t < \sum_{k=1}^{p}|\hat{\beta}_k|$, maka $\widehat{\boldsymbol{\beta}}^{LASSO}$ akan mengecil mendekati nol atau tepat nol. Jika nilai $t \geq \sum_{k=1}^{p}|\hat{\beta}_k|$ maka hasil estimasi parameter dengan LASSO akan sama dengan nilai estimasi parameter dengan metode *Ordinary Least Squares* (OLS) (Wirdiastuti dkk., 2023). Persamaan (2.15) juga dapat dituliskan dalam persamaan lagrange sebagaimana dituliskan pada Persamaan (2.16).

$$\widehat{\boldsymbol{\beta}}^{LASSO} = \arg\min\left\{\sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{k=1}^{p} x_{ki}\beta_k\right)^2 + \lambda\sum_{k=1}^{p}|\beta_k|\right\} \tag{2.16}$$

Nilai λ pada Persamaan (2.16) disebut sebagai variabel yang berkorespondensi satu-satu dengan t. Artinya untuk setiap nilai $t \geq 0$ yang menghasilkan solusi $\widehat{\boldsymbol{\beta}}^{LASSO}$ terdapat $\lambda \geq 0$ sedemikian sehingga menghasilkan solusi $\widehat{\boldsymbol{\beta}}^{LASSO}$ (Frans L dkk., 2022). Adapun hubungan antara nilai λ dan t dituliskan pada Persamaan (2.17) (Wantoro, 2017).

$$\lambda = \frac{t}{\sum_{k=1}^{p}|\hat{\beta}_k|} \tag{2.17}$$

Nilai λ didapatkan dengan menggunakan metode *k-fold Cross-Validation* (CV), yaitu dengan membagi data secara acak menjadi k-kelipatan (*fold*) yang berukuran sama. Diantara k-*fold*, satu *fold* disimpan sebagai data *test* untuk menguji model. Kemudian sisa $k-1$ *fold* akan digunakan sebagai data *train* untuk menyesuaikan model dengan nilai dari parameter penyusutan. Nilai k-*fold* CV untuk masing-masing λ didefinisikan pada Persamaan (2.18) (Hidayah & Adnan, 2021).

$$k - CV_{(\lambda)} = \frac{1}{k}\sum_{i=1}^{k}\left(y_i - \hat{y}_{i(\lambda)}^{-k(i)}\right)^2 \tag{2.18}$$

$\hat{y}_{i(\lambda)}^{-k(i)}$ adalah nilai dugaan y pada saat *fold* ke-k tidak digunakan dalam pendugaan
r (Bahmid, 2018). Nilai λ optimal merupakan λ yang memiliki nilai k-*fold*
num (Hidayah & Adnan, 2021).

Metode LASSO dapat diterapkan pada regresi logistik. Penerapan LASSO pada regresi logistik dilakukan dengan memasukkan pinalti LASSO pada fungsi log-*likelihood* dalam pendugaan parameter model, sehingga estimasi parameter regresi logistik dengan LASSO diperoleh Persamaan (2.19) (Wantoro, 2017).

$$\widehat{\boldsymbol{\beta}}^{LASSO} = \arg\max\left\{L(\boldsymbol{\beta}) - \lambda\sum_{k=1}^{p}|\beta_k|\right\} \tag{2.19}$$

$L(\boldsymbol{\beta})$ merupakan fungsi *log-likelihood* untuk regresi logistik biner sesuai pada Persamaan (2.8).

2.6 Algoritma *Cyclic Coordinate Desent*

Untuk mendapatkan hasil dari estimasi parameter regresi logistik dengan LASSO maka diperlukan suatu algoritma, yaitu algoritma *Cyclic Coordinate Desent* (CCD). Tujuan dari algoritma CCD adalah membentuk nilai parameter baru $\left(\beta_k^{baru}\right)$ yang dapat memaksimumkan nilai *penalized log-likelihood* seperti pada Persamaan (2.19). Nilai dari parameter baru yang terbentuk ditampilkan pada Persamaan (2.20).

$$\beta_k^{baru} = \begin{cases} \beta_k - \Delta_k\,, & jika\;\Delta v_k < -\Delta_k \\ \beta_k + \Delta v_k\,, & jika - \Delta_k \le \Delta v_k \le \Delta_k \\ \beta_k + \Delta_k\,, & jika\;\Delta v_k > \Delta_k \end{cases} \tag{2.20}$$

Nilai β_k adalah penduga parameter awal, Δ_k adalah hampiran penduga β_k, dan Δv_k adalah nilai penduga terbaru (Ramadhan, 2017).

