PEMODELAN *GEOGRAPHICALLY WEIGHTED REGRESSION* DENGAN FUNGSI PEMBOBOT *KERNEL* PADA DATA MULTIKOLINEARITAS

SKRIPSI

DIAH LESTARI

H051191034

PROGRAM STUDI STATISTIKA DEPARTEMEN STATISTIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS HASANUDDIN

MAKASSAR

2023

PEMODELAN *GEOGRAPHICALLY WEIGHTED REGRESSION* DENGAN FUNGSI PEMBOBOT *KERNEL* PADA DATA MULTIKOLINEARITAS

SKRIPSI

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

DIAH LESTARI

H051191034

PROGRAM STUDI STATISTIKA DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS HASANUDDIN
MAKASSAR
AGUSTUS 2023

LEMBAR PERNYATAAN KEOTENTIKAN

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

Pemodelan *Geographically Weighted Regression* dengan Fungsi Pembobot *Kernel* pada Data Multikolinearitas

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

Makassar, 11 Agustus 2023

Diah Lestari

NIM H051191034

PEMODELAN *GEOGRAPHICALLY WEIGHTED REGRESSION* DENGAN FUNGSI PEMBOBOT *KERNEL* PADA DATA MULTIKOLINEARITAS

Disetujui Oleh:

Pembimbing Utama	**Pembimbing Pertama**
Drs. Raupong, M.Si.	**Andi Kresna Jaya, S.Si., M.Si.**
NIP. 19621015 198810 1 001	**NIP. 19731228 200003 1 001**

Ketua Program Studi

Dr. Anna Islamiyati, S.Si., M.Si.

NIP. 19770808 200501 2 002

Pada 11 Agustus 2023

HALAMAN PENGESAHAN

Skripsi ini diajukan oleh :

Nama	: Diah Lestari
NIM	: H051191034
Program Studi	: Statistika
Judul Skripsi	: Pemodelan *Geographically Weighted Regression* dengan Fungsi Pembobot *Kernel* pada Data Multikolinearitas

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

UNIVERSITAS HASANUDDIN

DEWAN PENGUJI

1. Ketua : Drs. Raupong, M.Si. (..............................)
2. Sekretaris : Andi Kresna Jaya, S.Si., M.Si. (..............................)
3. Anggota : Siswanto, S.Si., M.Si. (..............................)
4. Anggota : Dr. Nirwan, M.Si. (..............................)

Ditetapkan di : Makassar

Tanggal : 11 Agustus 2023

KATA PENGANTAR

Assalamu'alaikum Warahmatullahi Wabaraktuh

Puji syukur penulis panjatkan kepada Allah *Subhanahu Wa Ta'ala* atas segala limpahan rahmat, nikmat, dan karunia-Nya sehingga penulis dapat menyelesaikan penyusunan skripsi ini. Shalawat dan salam senantiasa tercurahkan kepada baginda Rasulullah *Shallallahu '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 skripsi yang berjudul "**Pemodelan *Geographically Weighted Regression* dengan Fungsi Pembobot *Kernel* pada Data Multikolinearitas"** 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.

Penulis menyadari bahwa dalam penyelesaian skripsi ini tidak lepas dari bantuan dan dorongan dari berbagai pihak yang senantiasa turut membantu dalam bentuk moril maupun materil sehingga dengan segala keterbatasan kemampuan dan pengetahuan, penulis dapat menyelesaikan skripsi ini. Oleh karena itu, penulis menyampaikan ucapan terima kasih yang setulus-tulusnya serta penghargaan yang setinggi-tingginya kepada Ayahanda **Suardi (Alm)** dan Ibunda **Sariyem** yang telah memberikan dukungan penuh, pengorbanan luar biasa, limpahan cinta dan kasih sayang, kesabaran hati, serta dengan ikhlas telah menemani setiap langkah penulis dengan doa dan restu mulianya. Ucapan terima kasih juga penulis haturkan kepada Kakak tersayang **Deni Setyadi** yang senantiasa mendengarkan segala keluh kesah penulis walau terkadang tingkahnya menyebalkan. Terima kasih juga kepada Kakak ipar penulis **Ayu Eka Ramadhani, S.M.** yang selalu menghibur dan memberi semangat kepada penulis dalam menyelesaikan skripsi ini, serta keluarga besar penulis, terima kasih atas doa mulia dan dukungannya selama ini.

Penghargaan yang tulus dan ucapan terima kasih dengan penuh keikhlasan dan ketulusan 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.**, selaku Ketua Departemen Statistika telah meluangkan waktunya untuk memberikan masukan dan motivasi.
4. **Bapak Dr. Raupong, M.Si.**, selaku Pembimbing Utama sekaligus Penasehat Akademik penulis yang dengan penuh kesabaran telah meluangkan waktu dan pemikirannya untuk senantiasa memberikan arahan, dorongan semangat, dan motivasi kepada penulis dari awal hingga selesainya penulisan tugas akhir ini.
5. **Bapak Andi Kresna Jaya, S.Si., M.Si.** selaku Pembimbing Pertama penulis yang telah ikhlas meluangkan waktunya ditengah kesibukan dan sabar dalam memberikan arahan dan masukan bagi penulis.
6. **Bapak Siswanto, S.Si., M.Si.** dan **Bapak Dr. Nirwan, M.Si.**, selaku Tim Dosen Penguji yang telah ikhlas meluangkan waktu dalam memberikan saran serta kritikan yang membangun dalam penyempurnaan tugas akhir ini.
7. Segenap **Dosen Pengajar dan Staf Departemen Statistika** yang telah membekali ilmu dan kemudahan kepada penulis dalam berbagai hal selama menempuh pendidikan sarjana di Departemen Statistika.
8. Sahabat tercinta penulis sejak awal perkuliahan, **Andini We Tenri Maharani**, **Fadhila Febriyanti Najamuddin**, **Nurul Hikmah**, dan **Wahyu Dwi Rahmawati** yang telah menjadi sahabat terbaik serta senantiasa memberikan semangat, mendengarkan segala keluh kesah penulis dalam hal apapun, dan menemani perjalanan suka dan duka penulis selama ini.
9. Sahabat terbaik penulis sejak di bangku SMA, **Ainun Ismaya Ismail**, **Andi Raihana Afiyah**, **Diva Adishya W**, dan **Faradias Izza Azzahra Faisal**, serta keluarga besar **XII MIPA 1 (*Courvarien*) MAN 2 Kota Makassar** yang senantiasa memberi semangat dan menemani perjuangan pendidikan penulis.
10. Sahabat seperjuangan di Statistika 2019, **A. Ahmad Qeis Tenridapi**, **Agus Hermawan, S.Si.**, **Alfiyah Salsa Dila Sabir**, **Alya Safira Irtiqa Miolo, S.Si.**, **Amalia Mentari Djalumang**, **Muh. Nur Iskandar Zulkarnain**, **Muhammad Rayhan Rifaldi, S.Si.**, **Muhammad Syamsul Bahri**, **Muhammad Yusran, S.Si.**, **Muharti Apriana**, **Muliana**, **Nurazizah**, **Nur Aisyah**, **Nursyahfika**, **Sapriadi Rasyid**, **Ummul Auliyah Syam, S.Si.**,

Terima kasih atas kebersamaan, kebahagiaan, kesedihan, serta kebaikannya menjadi sosok guru bagi penulis. Terima kasih telah mengukir kenangan indah bersama penulis selama masa perkuliahan.

