SKRIPSI

ESTIMASI PARAMETER MODEL *GEOGRAPHYCALLY WEIGHTED LOGISTIC REGRESSION PRINCIPAL COMPONENT ANALYSIS* DENGAN MAKSIMUM *LIKELIHOOD*

Disusun dan diajukan oleh

SITI IHZA ARSELLA KASIM

H051171017

PROGRAM STUDI STATISTIKA DEPARTEMEN STATISTIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS HASANUDDIN

MAKASSAR

2021

ESTIMASI PARAMETER MODEL *GEOGRAPHYCALLY WEIGHTED LOGISTIC REGRESSION PRINCIPAL COMPONENT ANALYSIS* DENGAN MAKSIMUM *LIKELIHOOD*

SKRIPSI

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

SITI IHZA ARSELLA KASIM

H051171017

PROGRAM STUDI STATISTIKA DEPARTEMEN STATISTIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS HASANUDDIN

MAKASSAR

2021

LEMBAR PERNYATAAN KEASLIAN

Yang bertanda tangan dibawah ini:

Nama	: Siti Ihza Arsella Kasim
NIM	: H051171017
Program Studi	: Statistika
Jenjang	: Sarjana (S1)

Menyatakan dengan ini bahwa karya tulis saya yang berjudul

ESTIMASI PARAMETER MODEL *GEOGRAPHICALLY WEIGHTED LOGISTIC REGRESSION PRINCIPAL COMPONENT ANALYSIS* DENGAN MAKSIMUM *LIKELIHOOD*

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

Apabila dikemudian hari terbukti atau dapat dibuktikan bahwa sebagian atau keseluruhan skripsi ini hasil karya orang lain, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Makassar, 19 Agustus 2021

SITI IHZA ARSELLA KASIM
NIM. H051171017

ESTIMASI PARAMETER MODEL *GEOGRAPHYCALLY WEIGHTED LOGISTIC REGRESSION PRINCIPAL COMPONENT ANALYSIS* DENGAN MAKSIMUM *LIKELIHOOD*

Disetujui Oleh:

Pembimbing Utama,	**Pembimbing Pertama,**
Andi Kresna Jaya, S.Si., M.Si.	**Anisa, S.Si., M.Si.**
NIP. 19731228 200003 1 001	**NIP. 19730227 199802 2001**

Ketua Departemen Statistika

Dr. Nurtiti Sunusi, S.Si., M.Si.
NIP. 19720117 199703 2002

Pada Tanggal : 19 Agustus 2021

LEMBAR PENGESAHAN

ESTIMASI PARAMETER MODEL *GEOGRAPHYCALLY WEIGHTED LOGISTIC REGRESSION PRINCIPAL COMPONENT ANALYSIS* DENGAN MAKSIMUM *LIKELIHOOD*

Disusun dan diajukan oleh

SITI IHZA ARSELLA KASIM
H051171017

Telah dipertahankan dihadapan Panitia Ujian yang dibentuk dalam rangka Penyelesaian Studi Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin
pada tanggal 19 Agustus 2021
dan dinyatakan telah memenuhi syarat kelulusan.

Menyetujui,

Pembimbing Utama,	**Pembimbing Pertama,**
Andi Kresna Jaya, S.Si., M.Si.	**Anisa, S.Si., M.Si.**
NIP. 19731228 200003 1 001	**NIP. 19730227 199802 2001**

Ketua Departemen Statistika

Dr. Nurtiti Sunusi, S.Si., M.Si.
NIP. 19720117 199703 2002

KATA PENGANTAR

Assalamu'alaikum Warahmatullahi Wabarakatuh.

Alhamdulillahi robbil'alamin, puji syukur kepada Allah *Subhanahu Wa Ta'ala* atas segala limpahan rahmat, nikmat, dan hidayah yang diberikan kepada penulis sehingga dapat menyelesaikan penulisan skripsi dengan judul **"Estimasi Parameter Model *Geographically Weighted Logistic Regression Principal Component Analysis* dengan Maksimum *Likelihood*"** sebagai salah satu syarat untuk memperoleh gelar Sarjana Sains pada Program Studi Statistika Jurusan Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam.

Salam dan sholawat *InsyaAllah* senantiasa tercurah kepada **Nabi Muhammad *Shallallahu'alaihi Wasallam***, sang kekasih tercinta yang telah memberikan petunjuk cinta dan kebenaran dalam kehidupan.

Dalam penyelesaian skripsi ini, penulis telah melewati perjuangan panjang dan pengorbanan yang tidak sedikit. Namun berkat rahmat dan izin-Nya serta dukungan dari berbagai pihak yang turut membantu baik moril maupun material sehingga akhirnya tugas akhir ini dapat terselesaikan. Oleh karena itu, penulis menyampaikan ucapan terima kasih yang setinggi-tingginya dan penghargaan yang tak terhingga kepada Ayahanda ***Muhammad Kasim*** dan Ibunda tercinta ***Halapa Arif*** yang telah membesarkan dan mendidik penulis dengan penuh kesabaran dan dengan limpahan cinta dan kasih saying serta telah memberikan dukungan dan doa kepada penulis yang tak pernah habis. Dan kakak tersayang **Abdi Dzul Ikram** yang menjadi motivator dan penyemangat untuk segera menyelesaikan masa studi penulis.

Ucapan terima kasih yang juga penulis ucapkan kepada:

1. **Ibu Prof. Dr. Dwia Aries Tina Palubuhu, MA,** 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. Nurtiti Sunusi, S.Si., M.Si.** selaku Ketua Departemen Statistika yang telah seperti orang tua sendiri. Segenap dosen pengajar dan staf **Departemen Statistika** yang telah membekali ilmu dan kemudahan kepada

penulis dalam berbagai hal selama menjadi mahasiswa di Departemen Statistika,

4. **Bapak Andi Kresna Jaya, S.Si., M.Si.,** selaku Pembimbing Utama penulis yang telah meluangkan waktu dan pemikirannya untuk memberikan arahan, pengetahuan dan bimbingan kepada penulis mulai dari awal hingga selesainya penulisan tugas akhir ini. **Ibu Anisa, S.Si., M.Si.,** selaku Pembimbing Pertama penulis yang telah meluangkan waktunya ditengah kesibukan untuk memberikan arahan bagi penulis.
5. **Ibu Dr. Erna Tri Herdiani, S.Si., M.Si.,** selaku penguji sekaligus penasehat akademik penulis yang telah meluangkan waktu dan pemikirannya untuk memberikan arahan dan saran sehingga penulis dapat menyelesaikan tugas akhir ini. **Bapak Siswanto, S.Si., M.Si.,** selaku penguji yang telah memberikan saran dan pertanyaan yang tak terhingga nilainya yang bersifat membangun sehingga penulis dapat menuntaskan skripsi ini.
6. dalam penyempurnaan penyusunan tugas akhir ini.
7. Spesial untuk sahabat tercinta penulis, **Siti Nur Azizah, S.Kom., Nurfadlia, Farda Nurilmi**, bermula dari lomba kemudian atas kesamaan impian dan cita-cita hingga menjadi akrab. Suatu rasa syukur yang sangat besar bisa bertemu kalian. Terima kasih atas segalanya. **Man Jadda Wajada**.
8. Spesial untuk sahabat tersayang penulis, **Anugrah Julia, Adelvia, Rosnita Sari, Kak Sulis** sebagai tempat istirahat bagi penulis dari kepenatan dunia skripsi, sebagai tempat diskusi, curhat, dan teman jalan-jalan. Terima kasih atas kebersamaan, kebahagiaan, kebaikan, dan inspirasinya selama ini.
9. Spesial untuk sahabat seperjuangan selama kuliah, grup "**SEMANGAT BELAJAR" Musdalifah, S.Si., Nurkamalia, S.Si., Fitri, S.Si., Aqilah Salsabila Rahman., S.Si,** dan **Nur Alya Tussa'ada**. Terima kasih atas segala kebahagiaan, kebersamaan, kebaikan, dan nasihat-nasihat yang tidak akan pernah terlupakan oleh penulis.
10. Spesial untuk sahabat terbaik penulis, **Riska Rasyid, S.Si**, **Nur Aprillia Dzulhijjah,** dan **Fakhriyyah Dj Junus** yang selalu berhasil menenangkan

pikiran penulis di saat *overthinking*. Terima kasih atas nasihat, motivasi, saran, dan kesabarannya selama ini.