Nilai Δ_k dimisalkan dalam $[1, \infty)$ atau dapat pula menggunakan Persamaan (2.21).

$$\Delta_k = \max\left(2|\beta_k|, \frac{\Delta_k}{2}\right) \tag{2.21}$$

Nilai interval $(\beta_k - \Delta_k, \beta_k + \Delta_k)$ berubah-ubah tergantung nilai penduga awal dan nilai Δv_k sebelumnya. Adapun nilai Δv_k dapat diperoleh dengan Persamaan (2.22).

$$\Delta v_k = -\frac{S_k(\beta) - \lambda s}{Q(\beta_k, \Delta_k)} \tag{2.22}$$

$S_k(\beta)$ adalah turunan pertama fungsi *log-likelihood* yang ditampilkan pada ...n (2.10). $Q(\beta_k, \Delta_k)$ adalah turunan kedua fungsi *log-likelihood* yang ...an pada Persamaan (2.13), dan nilai $s = \beta_k/|\beta_k|$.

Adapun prosedur dari algoritma CCD adalah sebagai berikut (Wantoro, 2017).

1. Menentukan nilai awal $\beta_k = 0$ dan $\Delta_k = 1$ untuk $k = 1,2, \dots, p$
2. Menghitung nilai Δv_k dengan ketentuan:
   a. Apabila $\beta_k = 0$, maka $s = 1$, kemudian hitung Δv_k menggunakan Persamaan (2.22). apabila $\Delta v_k < 0$ maka $s = -1$ kemudian hitung Δv_k. Namun apabila $\Delta v_k \geq 0$ maka $\Delta v_k = 0$
   b. Apabila $\beta_k \neq 0$, maka $s = \frac{\beta_k}{|\beta_k|}$, kemudian hitung Δv_k menggunakan Persamaan (2.22). apabila $s(\beta_k + \Delta v_k) \leq 0$ maka $\Delta v_k = -\beta_k$.
3. Menghitung nilai Δ_k menggunakan Persamaan (2.21)
4. Menghitung nilai β_k^{baru} menggunakan Persamaan (2.20)
5. Mengulangi Langkah 2 hingga Langkah 4 hingga mencapai kondisi konvergen.

2.7 Uji Heterogenitas Spasial

Heterogenitas spasial diartikan sebagai suatu kondisi dimana suatu model regresi tidak dapat dijelaskan hubungan antara variabel respon dan variabel prediktor dalam model dikarenakan karakteristik antar wilayah pengamatan yang bervariasi secara spasial (Wulandari, 2018). Uji heterogenitas spasial diperlukan untuk mengetahui adanya keragaman spasial pada data. Adanya heterogenitas spasial menunjukkan bahwa data dapat dimodelkan dengan metode *geographically weighted logistic regression* (Solekha & Qudratullah, 2022). Uji heterogenitas spasial dapat dilakukan dengan menggunakan uji *Breusch-Pagan* dengan hipotesis sebagai berikut.

H_0 : tidak terdapat heterogenitas spasial

H_1 : terdapat heterogenitas spasial

Adapun statistik uji yang digunakan adalah:

$$BP = \frac{1}{2}\boldsymbol{f}^T Z(Z^T Z)^{-1} Z^T \boldsymbol{f} \sim \chi^2_{p,\alpha} \tag{2.23}$$

Keterangan:

$i = 1,2, \dots, n$ adalah banyaknya observasi.

$_2, \dots, f_n)^T, f_i = \left(\frac{e_i^2}{\sigma_e^2} - 1\right)$, dengan $e_i = y_i - \hat{y}_i$

matriks berukuran $n \times (p+1)$ dari variabel prediktor yang sudah

kan, dengan p adalah banyaknya variabel prediktor.

Hipotesis H_0 ditolak apabila nilai BP lebih besar dari $\chi^2_{p,\alpha}$ atau $p - value$ lebih kecil dari α (Reskia & Harison, 2022).