10. Seluruh teman **Posko 21 Desa Toddotoa KKN PUPR UNHAS Gel. 108**, yaitu **Adi Jahyadi**, **Danti Fitriani Ohorella**, **Diva Adishya W**, **Fachrial Novri Triyandi**, **Friselia Tesya Anjelin**, **Kak Irvan Tandililing**, **M. Ridzuan**, dan **Riswanda Rahmayani** yang telah menjadi keluarga selama 40 hari dan senantiasa memberikan semangat dan dukungan kepada penulis. Semoga tali silaturahmi kita tetap terjalin ke depannya.
11. Teman terbaik penulis selama di bangku SMP, **Aulia Meidina Pawindru**, **Kurniawati Rachmat**, dan **A. Muh. Akil Aminuar**. Terima kasih untuk senantiasa memberi semangat kepada penulis.
12. Sepupu tercinta penulis, **Refina Dwi Anjarwati (Almh)** dan **Mbak Renita Dyah R**, terima kasih atas semangat dan motivasinya kepada penulis.
13. Teman seperjuangan di **Statistika 2019**. Terima kasih atas ilmu, kebersamaan, suka dan duka dalam menjalani perkuliahan di Departemen Statistika. Terima kasih sudah menerima kehadiran penulis. Kita hebat dan kita luar biasa bisa bertahan hingga sejauh ini.
14. **Keluarga Besar Karang Taruna Cipto Rukun**, terima kasih atas pengalaman dan kegiatan positif yang telah dilakukan sejak tahun lalu. Hal ini sangat bermanfaat sebagai selingan penulis dalam mengerjakan tugas akhir.
15. Kepada seluruh pihak yang tidak dapat penulis sebutkan satu persatu, terima kasih setinggi-tingginya untuk segala dukungan, partisipasi, dan apresiasi yang diberikan kepada penulis.

Penulis menyadari bahwa masih banyak kekurangan dalam penyusunan skripsi ini, untuk itu dengan segala kerendahan hati penulis memohon maaf. Akhir kata, semoga tulisan ini dapat memberikan manfaat untuk berbagai pihak.

Makassar, 11 Agustus 2023

Diah Lestari

PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIK

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

Nama	: Diah Lestari
NIM	: H051191034
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:

"Pemodelan *Geographically Weighted Regression* dengan Fungsi Pembobot *Kernel* pada Data Multikolinearitas"

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 tanggal 11 Agustus 2023.

Yang menyatakan,

(Diah Lestari)

ABSTRAK

Kondisi pengamatan di suatu lokasi dengan lokasi lainnya itu tidak selalu sama yang disebut heterogenitas spasial. Oleh karena itu, dalam menjelaskan hubungan-hubungan antara variabel respon dan dua/lebih variabel prediktor antar satu lokasi dengan lokasi lain dapat dilakukan dengan pendekatan regresi spasial. Selain masalah heterogenitas spasial, regresi berganda yang mempunyai banyak variabel prediktor terkadang menimbulkan masalah multikolinearitas karena terjadi korelasi yang tinggi antar variabel prediktornya. Salah satu cara untuk mengatasi multikolinearitas pada model regresi spasial adalah dengan menggunakan kombinasi metode *Geographically Weighted Regression Principal Component Analysis* (GWRPCA). Penelitian ini bertujuan untuk mengetahui model kombinasi linear yang terbentuk dari analisis PCA dan nilai komponen utamanya serta membentuk model GWRPCA yang menjelaskan faktor-faktor yang mempengaruhi Angka Kematian Bayi (AKB) di Provinsi Sulawesi Selatan. Berdasarkan analisis PCA, diperoleh variabel-variabel yang mempengaruhi Angka Kematian Bayi (AKB) di Sulawesi Selatan tahun 2020 dapat diwakili oleh 2 komponen utama yaitu variabel P_{c1} yang mampu menerangkan total varian data sebesar 74.6807% dan variabel P_{c2} yang mampu menerangkan total varian data sebesar 22.9734%. Pembobotan menggunakan fungsi *fixed gaussian kernel* melalui pemilihan *bandwidth* optimum sebesar 440.9749 dengan kriteria *Cross Validation* (CV) minimum sebesar 229.3496. GWRPCA merupakan model terbaik untuk memodelkan AKB di Sulawesi Selatan dibandingkan model RPCA karena memiliki nilai *Akaike Information Criterion* (AIC) dan *Bayesian Information Criterion* (BIC). Model GWRPCA menunjukkan bahwa parameter yang dihasilkan dari setiap model masing-masing kabupaten/kota dominan menghasilkan parameter bernilai negatif yang menunjukkan bahwa setiap peningkatan komponen utama akan menurunkan angka kematian bayi di kabupaten/kota di Sulawesi Selatan.

Kata Kunci: Angka Kematian Bayi, *Geographically Weighted Regression*, Multikolinearitas, *Principal Component Analysis*

ABSTRACT

The observation conditions in one location compared to another are not always the same, a phenomenon known as spatial heterogeneity. Therefore, to elucidate the relationships between response variables and two or more predictor variables across different locations, a spatial regression approach can be applied. Apart from the issue of spatial heterogeneity, multiple regression with numerous predictor variables can sometimes lead to multicollinearity problems due to high correlation among the predictors. One way to address multicollinearity in spatial regression models is by employing a combination of Geographically Weighted Regression Principal Component Analysis (GWRPCA) methods. This research aims to is to understand the linear combination model formed from PCA analysis and its principal component values, and to construct a GWRPCA model that explains the influencing factors of Infant Mortality Rate (IMR) in the South Sulawesi Province. Based on PCA analysis, the variables influencing the IMR in South Sulawesi for the year 2020 are represented by two principal components: the variable P_{c1}, which accounts for 74.6807% of the total data variance, and the variable P_{c2}, which accounts for 22.9734% of the total data variance. Weights are assigned using a fixed gaussian kernel function, with an optimal bandwidth selection of 440.9749 based on the minimum Cross Validation (CV) criterion of 229.3496. GWRPCA proves to be a superior model for representing IMR in South Sulawesi compared to the RPCA model, as evidenced by lower values of the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). The GWRPCA model indicates that the parameters derived for each district or city predominantly yield negative values. This suggests that an increase in the principal components will lead to a decrease in infant mortality rates in the districts and cities of South Sulawesi.

Keywords: *Infant Mortality Rate, Geographically Weighted Regression, Multicollinearity, Principal Component Analysis*

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

DAFTAR LAMPIRAN

BAB I
PENDAHULUAN

1.1 Latar Belakang

Pembentukan model dapat menunjukkan pola hubungan antara variabel respon dan dua/lebih variabel prediktor dan dapat dilakukan dengan pendekatan analisis regresi berganda. Akan tetapi, model yang terbentuk dengan pendekatan regresi berganda masih bersifat global karena diterapkan pada seluruh lokasi pengamatan (Erdkhadifa, 2021). Padahal, telah diketahui bahwa kondisi pengamatan di suatu lokasi dengan lokasi lainnya itu tidak selalu sama atau dalam statistika disebut heterogenitas spasial. Oleh karena itu, dalam menjelaskan hubungan-hubungan antara variabel respon dan dua/lebih variabel prediktor antar satu lokasi dengan lokasi lain dapat dilakukan dengan pendekatan regresi spasial.

Salah satu model regresi spasial yang umum digunakan untuk mengatasi heterogenitas spasial adalah *Geographically Weighted Regression* (GWR). Analisis GWR memodelkan hubungan antara variabel respon dan variabel prediktor dengan mempertimbangkan bobot yang dihitung berdasarkan jarak dari satu lokasi dengan lokasi lainnya. Pembobotan ini membuat estimasi parameter dalam model GWR menjadi lebih bervariasi secara spasial yang bertujuan untuk memperoleh informasi yang lebih mendalam untuk setiap titik pengamatan yang diteliti. Estimasi parameter pada metode GWR memerlukan adanya matriks pembobot berdasarkan kedekatan dengan lokasi pengamatan. Fungsi pembobot *kernel* adalah fungsi pembobot yang umum digunakan dalam membentuk matriks pembobot.

Selain masalah heterogenitas, regresi berganda yang mempunyai banyak variabel prediktor terkadang menimbulkan masalah multikolinearitas karena terjadi korelasi yang tinggi antar variabel prediktornya. Multikolinearitas dapat diatasi dengan banyak metode, salah satunya dengan menggunakan metode *Principal Component Analysis* (PCA) yang bertujuan untuk mereduksi dimensi suatu data tanpa mengurangi karakteristik data tersebut secara signifikan. Oleh karena itu, masalah multikolinearitas yang terjadi pada data spasial bisa diatasi dengan metode kombinasi GWRPCA.

