

**PERFORMA *BANDWIDTH CROSS VALIDATION* DAN
AKAIKE INFORMATION CRITERION CORRECTED
PADA MODEL *GEOGRAPHICALLY WEIGHTED*
*REGRESSION***

**(Studi Kasus: Jumlah Penduduk Kabupaten/Kota di
Provinsi Sulawesi Selatan Tahun 2018)**

SKRIPSI



RAYHANNA AULIYA AMIN

H12116302

**PROGRAM STUDI STATISTIKA DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS HASANUDDIN**

MAKASSAR

2021

**PERFORMA BANDWIDTH CROSS VALIDATION DAN
AKAIKE INFORMATION CRITERION CORRECTED PADA
MODEL GEOGRAPHICALLY WEIGHTED REGRESSION
(Studi Kasus: Jumlah Penduduk Kabupaten/Kota di Provinsi
Sulawesi Selatan Tahun 2018)**

SKRIPSI

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

RAYHANNA AULIYA AMIN

H 121 16 302

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

2021

LEMBAR PENGESAHAN SKRIPSI

**PERFORMA BANDWIDTH CROSS VALIDATION DAN AKAIKE
INFORMATION CRITERION CORRECTED PADA MODEL
GEOGRAPHICALLY WEIGHTED REGRESSION**

**(Studi Kasus: Jumlah Penduduk Kabupaten/Kota di Provinsi Sulawesi
Selatan Tahun 2018)**

Disusun dan diajukan oleh:

RAYHANNA AULIYA AMIN

H12116302


Telah dipertahankan di hadapan Panitia Ujian yang dibentuk dalam rangka
Penyelesaian Studi Program Sarjana Program Studi Statistika Fakultas
Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin
pada tanggal 04 Januari 2021
dan dinyatakan telah memenuhi syarat kelulusan

Menyetujui,

Pembimbing Utama


Andi Kresna Jawa, S.Si., M.Si.
NIP. 19731228 200003 1 001

Pembimbing Pertama


Sitti Sahriman, S.Si., M.Si.
NIP. 19881018 201504 2 002

Ketua Program Studi


Dr. Nurtiti Sunusi, S.Si., M.Si.
NIP. 19720117 199703 2 002



iii

PERNYATAAN KEASLIAN

Yang bertanda tangan di bawah ini:

Nama : Rayhanna Auliya Amin
NIM : H12116302
Program Studi : Statistika
Jenjang : S1

Menyatakan dengan ini bahwa karya tulisan saya berjudul

**PERFORMA BANDWIDTH CROSS VALIDATION DAN AKAIKE
INFORMATION CRITERION CORRECTED PADA MODEL
GEOGRAPHICALLY WEIGHTED REGRESSION**

**(Studi Kasus: Jumlah Penduduk Kabupaten/Kota di Provinsi Sulawesi
Selatan Tahun 2018)**

Adalah karya tulisan saya sendiri dan bukan merupakan pengambilan alihan tulisan orang lain bahwa Skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri.

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

Makassar, 04 Januari 2021

 Yang Menyatakan

Rayhanna Auliya Amin

KATA PENGANTAR

Segala puji bagi Allah *Subhanahu Wa Ta'ala* pemilik alam semesta yang senantiasa melimpahkan rahmat, hidayah, dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi dengan judul “Performa *Bandwidth Cross Validation* dan *Akaike Information Criterion Corrected* pada Model *Geographically Weighted Regression* (Studi Kasus: Jumlah Penduduk Kabupaten/Kota di Provinsi Sulawesi Selatan Tahun 2018). Shalawat dan salam senantiasa tercurah kepada Rasulullah *Shallallahu Alaihi Wasallam* suri tauladan terbaik sepanjang masa, kepada keluarga beliau serta para sahabat beliau yang senantiasa berjuang di jalan Allah *Subhanahu Wa Ta'ala*. Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar sarjana pada Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin. Penulis berharap skripsi ini dapat memberikan tambahan pengetahuan bagi para pembelajar statistika.