2.8 Matriks Pembobot Spasial

Pembobotan dalam analisis spasial merupakan hal yang penting karena nilai pembobotan mewakili letak data observasi satu terhadap data observasi yang lain (Wulandari, 2018). Pembobot berperan untuk memberikan pendugaan parameter yang berbeda di setiap lokasi pengamatan (Fadliana & Darajat, 2021). Penaksiran parameter pada suatu lokasi pengamatan ke-i akan lebih dipengaruhi oleh lokasi pengamatan yang dekat daripada lokasi pengamatan lain yang letaknya lebih jauh (Maulana dkk., 2019).

Matriks pembobot merupakan matriks diagonal yang berukuran $(n \times n)$ yang elemen diagonalnya merupakan nilai pembobot untuk lokasi pengamatan ke-i. Matriks pembobot dituliskan sebagai berikut (Pratiwi dkk., 2020).

$$W(u_i, v_i) = \begin{bmatrix} w_{i1} & 0 & \cdots & 0 \\ 0 & w_{i2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_{in} \end{bmatrix} \tag{2.24}$$

Matriks pembobot tersebut perlu dihitung untuk setiap lokasi pengamatan. Besarnya nilai pembobot ditentukan oleh fungsi pembobot. Fungsi pembobot yang dapat digunakan adalah fungsi pembobot *fixed gaussian* yang diperoleh dengan Persamaan (2.25) (Permai dkk., 2021).

$$w_{ij} = \exp\left\{-\frac{1}{2}\left(\frac{d_{ij}}{b}\right)^2\right\} \tag{2.25}$$

Keterangan:

w_{ij} : nilai pembobot pada lokasi ke-i terhadap lokasi ke-j

d_{ij} : jarak antara lokasi ke-i dan lokasi ke-j

b : nilai *bandwidth*

Nilai *bandwith* yang digunakan pada persamaan (2.25) merupakan nilai *bandwidth* optimum. Pemilihan nilai *bandwidth* yang optimum menjadi hal yang karena akan mempengaruhi ketepatan hasil estimasi. Nilai *bandwidth* dapat dilakukan dengan menggunakan nilai *Cross Validation* (CV) yang pada Persamaaan (2.26) (Meimela, 2021).

$$CV = \sum_{i=1}^{n} [y_i - \hat{y}_{\neq i}(b)] \tag{2.26}$$

Nilai $\hat{y}_{\neq i}(b)$ merupakan nilai taksiran y_i dengan menggunakan *bandwidth* b dengan pengamatan di lokasi ke-i tidak dimasukkan dalam perhitungan, yaitu dengan membuat bobot lokasi ke-i bernilai nol. *bandwidth* optimum akan diperoleh apabila nilai CV yang dihasilkan minimum (Wulandari, 2018).

2.9 *Geographically Weighted Logistic Regression*

Geographically weighted logistic regression (GWLR) merupakan pengembangan dari regresi logistik biner yang diterapkan pada data spasial, yaitu data yang mempertimbangkan unsur lokasi. Pemodelan data spasial dengan regresi logistik biasa akan menyebabkan model menjadi tidak valid akibat adanya hubungan antara data variabel respon dan lokasi pengamatan yang menyebabkan nilai estimasi parameter tiap lokasi pengamatan berbeda (Pratiwi dkk., 2020). Pada model GWLR, estimasi parameter untuk setiap pengamatan di lokasi ke-i dilakukan dengan menggunakan fungsi pembobot sesuai jarak lokasi lain terhadap lokasi ke-i (Basu dkk., 2023). Adapun model GWLR untuk setiap lokasi ke-i dituliskan pada Persamaan (2.27).

$$\begin{aligned}\pi(u_i, v_i) &= \frac{\exp\,(\beta_0(u_i, v_i) + \beta_1(u_i, v_i)x_{i1} + \beta_2(u_i, v_i)x_{i2} + \cdots + \beta_p(u_i, v_i)x_{ip})}{1 + \exp\,(\beta_0(u_i, v_i) + \beta_1(u_i, v_i)x_{i1} + \beta_2(u_i, v_i)x_{i2} + \cdots + \beta_p(u_i, v_i)x_{ip)})} \\ &= \frac{\exp(\boldsymbol{\beta}^T(u_i, v_i)\,\boldsymbol{x}_i)}{1 + \exp(\boldsymbol{\beta}^T(u_i, v_i)\,\boldsymbol{x}_i)}\end{aligned} \tag{2.27}$$

Keterangan:

$\boldsymbol{x}_i^T = (1 \quad x_{i1} \quad x_{i2} \quad \cdots \quad x_{ip})$

$\boldsymbol{\beta}^T(u_i, v_i) = (\beta_0(u_i, v_i) \quad \beta_1(u_i, v_i) \quad \beta_2(u_i, v_i) \quad \cdots \quad \beta_p(u_i, v_i))$