Heterogenitas spasial dapat dilihat salah satunya pada fenomena kesehatan. Masyarakat pedesaan cenderung untuk tidak berobat ke tenaga kesehatan dan memilih berobat pada dukun desa. Sebaliknya, masyarakat perkotaan cenderung untuk berobat ke tenaga kesehatan walau hanya sakit biasa. Kesehatan merupakan salah satu kebutuhan dasar manusia sehingga kesehatan termasuk dalam hak asasi manusia bagi setiap warga negara yang dilindungi oleh undang-undang (Ardinata, 2020). Usia bayi adalah usia yang paling rentan mengalami gangguan kesehatan dan bisa berakibat fatal apabila tidak diberi penanganan bahkan dapat mengakibatkan kematian pada bayi (Lengkong dkk., 2020). Angka Kematian Bayi (AKB) adalah banyaknya kematian bayi sebelum mencapai umur 1 tahun per 1000 kelahiran hidup pada periode waktu yang sama (Prahutama dkk., 2017).

Angka kematian bayi merupakan salah satu indikator penting dalam menentukan tingkat derajat kesehatan masyarakat dan keberhasilan pembangunan suatu daerah. Angka kematian bayi dipengaruhi oleh faktor endogen dan faktor eksogen. Faktor endogen meliputi pemberian Tablet Tambah Darah (TTD) kepada ibu hamil dan persalinan yang ditolong tenaga kesehatan serta faktor eksogen meliputi pemberian ASI eksklusif dan vitamin A kepada bayi. Berdasarkan data Kementerian Kesehatan Republik Indonesia tahun 2020, Sulawesi Selatan masuk dalam sepuluh provinsi dengan angka kematian bayi tertinggi di Indonesia dan tertinggi diantara semua provinsi di Pulau Sulawesi. Adapun daerah dengan angka kematian bayi tertinggi adalah Kabupaten Enrekang sedangkan angka kematian bayi terendah adalah Kota Makassar. Terjadi perbandingan kejadian kematian pada daerah pedesaan dan perkotaan sehingga bisa diatasi dengan regresi spasial.

Beberapa penelitian yang telah menggunakan model GWR pada data yang mengalami multikolinearitas dengan pembobot *kernel* antara lain Nurmalita Sari (2016) melakukan pemodelan GWRPCA menggunakan data pendapatan asli daerah di Jawa Tengah dengan penentuan fungsi pembobot berdasarkan nilai AIC yang menghasilkan fungsi pembobot terbaik menggunakan *fixed gaussian kernel* dan Musdalifah (2021) melakukan penelitian pada data PDRB Sulawesi Selatan yang mengalami multikolinearitas dengan menggunakan analisis komponen utama dan regresi terboboti geografis dengan batasan masalah menggunakan fungsi pembobot *adaptive gaussian kernel*.

Berdasarkan uraian sebelumnya, peneliti tertarik untuk mengkaji hal tersebut dengan judul "**Pemodelan *Geographically Weighted Regression* dengan Fungsi Pembobot *Kernel* pada Data Multikolinearitas**" menggunakan data Angka Kematian Bayi (AKB) Provinsi Sulawesi Selatan tahun 2020.

1.2 Rumusan Masalah

1. Berapakah variabel baru yang terbentuk dengan menggunakan metode *Principal Component Analysis* dan bagaimana model kombinasi linear yang terbentuk dari variabel-variabel prediktor di Provinsi Sulawesi Selatan?
2. Bagaimana pemodelan *Geographically Weighted Regression Principal Component Analysis* pada data angka kematian bayi di Provinsi Sulawesi Selatan yang mengalami multikolinearitas dengan fungsi pembobot *kernel*?

1.3 Batasan Masalah

1. Estimasi parameter yang digunakan pada penelitian ini adalah *Weighted Least Square* (WLS).
2. Perhitungan jarak antar lokasi pengamatan menggunakan jarak *Euclidean*.
3. Setiap titik lokasi pengamatan didasarkan pada letak kantor kepala daerah kabupaten/kota.
4. Pemilihan model terbaik menggunakan nilai *Akaike Information Criterion* (AIC) dan *Bayesian Information Criterion* (BIC).
5. Penelitian ini diaplikasikan pada data angka kematian bayi, jumlah bayi diberi ASI eksklusif, jumlah bayi diberi vitamin A, persentase ibu hamil yang mendapatkan tablet tambah darah, dan jumlah persalinan ditolong tenaga kesehatan di Provinsi Sulawesi Selatan tahun 2020 berdasarkan 24 kabupaten/kota.

1.4 Tujuan Penelitian

1. Memperoleh jumlah variabel baru yang terbentuk dengan menggunakan metode *Principal Component Analysis* dan memperoleh model kombinasi linear yang terbentuk dari variabel-variabel prediktor di Provinsi Sulawesi Selatan.
2. Memperoleh pemodelan *Geographically Weighted Regression Principal Component Analysis* pada data angka kematian bayi di Provinsi Sulawesi Selatan yang mengalami multikolinearitas dengan fungsi pembobot *kernel*.

1.5 Manfaat Penelitian

Manfaat dalam penelitian ini diharapkan dapat menambah pemahaman teoritis serta praktis bagi peneliti dan pembaca terkait pemodelan *Geographically Weighted Regression Principal Component Analysis* pada data angka kematian bayi di Provinsi Sulawesi Selatan. Hasil analisis data angka kematian bayi Provinsi Sulawesi Selatan diharapkan dapat menjadi bahan informatif dan advokasi ke pemerintahan terkhusus Dinas Kesehatan tingkat provinsi maupun kabupaten/kota agar dilakukan tindakan *preventing* terkait jenis kebijakan yang akan dilakukan oleh para pembuat keputusan (*decision making*) demi mengurangi angka kematian pada bayi baik itu berupa sosialisasi kepada para ibu maupun pelayanan kesehatan terhadap bayi.

BAB II
TINJAUAN PUSTAKA

2.1 Analisis Regresi

Analisis regresi merupakan metode statistik yang bertujuan untuk menguji hubungan sebab akibat antara variabel respon dan variabel prediktor. Analisis regresi digunakan untuk mengetahui variabel respon (y) dapat diprediksikan melalui variabel prediktor (x) secara parsial maupun simultan (Ayuni & Fitrianah, 2019). Model regresi yang melibatkan satu variabel prediktor (x) dengan satu variabel respon (y) disebut regresi linier sederhana. Bentuk umum dari model regresi linier sederhana adalah sebagai berikut (Suyono, 2018):

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \qquad , i = 1,2, \dots , n \tag{2.1}$$

Model regresi linier dapat melibatkan lebih dari satu variabel prediktor disebut model regresi linier berganda. Secara umum, jika terdapat k variabel prediktor, yakni $x_1, x_2, \dots , x_k$, maka model regresi liniernya adalah sebagai berikut:

$$y_i = \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij} + \varepsilon_i \tag{2.2}$$

dengan

y_i : variabel respon pada pengamatan ke-i

x_{ij} : variabel prediktor ke-j pada pengamatan ke-i

β_0 : parameter intersep (nilai ketika setiap variabel prediktor bernilai 0)

β_j : parameter koefisien regresi

ε_i : residual pada pengamatan ke-i

Pada model regresi linier, hubungan antara variabel prediktor dan variabel respon dianggap sama pada setiap lokasi geografis (Lutfiani dkk., 2019). Persamaan (2.2) dalam bentuk matriks ditulis sebagai berikut:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

atau

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \tag{2.3}$$

Parameter $\beta_0, \beta_1, \ldots, \beta_j$ adalah parameter-parameter yang nilainya tidak diketahui sehingga perlu dilakukan estimasi. *Ordinary Least Square* (OLS) atau metode kuadrat terkecil adalah salah satu metode estimasi parameter yang sering digunakan dengan tujuan untuk meminimumkan jumlah kuadrat residual. Berdasarkan persamaan (2.3), maka didapatkan persamaan sebagai berikut:

$$\boldsymbol{\varepsilon} = \boldsymbol{Y} - \boldsymbol{X\beta} \tag{2.4}$$

Selanjutnya, jumlah kuadrat residual diturunkan terhadap $\boldsymbol{\beta}$ dan disamakan dengan 0, sehingga diperoleh estimasi dengan metode OLS sebagai berikut:

$$\hat{\boldsymbol{\beta}} \quad = (\boldsymbol{X'X})^{-\mathbf{1}}\boldsymbol{X'Y} \tag{2.5}$$

2.2 Pengujian Statistika

2.2.1 Uji Asumsi Klasik

Uji Asumsi klasik adalah analisis yang dilakukan sebagai syarat untuk memastikan model regresi linier tidak terdapat masalah-masalah asumsi klasik. Jadi, regresi linier mengasumsikan bahwa terdapat hubungan linier antara kedua variabel. Jika hubungannya tidak linier, maka model regresi linier tersebut bukan merupakan metode yang ideal untuk analisis penelitian sehingga diperlukan suatu modifikasi pada variabel atau analisis tersebut (Mardiatmoko, 2020). Model regresi yang baik pada penelitian ini harus memenuhi asumsi klasik yang mencakup uji normalitas dan uji multikolinearitas.

