PEMODELAN *MIXED GEOGRAPHICALLY WEIGHTED REGRESSION* DENGAN METODE *LEAST ABSOLUTE SHRINKAGE AND SELECTION OPERATOR*

SKRIPSI

ABDUL RAHMAN

H051171007

PROGRAM STUDI STATISTIKA DEPARTEMEN STATISTIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS HASANUDDIN

MAKASSAR

AGUSTUS 2023

PEMODELAN *MIXED GEOGRAPHICALLY WEIGHTED REGRESSION* DENGAN METODE *LEAST ABSOLUTE SHRINKAGE AND SELECTION OPERATOR*

SKRIPSI

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

ABDUL RAHMAN

H051171007

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 *Mixed Geographically Weighted Regression* Dengan Metode *Least Absolute Shrinkage And Selection Operator*

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

Makassar, 18 Agustus 2023

METERAI TEMPEL 10000 SEPULUH RIBU RUPIAH 280AKX621751

Abdul Rahman

NIM. H051171007

PEMODELAN *MIXED GEOGRAPHICALLY WEIGHTED REGRESSION* DENGAN METODE *LEAST ABSOLUTE SHRINKAGE AND SELECTION OPERATOR*

Disetujui oleh:

Pembimbing Utama	**Pembimbing Pertama**
Andi Kresna Jaya, S.Si., M.Si.	**Anisa, S.Si., M.Si.**
NIP. 197312282000031001	**NIP. 197302271998022001**

Ketua Departemen Statistika

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

NIP. 197708082005012002

Pada Tanggal: 18 Agustus 2023

HALAMAN PENGESAHAN

Skripsi ini diajukan oleh:

Nama	: Abdul Rahman
NIM	: H051171007
Program Studi	: Statistika
Judul Skripsi	: Pemodelan *Mixed Geographically Weighted Regression* Dengan Metode *Least Absolute Shrinkage And Selection Operator*

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.

DEWAN PENGUJI

			Tanda Tangan
1.	Ketua	: Andi Kresna Jaya, S.Si.M.Si.	(……………)
2.	Sekretaris	: Anisa, S.Si., M.Si.	(……………)
3.	Anggota	: Drs. Raupong, M.Si.	(……………)
4.	Anggota	: Siswanto, S.Si., M.Si.	(……………)

Ditetapkan di: Makassar

Tanggal : 18 Agustus 2023

KATA PENGANTAR

Assalamu'alaikum Warahmatullahi Wabarakatuh

Puji syukur penulis panjatkan kepada **Allah** ***Subhanahu Wa Ta'ala*** atas segala limpahan rahmat, hidayah, 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** ***Mixed Geographically Weighted Regression*** **Dengan Metode** ***Least Absolute Shrinkage And Selection Operator*****"** 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 **Tallasa Sibali** dan Ibunda **Salawati** 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. Untuk Kakakku tercinta, **Muhammad Nur** yang senantiasa memberikan semangat bagi penulis dalam menyelesaikan skripsi ini, Bibi **Hj. Pujiyati, S.Ag., M.Ag., Hj. Juniarti, S.E,** dan Almarhum Paman **Ibrahim, S.Pd., M.Pd** yang penulis sayangi, serta keluarga besar penulis, terima kasih atas do'a 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, segenap dosen pengajar, dan staf Departemen Statistika yang telah memberikan ilmu dan pengetahuan serta bantuan-bantuan kepada penulis dalam berbagai hal selama menjadi mahasiswa di Departemen Statistika.
4. Bapak **Andi Kresna Jaya, S.Si., M.Si.**, selaku Pembimbing Akademik sekaligus Pembimbing utama penulis yang dengan penuh kesabaran telah meluangkan waktu dan pemikirannya untuk senantiasa memberikan arahan, dorongan semangat, dan motivasi kepada penulis hingga selesainya penulisan tugas akhir ini.
5. Ibu **Anisa, S.Si., M.Si.,** selaku dosen pembimbing pertama yang telah meluangkan waktu dan pemikirannya untuk memberikan masukan, arahan, dan motivasi kepada penulis mulai dari awal hingga selesainya penulisan tugas akhir ini.
6. Bapak **Drs Raupong, M.Si.,** dan **Siswanto, S.Si., M.Si.,** selaku Tim Penguji yang telah meluangkan waktu dalam memberikan motivasi serta kritikan yang membangun kepada penulis dalam penyempurnaan tugas akhir ini.
7. Segenap Dosen Pengajar dan Staf yang telah memberikan ilmu dan kemudahan kepada penulis dalam berbagai hal selama menempuh pendidikan sarjana di Departemen Statistika.
8. Kakak-kakak, teman-teman, dan adik-adik segenap anggota **Keluarga Mahasiswa FMIPA Unhas**, keluarga **Himatika FMIPA Unhas** dan **Himastat FMIPA Unhas**, terkhusus **DISKRIT 2017,** terima kasih atas ilmu yang tidak didapatkan di bangku perkuliahan dan banyak belajar betapa pentingnya sebuah proses dalam mencapai sessuatu yang diinginkan.
9. Teman seperjuangan **MIPA 2017, dan Statistika 2017,** Terima kasih atas motivasi, segala cerita suka dan dukanya, serta selalu menemani penulis sehingga masa perkuliahan dapat dijalani lebih bermakna dan berwarna.
10. Serta kepada seluruh pihak yang tidak dapat penulis sebutkan satu persatu, terima kasih untuk semuanya. Semoga segala dukungan dan partisipasi yang diberikan kepada penulis bernilai ibadah di sisi **Allah** ***Subhanahu Wa Ta'ala*****.**

Penulis berharap tugas akhir ini dapat memberikan tambahan pengetahuan baru bagi para pembelajar statistika. Penulis menyadari bahwa dalam penulisan tugas akhir ini masih banyak kekurangan. Oleh karena itu, dengan segala kerendahan hati penulis memohon maaf. Akhir kata, semoga apa yang telah kita lakukan hari ini dapat membuat kita selangkah lebih maju dari hari-hari sebelumnya dan mudah-mudahan tugas akhir ini bermanfaat bagi orang-orang yang berkepentingan. *Aamiin Yaa Rabbal Alamin*.

Makassar, 18 Agustus 2023

Abdul Rahman

PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIK

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

Nama : Abdul Rahman

NIM : H051171007

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 *Mixed Geographically Weighted Regression* Dengan Metode *Least Absolute Shrinkage And Selection Operator*"

beserta perangkat yang ada (jika diperlukan). Terkait dengan hal di atas, maka pihak universitas berhak menyimpan, mengalih-media/fortmat-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 18 Agustus 2023

Yang menyatakan,

(Abdul Rahman)

ABSTRAK

Mixed Geographically Weigted Regression (MGWR) merupakan model kombinasi antara model regresi global dengan GWR yang mempertimbangkan sebagian variabel prediktor bersifat global dan sebagian lainnya bersifat lokal. Model MGWR tidak dapat mengatasi data yang mengalami multikolinearitas yaitu terjadinya korelasi antar variabel prediktor, sehingga untuk mengatasi hal tersebut digunakan model MGWR dengan metode Lasso. Lasso melakukan estimasi parameter dengan algoritma LARS yang menyusutkan koefisien hingga nol. Model MGWR dengan metode Lasso dapat diterapkan pada data Produk Domestik Regional Bruto (PDRB) di Sulawesi Selatan. Nilai PDRB dapat bervariasi secara spasial, artinya nilai PDRB di suatu wilayah yang berdekatan dengan wilayah yang memiliki nilai PDRB tinggi akan memiliki kecenderungan yang sama. Penelitian ini bertujuan untuk menentukan estimasi parameter model MGWR dengan metode Lasso kemudian diterapkan pada data PDRB Sulawesi Selatan tahun 2019. Data PDRB Sulawesi Selatan mengalami multikolinearitas, sehingga diselesaikan menggunakan model MGWR dengan metode Lasso. Hasil estimasi diperoleh variabel yang berpengaruh signifikan secara global adalah x_3, x_4, x_6 dan x_7, sementara variabel yang berpengaruh signifikan secara lokal adalah x_1 dan x_2. Nilai AIC model MGWR dengan metode Lasso sebesar 113.4487 lebih kecil dari model MGWR, sehingga penggunaan metode Lasso pada model MGWR menjadikan model lebih baik dalam mengatasi masalah multikolinearitas.

Kata Kunci: *Mixed Geographically Weighted Regression*, Multikolinearitas, Lasso, LARS, PDRB.

ABSTRACT

Mixed Geographically Weigted Regression (MGWR) is a combination model between the global regression model and GWR which considers some predictor variables are global and some are local. In MGWR modeling, data is often found to experience multicollinearity, namely the correlation between predictor variables, to overcome this, the MGWR model is used with the Lasso method. Lasso performs parameter estimation with the LARS algorithm that shrinks the coefficient to zero. The MGWR model can be applied to Gross Regional Domestic Product (GRDP) data in South Sulawesi. GRDP values can vary spatially, meaning that GRDP values in an area adjacent to an area that has a high GRDP value will have the same tendency. This study aims to determine the parameter estimation of the MGWR model with the Lasso method and then apply it to South Sulawesi GRDP data in 2019. South Sulawesi GRDP data experienced multicollinearity, so it was solved using the MGWR model with the Lasso method. The estimation results obtained variables that have a significant effect globally are x_3, x_4, x_6, and x_7. Meanwhile, the variables that demonstrate a significant effect locally are x_1 and x_2. The AIC value of the MGWR model with the Lasso method of 113.4487 is smaller than the MGWR model, so utilization of the Lasso method in the MGWR model makes the model better in overcoming multicollinearity problems.