Kemudian, dengan melakukan transformasi logit pada Persamaan (2.27), maka diperoleh Persamaan (2.28) (Solekha & Qudratullah, 2022).

$$\begin{aligned}logit[\pi(u_i, v_i)] &= \ln\left(\frac{\pi(u_i, v_i)}{1 - \pi(u_i, v_i)}\right) \\ &= \ln\left(\frac{\dfrac{\exp(\boldsymbol{\beta}^T(u_i, v_i)\,\boldsymbol{x}_i)}{1 + \exp(\boldsymbol{\beta}^T(u_i, v_i)\,\boldsymbol{x}_i)}}{1 - \dfrac{\exp(\boldsymbol{\beta}^T(u_i, v_i)\,\boldsymbol{x}_i)}{1 + \exp(\boldsymbol{\beta}^T(u_i, v_i)\,\boldsymbol{x}_i)}}\right)\end{aligned}$$

$$= \ln\left(\frac{\dfrac{\exp(\boldsymbol{\beta}^T(u_i,v_i)\,\boldsymbol{x}_i)}{1+\exp(\boldsymbol{\beta}^T(u_i,v_i)\,\boldsymbol{x}_i)}}{\dfrac{1+\exp(\boldsymbol{\beta}^T(u_i,v_i)\,\boldsymbol{x}_i)}{1+\exp(\boldsymbol{\beta}^T(u_i,v_i)\,\boldsymbol{x}_i)} - \dfrac{\exp(\boldsymbol{\beta}^T(u_i,v_i)\,\boldsymbol{x}_i)}{1+\exp(\boldsymbol{\beta}^T(u_i,v_i)\,\boldsymbol{x}_i)}}\right)$$

$$= \ln\left(\frac{\dfrac{\exp(\boldsymbol{\beta}^T(u_i,v_i)\,\boldsymbol{x}_i)}{1+\exp(\boldsymbol{\beta}^T(u_i,v_i)\,\boldsymbol{x}_i)}}{\dfrac{1}{1+\exp(\boldsymbol{\beta}^T(u_i,v_i)\,\boldsymbol{x}_i)}}\right)$$

$$= \ln(\exp(\boldsymbol{\beta}^T(u_i,v_i)\,\boldsymbol{x}_i))$$

$$= \boldsymbol{\beta}^T(u_i,v_i)\,\boldsymbol{x}_i$$

$$= \beta_0(u_i,v_i) + \sum_{k=1}^{p} \beta_k(u_i,v_i)x_{ik} \tag{2.28}$$

2.10 Metode *Bootstrap*

Metode *bootstrap* merupakan metode yang dapat digunakan untuk menaksir nilai *standard error* dari suatu estimator. Metode *bootstrap* merupakan metode berbasis *resampling* atau pengambilan kembali data dengan pengembalian dari data hasil observasi, dengan prosedur tersebut diulang sebanyak B kali. Adapun nilai dugaan *standard error* dengan metode *bootstrap* dapat dihitung dengan Persamaaan (2.29) (Febriani dkk., 2019).

$$\widehat{SE}_B = \left[\frac{\sum_{b=1}^{B}\left(\hat{\theta}_b^* - \bar{\theta}^*\right)^2}{B-1}\right]^{\frac{1}{2}} \tag{2.29}$$

Nilai $\hat{\theta}_b^*$ merupakan penduga parameter *bootstrap* saat $b = 1,2,\dots,B$, nilai $\bar{\theta}^*$ merupakan nilai rata-rata penduga parameter *bootstrap*, dengan $\bar{\theta}^* = \sum_{b=1}^{B} \hat{\theta}_b^*/B$, dan nilai $\widehat{SE}_B$ merupakan nilai dugaan *standard error* dengan metode *bootstrap* (Gultom dkk., 2022).