11. Teman-teman **UKM KPI Unhas,** terkhusus **Nihar Nurkhalifa, Nur Naningsih, Munawwara Ildana,** terima kasih telah menjadi teman baik bagi penulis, kakak **Mohd. Riswan Bin Jamal, S.IP,** kakak **Muh. Amri Arafah, S.T,** kakak **Muh. Farid Sulaiman, S.T**, kakak **Vietgar Membalik, S.P.,** kakak **Dwi Nining Lestari, S.Kel.,** terima kasih atas segala wejangan, kebersamaan, dan kebaikannya selama ini serta menjadi inspirator bagi penulis.
12. Teman-teman **Paguyuban KSE Unhas,** terkhusus ketum **Ratnah, Enca, Risma,** terima kasih atas kebaikan, kebersamaan, canda tawa selama di **KSE.**
13. kakak **Reski Amalah, S.Si,** dan kakak **Rayhanna Auliya Amin, S.Si,** terima kasih atas segala bentuk bantuan selama penulis mengerjakan tugas akhir ini.
14. Teman-teman **Statistika 2017**, terima kasih atas kebersamaan, suka dan duka selama menjalani pendidikan di Dapartemen Statistika.
15. Kepada seluruh pihak yang tidak dapat penulis sebutkan satu persatu, terima kasih setinggi-tingginya untuk segala dukungan dan partisipasi yang diberikan kepada penulis semoga bernilai ibadah di sisi **Allah *Subhanahu Wa Ta'ala.***

Penulis berharap skripsi ini dapat memberikan tambahan pengetahuan baru bagi para pembelajar statistika. Penulis menyadari bahwa dalam penulisan tugas akhir ini masih banyak terdapat kekurangan. Oleh karena itu, dengan segala kerendahan hati penulis memohon maaf. Akhir kata, semoga dapat bermanfaat bagi pihak-pihak yang berkepentingan. *Aamiin Yaa Rabbal Alamin.*

Makassar, 19 Agustus 2021

Siti Ihza Arsella Kasim

PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS

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

Nama	: Siti Ihza Arsella Kasim
NIM	: H051171017
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:

"ESTIMASI PARAMETER MODEL *GEOGRAPHICALLY WEIGHTED LOGISTIC REGRESSION PRINCIPAL COMPONENT ANALYSIS* DENGAN MAKSIMUM *LIKELIHOOD*"

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 memublikasikan 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 pada tanggal 19 Agustus 2021

Yang menyatakan

Siti Ihza Arsella Kasim

ABSTRAK

Geographically Weighted Logistic Regression (GWLR) merupakan data spasial yang memiiki variabel respon berdistribusi Bernoulli. Dalam pengaplikasian model GWLR terkadang ditemukan adanya multikolinearitas pada data pengamatan. Multikolinearitas yang tidak teratasi pada data spasial dapat mempengaruhi interpretasi koefisien regresi sehingga untuk mengatasi hal tersebut digunakan model *Geographically Weighted Logistic Regression Principal Component Analysis* (GWLRPCA) dengan gabungan PCA sebagai metode untuk mengatasi masalah multikolinearitas. Tujuan penelitian ini untuk mengetahui estimasi parameter model GWLRPCA serta memodelkan GWLRPCA pada data Indeks Pembangunan Kesehatan Masyarakat (IPKM) di Provinsi Sulawesi Selatan Tahun 2018. Pada penelitian ini, estimasi parameter dilakukan menggunakan *maximum likelihood estimation* dengan algoritma *fisher scoring.* Hasil penelitian menunjukkan variabel yang memberikan pengaruh signifikan terhadap peningkatan probabilitas IPKM Kabupaten/Kota masuk kategori diatas rata-rata nasional adalah cakupan kepemilikan JPK dan akses sanitasi. Secara keseluruhan tingkat akurasi model GWLRPCA dalam mengklasifikasikan IPKM Kabupaten/Kota terhadap IPKM Nasional adalah sebesar 79,17%. Terdapat 22 kabupaten/kota di Provinsi Sulawesi Selatan yang signifikan berpengaruh secara lokal terhadap IPKM Nasional.

Kata Kunci: GWLRPCA, IPKM, MLE, *Fisher Scoring, Fixed Gaussian Kernel*

ABSTRACT

Geographically Weighted Logistic Regression (GWLR) is spatial data that has a Bernoulli distribution of response variables. In the application of the GWLR model, it is sometimes found that there is multicollinearity in the observation data. Unresolved multicollinearity can affect the interpretation of the regression coefficients, so to overcome this problem, the Geographically Weighted Logistic Regression Principal Component Analysis (GWLRPCA) model with a combination of PCA is used as a method to overcome the multicollinearity problem. The purpose of this study is to determine the parameter estimation of the GWLRPCA model and to model the GWLRPCA on the Public Health Development Index (IPKM) data in South Sulawesi Province in 2018. In this study, parameter estimation is carried out using maximum likelihood estimation with fisher scoring algorithm. The results showed that the variables that had a significant effect on increasing the probability of Regency/City IPKM entering the category above the national average were the coverage of JPK ownership and access to sanitation. Overall, the level of accuracy of the GWLRPCA model in classifying the Regency/City IPKM against the National IPKM is 79.17%. There are 22 districts/cities in South Sulawesi Province which have a significant local influence on the National IPKM.

Keywords: GWLRPCA, IPKM, MLE, Fisher Scoring, Fixed Gaussian Kernel

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

DAFTAR LAMPIRAN

BAB I
PENDAHULUAN

1.1 Latar Belakang

Salah satu indikator kemajuan suatu negara adalah tercapainya pembangunan kesejahteraan bagi masyarakat diberbagai aspek bidang sosial, pertahanan, kelembagaan, politik, pendidikan dan teknologi, serta infrastruktur dan kesehatan (Alexander, 1994). Kesehatan sebagai salah satu faktor yang berperan penting dalam investasi pembangunan sumber daya manusia berkualitas. Indeks yang dapat mengukur pembangunan kesehatan diperlukan guna menjabarkan derajat kesehatan dengan baik. Salah satu indeks yang dimaksud adalah Indeks Pembangunan Kesehatan Masyarakat (IPKM). IPKM bermanfaat untuk menentukan peringkat kabupaten/kota dalam mencapai keberhasilan pembangunan kesehatan masyarakat dan sebagai rujukan dalam menentukan prioritas bantuan daerah, sehingga terwujudnya derajat kesehatan masyarakat yang merata dan tepat sasaran (Fathurrahman, 2019).

IPKM Indonesia mengalami peningkatan dari tahun 2013 sebesar 0.5404 menjadi 0.6087 di tahun 2018 (Kemenkes, 2019). Provinsi Bali menempati peringkat tertinggi IPKM, sebaliknya peringkat terendah adalah Provinsi Papua yang sebelumnya juga tidak mengalami peningkatan, sementara dilansir dari laman *tekape.co* untuk IPKM Provinsi Sulawesi Selatan berada di peringkat ke-16. Dilansir dari laman *info publik*, Kepala Badan Litbangkes Kemenkes, dr Siswanto, MHP., DTM mengatakan meskipun IPKM Indonesia mengalami peningkatan, namun kesenjangan nilai IPKM kabupaten/kota masih bervariasi antar provinsi, sehingga tetap perlu intervensi semua pihak.