Keywords: *Mixed Geographically Weighted Regression*, *Multicollinearity*, Lasso, LARS, GRDP.

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

DAFTAR LAMPIRAN

BAB I
PENDAHULUAN

1.1 Latar Belakang

Analisis regresi merupakan metode analisis statistik yang digunakan untuk memodelkan hubungan antara variabel respon dengan satu atau lebih variabel prediktor. Model analisis ini akan memberikan informasi yang reliabel untuk wilayah lokal jika tidak ada atau hanya ada sedikit keragaman antar wilayah lokal tersebut (Fotheringham dkk, 2002). Metode ini mengasumsikan bahwa antar pengamatan harus saling bebas atau variabel respon tidak dipengaruhi oleh letak geografis. Sementara pengamatan suatu lokasi sangat mungkin dipengaruhi oleh pengamatan di lokasi lainnya. Hal ini akan menyebabkan asumsi kebebasan antar pengamatan pada analisis regresi sulit terpenuhi.

Menurut Fotheringham (2002), *Geographically Weighted Regression* (GWR/Regresi Terboboti Geografis) menjadi salah satu solusi untuk mengatasi masalah asumsi kebebasan yang sulit terpenuhi pada analisis regresi. Sehingga nantinya dapat membentuk model persamaan analisis regresi yang nilai-nilai parameternya hanya berlaku pada tiap lokasi pengamatan dan berbeda dengan lokasi lainnya. GWR merupakan bagian dari analisis regresi spasial yang menggunakan pendekatan titik, dan unsur matriks pembobot untuk menghasilkan estimasi parameter lokal berdasarkan posisi atau jarak satu wilayah pengamatan dengan wilayah pengamatan lainnya. Semakin jauh suatu lokasi dari lokasi yang diduga model regresinya, semakin rendah bobot untuk data pada lokasi tersebut (Fotheringham, 2002).

Pada saat pengujian parameter prediktor GWR terdapat beberapa variabel yang tidak signifikan atau tidak mempunyai pengaruh lokasi, namun bila dikaji lebih lanjut ternyata terdapat variabel-variabel yang berpengaruh secara global (Apriyani, 2018). Hal ini menunjukkan bahwa dalam beberapa kejadian tidak semua parameter dalam model GWR memiliki pengaruh secara lokal, adakalanya beberapa parameter bersifat global. Nakaya, dkk. (2005) mengembangkan model Regresi Terboboti Geografis Campuran (*Mixed Geographically Weighted Regression*/RTGC) yang merupakan model kombinasi atau gabungan antara regresi global dan GWR dengan mempertimbangkan beberapa variabel prediktor memengaruhi variabel respon secara global dan variabel prediktor yang lainnya secara lokal sehingga terdapat dua parameter yaitu lokal dan global.

Penerapan MGWR telah banyak dilakukan, diantarnya oleh Wuryanti, dkk. (2013) pada pemodelan angka kematian balita di Kabupaten Bojonegoro tahun 2011 yang menyimpulkan MGWR sebagai model terbaik dibandingkan dengan regresi linear dan GWR dilihat dari nilai AIC terkecil. Darsyah, dkk. (2015) memodelkan MGWR pada tingkat kemiskinan di Jawa Tengah, Rohmah (2015) melakukan pemodelan MGWR untuk memetakan potensi produk sapi potong di Jawa Timur tahun 2012, dan Suritman (2020) memodelkan MGWR yang mengandung multikolinearitas dengan metode regresi ridge.

Model GWR dan MGWR tidak dapat mengatasi kasus multikolinearitas yang merupakan masalah akibat terjadinya korelasi antara satu atau lebih peubah prediktor dengan peubah prediktor yang lainnya. Adanya multikolinearitas dapat menyebabkan estimasi parameter dari model yang dihasilkan memiliki variansi yang besar, sehingga dapat menyebabkan terjadinya kesalahan dalam menginterpretasikan parameter. Multikolinearitas pada regresi spasial dapat diatasi dengan menggunakan metode *Least Absolute Shrinkage and Selection Operator* (Lasso). Lasso melakukan estimasi parameter dengan algoritma LARS (*Least Angle Regression*) yang menyusutkan koefisien dugaaan sampai ke nol. Penerapan metode Lasso pada regresi spasial diharapkan dapat diperoleh estimasi parameter koefisien yang tidak bias dan efisien sehingga hasil prediksi pada model lebih akurat (Wheller dalam Ramadhan, 2013). Sehingga dengan penambahan Lasso pada model MGWR diharapkan dapat mengatasi masalah multikolinearitas yang terjadi pada model MGWR.

Model MGWR dapat diterapkan pada data Produk Domestik Regional Bruto (PDRB) Sulawesi Selatan. PDRB merupakan nilai tambah bruto seluruh barang dan jasa yang tercipta atau dihasilkan di wilayah domestik suatu negara yang timbul akibat berbagai aktivitas ekonomi dalam suatu periode tertentu (Badan Pusat Statistik, 2019). Menurut Handayani (2017), nilai PDRB di kabupaten/kota yang berdekatan dengan wilayah yang memiliki nilai PDRB tinggi akan memiliki kecenderungan nilai PDRB yang lebih tinggi pula. Hal ini menunjukkan bahwa nilai PDRB dapat bervariasi secara spasial karena kondisi geografi, budaya, dan kebijakan ekonomi yang ada di setiap wilayah.

Data Badan Pusat Statistik (BPS) menunjukkan bahwa perekonomian Sulawesi Selatan berdasarkan besaran PDRB atas dasar harga yang berlaku tahun 2019 mencapai 6.92%. Nilai PDRB mengalami penurunan sebesar 0.14% jika dibandingkan nilai PDRB tahun 2018 dengan nilai pada periode tersebut sebesar 7.06%. PDRB dapat digunakan sebagai ukuran maupun gambaran menyeluruh tentang kondisi perekonomian suatu daerah, karena merupakan suatu indikator ekonomi. Menurut Putra (2017), terjadinya kenaikan atau penurunan produksi barang dan jasa suatu daerah diindikasi dengan kenaikan atau penurunan PDRB dari suatu daerah tersebut.

Sehubungan dengan data PDRB Sulawesi Selatan yang telah diuraikan sebelumnya, data yang digunakan pada penelitian ini yaitu data produk domestik regional bruto kabupaten/kota di provinsi Sulawesi Selatan tahun 2019 dengan melibatkan lokasi geografis. Merujuk pada penelitian sebelumnya, Ramadhan (2013) yang melakukan perbandingan metode *Geogrhapically Weighted Lasso* (GWL) lokal dan global dalam mengatasi multikolinearitas pada model GWR, dan Miranti (2015) yang mendeteksi terjadinya multikolinearitas pada pemodelan prevalensi Malaria di Indonesia dalam model GWR kemudian melakukan estimasi parameter pada model GWR dengan metode Lasso. Oleh karena itu, pada penelitian ini penulis akan menentukan model MGWR menggunakan metode Lasso. Adapun judul penelitian ini, yaitu "**Pemodelan *Mixed Geographically Weighted Regression* Dengan Metode *Least Absolute Shrinkage and Selection Operator***".

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan sebelumnya, permasalahan dirumuskan sebagai berikut:

1. Bagaimana mengestimasi parameter pada model MGWR dengan menggunakan metode Lasso?
2. Bagaimana aplikasi MGWR dengan menggunakan metode Lasso pada data Produk Domestik Regional Bruto (PDRB) kabupaten/kota di Provinsi Sulawesi Selatan tahun 2019?

1.3 Batasan Masalah

Penelitian ini memiliki permasalahan yang dibatasi pada:

1. Fungsi pembobot yang digunakan adalah *Adaptive Gaussian Kernel.*
2. Data yang digunakan adalah data Produk Domestik Regional Bruto (PDRB) kabupaten/kota di Provinsi Sulawesi Selatan tahun 2019.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah:

1. Mendapatkan estimasi parameter model MGWR dengan metode Lasso.
2. Mendapatkan model MGWR dengan metode Lasso pada data Produk Domestik Regional Bruto (PDRB) kabupaten/kota di Provinsi Sulawesi Selatan tahun 2019.

1.5 Manfaat Penelitian

Pada penelitian ini, penulis berharap dapat memberikan manfaat berupa tambahan kepustakaan bagi pengguna ilmu statistika tentang penerapan model MGWR dengan menggunakan metode Lasso, serta pemahaman menganalisis data Produk Domestik Regional Bruto (PDRB) kabupaten/kota di Provinsi Sulawesi Selatan.