2.11 Uji Signifikansi Parameter Model *Geographically Weighted Logistic Regression*

Uji signifikansi parameter dilakukan untuk mengetahui parameter-parameter apa saja yang memiliki pengaruh yang signifikan terhadap variabel respon ah & Prastuti, 2023). Uji signifikansi parameter dilakukan dengan *wald* an hipotesis:

H_0 : $\beta_k(u_i, v_i) = 0$; $k = 1,2, \dots, p$ (variabel prediktor $\beta_k(u_i, v_i)$ tidak berpengaruh signifikan terhadap variabel respon)

H_1 : $\beta_k(u_i, v_i) \neq 0$ (variabel prediktor $\beta_k(u_i, v_i)$ berpengaruh signifikan terhadap variabel respon)

Statistik uji yang digunakan adalah:

$$W = \frac{\hat{\beta}_k(u_i, v_i)}{SE\left(\hat{\beta}_k(u_i, v_i)\right)} \tag{2.30}$$

Nilai $\hat{\beta}_k(u_i, v_i)$ adalah nilai estimasi parameter $\beta_k(u_i, v_i)$ pada lokasi ke-i dan $SE\left(\hat{\beta}_k(u_i, v_i)\right) = \sqrt{Var\left(\hat{\beta}_k(u_i, v_i)\right)}$. Adapun kriteria uji yang digunakan adalah H_0 ditolak apabila nilai $p - value$ lebih kecil dari α atau $|W|$ lebih besar dari $Z_{\alpha/2}$ (Aliu dkk., 2022).

2.12 Uji Kesesuaian Model *Geographically Weighted Logistic Regression*

Uji kesesuaian model GWLR dilakukan untuk melihat apakah faktor geografis memiliki pengaruh yang siginifikan atau tidak terhadap model (Maulidina & Oktora, 2020). Pengujian ini dilakukan dengan membandingkan nilai *deviance* model regresi logistik dan model GWLR dengan hipotesis:

H_0 : $\beta_k(u_i, v_i) = \beta_k$; $i = 1,2, \dots, n$; $k = 1,2, \dots, p$ (tidak ada perbedaan yang signifikan antara model regresi logistik dan model GWLR)

H_1 : paling sedikit ada satu $\beta_k(u_i, v_i) \neq \beta_k$ (Ada perbedaan yang signifikan antara model regresi logistik dan model GWLR)

Nilai *deviance* model regresi logistik adalah $D(\hat{\beta})$ dengan derajat bebas $df_1 = n$ dan nilai *deviance* model GWLR adalahi $D\left(\hat{\beta}(u_i, v_i)\right)$ dengan derajat bebas $df_2 = np$. Uji statistik yang digunakan adalah (Purnatirani, 2019):

$$F_{hitung} = \frac{D(\hat{\beta})/df_1}{D\left(\hat{\beta}(u_i, v_i)\right)/df_2} \tag{2.31}$$

Keterangan:

$$-2\sum_{i=1}^{n}(\pi(x_i)logit[\pi(x_i)] + log[1 - \pi(x_i)]) \tag{2.32}$$

$$D\left(\hat{\beta}(u_i,v_i)\right) = 2\left[\sum_{i=1}^{n}\left(y_{1i}\ln\hat{\pi}(u_i,v_i) + (1-y_{0i})\ln\left(1-\hat{\pi}(u_i,v_i)\right)\right) - \left(n_1\ln n_1 + n_0\ln(n_0) - n\ln(n)\right)\right] \tag{2.33}$$

y_{1i} = nilai variabel respon berkategori 1 pada lokasi ke- i

y_{0i} = nilai variabel respon berkategori 0 pada lokasi ke- i

n_{1i} = banyak observasi y yang berkategori 1

n_0 = banyak observasi y yang berkategori 0

$n = n_{0i} + n_{1i}$

F_{hitung} akan mengikuti distribusi F dengan derajat bebas df_1 dan df_2. Kriteria uji yang digunakan adalah H_0 ditolak apabila nilai F_{hitung} lebih besar dari F_{α,df_1,df_2} atau $p-value$ lebih kecil dari α (Purnatirani, 2019). Penolakan H_0 menandakan bahwa faktor geografis memiliki pengaruh yang signifikan pada model GWLR (Maulidina & Oktora, 2020).

2.13 Interpretasi Model *Geographically Weighted Logistic Regression*

Interpretasi model GWLR dilakukan dengan menggunakan nilai *odds ratio* (OR). Nilai OR digunakan untuk melihat kecenderungan variabel prediktor yang signifikan memengaruhi variabel respon. Nilai OR menggambarkan perubahan kecenderungan setiap penambahan satu unit variabel prediktor (Rai & Ramadhan, 2018). Nilai OR pada model GWLR dapat dihitung menggunakan Persamaan (2.34) (Aliu dkk., 2022).

$$OR(u_i,v_i) = \exp\left(\hat{\boldsymbol{\beta}}(u_i,v_i)\right) \tag{2.34}$$

Apabila nilai $OR = 1$, berarti variabel prediktor dan variabel respon tidak memiliki hubungan. Kemudian apabila nilai $OR < 1$, berarti variabel prediktor memiliki hubungan negatif terhadap variabel respon. Nilai $OR > 1$ menyatakan variabel prediktor memiliki hubungan positif terhadap variabel respon (Findasari & Himayati, 2023).