Penulis menyadari bahwa dalam penyusunan skripsi ini tidak lepas dari hambatan dan masalah namun dapat terselesaikan berkat bantuan, dorongan, dan dukungan dari berbagai pihak. Oleh karena itu, penulis mengucapkan terima kasih dan penghargaan yang setinggi-tingginya kepada orang tua penulis, Ibunda tercinta dan terkasih **Israwati, S.Pd.AUD** dan Ayahanda tercinta dan tersayang **Muhammad Amin** (almarhum) yang senantiasa mencintai, mengasihi, melindungi, mendidik, menasehati, dan menjadi motivasi sepanjang hidup penulis untuk senantiasa menjadi pribadi yang lebih baik serta selalu mengiringi langkah penulis dalam taburan do'a paling ikhlas dan tulus. Ucapan terima kasih juga kepada kakak penulis, **Alif Hidayatullah Amin, S.Pd** atas doa, nasehat, motivasi, bantuan, dan do'a yang selalu diberikan kepada penulis.

Terima kasih yang sebesar-besarnya dan penghargaan yang tulus juga penulis ucapkan kepada:

1. **Ibu Prof. Dr. Dwia Aries Tina Pulubuhu, MA** selaku Rektor Universitas Hasanuddin beserta seluruh jajarannya.
2. **Bapak Dr. Eng. Amiruddin, M.Si.** selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin beserta seluruh jajarannya

3. **Ibu Dr. Nurtiti Sunusi, S.Si., M.Si.** selaku Ketua Departemen Statistika atas segala nasehat, ilmu, motivasi, dan bantuan yang senantiasa diberikan selama menjadi mahasiswa di Departemen Statistika.
4. **Bapak Andi Kresna Jaya, S.Si., M.Si.** selaku Pembimbing Utama dan **Ibu Sitti Sahrinan, S.Si., M.Si.** selaku Pembimbing Pertama yang dengan sabar senantiasa meluangkan waktu untuk membimbing dan memberi masukan serta motivasi dalam penyelesaian skripsi ini.
5. **Ibu Dr. Nurtiti Sunusi, S.Si., M.Si.** selaku Penguji dan Penasehat Akademik atas saran, nasehat, motivasi, dan dukungan yang telah diberikan selama menjadi mahasiswa dan atas waktu yang telah diluangkan untuk memberikan kritik dan saran yang membangun demi penyempurnaan skripsi ini. **Ibu Anisa, S.Si., M.Si.** selaku Penguji yang dengan sabar telah meluangkan waktu untuk memberikan kritik dan saran yang membangun demi penyempurnaan penulisan skripsi ini.
6. Seluruh **Dosen dan Staf Departemen Statistika** yang senantiasa berbagi ilmu, nasehat, dan motivasi selama menjadi mahasiswa di Departemen Statistika.
7. Sahabat-sahabat tercinta penulis, **Nurul Azizah, Resky Afriyani Syam, Rezki Septiani TRI, S.Tr.Gz, Suhartatik Kuntari, S.Pd,** dan **Amaliah Fauziah M.** yang menjadi tempat berbagi canda tawa terbaik, serta senantiasa memberikan motivasi, kebahagiaan, serta semangat dalam setiap keadaan.
8. Sahabat-sahabat terbaik sekaligus teman berbagi kamar selama menjadi mahasiswa, **Nurhidayani, S.E, Andi Nurul Anisa, S.Si, Wiwik Rabiatul Adawiah,** dan **Rizky Utami Nur, A.Md.T.** Terima kasih atas segala suka duka, motivasi, nasehat, dan pengalaman yang telah dibagi bersama.
9. Teman-teman **Statistika 2016** terkhusus kepada sahabat seperjuangan, **WKND 4EVER: Dewi Santika Upa P, S.Si, Reski Amalah, S.Si, Fitriatusakiah, S.Si, Rosdiana, S.Si, Dewi Rahma Ente, S.Si, Widya Nauli Amalia Puteri, S.Si, Jumrianti, Isnawati, Andi Riska Fitriani, Zhazha Alifkhamulki Ramdhani, S.Si, Rusydah Khaerati, S.Si, Ayu**

Riski Ramadani, Bunga Aprilia, S.Si, Halniati, Reski Ulandari, Agung Muh. Takdir, Fajar Affan, S.Si, Samsul Arifin, S.Si, Suritman, S.Si, Muhammad Jayzul Usrah, S.Si, Rizki Adiputra, S.Si. Terima kasih atas segala pengalaman, cerita, kebahagiaan, keluh kesah, canda tawa, dan kenangan yang telah dibagi bersama-sama.