BAB II
TINJAUAN PUSTAKA

2.1 Model Regresi Linear

Analisis regresi adalah metode analisis statistik yang digunakan untuk memodelkan hubungan kebergantungan yang mungkin ada antara variabel respon (Y) dengan variabel prediktor (x) (Draper dan Smith, 1992). Secara umum, model regresi dengan *p* variabel prediktor dan jumlah pengamatan n dapat dituliskan pada Persamaan (2.1).

$$y_i = \beta_0 + \sum_{k=1}^{p} \beta_k x_{ik} + \varepsilon_i \text{ , } i = 1,2,\dots,n \tag{2.1}$$

Analisis Regresi mengasumsikan bahwa estimasi dari parameter regresi bernilai sama untuk setiap lokasi pengamatan atau berlaku secara global (Drapper dan Smith, 1992). Model Persamaan (2.1) dapat dituliskan dalam bentuk matriks.

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \tag{2.2}$$

$\boldsymbol{Y}$ merupakan vektor variabel respon berukuran $(n \times 1)$, $\boldsymbol{X}$ merupakan vektor variabel prediktor berukuran $\big(n \times (p+1)\big)$, $\boldsymbol{\beta}$ adalah vektor parameter regresi berukuran $\big((p+1) \times 1\big)$, dan $\boldsymbol{\varepsilon}$ merupakan vektor galat model regresi berukuran $(n \times 1)$.

Persamaan (2.2) dapat juga ditulis:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \tag{2.3}$$

Estimasi parameter menggunakan metode kuadrat terkecil dengan fungsi kuadrat terkecil (*Ordinary Least Square*/OLS), sehingga didapatkan estimasi parameter pada Persamaan (2.4).

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y} \tag{2.4}$$

$\widehat{\boldsymbol{\beta}}$ adalah vektor estimasi parameter regresi berukuran $\big((p+1) \times 1\big)$, $\boldsymbol{X}$ adalah matriks variabel prediktor yang berukuran $(n \times (P+1))$, dan $\boldsymbol{Y}$ adalah matriks variabel respon berukuran $(n \times 1)$.

2.2 *Geographically Weighted Regression*

Geographically Weighted Regression (GWR) merupakan pengembangan dari model regresi linear klasik yang digunakan untuk memodelkan data yang memiliki pengaruh spasial (Rahmahdianti dalam Rahmi, 2021). Oleh karena itu, sebelum memodelkan GWR perlu melihat ada tidaknya pengaruh spasial pada data.

2.2.1 Pengujian Pengaruh Spasial

Salah satu karakteristik data spasial pada analisis spasial dengan pendekatan titik adalah terjadi heterogenitas spasial atau keragaman variansi yang terdapat di setiap lokasi (Anselin, 2009). Menurut Rakhmasanti dalam Widayaka (2016), heterogenitas spasial terjadi akibat adanya perbedaan antara satu wilayah dengan wilayah lainnya. Pengujian heterogenitas spasial dapat dilakukan dengan menggunakan statistik uji *Breusch-Pagan* (BP) dengan Hipotesis:

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

H_1 : ada ${\sigma_i}^2 \neq \sigma^2$,$i = 1,2, \dots, n$ (Terdapat heterogenitas spasial)

Statistik uji:

$$BP = \left(\frac{1}{2}\right)\boldsymbol{f}'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{f} \sim \chi^2{}_{(\alpha,p)} \tag{2.5}$$

$\boldsymbol{f}$ merupakan vektor berukuran $(n \times 1)$ dengan elemennya adalah $f_i = \left(\frac{{\varepsilon_i}^2}{\sigma^2} - 1\right)$, ε_i merupakan galat kuadrat untuk observasi ke-i, σ^2 merupakan ragam galat ε_i, dan $\boldsymbol{Z}$ merupakan matriks berukuran $(n \times (p + 1))$ berisi vektor dari **X** yang telah terstandardisasi untuk setiap lokasi dan p merupakan banyaknya variabel prediktor. Kriteria keputusan adalah tolak H_0 jika $BP > \chi^2_{(\alpha,p)}$ sehingga dapat disimpulkan terdapat heterogenitas spasial (Anselin, 2009 dalam Ramadhan, 2013).

2.2.2 Model *Geographically Weighted Regression*

Geographically Weighted Regression (GWR) adalah metode yang menghasilkan koefisien regresi yang bervariasi secara sistematis dalam ruang. Model GWR merupakan pengembangan dari model regresi dimana setiap parameter dihitung pada setiap titik lokasi (Fortheringram dkk, 2002). Pendekatan ini akan menghasilkan model regresi berbeda-beda untuk setiap lokasi, dimana semua amatan terlibat dalam penggunaan model regresi tetapi dengan bobot yang berbeda. Semakin jauh suatu lokasi dari lokasi yang diduga model regresinya, semakin rendah bobot untuk data pada lokasi tersebut (Djuraidah, 2020).

Model GWR dapat ditulis pada Persamaan (2.6).

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^{p} \beta_k(u_i, v_i)x_{ik} + \varepsilon_i \tag{2.6}$$

Untuk lokasi-$i = 1,2, \dots, n$.

y_i merupakan nilai amatan variabel respon pada lokasi ke-i, $\beta_0(u_i, v_i)$ adalah *intercept* pada pengamatan ke-i, $\beta_k(u_i, v_i)$ adalah koefisien regresi untuk variabel prediktor ke-k pada lokasi ke-i, (u_i, v_i) merupakan fungsi dari koordinat lintang (u_i) dan bujur (v_i), x_{ik} merupakan nilai prediktor ke-k dari lokasi-i, dan ε_i merupakan galat regresi pada lokasi-i.

Persamaan (2.6) dapat dituliskan kedalam bentuk matriks, yaitu:

$$\boldsymbol{Y} = \boldsymbol{X\beta}(u_i, v_i) + \boldsymbol{\varepsilon} \tag{2.7}$$

$\boldsymbol{Y}$ merupakan vektor variabel respon berukuran $(n \times 1)$, $\boldsymbol{X}$ merupakan vektor variabel prediktor berukuran $\left(n \times (p+1)\right)$, $\boldsymbol{\beta}(u_i, v_i)$ adalah vektor parameter regresi berukuran $((p+1) \times 1)$, dan $\boldsymbol{\varepsilon}$ merupakan vektor galat model regresi berukuran $(n \times 1)$.

2.2.3 Penentuan *Bandwith* Optimum

Bandwith merupakan radius (h) dari suatu lingkaran terhadap titik pusat lokasi. Pengamatan yang dilakukan pada sebuah titik lokasi i didalam radius akan mempengaruhi dalam estimasi parameter model di setiap lokasi. Hal tersebut dikarenakan pengamatan yang terletak dalam radius h dilakukan pemberian pembobot sesuai dengan fungsi yang digunakan. Menurut Mei (2005), pengamatan yang terletak di luar radius h mempunyai pembobot yang bernilai nol sehingga tidak mempengaruhi estimasi parameter.

Pemilihan *bandwith* menjadi sangat penting karena akan mempengaruhi dalam estimasi fungsi kernel maupun keakuratan model data, yang mengatur variasi dan bias model. Nilai *bandwith* yang ukurannya kecil akan mengakibatkan varians mengecil, yakni sedikit pengamatan yang berada pada radius h menyebabkan model yang diperoleh sangat kasar karena estimasi pengamatan yang digunakan sedikit. Sebaliknya, jika nilai *bandwith* ukurannya besar akan mengakibatkan varians membesar, maka akan menimbulkan bias yang semakin besar pula. Hal tersebut disebabkan banyak pengamatan yang ada dalam radius h sehingga model yang diperoleh terlalu halus karena pengamatan yang digunakan dalam estimasi terlalu banyak.

Menurut Fotheringham, dkk (2002), terdapat beberapa metode yang digunakan untuk memilih *bandwith* optimum, salah satunya adalah menggunakan *Cross Validation* (CV). *Bandwith* optimum adalah *bandwith* yang menghasilkan nilai CV minimum. Secara matematis nilai CV dapat dituliskan pada Persamaan (2.8).

$$CV(h) = \sum_{i=1}^{n} (y_i - \hat{y}_{\neq i}(h))^2 \tag{2.8}$$

$\hat{y}_{\neq i}(h)$ merupakan estimasi y_i dengan pengamatan di lokasi (u_i, v_i) dihilangkan dari proses estimasi, y_i merupakan pengamatan ke-i, dan n merupakan jumlah sampel.

2.2.4 Pemilihan Pembobot

Pada analisis spasial pemilihan pembobot sangat penting karena nilai pembobot mewakili lokasi data yang diamati satu dengan yang lain. Estimasi parameter titik spasial (u_i, v_i) mempunyai pengaruh lebih besar apabila titik-titiknya saling berdekatan pada lokasi i. Fungsi pembobot (w_{ij}) digunakan untuk memberikan estimasi hasil parameter yang berbeda disetiap lokasi pengamatan (Chasco dalam Harahap, 2022).

Estimasi parameter titik spasial (u_i, v_i) dipengaruhi oleh titik-titik yang saling berdekatan lokasinya daripada lokasi dengan tempat yang berjauhan. Oleh sebab itu, pemilihan pembobot spasial sangat penting dalam analisis spasial karena nilai pembobot mewakili letak data yang satu dengan yang lain sehingga ketepatan cara pembobotan dibutuhkan.