2.2.1.1 Uji Normalitas

Uji normalitas bertujuan untuk memastikan nilai residual terdistribusi secara normal atau tidak. Model regresi yang baik adalah model dengan nilai residual yang terdistribusi secara normal (Mardiatmoko, 2020). Terdapat beberapa metode yang dapat digunakan untuk menguji normalitas suatu model seperti uji *Shapiro-Wilk*, uji *Lilliefors*, dan uji *Kolmogorov Smirnov*. Berikut adalah tahapan pengujian *Shapiro-Wilk*:

a. Hipotesis

   H_0: data berdistribusi normal

   H_1: data tidak berdistribusi normal

b. Taraf Signifikansi

   $\alpha = 5\%$

c. Statistik Uji

$$W = \frac{1}{D}\left[\sum_{i=1}^{n} a_i(\varepsilon_{n-i+1} - \varepsilon_i)\right]^2 \tag{2.6}$$

dengan:

$$D = \sum_{i=1}^{n} (\varepsilon_i - \bar{\varepsilon})^2$$

Keterangan:

a_i = Koefisien uji *Shapiro-Wilk*

ε_{n-i+1} = Residual ke $n - i + 1$

ε_i = Residual ke-i

$\bar{\varepsilon}$ = Rata-rata residual

d. Kriteria Kesimpulan

Tolak H_0 jika $W_{hit} < W_{tabel(\alpha;n)}$ yang berarti data tidak berdistribusi normal dan sebaliknya terima H_0 jika $W_{hit} \geq W_{tabel(\alpha;n)}$ yang berarti data berdistribusi normal.

Data yang berdistribusi normal cenderung memiliki pola simetri di sekitar nilai rata-rata yang menandakan bahwa frekuensi nilai yang lebih dan kurang dari rata-rata akan seimbang atau dengan kata lain data cenderung terkonsentrasi di sekitar nilai rata-rata. Selain itu, distribusi normal sering digambarkan sebagai kurva lonceng dengan puncak kurva lonceng sebagai nilai rata-rata data. Semakin besar variansi data maka bentuk kurva akan semakin rendah dan distribusinya semakin lebar.

2.2.1.2 Uji Multikolinearitas

Uji multikolinearitas adalah uji yang digunakan untuk melihat ada atau tidaknya korelasi yang tinggi antara variabel-variabel prediktor dalam suatu model regresi linier. Suatu model regresi dikatakan mengalami multikolinearitas jika terjadi hubungan linier yang sempurna atau mendekati pada beberapa atau semua variabel prediktor dalam model regresi (Mardiatmoko, 2020). Gejala multikolinearitas pada model regresi dapat ditentukan berdasarkan nilai *tolerance* dan nilai *Variance Inflation Factor* (VIF). Jika diperoleh nilai VIF yang tinggi maka nilai *tolerance* akan rendah. Hal ini dikarenakan $\text{VIF} = \frac{1}{tolerance}$ sehingga menunjukkan kolinearitas yang tinggi (Widana & Muliani, 2020). Hipotesis yang digunakan untuk menguji multikolinearitas adalah sebagai berikut:

a. Hipotesis

H_0: data mengalami multikolinearitas

H_1: data tidak mengalami multikolinearitas

b. Taraf Signifikansi

$\alpha = 5\%$

c. Statistik Uji

$$\mathrm{VIF} = \frac{1}{1 - \mathrm{R}^2} \tag{2.7}$$

dengan:

$$\mathrm{R}^2 = \frac{JK_R}{JK_T} = \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})^2}$$

d. Kriteria Kesimpulan

Tolak H_0 jika $\mathrm{VIF} < 10$ yang berarti data tidak mengalami multikolinearitas dan sebaliknya terima H_0 jika $\mathrm{VIF} \geq 10$ yang berarti terjadi multikolinearitas.

2.2.2 Uji Asumsi *Principal Component Analysis*

Sebelum melakukan *Principal Component Analysis* (PCA), pengujian diawali dengan pembentukan matriks korelasi. Matriks ini bertujuan untuk memperoleh nilai kedekatan hubungan antar variabel prediktor. Selanjutnya, nilai kedekatan ini digunakan pada beberapa pengujian untuk melihat kesesuaian dengan nilai korelasi yang diperoleh dari analisis PCA. Berdasarkan hal tersebut, terdapat asumsi yang harus dipenuhi yaitu asumsi kelayakan sampel menggunakan uji *Kaiser-Meyer-Olkin Measure of Sampling Adequacy* (KMO MSA).

2.2.2.1 Uji *Kaiser-Meyer-Olkin*

Uji *Kaiser-Meyer-Olkin* (KMO) bertujuan untuk melihat kecukupan sampel yang dianalisis dengan membandingkan besarnya koefisien korelasi dan koefisien korelasi parsial. Adapun hipotesis uji KMO adalah sebagai berikut (Kasim, 2021):

a. Hipotesis

H_0: ukuran data cukup untuk melakukan analisis PCA

H_1: ukuran data tidak cukup untuk melakukan analisis PCA

b. Taraf Signifikansi

$\alpha = 5\%$

c. Statistik Uji

$$\text{KMO} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} r_{x_i x_j}^2}{\sum_{i=1}^{n}\sum_{j=1}^{n} r_{x_i x_j}^2 + \sum_{i=1}^{n}\sum_{j=1}^{n} a_{x_i x_j}^2} \tag{2.8}$$

dengan:

$$r_{ij}^2 = \frac{\sum_{k=1}^{n}(x_{ik} - \bar{x}_i)\left(x_{jk} - \bar{x}_j\right)}{\sqrt{\sum_{k=1}^{n}(x_{ik} - \bar{x}_i)^2 \sum_{k=1}^{n}(x_{ik} - \bar{x}_i)^2}}$$

Keterangan:

r_{ij}^2 : koefisien korelasi antara variabel i dan *j*

a_{ij}^2 : koefisien korelasi parsial antara variabel i dan *j*

x_{ik} : nilai variabel *i* pada observasi ke-*k*

x_{jk} : nilai variabel *j* pada observasi ke-*k*

$\bar{x}_i$: rata-rata dari variabel *i*

$\bar{x}_j$: rata-rata dari variabel *j*

n : jumlah observasi

d. Kriteria Kesimpulan

Tolak H_0 jika nilai KMO < 0.5 yang berarti ukuran data tidak cukup untuk melakukan analisis PCA dan sebaliknya terima H_0 jika nilai KMO ≥ 0.5 yang berarti ukuran data cukup untuk melakukan analisis PCA

Nilai KMO yang kecil mengindikasikan bahwa penggunaan analisis PCA harus dipertimbangkan kembali, karena korelasi antar variabel asal tidak dapat diterangkan oleh variabel lainnya. Nilai KMO yang berada dibawah 0.5 dinilai tidak cukup layak untuk digunakan dan nilai KMO dari rentang 0.5 hingga 1 dinyatakan bisa digunakan untuk analisis PCA (Delsen dkk., 2017).