Geographically Weighted Logistic Regression (GWLR) merupakan pengembangan dari *Geographically Weighted Regression* (GWR) dengan model regresi logistik yang berdistribusi Bernoulli. Model GWLR merupakan bentuk lokal dari regresi logistik yang memperhatikan faktor geografis suatu wilayah sehingga mampu menangkap karakteristik dari masing-masing lokasi (Kurnia, 2011). Estimasi parameter model GWLR dilakukan pada setiap lokasi

pengamatan dan menggunakan pembobot spasial yang ditentukan menggunakan fungsi pembobot. Fungsi pembobot merupakan fungsi jarak antar lokasi pengamatan dan bergantung pada *bandwidth.* Hasil nilai fungsi pembobot akan mewakili letak observasi satu dengan observasi lainnya.

Dalam pengaplikasian model GWLR terkadang ditemukan adanya multikolinearitas pada data pengamatan. Multikolinearitas adalah suatu keadaan adanya hubungan linier atau korelasi yang tinggi antara variabel prediktor di setiap lokasi pengamatan. Multikolinearitas yang tidak teratasi dapat mempengaruhi interpretasi koefisien regresi sehingga memungkinkan terjadi kesalahan pengambilan keputusan serta estimasi parameter menjadi tidak konsisten. Terdapat beberapa metode yang bisa digunakan dalam mengatasi multikolinearitas pada model GWLR tanpa penghapusan variabel prediktor yaitu *Principal Component Analysis* (PCA), *Partial Least Square* (PLS), *Lasso*, dan *Ridge* (Fitriyaningsih & Sutikno, 2015).

Penelitian ini akan menggunakan *Principal Component Analysis* (PCA) dalam mengatasi multikolinearitas. PCA merupakan salah satu teknik statistika untuk mengatasi gejala multikolinearitas dengan mentransformasikan variabel bebas yang saling berkorelasi menjadi variabel baru yang tidak berkorelasi. PCA menjelaskan struktur varians kovarians dari sekumpulan variabel baru yang saling bebas dan merupakan kombinasi linear dari variabel asli. Banyaknya komponen utama atau variabel baru yang terbentuk sama dengan banyaknya variabel asli.

Beberapa penelitian yang telah menggunakan model *Geographically Weighted Logistic Regression* (GWLR) pada data yang mengalami multikolinearitas antara lain oleh Yunus (2016) melakukan estimasi parameter model GWLR pada data yang mengandung multikolinieritas menggunakan *Partial Least Square* (PLS), dan Reski Amalah (2020) juga melakukan pemodelan *Geographically Weighted Logistic Regression* menggunakan metode *Rigde.* Berdasarkan sumber referensi yang telah ditelusuri, peneliti menyadari bahwa belum ditemukan pemodelan *Geographically Weighted Logistic Regression* menggunakan *Principal Component Analysis* (PCA) dalam mengatasi mutikolinearitas data, sehingga peneliti tertarik untuk mengkaji hal tersebut

dengan mengambil judul "***Estimasi Parameter Geographically Weighted Logistic Regression Principal Component Analysis dengan Maksimum Likelihood"*** menggunakan data Indeks Pembangunan Kesehatan Masyarakat (IPKM) Provinsi Sulawesi Selatan Tahun 2018.

1.2 Rumusan Masalah

1. Bagaimana estimasi parameter model *Geographically Weighted Logistic Regression Principal Component Analysis* dengan maksimum *likelihood*?
2. Bagaimana pemodelan *Geographically Weighted Logistic Regression Principal Component Analysis* dengan maksimum *likelihood* pada data Indeks Pembangunan Kesehatan Masyarakat di Provinsi Sulawesi Selatan Tahun 2018?

1.3 Batasan Masalah

Adapun batasan masalah pada penelitian ini yaitu menggunakan metode *Maximum Likelihood Estimation* (MLE) dan iterasi *fisher scoring* dalam mengestimasi parameter model *Geographically Weighted Logistic Regression Principal Component Analysis* (GWLRPCA).

1.4 Tujuan Penelitian

1. Untuk mengestimasi parameter model *Geographically Weighted Logistic Regression Principal Component Analysis* dengan maksimum *likelihood*.
2. Untuk memodelkan *Geographically Weighted Logistic Regression Principal Component Analysis* dengan maksimum *likelihood* pada data Indeks Pembangunan Kesehatan Masyarakat di Provinsi Sulawesi Selatan Tahun 2018?

1.5 Manfaat Penelitian

Manfaat dalam penelitian ini diharapkan dapat menambah pemahaman teoritis serta praktis bagi peneliti dan pembaca terkait pemodelan *Geographically Weighted Logistic Regression Principal Component Analysis* pada data Indeks Pembangunan Kesehatan Masyarakat (IPKM) di Provinsi Sulawesi Selatan Tahun 2018. Hasil analisis data IPKM Tahun 2018 Provinsi Sulawesi Selatan diharapkan

dapat menjadi bahan informatif dan advokasi ke pemerintahan terkhusus pemerintah provinsi maupun kabupaten/kota agar terpacu memperbaiki sumber daya melalui prioritas program kesehatan dan alokasi dana bantuan kesehatan dari pusat ke provinsi atau kabupaten/kota.

BAB II
TINJAUAN PUSTAKA

2.1 Regresi Logistik

Regresi logistik adalah analisis yang digunakan untuk data respon berkategori yang hanya memiliki dua kemungkinan nilai (dikotomus) misalnya ya/tidak atau sukses/gagal. Regresi logistik termasuk dalam model linier diperumum atau *Generalized Linier Models* (GLM). GLM merupakan pengembangan dari model linier klasik yang mengasumsikan variabel respon tidak harus mengikuti distribusi normal tetapi bisa termasuk dalam distribusi keluarga eksponensial (Pradita, 2011). Hasil pengamatan variabel acak Y respon hanya mempunyai dua kategori yaitu 0 dan 1, sehingga mengikuti distribusi Bernoulli. Adapun distribusi peluang regresi logistik dirumuskan sebagai berikut:

$$P\ (Y = y) = \ \pi^{y}(1-\pi)^{1-y}\ ; y = 0.1$$

Dalam regresi logistik, variabel respon Y dituliskan sebagai $y = \pi(x) + \varepsilon$, dengan ε mempunyai salah satu dari kemungkinan dua nilai, yaitu apabila nilai $y = 1$ maka $\varepsilon = 1 - \pi(x)$ dengan peluang $\pi(x)$ dan jika $y = 0$ maka $\varepsilon = -\pi(x)$ dengan peluang $1 - \pi(x)$. Menurut Agresti (2002), fungsi kepadatan peluang untuk Y dengan parameter $\pi(x)$ dapat dilihat pada persamaan berikut:

$$f(y) = \big(\pi(x)\big)^{y}(1-\ \pi(x))^{1-y}, y = 0{,}1 \qquad (2.1)$$

Menurut Hosmer & Lemashow (2000), model regresi logistik biner dapat dilihat pada persamaan berikut:

$$\pi(x) = \frac{\exp(\beta_0+\beta_1 x_1+\ \beta_2 x_2+\cdots+\beta_k x_k)}{1+\ \exp(\beta_0+\beta_1 x_1+\ \beta_2 x_2+\cdots+\beta_k x_k)}$$

$$\pi(x) = \frac{\exp(\boldsymbol{\beta}^T\boldsymbol{x})}{1+\ \exp(\boldsymbol{\beta}^T\boldsymbol{x})} \qquad (2.2)$$

Dengan $\boldsymbol{x} = (1, x_1, x_2\ , \ldots, x_k)^T$ dan $\boldsymbol{\beta}^T = (\beta_0, \beta_1, \beta_2, \ldots, \beta_k)$ serta β_0 merupakan konstanta dan β_j koefisien regresi logistik untuk variabel prediktor ke-j, dengan $j = 1, 2, \ldots, k$.