Untuk mendapatkan matriks pembobot pada lokasi i yang terletak pada koordinat (u_i, v_i) yaitu lintang dan bujur, terlebih dahulu dilakukan penentuan fungsi pembobot yang akan digunakan. Tetapi sebelum pembobot ditentukan, harus dihitung terlebih dahulu jarak lokasi (u_i, v_i) dengan lokasi (u_j, v_j) menggunakan jarak *euclidean* (Djuraidah, 2020):

$$d_{ij} = \sqrt{(u_i - u_j)^2 + (v_i - v_j)^2} \tag{2.9}$$

Pada umumnya pembobotan pada GWR dilakukan dengan menggunakan fungsi kernel untuk estimasi kuadrat-kuadrat terkecil pada model GWR (Fotheringham dkk, 2002). Berdasarkan *bandwith*, kernel dikelompokkan ke dalam dua jenis, yakni fungsi kernel dengan *bandwith* tetap *(fixed)* dan fungsi kernel dengan *bandwith* adaptif *(adaptive)*.

1. Fungsi kernel dengan *bandwith* tetap *(fixed)*
    a. Kernel *Gaussian*

    $$w_j(u_i, v_i) = exp\left(-\frac{1}{2}\left(\frac{d_{ij}}{h}\right)^2\right)$$

    b. Kernel *Bisquare*

    $$w_j(u_i, v_i) = \begin{cases} \sqrt{\left(1-\left(\frac{d_{ij}}{h}\right)^2\right)} & untuk\ d_{ij} \leq h \\ 0 & untuk\ d_{ij} > h \end{cases}$$

2. Fungsi kernel dengan *bandwith* adaptif *(adaptive)*
    a. Kernel *Gaussian*

    $$w_j(u_i, v_i) = exp\left(-\frac{1}{2}\left(\frac{d_{ij}}{h_i}\right)^2\right)$$

    b. Kernel *Bisquare*

    $$w_j(u_i, v_i) = \begin{cases} \sqrt{\left(1-\left(\frac{d_{ij}}{h_i}\right)^2\right)} & untuk\ d_{ij} \leq h_i \\ 0 & untuk\ d_{ij} > h_i \end{cases}$$

Pada penelitian ini menggunakan fungsi pembobot adaptif kernel. Hal ini disebabkan kemampuan fungsi kernel adaptif yang dapat disesuaikan dengan kondisi titik-titik pengamatan. Jika titik-titik lokasi pengamatan tersebar secara padat di sekitar lokasi pengamatan ke-i maka *bandwith* yang diperoleh relatif sempit. Sebaliknya jika titik-titik lokasi pengamatan memiliki jarak yang relatif jauh dari titik lokasi pengamatan ke-i maka *bandwith* yang diperoleh akan semakin luas (Wheeler, 2010). Terdapat beberapa fungsi kernel adaptif yang umum digunakan, salah satunya adalah *adaptive gaussian kernel* (Fotheringham dkk, 2002).

Fungsi *kernel adaptive gaussian* dituliskan pada Persamaan (2.10).

$$w_j(u_i, v_i) = exp\left(-\frac{1}{2}\left(\frac{d_{ij}}{h_i}\right)^2\right) \qquad (2.10)$$

dengan d_{ij} adalah jarak antara lokasi (u_i, v_i) ke lokasi (u_j, v_j) dan h adalah parameter penghalus *(bandwith)*.

2.2.5 Estimasi Parameter *Geographically Weighted Regression*

Parameter *Geographically Weighted Regression* (GWR) di estimasi dengan metode kuadrat terkecil terboboti (*Weighted Least Square* /WLS). Metode WLS merupakan pengembangan dari OLS yang digunakan untuk mengatasi masalah heteroskedastisitas dengan menggunakan unsur pembobotan pada proses estimasinya. Matriks pembobot untuk lokasi-i merupakan matriks diagonal dengan elemen diagonalnya adalah w_{ij}, untuk $j = 1,2, \ldots., n$. Pembobot w_{ij} pada lokasi (u_i, v_i) ditentukan oleh suatu fungsi jarak d_{ij} antara lokasi-i dengan lokasi lainnya (Djuraidah, 2020). Matriks pembobot untuk lokasi-i berukuran $(n \times n)$, yaitu:

$$\boldsymbol{W}(u_i, v_i)_{(n\times n)} = \begin{bmatrix} w_{i1}(u_i, v_i) & 0 & \cdots & 0 \\ 0 & w_{i2}(u_i, v_i) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_{in}(u_i, v_i) \end{bmatrix} \qquad (2.11)$$

$\boldsymbol{W}(u_i, v_i)$ merupakan matriks pembobot spasial untuk model GWR pada lokasi ke$-i$ dengan ukuran $(n \times n)$.

Estimasi parameter model GWR dengan menggunakan WLS untuk model Persamaan (2.6) pada lokasi ke$-i$, dapat dilihat pada Persamaan (2.12).

$$\sum_{j=1}^{n} w_{\boldsymbol{j}}(u_i, v_i)\varepsilon_j^2 = \sum_{i=1}^{n} w_j(u_i, v_i)\left[y_j - \beta_0(u_i, v_i) - \sum_{k=1}^{P} \beta_k(u_i, v_i)\, x_{jk}\right] \qquad (2.12)$$

Persamaan (2.12) dapat dituliskan kedalam bentuk matriks, yaitu:

$$\boldsymbol{\varepsilon}'\boldsymbol{W}(u_i, v_i)\boldsymbol{\varepsilon} = (\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta_l}(u_i, v_i))'\boldsymbol{W}(u_i, v_i)(\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta_l}(u_i, v_i)) \qquad (2.13)$$

dengan:

$$\boldsymbol{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \boldsymbol{\beta_l}(u_i, v_i) = \begin{bmatrix} \beta_0(u_i, v_i) \\ \beta_1(u_i, v_i) \\ \vdots \\ \beta_p(u_i, v_i) \end{bmatrix},$$

$\boldsymbol{W}(u_i, v_i) = diag[w_{i1}(u_i, v_i), w_{i2}(u_i, v_i), \ldots, w_{in}(u_i, v_i)]$, dan $\boldsymbol{\varepsilon}' = [\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n]$

Solusi dari Persamaan (2.13) adalah:

$$\begin{aligned}\boldsymbol{\varepsilon}'\boldsymbol{W}(u_i,v_i)\boldsymbol{\varepsilon} &= \left(\boldsymbol{Y}-\mathbf{X}\boldsymbol{\beta_l}(u_i,v_i)\right)'\boldsymbol{W}(u_i,v_i)\left(\boldsymbol{Y}-\mathbf{X}\boldsymbol{\beta_l}(u_i,v_i)\right)\\ &= \boldsymbol{Y}'\boldsymbol{W}(u_i,v_i)\boldsymbol{Y}-\mathbf{2}\boldsymbol{\beta_l}'(u_i,v_i)\mathbf{X}'\boldsymbol{W}(u_i,v_i)\boldsymbol{Y}\\ &\quad+\boldsymbol{\beta_l}'(u_i,v_i)\mathbf{X}'\boldsymbol{W}(u_i,v_i)\mathbf{X}\boldsymbol{\beta_l}(u_i,v_i)\end{aligned} \tag{2.14}$$

Untuk mendapatkan estimasi parameter $\boldsymbol{\beta_l}(u_i,v_i)$, maka dilakukan penurunan pada Persamaan (2.14) terhadap $\boldsymbol{\beta_l}(u_i,v_i)$ dan hasilnya disamadengankan dengan nol, sehingga didapatkan:

$$\left.\frac{\boldsymbol{\partial\varepsilon}'\boldsymbol{W}(u_i,v_i)\boldsymbol{\varepsilon}}{\boldsymbol{\partial\beta_l}(u_i,v_i)}\right|_{\boldsymbol{\beta_l}(u_i,v_i)=\widehat{\boldsymbol{\beta_l}}(u_i,v_i)}=0$$

$$\frac{\boldsymbol{\partial}\left(\boldsymbol{Y}'\boldsymbol{W}(u_i,v_i)\boldsymbol{Y}-\mathbf{2}\boldsymbol{\beta_l}'(u_i,v_i)\mathbf{X}'\boldsymbol{W}(u_i,v_i)\boldsymbol{Y}+\boldsymbol{\beta_l}'(u_i,v_i)\mathbf{X}'\boldsymbol{W}(u_i,v_i)\mathbf{X}\boldsymbol{\beta_l}(u_i,v_i)\right)}{\boldsymbol{\partial\beta_l}(u_i,v_i)}=0$$

$$\mathbf{2X}'\boldsymbol{W}(u_i,v_i)\boldsymbol{Y}+\mathbf{2X}'\boldsymbol{W}(u_i,v_i)\mathbf{X}\widehat{\boldsymbol{\beta_l}}(u_i,v_i)(u_i,v_i)=0$$

$$\mathbf{2X}'\boldsymbol{W}(u_i,v_i)\mathbf{X}\widehat{\boldsymbol{\beta_l}}(u_i,v_i)(u_i,v_i)=\mathbf{2X}'\boldsymbol{W}(u_i,v_i)\boldsymbol{Y}$$

$$\mathbf{X}'\boldsymbol{W}(u_i,v_i)\mathbf{X}\widehat{\boldsymbol{\beta_l}}(u_i,v_i)(u_i,v_i)=\mathbf{X}'\boldsymbol{W}(u_i,v_i)\boldsymbol{Y}$$