2.2.2.2 Uji *Measure of Sampling Adequacy*

Selain uji KMO, terdapat uji asumsi yang dilakukan untuk menunjukkan ukuran kecukupan data setiap variabel yaitu uji *Measure of Sampling Adequacy* (MSA). Nilai MSA berkisar antara 0 hingga 1. Apabila nilai MSA diatas 0.5 maka variabel dianggap dapat diprediksi dan dianalisis lebih lanjut. Akan tetapi, apabilai nilai MSA dibawah 0.5 maka variabel dianggap tidak dapat diprediksi dan dianalisis lebih lanjut sehingga variabel tersebut harus dieliminasi (Mudri &

Hardjomuljadi, 2019). Adapun bentuk hipotesis uji MSA adalah sebagai berikut (Puspitasari dkk., 2014):

a. Hipotesis

H_0: variabel telah memadai untuk dianalisis lebih lanjut

H_1: variabel belum memadai untuk dianalisis lebih lanjut

b. Taraf Signifikansi

$\alpha = 5\%$

c. Statistik Uji

$$\text{MSA} = \frac{\sum_{i=1}^{n} r_{x_i x_j}^2}{\sum_{i=1}^{n} r_{x_i x_j}^2 + \sum_{i=1}^{n} a_{x_i x_j}^2} \quad (2.9)$$

d. Kriteria Kesimpulan

Tolak H_0 jika nilai MSA < 0.5 yang berarti variabel belum memadai untuk dianalisis lebih lanjut dan sebaliknya terima H_0 jika nilai MSA ≥ 0.5 yang berarti variabel belum memadai untuk dianalisis lebih lanjut.

2.2.3 Uji Hipotesis Model *Geographically Weighted Regression*

Pengujian hipotesis pada model GWR terdiri dari dua uji, yaitu uji kesesuaian model dan uji signifikansi parameter model.

2.2.3.1 Uji Kesesuaian Model *Geographically Weighted Regression*

Uji kesesuaian model adalah uji yang dilakukan dengan menguji kesesuaian dari koefisien parameter secara serentak, yaitu dengan mengkombinasikan uji regresi linier berganda dengan model untuk data spasial (Utami dkk., 2017). Pengujian ini menggunakan uji *analysis of variance* (ANOVA) dengan hipotesis sebagai berikut (Caraka & Yasin, 2017):

a. Hipotesis

$H_0: \beta_1(u_i, v_i) = \beta_2(u_i, v_i) = \cdots = \beta_k(u_i, v_i) = 0$

(tidak ada perbedaan yang signifikan antara model regresi global dengan model GWR dalam memodelkan data)

$H_1: \exists \beta_j(u_i, v_i) \neq 0; j = 1,2,3, \dots, k$

(terdapat perbedaan yang signifikan antara model regresi global dengan model GWR dalam memodelkan data)

b. Taraf Signifikansi

$\alpha = 5\%$

c. Statistik Uji

$$F_{htg} = \frac{JKR_{OLS} - JKR_{GWR}/df_1}{JKR_{GWR}/df_2} \tag{2.10}$$

dengan:

JKR_{OLS} : jumlah kuadrat residual model regresi global

JKR_{GWR} : jumlah kuadrat residual model GWR

df_1 : derajat kebebasan model regresi global

df_2 : derajat kebebasan model GWR

d. Kriteria Kesimpulan

Tolak H_0 jika $F_{htg} \geq F_{\alpha;(df_1,df_2)}$ dan sebaliknya terima H_0 jika $F_{htg} < F_{\alpha;(df_1,df_2)}$

2.2.3.2 Uji Signifikansi Parameter *Geographically Weighted Regression*

Uji signifikansi parameter model adalah uji yang dilakukan untuk mengetahui parameter-parameter yang berpengaruh signifikan terhadap variabel respon secara parsial di setiap subjek pengamatan. Bentuk hipotesisnya adalah sebagai berikut (Utari dkk., 2019):

a. Hipotesis

$H_0: \beta_j(u_i, v_i) = 0; j = 1,2,3, \dots, k$

(tidak ada pengaruh variabel prediktor terhadap variabel respon dalam memodelkan data)

$H_1: \beta_j(u_i, v_i) \neq 0; j = 1,2,3, \dots, k$

(terdapat minimal satu pengaruh variabel prediktor terhadap variabel respon dalam memodelkan data)

b. Taraf Signifikansi

$\alpha = 5\%$

c. Statistik Uji

$$t_{hit} = \frac{\hat{\beta}_j(u_i, v_i)}{SE_{\hat{\beta}_j(u_i,v_i)}} \tag{2.11}$$

dengan:

$$SE_{\hat{\beta}_j(u_i,v_i)} = \sqrt{var\hat{\beta}_j(u_i, v_i)}$$

$$= \sqrt{CC^T\sigma^2}$$

$$C = (\boldsymbol{X}'\boldsymbol{W}(u_i, v_i)\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{W}(u_i, v_i)$$

$$\sigma^2 = \frac{\sum_{i=1}^{n}\left(\boldsymbol{Y}_i - \widehat{\boldsymbol{Y}}_i\right)^2}{df_2}$$

e. Kriteria kesimpulan

Tolak H_0 jika $|t_{hit}| \geq t_{\frac{\alpha}{2};df_2}$ dan sebaliknya terima H_0 jika $|t_{hit}| < t_{\frac{\alpha}{2};df_2}$

2.3 *Principal Component Analysis*

Principal Component Analysis (PCA) merupakan salah satu metode statistika yang bertujuan untuk mengatasi masalah multikolinearitas dengan mentransformasi sebagian besar variabel asal yang saling berkorelasi satu dengan yang lainnya menjadi satu set variabel baru yang lebih kecil, saling bebas dan tidak berkorelasi lagi. Jadi, PCA digunakan untuk mereduksi data sehingga lebih mudah dalam menginterpretasikan hasil analisis (Delsen dkk., 2017). Jika data dari variabel-variabel prediktor pengamatan tidak semuanya menggunakan satuan pengukuran yang sama maka perlu dilakukan standarisasi data terhadap variabel tersebut sebagai berikut:

$$Z_{ki} = \frac{(X_{ki} - \bar{X}_k)}{\sigma_k} \tag{2.12}$$

dengan:

X_{ki} : variabel prediktor ke-k pada lokasi ke-i

$\bar{X}_k$: mean variabel prediktor ke-k

σ_k : standar deviasi variabel prediktor ke-k

Misalkan terdapat vektor random $\boldsymbol{Z}' = (Z_1, Z_2, Z_3, \dots, Z_k)$ yang memiliki matriks korelasi dengan nilai eigen $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \cdots \geq \lambda_k$, maka komponen utama yang didefinisikan sebagai kombinasi linier dari $\boldsymbol{Z}$ variabel asal dengan jumlah k buah variabel asal adalah sebagai berikut (Maubanu & Kartiko, 2018):

$$\begin{gathered}
P_{c1} = \boldsymbol{e_1 Z}' = e_{11}\boldsymbol{Z}_1 + e_{21}\boldsymbol{Z}_2 + \cdots + e_{k1}\boldsymbol{Z}_k \\
P_{c2} = \boldsymbol{e_2 Z}' = e_{12}\boldsymbol{Z}_1 + e_{22}\boldsymbol{Z}_2 + \cdots + e_{k2}\boldsymbol{Z}_k \\
P_{c3} = \boldsymbol{e_3 Z}' = e_{13}\boldsymbol{Z}_1 + e_{23}\boldsymbol{Z}_2 + \cdots + e_{k3}\boldsymbol{Z}_k \\
\vdots \\
P_{ck} = \boldsymbol{e_k Z}' = e_{1k}\boldsymbol{Z}_1 + e_{2k}\boldsymbol{Z}_2 + \cdots + e_{kk}\boldsymbol{Z}_k
\end{gathered} \tag{2.13}$$