2.14 Evaluasi Performa Model

Evaluasi performa model dilakukan untuk mengetahui seberapa besar keakuratan model dalam mendeskripsikan data (Amalah, 2020). Metode yang populer digunakan dalam evaluasi performa model adalah menggunakan nilai *akaike's information criterian* (AIC). Semakin kecil nilai AIC suatu model, maka semakin baik pula model tersebut (Utama & Hajarisman, 2021). Nilai AIC dapat diperoleh dari Persamaan (2.35) (Permana, 2021).

$$AIC = -n \log\left(\frac{RSS}{n}\right) + 2p \tag{2.35}$$

RSS merupakan jumlah kuadrat galat, n merupakan banyak observasi, dan p merupakan banyak variabel prediktor.

2.15 *Unmet Need* KB

Unmet need KB diartikan sebagai PUS yang tidak ingin anak lagi atau ingin menunda kelahiran anak tetapi tidak menggunakan suatu alat kontrasepsi (Putri dkk., 2021). Pemahaman tentang *unmet need* KB sangat penting dalam menilai sejauh mana program KB dapat memenuhi kebutuhan masyarakat. Tingginya angka *unmet need* KB menjadi salah satu faktor yang menimbulkan banyaknya kehamilan yang tidak diharapkan, sehingga banyak pasangan yang menggugurkan kandungannya. Ini menjadi salah satu penyebab utama terjadinya kematian ibu dan bayi (Guspianto dkk., 2021). *Unmet need* KB merupakan permasalahan yang dipengaruhi oleh berbagai faktor, seperti karakteristik demografi, sosial ekonomi, dan akses pelayanan (Zia, 2019). Secara umum, faktor-faktor yang mempengaruhi *unmet need* KB adalah:

1) Pendidikan

Pendidikan dapat mempengaruhi pola pikir terhadap adat kebiasaan. Dengan pendidikan yang tinggi, seseorang dapat lebih mudah untuk menerima ide atau masalah baru seperti pembatasan paritas (banyak kelahiran hidup yang dipunyai oleh seorang perempuan). Pendidikan juga akan menungkatkan kesadaran wanita terhadap manfaat yang dapat dinikmati bila ia mempunyai paritas sedikit. Wanita ... rpendidikan lebih tinggi cenderung membatasi jumlah kelahiran ... gkan dengan yang tidak berpendidikan atau berpendidikan rendah (Marito,

2) Pekerjaan

Pekerjaan memiliki kaitan yang sangan erat dengan kehidupan sehari-hari. Dalam hal status pekerjaan, seorang ibu yang tidak bekerja mempunyai peluang menjadi *unmet need* dibandingkan ibu yang bekerja (Marito, 2021).

3) Penghasilan

Penghasilan yang tinggi dapat membawa dampak positif bagi keluarga karena seluruh kebutuhan dapat terpenuhi. Namun tidak demikian dengan keluarga yang memiliki penghasilan rendah. Penghasilan yang rendah akan mengakibatkan keluarga mengalami kerawanan dalam memenuhi kebutuhan kehidupannya, salah satunya adalah kesehatan. Pendapatan berbanding terbalik dengan peluang status *unmet need*. Semakin tinggi pendapatan maka peluang status *unmet need* semakin menurun. Sebaliknya, semakin rendah tingkat pendapatan maka peluang status *unmet need* semakin tinggi atau naik (Marito, 2021).

4) Larangan Suami

Faktor larangan suami merupakan faktor penyebab *unmet need* KB yang didasari oleh budaya sebagian besar masyarakat Indonesia, yaitu menjadikan laki-laki sebagai kepala keluarga sekaligus pihak yang bertanggung jawab dalam menentukan keikutsertaan wanita untuk menggunakan alat kontrasepsi. Istri yang tidak mendapatkan dukungan suami cenderung akan mengalami *unmet need* (Nabila & Nindya, 2021).