Sehingga estimasi parameter model GWR untuk setiap lokasi pengamatan, yaitu:

$$\widehat{\boldsymbol{\beta}}_{\boldsymbol{l}}(u_i,v_i)=[\boldsymbol{X}'\boldsymbol{W}(u_i,v_i)\boldsymbol{X}]^{-1}\boldsymbol{X}'\boldsymbol{W}(u_i,v_i)\boldsymbol{Y} \tag{2.15}$$

$\boldsymbol{X}$ adalah matriks variabel prediktor yang berukuran $(n\times(p+1))$, $\boldsymbol{Y}$ adalah matriks variabel respon berukuran $(n\times 1)$, $\widehat{\boldsymbol{\beta}}(u_i,v_i)$ adalah vektor estimasi parameter GWR berukuran $((p+1)\times 1)$, dan $\boldsymbol{W}(u_i,v_i)$ merupakan matriks pembobot spasial lokasi ke-i berukuran $(n\times n)$.

2.3 *Mixed Geographically Weighted Regression*

Mixed Geographically Weighted Regression (MGWR) merupakan penggabungan dari model GWR dengan model regresi linear. Hal ini disebabkan karena tidak semua variabel prediktor berpengaruh secara global, tetapi sebagian hanya berpengaruh secara lokal. Untuk menentukan variabel prediktor global dan lokal dari model GWR maka terlebih dahulu perlu dilakukan pengujian variabilitas spasial, dan kemudian melakukan uji hipotesis untuk mengetahui kesesuaian model MGWR yang terbentuk dan pengaruh signifikansi variabel prediktor global dan terhadap variabel respon pada model MGWR.

2.3.1 Uji Variabilitas Model *Geographically Weighted Regression*

Uji Variabilitas pada model GWR dilakukan untuk mengetahui ada atau tidaknya perbedaan pengaruh yang signifikan pada variabel prediktor antar lokasi pengamatan. Uji variabilitas akan diperoleh variabel yang bersifat lokal dan variabel yang bersifat global (Suritman, 2020). Hipotesis yang digunakan sebagai berikut:

$H_0 : \beta_{1k}(u_1, v_1) = \beta_{2k}(u_2, v_2) = \cdots = \beta_{np}(u_n, v_n) = 0$

untuk setiap $k = 1,2, \dots, p$ (Tidak ada perbedaan pengaruh yang signifikan dari variabel prediktor X_k antara satu lokasi dengan lokasi yang lainnya)

$H_1 :$ minimal ada satu $\beta_{ik}(u_i, v_i) \neq 0$ untuk setiap $k = 1,2, \dots, p$ dan $i = 1,2, \dots, n$ (Ada perbedaan pengaruh yang signifikan dari variabel prediktor X_k antara satu lokasi dengan lokasi yang lainnya)

Statistik uji:

$$F = \frac{V_k^2 \Big/ tr\left(\frac{1}{n}\boldsymbol{B}_k'\left[\boldsymbol{I} - \frac{1}{n}\boldsymbol{J}\right]\right)\boldsymbol{B}_k}{SSE(H_1) \big/ b_1} \sim F_{(\alpha, df_1, df_2)} \tag{2.16}$$

$\boldsymbol{J}$ adalah matriks berukuran $(n \times n)$ yang semua elemennya adalah 1, $\boldsymbol{I}$ adalah matriks identitas berukuran $(n \times n)$, dan $\boldsymbol{e}_k$ adalah vektor kolom berukuran $(p + 1)$ yang bernilai satu untuk elemen ke-k dan nol untuk lainnya.

dengan

$$V_k^2 = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{\beta}_k(u_i, v_i) - \frac{1}{n}\sum_{i=1}^{n}\hat{\beta}_k(u_i, v_i)\right)^2 = \frac{1}{n}\boldsymbol{\beta}_k'\left[\boldsymbol{I} - \frac{1}{n}\boldsymbol{J}\right]\boldsymbol{\beta}_k$$

$$\boldsymbol{\beta}_k(u_i, v_i) = \begin{bmatrix} \beta_0(u_i, v_i) \\ \beta_1(u_i, v_i) \\ \vdots \\ \beta_p(u_i, v_i) \end{bmatrix}, \boldsymbol{B}_k = \begin{bmatrix} \boldsymbol{e}_1'(\boldsymbol{X}'\boldsymbol{W}(u_1, v_1)\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{W}(u_1, v_1) \\ \boldsymbol{e}_2'(\boldsymbol{X}'\boldsymbol{W}(u_2, v_2)\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{W}(u_2, v_2) \\ \vdots \\ \boldsymbol{e}_k'(\boldsymbol{X}'\boldsymbol{W}(u_n, v_n)\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{W}(u_n, v_n) \end{bmatrix}$$

$$SSE(H_1) = \boldsymbol{Y}'(\boldsymbol{I} - \boldsymbol{S})'(\boldsymbol{I} - \boldsymbol{S})\boldsymbol{Y}, \boldsymbol{Y}' = [y_1, y_2, \dots, y_n]$$

$$b_t = tr([(\boldsymbol{I} - \boldsymbol{S})'(\boldsymbol{I} - \boldsymbol{S})]^t), t = 1,2. ,\boldsymbol{S} = \begin{bmatrix} \boldsymbol{x}_1'(\boldsymbol{X}'\boldsymbol{W}(u_1, v_1)\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{W}(u_1, v_1) \\ \boldsymbol{x}_2'(\boldsymbol{X}'\boldsymbol{W}(u_2, v_2)\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{W}(u_2, v_2) \\ \vdots \\ \boldsymbol{x}_k'(\boldsymbol{X}'\boldsymbol{W}(u_n, v_n)\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{W}(u_n, v_n) \end{bmatrix}$$

$$c_t = tr\left(\frac{1}{n}\boldsymbol{B}_k'\left[\boldsymbol{I} - \frac{1}{n}\boldsymbol{J}\right]\boldsymbol{B}_k\right)^t, t = 1,2$$

Nilai statistik uji F mengikuti distribusi F dengan derajat bebas df_1 dan df_2, $df_1 = \frac{{c_1}^2}{c_2}$ dan $df_2 = \frac{{b_1}^2}{b_2}$. Jika diberikan tingkat signifikansi sebesar α maka H_0

ditolak apabila nilai statistik uji $F \geq F_{\alpha,df_1,df_2}$, artinya variabel prediktor bersifat lokal. Sebaliknya apabila nilai statistik uji $F < F_{\alpha,df_1,df_2}$, artinya variabel prediktor bersifat global.

2.3.2 Model *Mixed Geographically Weighted Regression*

Model MGWR dengan p variabel prediktor dan q variabel prediktor diantaranya bersifat lokal dan global, dengan mengasumsikan bahwa intersep model bersifat lokal dapat dituliskan seperti pada Persamaan (2.17) (Rohmah, 2015):

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^{q} \beta_k(u_i, v_i)x_{ik} + \sum_{l=q+1}^{p} \beta_l x_{il} + \varepsilon_i \qquad (2.17)$$

y_i adalah variabel respon, $\beta_0(u_i, v_i)$ adalah intersep pada lokasi pengamatan ke-i, $\beta_k(u_i, v_i)$ adalah parameter variabel prediktor lokal, (u_i, v_i) merupakan titik koordinat lokasi ke-i, β_l adalah parameter variabel global, x_{ik} adalah variabel prediktor lokal, x_{il} adalah variabel prediktor global, dan ε_i adalah galat pada pengamatan lokasi ke-i.