Pada regresi logistik, $\pi(x)$ adalah fungsi nonlinear sehingga untuk mempermudah pendugaan parameter, maka $\pi(x)$ ditransformasi menggunakan transformasi logit, sehingga diperoleh (Hosmer & Lemeshow, 2000):

$$g(x) = \ln\left(\frac{\pi(x)}{1-\pi(x)}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k \qquad (2.3)$$

Maka model regresi logistik dapat dituliskan dalam bentuk persamaan berikut ini:

$$\pi(x) = \frac{\exp g(x)}{1+\exp g(x)} \qquad (2.4)$$

2.2 *Maximum Likelihood Estimation*

Estimasi parameter regresi logistik dilakukan dengan menggunakan metode *Maximum Likelihood Estimation* (MLE). Prinsip dari MLE adalah menemukan penduga β dengan memaksimumkan fungsi *likelihood* kemudian disamakan dengan nol. Langkah pertama yang dilakukan yaitu membentuk fungsi *likelihood*, dapat dilihat pada persamaan sebagai berikut:

$$l(\beta) = \prod_{i=1}^{n} f(y_i) = \prod_{i=1}^{n} \pi(x_i)^{y_i}(1 - \pi(x_i))^{1-y_i} \qquad (2.5)$$

Untuk mempermudah perhitungan, maka fungsi *likelihood* dimaksimumkan dalam bentuk $\ln l(\beta)$, dapat dituliskan sebagai berikut:

$$\ln l(\beta) = \ln \left(\prod_{i=1}^{n} \pi(x_i)^{y_i}(1 - \pi(x_i))^{1-y_i}\right)$$

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

$$\ln l(\beta) = \sum_{k=0}^{p}\left(\sum_{i=1}^{n} y_i x_{ik}\right)\beta_k - \sum_{i=1}^{n} \ln\left\{1 + \exp\left(\beta_0 + \sum_{k=1}^{p} \beta_k x_{ik}\right)\right\}$$

Untuk mendapatkan fungsi *likelihood* yang dimaksimumkan dalam bentuk $\ln l(\beta)$, selanjutnya menurunkan $\ln l(\beta)$ tersebut terhadap β_j dan hasil dari turunan tersebut disamakan dengan nol, sehingga dapat dinyatakan dalam bentuk persamaan berikut:

$$\frac{\partial \ \ln l(\beta)}{\partial \beta_j} = \frac{\partial \sum_{k=0}^{p}\left(\sum_{i=1}^{n} y_i x_{ik}\right)\beta_k - \sum_{i=1}^{n} \ln\left\{1 + \exp\left(\beta_0 + \sum_{k=1}^{p} \beta_k x_{ik}\right)\right\}}{\partial \beta_j}$$

$$= \sum_{i=1}^{n} y_i x_{ik} - \sum_{i=1}^{n} x_{ik}\, \pi(x_i) = 0$$

$$= \sum_{i=1}^{n} x_{ik} \left(y_i - \frac{\exp\left(\beta_0 + \sum_{k=1}^{p} \beta_k x_{ik}\right)}{1 + \exp\left(\beta_0 + \sum_{k=1}^{p} \beta_k x_{ik}\right)} \right) \tag{2.6}$$

Dari turunan pertama fungsi ln *likelihood* akan diperoleh turunan kedua fungsi ln *likelihood* pada persamaan berikut ini:

$$\frac{\partial^2 \ \ln l(\beta)}{\partial \beta^2{}_k} = -\sum_{i=1}^{n} x_{ik} \left(\frac{x_{ik}\left(\exp\left(\beta_0 + \sum_{k=1}^{p} \beta_k x_{ik}\right)\right)}{\left(1 + \exp\left(\beta_0 + \sum_{k=1}^{p} \beta_k x_{ik}\right)\right)^2} \right)$$

$$\frac{\partial^2 \ \ln l(\beta)}{\partial \beta^2{}_k} = -\sum_{i=1}^{n} x_{ik}{}^2 \, \pi(x_i) \left(1 - \pi(x_i)\right) \tag{2.7}$$

Pada persamaan (2.7) dapat dinyatakan dalam bentuk matriks sebagai berikut:

$$\frac{\partial^2 \ \ln l(\beta)}{\partial \beta^2{}_k} = -\boldsymbol{X^T \widehat{W} X} = 0$$

Keterangan:

$\widehat{\boldsymbol{W}} = diag\left(\pi(x_1)\left(1 - \pi(x_1)\right),\ \pi(x_2)\left(1 - \pi(x_2)\right), \dots,\ \pi(x_i)\left(1 - \pi(x_i)\right)\right)$

Selanjutnya, karena hasil dari turunan parsial bersifat nonlinier, maka proses perhitungan *Maximum Likelihood Estimation* (MLE) dapat diestimasi menggunakan pendekatan numerik yaitu metode *fisher scoring* (Agresti, 2002). Menurut Smyth (2002), *fisher-scoring* adalah salah satu bentuk pengembangan dari metode *Newton-Raphson* dengan mengganti matriks Hessian dengan $\boldsymbol{I}^{(t)}\left(\beta^{(t)}\right)$. Adapun bentuk penduga parameter menggunakan metode *fisher scoring* adalah sebagai berikut:

$$\boldsymbol{\beta}^{t+1} = \boldsymbol{\beta}^{(t)} + \boldsymbol{I}^{(t)-1}\left(\beta^{(t)}\right)\boldsymbol{S}^{(t)}\left(\beta^{(t)}\right) \tag{2.8}$$

Keterangan:

$\boldsymbol{\beta}^{t+1}$ dan $\boldsymbol{\beta}^{(t)}$: Vektor untuk $\boldsymbol{\beta}$ pada iterasi ke-*t* dan ke-*t* + 1

$\boldsymbol{I}^{(t)-1}\left(\beta^{(t)}\right)$: Matriks dari turunan kedua ln-*likelihood* terhadap $\beta^{(t)}$

$\boldsymbol{S}^{(t)}\left(\beta^{(t)}\right)$: Vektor gradien dari turunan pertama ln-*likelihood* terhadap $\beta^{(t)}$

Iterasi tersebut berhenti pada saat keadaan konvergen yaitu:

$$\left\| \boldsymbol{\beta}^{t+1} - \boldsymbol{\beta}^{(t)} \right\| \leq \varepsilon$$

Dengan ε adalah nilai bilangan positif kecil yaitu $\varepsilon = 0{,}001$.

2.3 Pengujian Heterogenitas Spasial

Adanya perbedaan karakteristik di setiap lokasi pengamatan tetapi memiliki hubungan yang cukup erat antar data spasial disebut efek spasial. Efek spasial yang terjadi antar wilayah pengamatan terdiri dari efek spasial tipe wilayah (autokorelasi spasial) dan efek spasial titik (heterogenitas spasial) (Anselin, 1992). Pada efek spasial titik, heterogenitas spasial terjadi akibat adanya perbedaan pengaruh variabel prediktor terhadap variabel respon antara satu wilayah dengan wilayah lainnya sehingga menghasilkan parameter regresi yang berbeda-beda di setiap lokasi pengamatan. Heterogenitas spasial dapat diuji menggunakan rumus *Breusch-Pagan* (BP) dengan hipotesis sebagai berikut:

H_0 : $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_n^2 = \sigma^2$ (Tidak terdapat heterogenitas spasial)

H_1 : minimal ada satu $\sigma_i^2 \neq \sigma^2$ (Terdapat heterogenitas spasial)

Statistik uji:

$$BP = \frac{1}{2}\boldsymbol{f}^T\boldsymbol{Z}(\boldsymbol{Z}^T\boldsymbol{Z})^{-1}\boldsymbol{Z}^T\boldsymbol{f} \sim \chi^2 \tag{2.9}$$

Elemen vektor $\boldsymbol{f}$ adalah

$$f_i = \frac{e_i^2}{\sigma^2} - 1$$

Keterangan:

Z : Matriks dengan elemen vektor yang sudah dinormal standarkan untuk setiap pengamatan.

e_i : Residual OLS untuk pengamatan ke-*i*

σ^2 : Ragam galat e_i

Kriteria keputusan adalah tolak H_0 jika BP $> \chi^2_{a;p}$. Dengan p adalah jumlah prediktor sehingga dapat disimpulkan bahwa terdapat heterogenitas spasial.