Pembentukan komponen utama dapat dilakukan dengan menggunakan matriks variansi kovariansi maupun matriks korelasi. Matriks variansi kovariansi adalah matriks yang elemen-elemennya terdiri atas variansi dan kovariansi sekumpulan variabel. Matriks variansi-kovariansi dari sampel dinotasikan dengan $\boldsymbol{S}$ dan dapat dinyatakan sebagai berikut:

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

Sedangkan, matriks korelasi sering digunakan untuk memperoleh nilai kedekatan hubungan antar variabel pengamatan. Nilai kedekatan ini dapat digunakan untuk melakukan beberapa pengujian dalam melihat kesesuaian dengan nilai korelasi yang diperoleh dari analisis komponen utama. Matriks korelasi dinotasikan dengan $\boldsymbol{R}$ dan dapat dinyatakan sebagai berikut (Kasim, 2021):

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

Nilai *eigen* atau *eigen value* adalah nilai yang menggambarkan seberapa besar pengaruh suatu variabel terhadap pembentukan karakteristik yang dinotasikan dengan λ. Misal, terdapat matriks **R** berukuran k × k dan sebuah vektor tak nol yang ditunjukkan oleh **x** yang disebut vektor *eigen* dari **R**, jika **Rx** adalah kelipatan skalar dari **x**, maka dapat dinyatakan sebagai berikut:

$$\mathbf{Rx} = \lambda \mathbf{x} \tag{2.16}$$

dengan λ adalah suatu nilai atau sebuah skalar. Skalar λ merupakan nilai *eigen* dan **x** merupakan vektor *eigen* dari matriks **R**. λ diperoleh berdasarkan matriks **R**, dari persamaan (2.16) ditambahkan matriks identitas **I** ke dalam persamaan menjadi:

$$(\mathbf{R} - \lambda \mathbf{I})\mathbf{x} = 0 \tag{2.17}$$

Selanjutnya, diperlukan solusi tak nol dari persamaan (2.17) agar diperoleh nilai *eigen*, maka diperoleh solusi tak nol sebagai berikut:

$$\det(\mathbf{R} - \lambda \mathbf{I}) = 0 \tag{2.18}$$

Persamaan (2.18) merupakan persamaan karakteristik dari matriks **R**. Apabila telah diperoleh nilai λ, maka vektor *eigen* dapat dihitung dengan mensubtitusikan nilai *eigen* yang diperoleh ke persamaan (2.17).

Apabila nilai *eigen* lebih besar dari 1 maka dipertahankan dalam model, sebaliknya jika nilai *eigen* lebih kecil dari 1 maka faktornya dikeluarkan dari model

(Delsen dkk., 2017). Sedangkan, nilai pada vektor *eigen* memberikan gambaran besarnya kontribusi masing- masing variabel pada kombinasi linier yang terbentuk. Sehingga, dapat diketahui variabel yang dominan berpengaruh pada masing-masing kombinasi linier.

Terdapat tiga cara yang bisa digunakan untuk mengetahui jumlah komponen utama (*principal component*). Pertama, dengan melihat persentase kumulatif variansi yang mampu menerangkan total variansi data sekitar 80% keatas. Kedua, dengan melihat nilai *eigen* yang bernilai lebih dari 1. Terakhir, dengan mengamati *scree plot* yang didasarkan pada pola penurunan yang tajam dari *scree plot*. Penurunan yang tajam atau terlihat seperti patahan siku pada *scree plot* menunjukkan bahwa terjadi perubahan nilai *eigen* yang besar.

2.4 *Geographically Weighted Regression*

2.4.1 Efek Spasial

Data spasial memuat suatu informasi lokasi yang memiliki ketergantungan antar lokasi. Data spasial menandakan bahwa terdapat ketergantungan antara pengukuran data dengan lokasi pengamatan atau sering disebut efek spasial. Efek spasial dibedakan menjadi dua yaitu dependensi spasial dan keragaman spasial. Jika terjadi dependensi spasial, maka penyelesaian dilakukan dengan pendekatan area. Sebaliknya, jika terjadi keragaman spasial maka dapat diselesaikan dengan menggunakan pendekatan titik (Fadlilah dkk., 2019).

Berdasarkan tipe data, pemodelan spasial dapat dibedakan menjadi pemodelan dengan pendekatan titik dan pendekatan area. Pemodelan dengan jenis pendekatan titik dapat menggunakan metode *Geographically Weighted Regression* (GWR) jika data berdistribusi normal, *Geographically Weighted* Poisson *Regression* (GWPR) jika data berdistribusi Poisson, dan *Generalized Space Time Autoregressive* (GSTAR) jika data berupa data runtun waktu. Sebaliknya, pemodelan dengan jenis pendekatan area dapat menggunakan metode *Mixed Regressive-Autoregressive* atau *Spatial Autoregressive Model* (SAR), *Spatial Error Model* (SEM), *Spatial Autogressive Moving Average* (SARMA), *Spatial Durbin Model* (SDM), dan data panel (Lutfiani dkk., 2019).

2.4.2 Heterogenitas Spasial

Heterogenitas spasial adalah kondisi apabila variabel prediktor memberikan respon yang berbeda pada lokasi yang berbeda dalam satu wilayah pengamatan (Tizona dkk., 2017). Metode yang sering digunakan untuk mendeteksi gejala heterogenitas spasial pada model regresi adalah uji *Breusch Pagan*. Hipotesis yang digunakan untuk menguji heterogenitas spasial adalah sebagai berikut:

f. Hipotesis

$H_0: \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_n^2 = \sigma^2$ (tidak terdapat pengaruh variabel prediktor terhadap variabel respon antar satu lokasi dengan lokasi lainnya)

$H_1: \exists\sigma_i^2 \neq \sigma^2$ (terdapat minimal satu pengaruh variabel prediktor terhadap variabel respon antar satu lokasi dengan lokasi lainnya)

g. Taraf Signifikansi

$\alpha = 5\%$

h. Statistik Uji

$$BP = \left(\frac{1}{2}\right)\boldsymbol{f}'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{f} \qquad (2.19)$$

dengan

$f_i = \left(\frac{\varepsilon_i^2}{\sigma^2} - 1\right)$

$\boldsymbol{Z}$ = matriks $n \times p + 1$ dari pengamatan dengan menambahkan kolom pertama berisi vektor 1

i. Kriteria Kesimpulan

Tolak H_0 jika $BP \geq \chi^2{}_{(\alpha,p)}$ dan sebaliknya terima H_0 jika $BP < \chi^2{}_{(\alpha,p)}$

2.4.3 *Geographically Weighted Regression*

Geographically Weighted Regression (GWR) adalah pengembangan dari model regresi global yang mempertimbangkan aspek lokasi (Marizal & Monalisa, 2022). Model GWR merupakan model regresi linier lokal yang menghasilkan estimasi parameter model yang bersifat lokal untuk setiap titik atau lokasi data tersebut dikumpulkan. Dalam model GWR, variabel respon diprediksi dengan variabel prediktor yang masing-masing koefisien regresinya bergantung pada lokasi pengamatan (Fadli dkk., 2018). Untuk n pengamatan dengan k variabel prediktor maka model GWR dapat ditulis sebagai berikut (Caraka & Yasin, 2017):

$$y_i = \beta_0(u_i, v_i) + \sum_{j=1}^{k} \beta_j(u_i, v_i)\, x_{ij} + \varepsilon_i \tag{2.20}$$

dengan

y_i : variabel respon pada pengamatan ke-i

$\beta_o(u_i, v_i)$: parameter intersep pada titik koordinat (*latitude* dan *longitude*) pengamatan ke-*i*