10. Keluarga Besar **KMF MIPA UNHAS, HIMASTAT FMIPA UNHAS, HIMATIKA FMIPA UNHAS**, terkhusus **ALGORITMA 2016**. Terima kasih atas ilmu dan pengalaman yang tidak dapat diperoleh di bangku perkuliahan.
11. Teman-teman **KKN Reguler Unhas Gelombang 102 Kecamatan Barebbo**, terkhusus kepada teman-teman **UJUNG ASPAL SAMAELO**. Terima kasih atas suka duka selama sebulan di posko dan setelahnya.
12. Serta semua pihak yang turut membantu dalam penyelesaian skripsi ini yang tidak dapat penulis ucapkan satu per satu. Terima kasih sebesar-besarnya dan semoga dapat bernilai ibadah.

Penulis menyadari bahwa dalam penulisan skripsi ini masih terdapat banyak kekurangan. Oleh karena itu, dengan segala kerendahan hati memohon maaf. Semoga skripsi ini dapat memberikan manfaat.

Makassar, 04 Januari 2021

Rayhanna Auliya Amin

ABSTRAK

Jumlah penduduk di suatu wilayah dapat dipengaruhi oleh faktor pengaruh jumlah penduduk yang ada di wilayah sekitarnya. Hal tersebut dinamakan efek spasial. Analisis statistik yang memodelkan hubungan antar variabel dan mengandung unsur spasial adalah *Geographically Weighted Regression* (GWR). Model GWR menggunakan pembobot berdasarkan lokasi setiap pengamatan sehingga model yang diperoleh berlaku hanya untuk lokasi tersebut. Penentuan pembobot bergantung pada nilai *bandwidth*. *Bandwidth* merupakan lingkaran dengan radius (h) dari titik pusat lokasi pengamatan yang digunakan sebagai dasar penentuan pembobot setiap lokasi pengamatan. Nilai *bandwidth* yang kecil akan mengakibatkan variansi yang besar sehingga model yang diperoleh *undersmoothed* dan sebaliknya. Oleh karena itu, pemilihan *bandwidth* optimum mempengaruhi ketepatan model yang dibentuk. Penelitian ini bertujuan untuk mengetahui perbandingan model GWR yang menggunakan metode *bandwidth Cross Validation* (CV) dan *Akaike Information Criterion Corrected* (AICc) pada data jumlah penduduk Kabupaten/Kota di Provinsi Sulawesi Selatan tahun 2018. Penelitian ini menggunakan fungsi pembobot *fixed Gaussian kernel*. Hasil analisis menjelaskan bahwa setiap kabupaten/kota memiliki model GWR yang berbeda-beda. Model GWR *bandwidth* CV lebih baik dalam menjelaskan data jumlah penduduk kabupaten/kota di Provinsi Sulawesi Selatan tahun 2018 karena memiliki nilai koefisien determinasi disesuaikan (R_a^2) lebih besar yaitu 98,249% dan nilai RMSE lebih kecil yaitu 26.859,385.

Kata Kunci: *Geographically Weighted Regression, Bandwidth, Cross Validation, Akaike Information Criterion Corrected, Jumlah Penduduk*

ABSTRACT

Total population of an area can be influenced by the factor of total population of the surrounding area. This is called the spatial effect. The statistical analysis that model the relationship between variables and contains spatial elements is Geographically Weighted Regression (GWR). The GWR model uses a weight based on the location of each observation so that the model obtained applies only to that location. The weighting determination depends on the bandwidth value. Bandwidth is a circle of a radius (h) from the center of the observation location which is used as the basis of determining the weight of each observation location. Small bandwidth values will result in large variances so that the model obtained is under smoothed and vice versa. Therefore, choosing the optimum bandwidth affects the accuracy of the model being formed. This study aims to determine the comparison of the GWR model using the Cross-Validation (CV) and Akaike Information Criterion Corrected (AICc) bandwidth method on total population data of districts/cities in South Sulawesi Province in 2018. This study uses a fixed Gaussian kernel weighting function. The results of the analysis show that each district/city has a different GWR model. The CV bandwidth GWR model is better at explaining the data onto total population of districts/cities in South Sulawesi Province in 2018 because it has a greater value of adjusted determination coefficient (R_a^2) is 98.249% and it has a smaller value of RMSE is 26,859.385.