2.4 Multikolinearitas

Istilah multikolinearitas digunakan pertama kali oleh Ragner Frisch pada tahun 1934, menyatakan bahwa multikolinearitas terjadi jika terdapat hubungan linier atau korelasi yang tinggi antara variabel prediktor dari model regresi. Multikolinearitas yang tidak teratasi dapat mempengaruhi interpretasi koefisien

regresi sehingga memungkinkan terjadi kesalahan pengambilan keputusan serta pengujian signifikansi parsial menjadi tidak signifikan (Yan dan Su, 2009). Salah satu cara untuk mengecek multikolinieritas yaitu dengan melihat nilai *Variance Inflation Factor* (VIF). Nilai VIF dapat dinyatakan dalam persamaan berikut (Yan dan Su, 2009):

$$VIF_k = \frac{1}{1-R_k^2} \tag{2.10}$$

Dengan R_k^2 merupakan koefisien determinasi antara x_k dengan variabel prediktor lainnya pada persamaan regresi. Apabila nilai VIF > 10, maka dapat disimpulkan bahwa data mengalami multikolinearitas (Gujarati, 2004).

2.5 Uji Asumsi Komponen Utama

Dalam penerapan analisis komponen utama terdapat asumsi yang harus dipenuhi berupa asumsi kecukupan data menggunakan uji *Kaiser-Meyer-Olkin* (KMO) dan asumsi korelasi multivariat menggunakan *Bartlett Test of Sphericity* (BTS) (Bilson, 2005). Tujuan pengujian KMO untuk melihat kecukupan sampel yang dianalisis dengan membandingkan besarnya koefisien korelasi dan koefisien korelasi parsial. Apabila nilai KMO berada diantara 0,5 sampai 1 maka dapat disimpulkan data yang digunakan dapat diselesaikan menggunakan analisis komponen utama (Delsen dkk, 2017). Adapun bentuk hipotesis KMO adalah sebagai berikut:

H_0 : Ukuran data cukup untuk analisis komponen utama

H_1 : Ukuran data tidak cukup untuk analisis komponen utama

Statistik Uji:

$$KMO = \frac{\sum_{i=1}^{p}\sum_{j=1}^{p} r_{ij}^2}{\sum_{i=1}^{p}\sum_{j=1}^{p} r_{ij}^2 + \sum_{i=1}^{p}\sum_{j=1}^{p} a_{ij}^2}, \; i = 1,2,3 \dots, p \text{ dan } j = 1,2,3 \dots, p \tag{2.11}$$

Dengan r_{ij}^2 adalah koefisien korelasi antara variabel *i* dan *j*, a_{ij}^2 adalah koefisien korelasi parsial antara variabel *i* dan *j*. Kriteria keputusan adalah terima H_0 apabila nilai KMO lebih besar dari 0,5.

Pengujian BTS bertujuan untuk mengetahui korelasi antar variabel prediktor. Adapun bentuk hipotesis BTS adalah sebagai berikut:

H_0 : Tidak terdapat korelasi antar variabel prediktor

H_1 : Terdapat korelasi antar variabel prediktor

Statistik Uji:

$$\chi^2(\boldsymbol{B}) = -\left[(n-1) - \frac{1}{6}(2p+5)\right] \times \ln|\boldsymbol{R}| \qquad (2.12)$$

Keterangan:

R : Matriks korelasi dari masing-masing variabel prediktor

n : Jumlah pengamatan

p : Jumlah variabel prediktor

Kriteria keputusan:

H_0 : diterima jika $\chi^2(\boldsymbol{B}) < \chi^2_{\left(\frac{p(p-1)}{2};\alpha\right)}$

H_0 : ditolak jika $\chi^2(\boldsymbol{B}) \geq \chi^2_{\left(\frac{p(p-1)}{2};\alpha\right)}$

2.6 Standarisasi Data

Standarisasi data adalah bentuk mentransformasikan variabel asal (X) ke angka baku (Z) (Suritman, 2020). Standarisasi data digunakan untuk menghilangkan kondisi buruk yang tidak menguntungkan akibat adanya multikolinearitas dalam data. Adapun rumus standarisasi data adalah sebagai berikut (Mariana, 2013):

$$Z_{pi} = \frac{X_{pi} - \bar{X}_p}{\sigma_p} \qquad (2.13)$$

Dengan X_{pi} adalah variabel respon ke-*p* pada lokasi ke-*i*, $\bar{X}_p$ adalah rata-rata variabel respon ke-*p*, dan σ_p adalah standar deviasi variabel respon ke-*p*.

2.7 Matriks Varians Kovarians dan Matriks Korelasi

Matriks varians kovarians dari sampel disimbolkan dengan **Σ** atau **S,** vektor mean sampel dari matriks varians kovarians dapat diperoleh dengan cara sebagai berikut (Anton, 2001):

$$\mathbf{S} = \frac{1}{n-1}\sum_{i=1}^{n}\left(\boldsymbol{x}_{ji} - \bar{\boldsymbol{x}}_j\right)\left(\boldsymbol{x}_{ji} - \bar{\boldsymbol{x}}_j\right)^T \qquad j = 1,2,\ldots,k$$

$$= \frac{1}{n-1} \sum_{i=1}^{n} \left[\begin{pmatrix} x_{1i} - \bar{x}_1 \\ x_{2i} - \bar{x}_2 \\ \vdots \\ x_{ki} - \bar{x}_k \end{pmatrix} (x_{1i} - \bar{x}_1 \quad x_{2i} - \bar{x}_2 \quad \dots \quad x_{ki} - \bar{x}_k) \right]$$

$$= \begin{bmatrix} \sum_{i=1}^{n} \frac{(x_{1i} - \bar{x}_1)^2}{n-1} & \sum_{i=1}^{n} \frac{(x_{1i} - \bar{x}_1)(x_{2i} - \bar{x}_2)}{n-1} & \cdots & \sum_{i=1}^{n} \frac{(x_{1i} - \bar{x}_1)(x_{ki} - \bar{x}_k)}{n-1} \\ \sum_{i=1}^{n} \frac{(x_{2i} - \bar{x}_2)(x_{1i} - \bar{x}_1)}{n-1} & \sum_{i=1}^{n} \frac{(x_{2i} - \bar{x}_2)^2}{n-1} & \cdots & \sum_{i=1}^{n} \frac{(x_{2i} - \bar{x}_2)(x_{ki} - \bar{x}_k)}{n-1} \\ \vdots & \vdots & & \vdots \\ \sum_{i=1}^{n} \frac{(x_{ki} - \bar{x}_k)(x_{1i} - \bar{x}_1)}{n-1} & \sum_{i=1}^{n} \frac{(x_{ki} - \bar{x}_k)(x_{2i} - \bar{x}_2)}{n-1} & \cdots & \sum_{i=1}^{n} \frac{(x_{ki} - \bar{x}_k)^2}{n-1} \end{bmatrix}$$

$$\mathbf{S} = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1k} \\ s_{21} & s_{22} & \cdots & s_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ s_{k1} & s_{k2} & \dots & s_{kk} \end{bmatrix} \tag{2.14}$$