$\beta_j(u_i, v_i)$: parameter koefisien regresi pada titik koordinat pengamatan ke-*i*

x_{ij} : variabel prediktor ke-j pada pengamatan ke-i

ε_i : residual pada pengamatan ke-i

2.4.4 Matriks Pembobot Spasial

Matriks pembobot spasial adalah matriks diagonal yang digunakan untuk mengetahui kedekatan atau hubungan spasial antar lokasi pengamatan (Wardani dkk., 2018). Pembobot bergantung pada jarak antar titik lokasi pengamatan. Elemen-elemen pada matriks diagonal merupakan fungsi pembobot dari setiap titik lokasi pengamatan. Matriks pembobot berfungsi untuk menentukan atau menaksir parameter yang berbeda pada setiap titik lokasi pengamatan (Lutfiani dkk., 2019). Pembentukan matriks pembobot spasial diperoleh dari perhitungan jarak *Euclidean* antara lokasi yang didasarkan pada derajat *latitude* dan *longitude* dengan rumus jarak *Euclidean* antara lokasi ke-i dan ke-j sebagai berikut (Yulianti dkk., 2022):

$$d_{ij} = \sqrt{\left(u_i - u_j\right)^2 + \left(v_i - v_j\right)^2} \tag{2.21}$$

Fungsi pembobot *kernel* dapat digunakan untuk menentukan matriks pembobot pada setiap lokasi pengamatan. Terdapat dua jenis fungsi *kernel* yaitu fungsi *adaptive kernel* dan fungsi *fixed kernel*. Selain itu, terdapat tiga jenis pembobot untuk setiap fungsi *kernel* yaitu pembobot *gaussian*, *bisquare* dan *tricube* (Aliu dkk., 2022). Perbedaan antara fungsi *adaptive kernel* dan *fixed kernel* adalah fungsi *adaptive kernel* memiliki *bandwidth* yang berbeda untuk setiap titik lokasi pengamatan sedangkan fungsi *fixed kernel* memiliki *bandwidth* yang sama pada setiap titik lokasi pengamatan. Hal ini dikarenakan kemampuan fungsi *adaptive kernel* yang dapat disesuaikan dengan kondisi titik-titik pengamatan (Tizona dkk., 2017). Persamaan fungsi pembobot *kernel* adalah sebagai berikut (Hapsery & Trishnanti, 2021):

1. *Adaptive gaussian kernel*

$$w_{ij} = \exp\left[-\frac{1}{2}\left(\frac{d_{ij}}{h_i}\right)^2\right] \quad ; i,j = 1,2,\dots,n \; ; \; i \neq j \tag{2.22}$$

2. *Fixed gaussian kernel*

$$w_{ij} = \exp\left[-\frac{1}{2}\left(\frac{d_{ij}}{h}\right)^2\right] \tag{2.23}$$

3. *Adaptive bisquare kernel*

$$w_{ij} = \begin{cases} \left[-\frac{1}{2}\left(\frac{d_{ij}}{h_i}\right)^2\right]^2 & , jika\ d_{ij} \leq h_i \\ 0 & , jika\ d_{ij} > h_i \end{cases} \tag{2.24}$$

4. *Fixed bisquare kernel*

$$w_{ij} = \begin{cases} \left[-\frac{1}{2}\left(\frac{d_{ij}}{h}\right)^2\right]^2 & , jika\ d_{ij} \leq h \\ 0 & , jika\ d_{ij} > h \end{cases} \tag{2.25}$$

5. *Adaptive tricube kernel*

$$w_{ij} = \begin{cases} \left[-\frac{1}{2}\left(\frac{d_{ij}}{h_i}\right)^3\right]^3 & , jika\ d_{ij} \leq h_i \\ 0 & , jika\ d_{ij} > h_i \end{cases} \tag{2.26}$$

6. *Fixed tricube kernel*

$$w_{ij} = \begin{cases} \left[-\frac{1}{2}\left(\frac{d_{ij}}{h}\right)^3\right]^3 & , jika\ d_{ij} \leq h \\ 0 & , jika\ d_{ij} > h \end{cases} \tag{2.27}$$

dengan:

d_{ij} : jarak *Euclidean* antara lokasi amatan i dengan lokasi *j*

h_i : lebar *bandwidth* pada lokasi pengamatan ke-i

2.4.5 *Bandwidth*

Pengertian *bandwidth* secara teoritis merupakan lingkaran dengan radius dari titik pusat lokasi yang digunakan sebagai dasar dalam menentukan bobot setiap pengamatan terhadap model regresi. Titik pengamatan yang terletak di dalam radius lingkaran masih dianggap berpengaruh terhadap model pada lokasi tersebut sehingga akan diberi bobot sesuai fungsi yang digunakan (Fadli dkk., 2018). *Bandwidth* (h) berfungsi sebagai pengontrol keseimbangan antar kemulusan (*smoothing*) dan kesesuaian fungsi terhadap data. Nilai *bandwidth* yang sangat kecil menandakan semakin sedikit pengamatan yang berada dalam radius sehingga menyebabkan variansi yang diperoleh semakin besar dan model yang diperoleh akan sangat kasar (*under smoothing*). Sebaliknya, nilai *bandwidth* yang besar menandakan semakin banyak pengamatan yang berada dalam radius sehingga mengakibatkan bias yang semakin besar dan model yang diperoleh sangat mulus (*over smoothing*). Pemilihan *bandwidth* sangat penting agar estimator yang diperoleh juga optimal (Tizona dkk., 2017). Salah satu cara yang digunakan untuk memperoleh *bandwidth* optimum yaitu dengan menghitung nilai *Cross Validation* (CV) yang secara matematis didefinisikan sebagai berikut:

$$CV = \sum_{i=1}^{n} \left(y_i - \hat{y}_{\neq i}(h)\right)^2 \tag{2.28}$$

dengan $\hat{y}_{\neq i}(h)$ adalah nilai estimasi y_i (*fitting value*) saat pengamatan di lokasi (u_i, v_i) dihilangkan dari proses estimasi. Nilai *bandwidth* optimum diperoleh dari nilai CV yang paling minimum (Caraka & Yasin, 2017).

2.4.6 Estimasi Parameter Model *Geographically Weighted Regression*

Estimasi parameter $\beta(u_i, v_i)$ pada lokasi ke-i dapat dilakukan dengan menggunakan salah satu metode yaitu *Weighted Least Square* (WLS). Metode WLS merupakan bentuk pengembangan dari metode *Ordinary Least Square* (OLS) yang meminimumkan jumlah kuadrat residual hanya saja pada metode WLS perlu memberikan pembobot yang berbeda pada setiap lokasi pengamatan (Caraka & Yasin, 2017). Misalkan pembobot untuk setiap lokasi ke-*i* adalah $\boldsymbol{w_i}(u_i, v_i)$ dengan $i = 1,2, \ldots, n$, maka parameter lokasi (u_i, v_i) diestimasi dengan menambahkan pembobot dan meminimumkan jumlah kuadrat residual berikut:

$$\sum_{j=1}^{n} \boldsymbol{w_{ij}}(u_i, v_i)\, \boldsymbol{\varepsilon_i}^2 = \sum_{j=1}^{n} \boldsymbol{w_{ij}}(u_i, v_i) \left(\boldsymbol{y_i} - \boldsymbol{\beta_0}(u_i, v_i) - \sum_{k=1}^{p} \boldsymbol{\beta_k}(u_i, v_i)\boldsymbol{x_{ik}}\right)^2 \tag{2.29}$$

Bentuk estimasi parameter model GWR untuk setiap lokasi pengamatan adalah:

$$\hat{\boldsymbol{\beta}}(u_i, v_i) = (\boldsymbol{X'W}(u_i, v_i)\boldsymbol{X})^{-1}\boldsymbol{X'W}(u_i, v_i)\boldsymbol{y} \tag{2.30}$$

Misalkan $\boldsymbol{x}_i' = (1, x_{i1}, x_{i2}, \dots, x_{ik})$ merupakan elemen baris ke-i pada matriks $\mathbf{X}$ dan $\widehat{\boldsymbol{\beta}}(u_i, v_i)$ adalah vektor estimasi parameter model GWR pada lokasi (u_i, v_i), maka diperoleh estimasi model GWR untuk setiap lokasi pengamatan sebagai berikut:

$$\widehat{\boldsymbol{y}}_i = \boldsymbol{x}_i' . \widehat{\boldsymbol{\beta}}(u_i, v_i)$$

$$\widehat{\boldsymbol{y}}_i = \boldsymbol{x}_i' . (\boldsymbol{X}'\boldsymbol{W}(u_i, v_i)\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{W}(u_i, v_i)\boldsymbol{y} \tag{2.31}$$

2.4.7 Penentuan Kebaikan Model

Penentuan kebaikan model bertujuan untuk mendapatkan model yang terbaik dalam menggambarkan kondisi variabel yang dianalisis (Erdkhadifa, 2021). Adapun metode yang digunakan untuk mendapatkan model terbaik, diantaranya dengan nilai AIC dan BIC.