Persamaan (2.17) dapat dituliskan dalam bentuk matriks, yaitu:

$$\boldsymbol{Y} = \boldsymbol{X}_l \boldsymbol{\beta}_l(u_i, v_i) + \boldsymbol{X}_g \boldsymbol{\beta}_g + \boldsymbol{\varepsilon} \qquad (2.18)$$

$\boldsymbol{Y}$ adalah matriks variabel respon berukuran $(n \times 1)$, $\boldsymbol{X}_l$ adalah matriks variabel prediktor bersifat lokal berukuran $(n \times (q+1))$, $\boldsymbol{X}_g$ adalah matriks variabel prediktor bersifat global berukuran $(n \times (p-q))$, $\boldsymbol{\beta}_l(u_i, v_i)$ adalah vektor estimasi parameter bersifat lokal yang berukuran $((q+1) \times 1)$, $\boldsymbol{\beta}_g$ adalah vektor estimasi parameter bersifat global berukuran $((p-q) \times 1)$, dan $\boldsymbol{\varepsilon}$ merupakan galat pengamatan berukuran $(n \times 1)$.

$$\boldsymbol{X}_l = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1q} \\ 1 & x_{21} & \cdots & x_{2q} \\ \vdots & \cdots & \ddots & \vdots \\ 1 & x_{n1} & \dots & x_{nq} \end{bmatrix}, \qquad \boldsymbol{X}_g = \begin{bmatrix} x_{1(q+1)} & x_{2(q+2)} & \cdots & x_{1p} \\ x_{2(q+1)} & x_{2(q+2)} & \cdots & x_{2p} \\ \vdots & \cdots & \ddots & \vdots \\ x_{n(q+1)} & x_{n(q+2)} & \dots & x_{np} \end{bmatrix}$$

$$\boldsymbol{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \boldsymbol{\beta}_l(u_i, v_i) = \begin{bmatrix} \beta_0(u_i, v_i) \\ \beta_1(u_i, v_i) \\ \vdots \\ \beta_q(u_i, v_i) \end{bmatrix}, \boldsymbol{\beta}_g = \begin{bmatrix} \beta_{q+1} \\ \beta_{q+2} \\ \vdots \\ \beta_p \end{bmatrix}, \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Estimasi parameter pada model MGWR dapat dilakukan dengan metode WLS seperti halnya pada model GWR (Yasin, 2013). Estimasi parameter model MGWR dilakukan dengan terlebih dahulu mengidentifikasi variabel global dan variabel lokal pada model MGWR.

Estimasi parameter untuk model MGWR yang bersifat global dan lokal menggunakan metode WLS. Estimasi parameter pada model MGWR yang bersifat lokal sebagai berikut:

$$\widehat{\boldsymbol{\beta}}_l(u_i,v_i) = (\boldsymbol{X_l}'\boldsymbol{W}(u_i,v_i)\boldsymbol{X_l})^{-1}\boldsymbol{X_l}'\boldsymbol{W}(u_i,v_i)\boldsymbol{Y} \tag{2.19}$$

$\boldsymbol{W}(u_i,v_i)$ adalah matriks pembobot yang sama pada model GWR berukuran n x n.

$$\boldsymbol{W}(u_i,v_i) = diag[w_{i1}(u_i,v_i), w_{i2}(u_i,v_i), \dots, w_{in}(u_i,v_i)]$$

Estimasi parameter pada model MGWR yang bersifat global

$$\widehat{\boldsymbol{\beta}}_g = \left[\boldsymbol{X_g}'(\boldsymbol{I}-\boldsymbol{S}_l)'(\boldsymbol{I}-\boldsymbol{S}_l)\boldsymbol{X_g}\right]^{-1}\boldsymbol{X_g}'(\boldsymbol{I}-\boldsymbol{S}_l)'(\boldsymbol{I}-\boldsymbol{S}_l)\boldsymbol{Y} \tag{2.20}$$

dengan:

$$\boldsymbol{S}_l = \begin{bmatrix} \left[x_{i1}'\boldsymbol{W}(u_1,v_1)\boldsymbol{X_g}\right]^{-1}\boldsymbol{X_l}'\boldsymbol{W}(u_1,v_1) \\ \left[x_{i2}'\boldsymbol{W}(u_2,v_2)\boldsymbol{X_g}\right]^{-1}\boldsymbol{X_l}'\boldsymbol{W}(u_2,v_2) \\ \vdots \\ \left[x_{in}'\boldsymbol{W}(u_n,v_n)\boldsymbol{X_g}\right]^{-1}\boldsymbol{X_l}'\boldsymbol{W}(u_n,v_n) \end{bmatrix}$$

Estimator $\widehat{\boldsymbol{\beta}}_l(u_i,v_i)$ dan $\widehat{\boldsymbol{\beta}}_g$ merupakan estimator tak bias untuk $\boldsymbol{\beta}_l(u_i,v_i)$ dan $\boldsymbol{\beta}_g$ **(**Puhardi dalam Yasin, 2013).

2.3.3 Uji Hipotesis Model MGWR

Pengujian hipotesis dari model MGWR dilakukan dengan cara membandingkan antara model MGWR dengan model regresi global (Leung dkk dalam Suritman, 2020). Pengujian hipotesis yang dilakukan meliputi uji kesesuaian model (*goodness of fits*) MGWR dan regresi global, pengujian serentak untuk parameter variabel global dan lokal, dan pengujian parsial pada setiap model MGWR.

2.3.3.1 Uji Kesesuaian Model

Uji kesesuaian model dilakukan dengan membandingkan model MGWR dengan model regresi global untuk mengetahui faktor geografis yang signifikan (Darsyah, 2015). Hipotesis pengujian yang dilakukan sebagai berikut:

H_0 : $\beta_k(u_i, v_i) = \beta_k$ untuk setiap $k = 0,1,2, \dots, q$ dan $i = 1,2, \dots, n$

(Model MGWR tidak berbeda dengan Model Regresi Global)

H_1 : minimal ada satu $\beta_k(u_i, v_i) \neq \beta_k$ untuk setiap $k = 1,2, \dots, q$

(Model MGWR berbeda dengan Model Regresi Global)

Statistik uji:

$$F_1 = \frac{\boldsymbol{Y}'[(\boldsymbol{I}-\boldsymbol{H}) - (\boldsymbol{I}-\boldsymbol{S})'(\boldsymbol{I}-\boldsymbol{S})]\boldsymbol{Y}\big/v_1}{\boldsymbol{Y}'(\boldsymbol{I}-\boldsymbol{S})'(\boldsymbol{I}-\boldsymbol{S})\boldsymbol{Y}\big/u_1} \tag{2.21}$$

dengan

$$\boldsymbol{S} = \boldsymbol{S}_l + (\boldsymbol{I}-\boldsymbol{S}_l)\boldsymbol{X}_g\left[\boldsymbol{X}_g'(\boldsymbol{I}-\boldsymbol{S}_l)'(\boldsymbol{I}-\boldsymbol{S}_l)\right]^{-1}\boldsymbol{X}_g'(\boldsymbol{I}-\boldsymbol{S}_l)'(\boldsymbol{I}-\boldsymbol{S}_l)$$

$$u_t = tr([(\boldsymbol{I}-\boldsymbol{S})'(\boldsymbol{I}-\boldsymbol{S})]^t), t = 1,2$$

$$v_t = tr([(\boldsymbol{I}-\boldsymbol{H})(\boldsymbol{I}-\boldsymbol{S})'(\boldsymbol{I}-\boldsymbol{S})]^t), t = 1,2$$

$$\boldsymbol{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{X}'$$

Nilai statistik uji F_1 mengikuti distribusi F dengan derajat bebas df_1 dan df_2. Jika diberikan tingkat signifikansi sebesar α maka H_0 ditolak apabila nilai statistik uji $F_1 < F_{\alpha, df_1, df_2}$, dengan $df_1 = \frac{{v_1}^2}{v_2}$ dan $df_2 = \frac{{u_1}^2}{u_2}$.

2.3.3.2 Uji Serentak Parameter Model

Uji hipotesis serentak bertujuan untuk mengetahui bagaimana signifikansi variabel-variabel prediktor pada model MGWR (Wuryanti, 2013). Pengujian secara serentak dilakukan pada parameter variabel prediktor global dan lokal. Untuk pengujian hipotesis serentak pada parameter variabel prediktor global menggunakan hipotesis sebagai berikut:

H_0 : $\beta_{q+1} = \beta_{q+2} = \cdots = \beta_P = 0$ (Tidak ada pengaruh variabel global secara serentak terhadap variabel respon)

H_1 : minimal ada satu $\beta_k \neq 0$ untuk setiap $k = q+1, q+2, \dots, p$ (Ada pengaruh variabel global secara serentak terhadap variabel respon)

Statistik uji:

$$F_2 = \frac{\boldsymbol{Y}'[(\boldsymbol{I}-\boldsymbol{S}_l)'(\boldsymbol{I}-\boldsymbol{S}_l) - (\boldsymbol{I}-\boldsymbol{S})'(\boldsymbol{I}-\boldsymbol{S})]\boldsymbol{Y}/r_1}{\boldsymbol{Y}'(\boldsymbol{I}-\boldsymbol{S})'(\boldsymbol{I}-\boldsymbol{S})\boldsymbol{Y}/u_1} \sim F_{(\alpha, df_1, df_2)} \tag{2.22}$$

dengan

$$r_t = tr([(\boldsymbol{I}-\boldsymbol{S}_l)'(\boldsymbol{I}-\boldsymbol{S}_l) - (\boldsymbol{I}-\boldsymbol{S})'(\boldsymbol{I}-\boldsymbol{S})]^t), t = 1,2$$

Nilai statistik uji F_2 mengikuti distribusi F dengan derajat bebas df_1 dan df_2. Jika diberikan tingkat signifikansi sebesar α maka H_0 ditolak apabila nilai statistik uji $F_2 \geq F_{\alpha,df_1,df_2}$, dengan $df_1 = \frac{{r_1}^2}{r_2}$ dan $df_2 = \frac{{u_1}^2}{u_2}$.