Adapun matriks korelasi ($\boldsymbol{R}$) digunakan untuk mengukur keeratan hubungan linear antara variabel yang dinyatakan dalam varians s_{ii} dan kovarians s_{ij} yaitu sebagai berikut:

$$\boldsymbol{R} = \frac{s_{ij}}{\sqrt{s_{ii}}\sqrt{s_{jj}}}$$

Matris korelasi adalah matriks simetri berukuran $k \times k$ yang dapat dinyatakan dalam persamaan sebagai berikut:

$$\boldsymbol{R} = \begin{pmatrix} \frac{s_{11}}{\sqrt{s_{11}}\sqrt{s_{11}}} & \frac{s_{12}}{\sqrt{s_{11}}\sqrt{s_{22}}} & \cdots & \frac{s_{1k}}{\sqrt{s_{11}}\sqrt{s_{kk}}} \\ \frac{s_{21}}{\sqrt{s_{22}}\sqrt{s_{11}}} & \frac{s_{22}}{\sqrt{s_{22}}\sqrt{s_{22}}} & \cdots & \frac{s_{2k}}{\sqrt{s_{22}}\sqrt{s_{kk}}} \\ \vdots & \vdots & & \vdots \\ \frac{s_{k1}}{\sqrt{s_{kk}}\sqrt{s_{11}}} & \frac{s_{k2}}{\sqrt{s_{kk}}\sqrt{s_{22}}} & \cdots & \frac{s_{kk}}{\sqrt{s_{kk}}\sqrt{s_{kk}}} \end{pmatrix}$$

$$\boldsymbol{R} = \begin{pmatrix} 1 & \rho_{12} & \cdots & \rho_{1k} \\ \rho_{21} & 1 & \cdots & \rho_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{k1} & \rho_{k2} & \dots & 1 \end{pmatrix} \tag{2.15}$$

2.8 Nilai Eigen dan Vektor Eigen

Nilai eigen adalah nilai karakteristik dari suatu matriks berukuran $n \times n$ sementara vektor eigen adalah vektor kolom bukan nol yang bila dikalikan dengan

suatu matriks berukuran $n \times n$ akan menghasilkan vektor lain yang berupa kelipatan dari vektor eigen itu sendiri. Apabila $\boldsymbol{A}$ adalah matriks berukuran $n \times n$ maka sebuah vektor tak nol $\boldsymbol{e}$ pada $\mathbb{R}^n$ disebut vektor eigen dari $\boldsymbol{A}$, jika $\boldsymbol{Ae}$ adalah sebuah kelipatan skalar dari λ maka berlaku:

$$\boldsymbol{Ae} = \lambda \boldsymbol{e}$$

Skalar λ disebut nilai eigen dari $\boldsymbol{A}$ dan $\boldsymbol{e}$ disebut vektor eigen yang bersesuaian dengan λ. Besar nilai eigen dari matriks $\boldsymbol{A}$ berukuran $n \times n$ dapat dihitung menggunakan persamaan sebagai berikut:

$$\det(\boldsymbol{A} - \lambda \boldsymbol{I}) = 0 \tag{2.16}$$

Persamaan (2.16) disebut dengan persamaan karakteristik.

2.9 *Principal Component Analysis* (PCA)

Principal Component Analysis (PCA) atau analisis komponen utama merupakan salah satu teknik statistika untuk mengatasi gejala multikolinearitas dengan mentransformasikan variabel bebas yang saling berkorelasi menjadi variabel baru yang tidak berkorelasi. Menurut Johnson (2002), PCA menjelaskan struktur varians kovarians dari sekumpulan variabel baru yang saling bebas serta merupakan kombinasi linear dari variabel asal. Misalkan terdapat vektor random $\boldsymbol{X}^T = (X_1, X_2, X_3, \dots X_k)$ memiliki matriks varian-kovarian atau matriks korelasi dengan nilai-nilai eigen $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k$, maka komponen utama yang merupakan kombinasi linear dari $\boldsymbol{X}$ variabel asal dan $\boldsymbol{e}$ sebagai vektor eigen didefinisikan sebagai berikut:

$$\begin{aligned}
\boldsymbol{w}_1 &= \boldsymbol{X^T e_1} = \boldsymbol{X}_1 e_{11} + \boldsymbol{X}_2 e_{21} + \cdots + \boldsymbol{X}_k e_{k1} \\
\boldsymbol{w}_2 &= \boldsymbol{X^T e_2} = \boldsymbol{X}_1 e_{12} + \boldsymbol{X}_2 e_{22} + \cdots + \boldsymbol{X}_k e_{k2} \\
&\vdots \qquad\qquad \vdots \qquad\qquad \vdots \\
\boldsymbol{w}_k &= \boldsymbol{X^T e_k} = \boldsymbol{X}_1 e_{1k} + \boldsymbol{X}_2 e_{2k} + \cdots + \boldsymbol{X}_k e_{kk}
\end{aligned}$$

Dengan matriks sebagai berikut:

$$\boldsymbol{w} = \boldsymbol{X^T e} = [\boldsymbol{X}_1 \quad \boldsymbol{X}_2 \quad \dots \quad \boldsymbol{X}_k] \begin{bmatrix} e_{11} & e_{12} & \cdots & e_{1k} \\ e_{21} & e_{22} & \cdots & e_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ e_{k1} & e_{k2} & \cdots & e_{kk} \end{bmatrix}$$

Kriteria pemilihan komponen utama menggunakan proporsi kumulatif varians terhadap total variansi. Proporsi varians yang dijelaskan komponen utama ke-*i* dapat dihitung menggunakan persamaan sebagai berikut (Johnson, 2002):

$$\frac{\lambda_i}{\sum_{r=1}^{k}\lambda_r} \times 100\% = \frac{\lambda_i}{\lambda_1+\lambda_2+\cdots+\lambda_k} \times 100\%, \text{ untuk } i = 1,2,\ldots,p \qquad (2.17)$$

Menurut Bilson (2005), jumlah komponen utama dapat ditentukan dengan melihat proporsi kumulatif varians dan mampu menerangkan total variansi data sekitar 70% sampai 80%. Selain itu, jumlah komponen utama juga dapat diketahui dengan melihat pola penurunan nilai eigen pada *scree plot.* Penurunan yang tajam pada *scree plot* menunjukkan perubahan nilai eigen yang besar. *Scree plot* adalah plot antara nilai eigen dengan banyaknya komponen utama yang terbentuk.

2.10 Penentuan *Bandwidth*

Secara teoritis, *bandwidth* merupakan lingkaran dengan radius *h* dari titik lokasi yang digunakan sebagai dasar untuk menentukan bobot setiap pengamatan terhadap model regresi pada lokasi tersebut. Besarnya nilai pembobot bergantung pada pemilihan *bandwidth.* Salah satu metode untuk menentukan *bandwidth* optimum adalah *Cross Validation* (CV). *Bandwidth* optimum adalah *bandwidth* yang menghasilkan nilai CV minimum. Secara matematis CV dapat dituliskan sebagai berikut (Fotheringham, 2002):

$$CV = \sum_{i=1}^{n}(y_i - \hat{y}_{\neq i}(h))^2 \qquad (2.18)$$

Dengan $\hat{y}_{\neq i}(h)$ adalah penduga y_i dengan pengamatan di lokasi (u_i, v_i) dihilangkan dari proses prediksi.