Key Words: Geographically Weighted Regression, Bandwidth, Cross Validation, Akaike Information Criterion Corrected, Total Population

DAFTAR ISI

HALAMAN SAMPUL	i
HALAMAN JUDUL	ii
LEMBAR PENGESAHAN SKRIPSI	iii
PERNYATAAN KEASLIAN	iv
KATA PENGANTAR	v
ABSTRAK	viii
ABSTRACT	ix
DAFTAR ISI	x
DAFTAR GAMBAR	xii
DAFTAR TABEL	xiii
DAFTAR LAMPIRAN	xiv
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	3
1.3. Batasan Masalah	3
1.4. Tujuan Penelitian	4
1.5. Manfaat Penelitian	4
BAB II TINJAUAN PUSTAKA	5
2.1. Model Regresi Linier Berganda	5
2.2. Pendugaan Parameter Model Regresi Linier Berganda	6
2.3. Heterogenitas Spasial	6
2.4. Model <i>Geographically Weighted Regression</i>	7
2.5. Fungsi Pembobot.....	8
2.6. <i>Bandwidth</i>	8
2.6.1. <i>Golden Section Search</i>	9
2.6.2. <i>Cross Validation</i>	9
2.6.3. <i>Akaike Information Criterion Corrected</i>	10
2.7. Pendugaan Parameter Model <i>Geographically Weighted Regression</i>	11
2.8. Pengujian Kesesuaian Model <i>Geographically Weighted Regression</i>	12
2.9. Pengujian Parameter Model <i>Geographically Weighted Regression</i>	13
2.10. Pemilihan Model Terbaik	14
2.11. Jumlah Penduduk	15
BAB III METODOLOGI PENELITIAN	17
3.1. Sumber Data	17
3.2. Variabel Penelitian.....	17
3.3. Metode Analisis	17

BAB IV HASIL DAN PEMBAHASAN.....	21
4.1. Deskripsi Data.....	21
4.2. Pengujian Heterogenitas Spasial.....	25
4.3. Matriks <i>Euclidean</i>	27
4.4. Pemodelan <i>Geographically Weighted Regression</i> dengan <i>Bandwidth Cross Validation</i>	28
4.4.1. <i>Bandwidth Cross Validation</i>	28
4.4.2. Matriks Pembobot dengan Fungsi Pembobot <i>Fixed Gaussian Kernel</i> Menggunakan <i>Bandwidth Cross Validation</i>	30
4.4.3. Pendugaan Parameter Model <i>Geographically Weighted Regression</i> dengan Menggunakan <i>Bandwidth Cross Validation</i>	32
4.4.4. Pengujian Kesesuaian Model <i>Geographically Weighted Regression</i> yang Menggunakan <i>Bandwidth Cross Validation</i>	33
4.4.5. Pengujian Parameter Model <i>Geographically Weighted Regression</i> yang Menggunakan <i>Bandwidth Cross Validation</i>	34
4.5. Pemodelan <i>Geographically Weighted Regression</i> dengan <i>Bandwidth Akaike Information Criterion Corrected</i>	35
4.5.1. <i>Bandwidth Akaike Information Criterion Corrected</i>	36
4.5.2. Matriks Pembobot dengan Fungsi Pembobot <i>Fixed Gaussian Kernel</i> Menggunakan <i>Bandwidth Akaike Information Criterion Corrected</i>	37
4.5.3. Pendugaan Parameter Model <i>Geographically Weighted Regression</i> dengan Menggunakan <i>Bandwidth Akaike Information Criterion Corrected</i>	40
4.5.4. Pengujian Kesesuaian Model <i>Geographically Weighted Regression</i> yang Menggunakan <i>Bandwidth Akaike Information Criterion Corrected</i>	41
4.5.5. Pengujian Parameter Model <i>Geographically Weighted Regression</i> yang Menggunakan <i>Bandwidth Akaike Information Criterion Corrected</i>	42
4.6. Pemilihan Model Terbaik	43
4.7. Peta Wilayah Variabel yang Signifikan	44
BAB V KESIMPULAN DAN SARAN	49
5.1. Kesimpulan	49
5.2. Saran.....	49
DAFTAR PUSTAKA	50
LAMPIRAN	52