Uji hipotesis serentak selanjutnya yakni pada parameter variabel prediktor lokal. Hipotesis yang digunakan sebagai berikut:

H_0 : $\beta_1(u_i, v_i) = \beta_2(u_i, v_i) = \cdots = \beta_q(u_i, v_i) = 0$ untuk setiap $i = 1,2, \ldots, n$

(Tidak ada pengaruh variabel lokal terhadap variabel respon)

H_1 : minimal ada satu $\beta_k(u_i, v_i) \neq 0$ untuk setiap $k = 1,2, \ldots, q$

(Ada pengaruh variabel lokal terhadap variabel respon)

Statistik uji:

$$F_3 = \frac{\boldsymbol{Y}'\left[\left(\boldsymbol{I} - \boldsymbol{S}_g\right)'\left(\boldsymbol{I} - \boldsymbol{S}_g\right) - (\boldsymbol{I} - \boldsymbol{S})'(\boldsymbol{I} - \boldsymbol{S})\right]\boldsymbol{Y}/t_1}{\boldsymbol{Y}'(\boldsymbol{I} - \boldsymbol{S})'(\boldsymbol{I} - \boldsymbol{S})\boldsymbol{Y}/u_1} \sim F_{(\alpha,df_1,df_2)} \tag{2.23}$$

dengan

$$t_t = tr\left(\left[\left(\boldsymbol{I} - \boldsymbol{S}_g\right)'\left(\boldsymbol{I} - \boldsymbol{S}_g\right) - (\boldsymbol{I} - \boldsymbol{S})'(\boldsymbol{I} - \boldsymbol{S})\right]^t\right), t = 1,2$$

$$\boldsymbol{S}_g = \mathbf{X}_{\boldsymbol{g}}\left(\mathbf{X}'_{\boldsymbol{g}}\mathbf{X}_{\boldsymbol{g}}\right)^{-\mathbf{1}}\mathbf{X}'_{\boldsymbol{g}}$$

Nilai statistik uji F_3 mengikuti distribusi F dengan derajat bebas df_1 dan df_2. Jika diberikan tingkat signifikansi sebesar α maka H_0 ditolak apabila nilai statistik uji $F_3 \geq F_{\alpha,df_1,df_2}$, dengan $df_1 = \frac{{t_1}^2}{t_2}$ dan $df_2 = \frac{{u_1}^2}{u_2}$.

2.3.3.3 Uji Parsial Parameter Model

Uji parsial bertujuan untuk mengetahui variabel prediktor global dan lokal yang berpengaruh signifikan terhadap respon pada model MGWR (Purhadi dkk dalam Suritman, 2020). Pengujian ini dilakukan pada parameter variabel global dan lokal. Pengujian parsial pada parameter variabel prediktor global menggunakan hipotesis sebagai berikut

H_0 : $\beta_k = 0$ untuk setiap $k = q + 1, q + 2, \ldots, p$ (Variabel global X_k tidak signifikan)

H_1 : $\beta_k \neq 0$ (Variabel global X_k signifikan)

Statistik uji:

$$T_{g\_hit} = \frac{\hat{\beta}_k}{\hat{\sigma}\sqrt{g_{kk}}} \sim T_{\left(\alpha/_{2},df\right)} \tag{2.24}$$

dengan

g_{kk} = Elemen diagonal ke k dari hasil perkalian matriks $\boldsymbol{GG'}$

$\boldsymbol{G} = \left[X_g'(\boldsymbol{I}-\boldsymbol{S}_l)'(\boldsymbol{I}-\boldsymbol{S}_l)\boldsymbol{X}_g\right]^{-1}X_g'(\boldsymbol{I}-\boldsymbol{S}_l)'(\boldsymbol{I}-\boldsymbol{S}_l)$

$\hat{\sigma}^2 = \dfrac{\boldsymbol{Y}'(\boldsymbol{I}-\boldsymbol{S})'(\boldsymbol{I}-\boldsymbol{S})\boldsymbol{Y}}{tr\left((\boldsymbol{I}-\boldsymbol{S})'(\boldsymbol{I}-\boldsymbol{S})\right)}$

Pada taraf signifikansi sebesar α, maka dapat diambil keputusan tolak H_0 jika $\left|T_{g\_hit}\right| > T_{\alpha/_{2},df}$. Dimana $df = \frac{{u_1}^2}{u_2}$.

Uji hipotesis parsial selanjutnya pada ditunjukkan untuk mengetahui variabel lokal yang berpengaruh signifikan terhadap respon pada model MGWR. Pengujian signifikansi suatu variabel lokal digunakan hipotesis yaitu:

$H_0 : \beta_k(u_i, v_i) = 0$ untuk setiap $k = 1,2, \dots, q$ dan $i = 1,2, \dots, n$ (Variabel lokal X_k pada lokasi ke-i tidak signifikan)

$H_1 : \beta_k(u_i, v_i) \neq 0$ (Variabel lokal X_k pada lokasi ke-i signifikan)

Statistik uji:

$$T_{l\_hit} = \frac{\hat{\beta}_k(u_i, v_i)}{\hat{\sigma}\sqrt{m_{kk}}} \sim T_{(\alpha/_{2},df)} \tag{2.25}$$

dengan

m_{kk} = Elemen diagonal ke k dari hasil perkalian matriks $\boldsymbol{MM'}$

$\boldsymbol{M} = [\boldsymbol{X}_l'\boldsymbol{W}(u_i, v_i)\boldsymbol{X}_l]^{-1}\boldsymbol{X}_l'\boldsymbol{W}(u_i, v_i)\left(\boldsymbol{I}-\boldsymbol{X}_g\boldsymbol{G}\right)$

$\hat{\sigma}^2 = \dfrac{\boldsymbol{Y}'(\boldsymbol{I}-\boldsymbol{S})'(\boldsymbol{I}-\boldsymbol{S})\boldsymbol{Y}}{tr\left((\boldsymbol{I}-\boldsymbol{S})'(\boldsymbol{I}-\boldsymbol{S})\right)}$

Pada taraf signifikansi sebesar α, maka dapat diambil keputusan tolak H_0 jika $\left|T_{l\_hit}\right| > T_{\alpha/_{2},df}$, dengan $df = \frac{{u_1}^2}{u_2}$.

2.4 Multikolinearitas

Multikolineritas merupakan suatu kondisi terjadinya hubungan linear atau korelasi yang tinggi antara masing-masing variabel prediktor dalam model regresi. Menurut Gujarati (2004), deteksi multikolinieritas bertujuan untuk mendeteksi adanya korelasi antar variabel prediktor pada model regresi. Jika variabel prediktor saling berkorelasi, maka mengakibatkan *standard error* (kesalahan baku) akan semakin besar seiring meningkatnya korelasi antara variabel prediktor dan juga akan mengakibatkan variabel-variabel ini tidak orthogonal. Variabel orthogonal adalah variabel prediktor yang nilai korelasi antar sesama variabel prediktor sama dengan nol. Untuk mengetahui ada atau tidak multikolinearitas, dapat dilihat nilai *Variance Inflation Factor* (VIF). Nilai VIF menunjukkan terjadi multikolinearitas pada data apabila terdapat variabel yang memiliki nilai VIF>10 (Putri, 2011).

Nilai *Variance Inflation Factor* (VIF) dapat dihitung dengan rumus:

$$VIF_k = \frac{1}{(1 - {R_k}^2)} \tag{2.26}$$

$$1 - {R_k}^2 = TOL, k = 1,2, \dots, p \tag{2.27}$$

${R_k}^2$ merupakan koefisien determinasi yang diperoleh dengan meregresikan variabel prediktor x_k dengan $p - 1$ variabel x lainnya. Nilai VIF yang semakin besar menunjukkan multikolineritas yang lebih kompleks. Jika nilai VIF < 10, maka secara signifikan dapat disimpulkan bahwa tidak terdapat multikolineritas (Widayaka, 2016).

2.5 *Least Absolute Shrinkage and Selection Operator*

Least Absolute Shrinkage and Selection Operator (Lasso) merupakan metode regresi hasil dari pengembangan regresi ridge yang mampu untuk mengatasi masalah muktikolinearitas. Metode ini pertama kali diperkenalkan oleh Robert Tibshirani (1996). Regresi Lasso digunakan untuk *shrinkage* yaitu menyusutkan koefisien regresi mendekati angka nol atau menjadi tepat nol dan *selection operator* yaitu menyeleksi variabel-variabel prediktor sehingga menghasilkan model dengan variabel terbaik (Robbani, 2019).

Estimasi koefisien regresi lasso ($\hat{\beta}^{Lasso}$) diperoleh dengan cara meminimumkan jumlah kuadrat galat dengan fungsi kendala $\sum_{k=1}^{p} |\beta_k|) \leq t$, Persamaannya sebagai berikut (Tibshirani 1996):

$$\hat{\beta}^{Lasso} = argmin \left\{ \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{k=1}^{p} \chi_{ik}\beta_k \right)^2 + \lambda \sum_{k=1}^{p} |\beta_k| \right\} \tag{2.28}$$

y_i menyatakan variabel respon pengamatan ke-i, β_0 merupakan konstanta, β_k merupakan koefisien dari variabel prediktor ke-k, χ_{ik} merupakan variabel prediktor, n menyatakan banyaknya observasi dan p merupakan banyaknya variabel prediktor dalam model.