2.11 Fungsi Pembobot

Pemilihan pembobot pada analisis spasial sangat penting dikarenakan nilai pembobot mewakili letak data observasi satu dengan lainnya. Fungsi pembobot spasial berbasis pada kedekatan lokasi pengamatan ke-*i* dengan lokasi pengamatan lainnya sehingga lokasi yang dekat dari nilai ketetapan memiliki pembobot yang jauh lebih besar. Nilai pembobot regresi spasial berbeda disetiap pengamatan karena dipengaruhi oleh kedekatan lokasi sebaliknya tiap pengamatan pada regresi global memiliki matriks pembobot yang sama. Perbedaan nilai pembobot regresi

spasial menghasilkan penaksiran parameter yang berbeda disetiap lokasi pengamatan.

Matriks pembobot pada data spasial berukuran $(n \times n)$ yang elemen-elemen diagonalnya menunjukkan pembobot geografis pada lokasi ke-*i*. Matriks pembobot tersebut dihitung untuk setiap lokasi pada pengamatan ke-*i* (Huang *et al*., 2010). Terdapat beberapa fungsi yang dapat digunakan untuk menentukan besarnya pembobot pada masing-masing lokasi yang berbeda, diantaranya menggunakan fungsi *fixed gaussian kernel*. Adapun persamaan fungsi *fixed gaussian kernel* adalah sebagai berikut:

$$w_{ij} = \exp\left[-\frac{1}{2}\left(\frac{d_{ij}}{h}\right)^2\right] \tag{2.19}$$

Keterangan:

w_{ij} : Nilai bobot dari pengamatan pada lokasi ke-*i*

d_{ij} : Jarak *euclidean* antara lokasi ke-*i* dengan lokasi ke-*j*

h : Nilai *bandwidth*

Jarak *euclidean* (d_{ij}) adalah jarak antara titik lokasi pengamatan ke-*i* dengan titik lokasi pengamatan ke-*j*. Adapun perhitungan rumus jarak *euclidean* (d_{ij}) adalah sebagai berikut:

$$d_{ij} = \sqrt{(u_i - u_j)^2 + (v_i - v_j)^2} \tag{2.20}$$

Dengan u_i menunjukkan *latitude* pada lokasi ke-*i*, u_j *latitude* pada lokasi ke-*j*, v_i *longitude* pada lokasi ke-*i*, dan v_j *longitude* pada lokasi ke-*j*.

2.12 Pembobot *Geographically Weighted Logistic Regression* (GWLR)

Pendugaan parameter GWLR dilakukan dengan memberikan pembobot yang berbeda untuk setiap lokasi pengamatan. Setiap parameter pada model GWLR dihitung berdasarkan titik lokasi geografis sehingga menghasilkan nilai parameter yang berbeda antar satu lokasi dengan lokasi lainnya. Adapun matriks pembobot spasial model GWLR dengan dimensi $(n \times n)$ dinyatakan sebagai berikut (Fotheringham *et al*., 2002):

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

2.13 Model *Geographically Weighted Logistic Regression* (GWLR)

Geographically Weighted Logistic Regression merupakan salah satu metode regresi logistik yang mempertimbangkan faktor spasial sehingga menghasilkan nilai parameter bagi masing-masing titik atau lokasi. Metode ini dikembangkan dari metode GWR yang digunakan untuk memprediksi atau menduga model dari kumpulan data yang memiliki peubah respon biner melalui model logistik (Pravitasary dkk, 2015). Pada model GWLR, lokasi geografis dimasukkan ke dalam model melalui fungsi pembobot. Adapun bentuk model *Geographically Weighted Logistic Regression* (GWLR) adalah sebagai berikut:

$$\pi(x_i) = \frac{\exp(\beta_0(u_i, v_i) + \beta_1(u_i, v_i)x_{i1} + \beta_2(u_i, v_i)x_{i2} + \cdots + \beta_k(u_i, v_i)x_{ik})}{1 + \exp(\beta_0(u_i, v_i) + \beta_1(u_i, v_i)x_{i1} + \beta_2(u_i, v_i)x_{i2} + \cdots + \beta_k(u_i, v_i)x_{ik})}$$

$$= \frac{\exp\ (\boldsymbol{\beta}^T(u_i,v_i)\ \boldsymbol{x}_i)}{1+\exp\ (\boldsymbol{\beta}^T(u_i,v_i)\ \boldsymbol{x}_i)}, i = 1,2 \dots, n\ ; k = 0, 1, 2, \dots, p \qquad (2.21)$$

Keterangan:

$\boldsymbol{x}_i$: $(1, x_{i1}, x_{i2}, \dots\ x_{ik})^T$

$\boldsymbol{\beta}^T(u_i, v_i)$: $\left(\beta_0(u_i, v_i),\ \beta_1(u_i, v_i), \dots, \beta_k(u_i, v_i)\right)$

(u_i, v_i) : Koordinat lokasi pengamatan ke-i

Bentuk logit untuk GWLR adalah:

$$g(x) = \ln\left(\frac{\pi(x)}{1-\pi(x)}\right) = \beta_0(u_i, v_i) + \sum_{k=1}^{p} \beta_k(u_i, v_i)x_{ik} \qquad (2.22)$$

Pengujian parameter model GWLR digunakan untuk mengetahui parameter yang berpengaruh signifikan terhadap model. Pengujian dilakukan secara serentak dan parsial. Pengujian secara serentak menggunakan uji rasio *likelihood* dengan hipotesisnya sebagai berikut:

$H_0 : \beta_1(u_i, v_i) = \beta_2(u_i, v_i) = \cdots = \beta_k(u_i, v_i) = 0$ (tidak terdapat pengaruh antara variabel prediktor terhadap variabel respon)

H_1 : paling tidak terdapat satu $\beta_k(u_i, v_i) \neq 0$ (terdapat pengaruh paling sedikit satu variabel prediktor terhadap variabel respon)

Statistik Uji:

$$G^2\ (u_i, v_i) = 2[\sum_{i=1}^{n}(y_{1i} \ln \hat{\pi}(x_i) + \ y_{0i} \ln(1 - \hat{\pi}(x_i)) - \sum_{i=1}^{n}(n_{1i} \ln(n_{1i}) + \ n_{0i} \ln(n_{0i}) + n \ln(n))] \quad (2.23)$$

Dengan y_{0i} menunjukkan nilai variabel respon berkategori nol pada lokasi ke-*i*, y_{1i} menunjukkan nilai variabel respon berkategori satu pada lokasi ke-*i*, n_{0i} adalah jumlah data yang berkategori nol, n_{1i} adalah jumlah data yang berkategori satu, dan n adalah total data.

Statistik uji $G^2(u_i, v_i)$ mendekati distribusi *chi-square.* Kriteria pengujiannya adalah tolak H_0 jika nilai $G^2 > \chi^2_{a;p}$ dengan p adalah jumlah variabel prediktor dan nilai $\chi^2_{a;p}$ dapat diperoleh dari tabel *chi-square.* Selanjutnya dilakukan parameter model GWLR secara parsial dengan hipotesis sebagai berikut:

H_0 : $\beta_k(u_i, v_i) = 0$

H_1 : paling tidak terdapat satu $\beta_k(u_i, v_i) \neq 0, \quad k = 1,2, \dots, p\ ; i = 1,2, \dots, n$

Statistik uji:

$$W = \frac{\hat{\beta}_k(u_i,v_i)}{SE\left(\hat{\beta}_k(u_i,v_i)\right)} \quad (2.24)$$

Dengan $SE\ \left(\hat{\beta}_k(u_i, v_i)\right) = \sqrt{var\ \left(\hat{\beta}_k(u_i, v_i)\right)}$ dan kriteria ujinya yaitu tolak H_0 jika $|W| > \ Z_{\alpha/2}$ (Hosmer dan Lemeshow, 2000).