2.4.7.1 *Akaike Information Criterion*

Uji *Akaike Information Criterion* (AIC) adalah metode yang digunakan untuk mengevaluasi kualitas model dengan menggunakan pendekatan *model selection* yang ditemukan oleh Hirotsugu Akaike. AIC memberikan nilai skor pada setiap model yang memperhitungkan seberapa baik model cocok dengan data yang digunakan. Model dengan nilai AIC terkecil menandakan bahwa model tersebut merupakan model terbaik diantara model lainnya. Secara matematis, AIC dapat dituliskan sebagai berikut (Agustini dkk., 2018):

$$\text{AIC} = -2\log L(\hat{\theta}) + 2k \tag{2.32}$$

dengan:

$\log L(\hat{\theta})$: log-Likelihood pada titik maksimum dari model

k : jumlah parameter

2.4.7.2 *Bayesian Information Criterion*

Uji *Bayesian Information Criterion* (BIC) adalah metode yang umum digunakan untuk mengevaluasi kualitas model selain metode AIC. Dalam BIC, model-model yang diuji dibandingkan dengan memperhitungkan nilai Likelihood data dan jumlah parameter dalam model. Model dengan jumlah parameter yang lebih sedikit akan diberikan preferensi dibanding model dengan jumlah parameter yang lebih banyak. Model dengan nilai BIC terkecil menandakan bahwa model tersebut merupakan model terbaik diantara model lainnya. Secara matematis, BIC dapat dituliskan sebagai berikut:

$$\mathrm{BIC} = -2\log L(\hat{\theta}) + \log(n)\,k \tag{2.33}$$

dengan:

$\log L(\hat{\theta})$: log-Likelihood pada titik maksimum dari model

n : jumlah pengamatan

k : jumlah parameter

2.5 *Geographically Weighted Regression Principal Component Analysis*

Geographically Weighted Regression Principal Component Analysis (GWRPCA) merupakan inovasi penggabungan metode GWR dan PCA. Pertama, apabila terdapat gejala multikolinearitas pada model maka dilakukan analisis PCA dengan mereduksi variabel prediktor yang saling berkorelasi. Lalu, metode GWR dilakukan jika terdeteksi bahwa model mengalami heterogenitas spasial sehingga terdapat perbedaan karakteristik suatu wilayah dengan wilayah lainnya. Model GWRPCA dapat dinyatakan sebagai berikut (Sari dkk., 2016):

$$y_i = \beta_0(u_i, v_i) + \sum_{j=1}^{l} \beta_j(u_i, v_i)\,P_{cij} + \varepsilon_i \tag{2.34}$$

dengan:

y_i : variabel respon pada pengamatan ke-i

$\beta_o(u_i, v_i)$: parameter intersep pada titik koordinat (*latitude* dan *longitude*) pengamatan ke-*i*

$\beta_j(u_i, v_i)$: parameter koefisien regresi pada titik koordinat pengamatan ke-*i*

P_{cij} : variabel komponen utama ke-j pada pengamatan ke-i

ε_i : residual pada pengamatan ke-i

2.6 Angka Kematian Bayi

Kematian bayi adalah kejadian kematian yang terjadi pada rentang waktu setelah bayi lahir sampai bayi belum berusia tepat satu tahun. Terdapat banyak faktor yang menjadi penyebab kematian bayi. Secara garis besar, kematian bayi dibedakan menjadi dua menurut sisi penyebabnya yaitu endogen dan eksogen. Kematian bayi endogen atau secara umum disebut kematian *neonatal* merupakan kejadian kematian bayi yang terjadi pada bulan pertama sejak dilahirkan, hal ini umumnya disebabkan dari faktor-faktor yang dibawa anak sejak lahir atau diperoleh dari orang tua selama kehamilan. Sebaliknya, kematian bayi eksogen atau secara umum disebut kematian *post neonatal* merupakan kejadian kematian bayi yang terjadi saat bayi berusia lebih dari satu bulan hingga menjelang usia satu tahun, kematian eksogen umumnya disebabkan oleh faktor-faktor yang berhubungan dengan pengaruh lingkungan luar (Hajarisman dkk., 2013).

Tingkat kesehatan biasanya diukur dari angka kematian bayi dan angka harapan hidup. Tingginya angka kematian bayi merupakan indikator rendahnya kesehatan lingkungan dan masyarakat. Angka Kematian Bayi (AKB) adalah banyaknya kematian bayi sebelum mencapai umur 1 tahun per 1000 kelahiran hidup pada periode waktu yang sama (Prahutama dkk., 2017). Angka kematian bayi secara matematis dituliskan sebagai berikut:

$$\text{AKB} = \frac{\text{jumlah kematian bayi}}{\text{jumlah kelahiran bayi}} \times 1000$$

Beberapa faktor endogen diduga bisa mempengaruhi naik turunnya kematian bayi seperti pemberian Tablet Tambah Darah (TTD) kepada ibu hamil dan persalinan yang ditolong tenaga kesehatan. Suplementasi zat besi berupa Tablet Tambah Darah (TTD) dapat membantu mencegah atau mengobati anemia pada ibu hamil. Anemia pada ibu hamil dapat berdampak negatif pada pertumbuhan dan perkembangan bayi serta meningkatkan risiko komplikasi selama kehamilan, persalinan, dan pasca persalinan (Millah, 2019) .

Sedangkan, tenaga kesehatan yang terlatih memiliki pengetahuan dan keahlian untuk mengidentifikasi dan mengatasi masalah yang mungkin timbul selama persalinan. Mereka mampu memberikan perawatan yang sesuai dan segera dalam menghadapi keadaan darurat seperti perdarahan pasca persalinan, asfiksia neonatal (masalah pernapasan pada bayi yang baru lahir), dan masalah serius lainnya yang dapat mengancam nyawa bayi (Wahyuni dkk., 2022).

Adapun beberapa faktor eksogen yang diduga bisa mempengaruhi naik turunnya kematian bayi seperti pemberian ASI eksklusif dan pemberian vitamin A kepada bayi. ASI adalah sumber makanan yang paling sesuai dan lengkap untuk bayi. ASI mengandung nutrisi yang tepat dan proporsi yang sesuai untuk mendukung pertumbuhan dan perkembangan optimal bayi. Bayi yang diberi ASI eksklusif memiliki risiko yang lebih rendah untuk mengalami kekurangan gizi dan masalah pertumbuhan yang dapat berkontribusi pada angka kematian bayi (Fitri & Shofiya, 2020).

Sedangkan, pemberian vitamin A kepada bayi secara rutin dapat mencegah defisiensi vitamin A dan masalah kesehatan terkait, sehingga mengurangi risiko kematian bayi (Mastin & Roosita, 2015). Bayi yang mengalami kekurangan vitamin A dapat mengalami gangguan pertumbuhan, masalah penglihatan, dan meningkatkan risiko terjadinya infeksi. Selain itu, vitamin A telah terbukti efektif dalam mengurangi risiko penyakit infeksi serius pada bayi, seperti pneumonia dan diare berat.

Angka kematian bayi merupakan salah satu indikator penting dalam menentukan tingkat derajat kesehatan masyarakat dan keberhasilan pembangunan suatu daerah. Berbagai upaya yang dilakukan untuk menurunkan AKB dan dinilai mempunyai dampak yang cukup bagus seperti mengupayakan persalinan dilakukan oleh tenaga kesehatan di fasilitas kesehatan serta menjamin tersedianya pelayanan kesehatan sesuai standar pada kunjungan bayi baru lahir (Lengkong dkk., 2020).