Nilai t merupakan suatu besaran yang mengontrol besarnya penyusutan pada koefisien regresi Lasso. Nilai t dinamakan parameter tuning dengan nilai $t \geq 0$. Misalkan diketahui β_k merupakan estimasi OLS dengan nilai t_0 di denifisikan $t_0 = \sum_{k=1}^{p} |\hat{\beta}_k|$, maka jika nilai $t < t_0$ akan menyebabkan koefisien OLS akan menyusut ke arah nol, dan memungkinkan beberapa koefisien tepat nol. Jika nilai $t \geq t_0$, maka koefisien regresi Lasso memberikan hasil yang sama dengan koefisien OLS.

Estimasi koefisien Lasso diperoleh dengan menentukan batas yang dibakukan yaitu $s = t / \sum_{k=1}^{p} |\widehat{\beta_k^o}|$ dengan $t = \sum_{k=1}^{p} |\hat{\beta}_k|$ dan $\widehat{\beta_k^o}$ adalah estimasi parameter untuk model penuh (Dewi dalam Ramadhan, 2013). Kemudian λ pada Persamaan (2.28) menurut (Tibshirani dalam Robbani, 2019), disebut sebagai parameter tuning yang berkorespondensi satu-satu dengan s artinya untuk setiap nilai $t \geq t_0$ yang menghasilkan solusi $\widehat{\beta_J}^{Lasso}$ terdapat $\lambda \geq 0$ sedemikian sehingga menghasilkan solusi $\widehat{\beta_J}^{Lasso}$ juga. Solusi regresi Lasso tidak memiliki solusi eksplisit karena pada fungsi kendala regresi Lasso berbentuk fungsi mutlak yang tidak dapat diturunkan pada titik beloknya, sehingga prosedur yang dapat digunakan untuk mengatasi hal tersebut dengan menggunakan pemrograman kuadratik (Hastie dkk dalam Yulita, 2016).

Algoritma LARS

LARS adalah metode seleksi model yang algoritmanya hasil modifikasi dari LAR (*Least Angle Regression*). LAR merupakan sebuah algoritima untuk menghasilkan model linear yang ditemukan oleh Efron dkk pada tahun 2002. Algortima LAR membutuhkan p langkah untuk mendapatkan estimasi koefisien OLS, namun dengan dimodifikasinya LAR menjadi LARS maka dapat meghasilkan estimasi koefisien metode Lasso. Hal ini dikarenakan algoritma hasil modifikasi (LARS) memiliki langkah yang lebih efisien dibanding dengan metode Lasso (Robbani dkk, 2019). Algoritma LARS di definiskan sebagai berikut:

1. Memulai dengan semua koefisien parameter sama dengan nol $\beta_1, \beta_2, \dots., \beta_p = 0$, sehingga galat $\varepsilon = y$.
2. Mencari variabel prediktor yang paling berkorelasi dengan ε.
3. Menduga koefisien nilai β_k untuk x_{ik} yang paling berkorelasi dengan ε.

4. Menghitung galat $\varepsilon = y - \hat{y}$ dengan variabel prediktor x_k yang masuk ke dalam model.
5. Menghitung korelasi parsial antara peubah prediktor yang tersisa dengan galat terbaru.
6. Mengulang langkah ke-3 sampai langkah ke-5 hingga semua variabel prediktor masuk kedalam model dan berhenti jika korelasi $(y, x_{ik}) = 0$.

Tahapan yang dilakukan dalam estimasi parameter Lasso model MGWR sebagai berikut:

1. Menghitung *bandwith* yang optimum dengan metode *Cross Validation* (CV).
2. Menghitung matriks pembobot $\boldsymbol{W}$ berukuran $n \times n$.
3. Untuk setiap lokasi $i = 1,2, \dots, n$
   a) $\boldsymbol{W}^{\frac{1}{2}}(i) = sqrt\big(diag(\boldsymbol{W}(i))\big)$
   b) $\mathbf{X}_w = \boldsymbol{W}^{\frac{1}{2}}(i)\mathbf{X}$ dan $\mathbf{Y}_w = \boldsymbol{W}^{\frac{1}{2}}(i)\mathbf{Y}$ menggunakan akar kuadrat dari pembobot kernel $\boldsymbol{W}(i)$ di setiap lokasi.
   c) Panggil algoritma LARS $(\mathbf{X}_w, \mathbf{Y}_w)$ pada *software* R.
4. Estimasi parameter akhir Lasso sesuai dengan CV berdasarkan fraksi dari nilai penyusutan (s).

Pada tahap penyelesaian Lasso dengan menggunakan algoritma LARS, akan muncul plot pergerakan variabel-variabel prediktor dengan parameter tuning bentuk standar (s). Parameter tuning merupakan parameter *shrinkage* (s) yang digunakan sebagai batasan Lasso untuk estimasi parameter Lasso yang berpengaruh signifikan terhadap variabel respon (Ramadhan, 2013). Estimasi parameter *shrinkage* Lasso dilakukan dengan menggunakan metode *cross validation* (CV) (Efron dkk, 2004).

Parameter *shrinkage* Lasso dapat didefinisikan seperti pada Persamaan (2.29).

$$s = \frac{\sum_{k=1}^{p} \left|\hat{\beta}_k\right|}{\sum_{k=1}^{p} \left|\widehat{\beta^{o}}_k\right|} \tag{2.29}$$

s menyatakan parameter penyusutan (*shrinkage)* yang memiliki nilai 0 sampai 1.

2.6 Pemilihan Model Terbaik

Salah satu metode yang dapat digunakan untuk pemilihan model terbaik adalah *Akaike Information Criterion* (AIC) yang didefinisikan pada Persamaan (2.30).

$$AIC = 2n \ln(\hat{\sigma}) + n \ln(2\pi) + n + tr(\boldsymbol{S}) \qquad (2.30)$$

dengan $\hat{\sigma}$ merupakan nilai estimator standar deviasi dari bentuk galat dan **S** merupakan matriks proyeksi dari model. Pemilihan model terbaik dilakukan dengan menentukan nilai AIC terkecil (Wuryanti, dkk. 2013).

2.7 Produk Domestik Regional Bruto

Produk Domestik Regional Bruto (PDRB) merupakan nilai tambah bruto seluruh barang dan jasa yang tercipta atau dihasilkan di wilayah domestik suatu negara yang timbul akibat berbagai aktivitas ekonomi dalam suatu periode tertentu tanpa memperhatikan apakah faktor produksi yang dimiliki *residen* atau *non-residen* (Badan Pusat Statistik, 2019).

PDRB dapat disusun dengan melalui 3 (tiga) pendekatan yaitu pendekatan produksi, pengeluaran, dan pendapatan yang disajikan atas dasar harga konstan dan harga berlaku. PDRB harga berlaku bertujuan untuk melihat struktur perekonomian yang menunjukkan kemampuan sumber daya ekonomi yang dihasilkan oleh suatu daerah. Struktur perekonomian pada suatu daerah dapat dipengaruhi oleh beberapa faktor yang dapat menyebabkan terjadinya kenaikan pertumbuhan PDRB diluar tiga pendekatan. Selain dilihat dari sumber pendukung asli perekonomian, nilai PDRB tidak lepas dari kualitas sumber daya manusia.

Penelitian mengenai faktor-faktor yang mempengaruhi PDRB telah banyak dilakukan, diantaranya oleh Hikmah (2012) yang memodelkan tingkat produk domestik regional bruto kabupaten/kota Jawa Barat selama periode 2005-2009 menggunakan variabel PAD, jumlah penduduk, total belanja daerah dan Indeks Pembangunan Manusia dengan spasial data panel. Sumardi (2016) melakukan pemodelan PDRB ADHK di Indonesia tahun 2013 menggunakan variabel luas wilayah, PMA, PMDN, jumlah tenaga kerja, angka harapan hidup, angka melek huruf, rata-rata lama sekolah, upah minimum provinsi, PAD, dan dana alokasi umum dengan menggunakan *Bayesian Model Averaging*. Handayani (2017) memodelkan PDRB pada 35 kabupaten/kota di Jawa Tengah tahun 2011-2015 menggunakan variabel tenaga kerja, PAD, UMK, dan IPM dengan Regresi Panel Terboboti Geografis. Kasyfurrahman (2020) memodelkan PDRB dengan 13 variabel yakni jumlah petani, jumlah pertambangan, jumlah industri, jumlah pelanggan air bersih, jumlah perusahaan kontruksi, jumlah hotel dan penginapan, jumlah kendaraan umum, pendapatan asli daerah, jumlah penduduk bekerja, angka harapan hidup, harapan lama sekolah, rata-rata lama sekolah, dan pengeluaran perkapita menggunakan metode *Geographically Weighted Lasso*.

Berdasarkan beberapan penelitian tersebut, pada penelitian ini akan menggunakan tujuh variabel respon yang diduga mempengaruhi PDRB kabupaten/kota di Sulawesi Selatan, yaitu Indeks Pembangunan Manusia (IPM), angka harapan hidup, rata-rata lama sekolah, pengeluaran perkapita, penduduk miskin, rumah tangga menggunakan listrik, dan angka melek huruf.