2.14 Interpretasi Model *Geographically Weighted Logistic Regression* (GWLR)

Interpretasi parameter pada model regresi menunjukkan hubungan pada variabel prediktor terhadap variabel respon. Menurut Hosmer & Lemeshow (2000), cara menginterpretasi parameter regresi dari variabel respon kategorik adalah dengan *odds ratio.* Hubungan antara *odds ratio* dan koefisien regresi pada model GWLR adalah sebagai berikut:

$$\text{OR}\ (u_i, v_i) = \frac{\frac{\pi(u_i,v_i)(1)}{1-\pi(u_i,v_i)(1)}}{\frac{\pi(u_i,v_i)(0)}{1-\pi(u_i,v_i)(0)}} = \exp(\hat{\beta}_k(u_i, v_i)) \qquad (2.25)$$

Setiap kenaikan satu satuan x_k maka nilai π akan menurunkan atau meningkatkan sebesar eksponensial $\hat{\beta}_k$.

2.15 Kriteria Kategorisasi

Statistik deskriptif untuk masing-masing variabel prediktor dalam penelitian ini menggunakan kategorisasi dengan tiga kategori yaitu rendah, sedang, dan tinggi. Adapun rumus kategorisasi variabel prediktor yang digunakan adalah sebagai berikut:

Tabel 2. 1. Kriteria Kategorisasi

Kategorisasi	
Rendah	$X < \bar{X} - \sigma$
Sedang	$\bar{X} - \sigma \leq X < \bar{X} + \sigma$
Tinggi	$\bar{X} + \sigma \leq X$

Sumber: Setiawati, 2014

Dengan $\bar{X}$ adalah rata-rata (*mean*) dan σ adalah standar deviasi.

2.16 Ketepatan Klasifikasi (KK)

Ketepatan Klasifikasi adalah suatu evaluasi untuk melihat probabilitas kesalahan klasifikasi yang dilakukan oleh suatu fungsi klasifikasi. Nilai ketepatan klasifikasi diperoleh dengan membandingkan nilai prediksi ($\hat{Y}_i$) dengan nilai observasi (Y_i). Adapun tabel ketepatan klasifikasi GWLR disajikan pada Tabel 2.2 sebagai berikut:

Tabel 2. 2. Ketepatan Klasifikasi

Lokasi ke-i	**Kabupaten/Kota**	$\boldsymbol{Y_i}$	$\boldsymbol{\hat{\pi}}(u_i, v_i)$	$\boldsymbol{\hat{Y}_i}$	**KK**(u_i, v_i)
1					
2					
⋮	⋮	⋮	⋮	⋮	⋮
n					
$\sum_{i=1}^{n} \mathbf{KK}(\boldsymbol{u_i}, \boldsymbol{v_i})$					

Sumber: Yunus, 2016

Keterangan:

$$\hat{Y}_i = \begin{cases} 1, & \text{jika } \hat{\boldsymbol{\pi}}(u_i, v_i) \geq 0.5 \\ 0, & \text{jika } \hat{\boldsymbol{\pi}}(u_i, v_i) < 0.5 \end{cases}$$

Nilai ketepatan klasifikasi adalah sebagai berikut,

$$\text{KK}(u_i, v_i)_i = \begin{cases} 1, \text{jika } Y_i = 1 \text{ dan } \hat{Y}_i = 1 \text{ atau } Y_i = 0 \text{ dan } \hat{Y}_i = 0 \\ 0, \quad \text{untuk yang lainnya} \end{cases}$$

Perhitungan Ketepatan Klasifikasi (KK) sebagai berikut:

$$\text{KK} = \frac{\sum_{i=1}^{24} \text{KK}(u_i, v_i)}{n} \times 100\% \tag{2.26}$$

2.17 Indeks Pembangunan Kesehatan Manusia (IPKM)

Kesehatan merupakan hak asasi manusia dan salah satu unsur kesejahteraan yang harus diwujudkan sesuai dengan cita-cita bangsa Indonesia sebagaimana dimaksud dalam Pancasila dan Undang-Undang Dasar Negara Republik Indonesia Tahun 1945 (Sholihah, 2015). Upaya peningkatan derajat kesehatan sangat penting untuk dilakukan secara terus menerus dan berkesinambungan. Hendrik L. Blum mengungkapkan ada empat faktor yang mempengaruhi status derajat kesehatan antara lain lingkungan, perilaku hidup sehat, pelayanan kesehatan, dan genetik (keturunan). Faktor-faktor tersebut tidak dapat diukur secara langsung sehingga membutuhkan beberapa indikator. IPKM adalah indikator yang menggambarkan keberhasilan kesehatan masyarakat sehingga dapat diketahui karakteristik kesehatan kabupaten/kota. Dalam penelitian ini menggunakan lima variabel yang mencakup faktor status derajat kesehatan menurut Hendrik L. Blum. Penjelasan terkait faktor tersebut adalah sebagai berikut:

1. Lingkungan

Lingkungan umumnya digolongkan menjadi dua kategori yaitu berhubungan dengan aspek fisik dan sosial. Faktor lingkungan yang dikaitkan dengan aspek fisik seperti sampah, air, udara, tanah, iklim, dan perumahan. Sedangkan lingkungan sosial merupakan hasil interaksi antar manusia seperti kebudayaan, kepercayaan, pendidikan, dan ekonomi. Faktor lingkungan yang dimaksud menekankan pada kondisi lingkungan yang sehat. Keadaan lingkungan yang sehat dapat tercipta dengan terwujudnya kesadaran masyarakat untuk

berperilaku hidup bersih dan sehat (PHBS). Indikator-indikator yang digunakan untuk mengukur faktor kesehatan pada IPKM yaitu berupa akses sanitasi, akses air bersih, dan kepadatan penduduk.

2. Perilaku Hidup Sehat

Perilaku hidup bersih dan sehat masyarakat merupakan salah satu faktor yang menentukan derajat kesehatan untuk mendukung peningkatan kesehatan masyarakat. Indikator yang digunakan untuk mengukur perilaku hidup sehat masyarakat antara lain merokok, perilaku cuci tangan yang benar, perilaku buang air besar (BAB) yang benar yaitu BAB di jamban, aktivitas fisik yang cukup, serta perilaku hidup bersih dan sehat (PHBS) yang benar.

3. Pelayanan Kesehatan

Keberadaan fasilitas kesehatan sangat menentukan dalam pelayanan pemulihan kesehatan, pencegahan terhadap penyakit, pengobatan, dan keperawatan terhadap kelompok dan masyarakat yang memerlukan pelayanan kesehatan. Bentuk pelayanan kesehatan tidak hanya terbatas pada fasilitas pelayanan saja, akan tetapi juga meliputi tenaga kesehatan. Indikator-indikator yang dapat digunakan untuk mengukur variabel pelayanan kesehatan antara lain: persalinan ditolong oleh tenaga kesehatan, kecukupan jumlah dokter, kecukupan jumlah posyandu, kecukupan jumlah bidan, dan kepemilikan jaminan pelayanan kesehatan.

4. Genetik (Keturunan)

Nasib suatu bangsa ditentukan oleh kualitas generasi mudanya. Oleh karena itu, untuk menyokong pembangunan bangsa perlu untuk meningkatkan kualitas generasi muda. Genetik (keturunan) merupakan faktor yang telah ada dalam diri manusia yang dibawa sejak lahir, misalnya dari golongan penyakit keturunan seperti diabetes melitus dan asma brohenial. Selain itu, faktor keturunan juga dapat dikaji dari kondisi balita dan ibu hamil. Masa kehamilan dan balita sangat menentukan perkembangan otak anak. Indikator yang digunakan untuk mengetahui derajat kesehatan yang dipengaruhi oleh faktor keturunan antara lain prevalensi balita sangat pendek dan pendek, prevalensi obesitas sentral, hipertensi, dan diabetes melitus.