DAFTAR GAMBAR

Gambar 4.1. Grafik Data Jumlah Penduduk (y) Kabupaten/Kota di Provinsi Sulawesi Selatan Tahun 2018.....	21
Gambar 4.2. Grafik Data Jumlah Fasilitas Kesehatan (x_1) Kabupaten/Kota di Provinsi Sulawesi Selatan Tahun 2018	22
Gambar 4.3. Grafik Data Jumlah Tenaga Kesehatan (x_2) Kabupaten/Kota di Provinsi Sulawesi Selatan Tahun 2018	22
Gambar 4.4. Grafik Data Jumlah Balita Kurang Gizi (x_3) Kabupaten/Kota di Provinsi Sulawesi Selatan Tahun 2018	23
Gambar 4.5. Grafik Data Jumlah Klinik KB & Pos Pelayanan KB Desa (x_4) Kabupaten/Kota di Provinsi Sulawesi Selatan Tahun 2018	23
Gambar 4.6. Grafik Data Jumlah Kasus Penyakit (x_5) Kabupaten/Kota di Provinsi Sulawesi Selatan Tahun 2018	24
Gambar 4.7. Peta Wilayah Variabel Signifikan pada Model GWR yang Menggunakan <i>Bandwidth CV</i>	45
Gambar 4.8. Peta Wilayah Variabel Signifikan pada Model GWR yang Menggunakan <i>Bandwidth AICc</i>	47

DAFTAR TABEL

Tabel 4.1.	Statistik Deskriptif Data Jumlah Penduduk Kabupaten/Kota di Provinsi Sulawesi Selatan Tahun 2018	25
Tabel 4.2.	Jarak <i>Euclidean</i> Kabupaten Kepulauan Selayar, Kabupaten Luwu Utara, dan Kota Makassar Terhadap Lokasi Pengamatan Lain	27
Tabel 4.3.	Nilai <i>Bandwidth</i> dengan Metode CV	29
Tabel 4.4.	Pembobot <i>Fixed Gaussian Kernel</i> Menggunakan <i>Bandwidth CV</i>	31
Tabel 4.5.	Pendugaan Parameter Model GWR yang Menggunakan <i>Bandwidth CV</i>	32
Tabel 4.6.	Hasil pengujian Kesesuaian Model GWR yang Menggunakan <i>Bandwidth CV</i>	34
Tabel 4.7.	Hasil Pengujian Parameter Model GWR yang Menggunakan <i>Bandwidth CV</i>	35
Tabel 4.8.	Nilai <i>Bandwidth</i> dengan Metode AICc	37
Tabel 4.9.	Pembobot <i>Fixed Gaussian Kernel</i> Menggunakan <i>Bandwidth AICc</i>	38
Tabel 4.10.	Pendugaan Parameter Model GWR yang Menggunakan <i>Bandwidth AICc</i>	40
Tabel 4.11.	Hasil pengujian Kesesuaian Model GWR yang Menggunakan <i>Bandwidth AICc</i>	42
Tabel 4.12.	Hasil Pengujian Parameter Model GWR yang Menggunakan <i>Bandwidth AICc</i>	43
Tabel 4.13.	Perbandingan Nilai R_a^2 dan RMSE Model GWR <i>Bandwidth CV</i> dan AICc	43
Tabel 4.14.	Pengelompokan Wilayah Berdasarkan Variabel Signifikan pada Model GWR yang Menggunakan <i>Bandwidth CV</i>	46
Tabel 4.15.	Pengelompokan Wilayah Berdasarkan Variabel Signifikan pada Model GWR yang Menggunakan <i>Bandwidth AICc</i>	48

DAFTAR LAMPIRAN

Lampiran 1. Data Jumlah Penduduk Kabupaten/Kota di Provinsi Sulawesi Selatan Tahun 2018	53
Lampiran 2. Letak Astronomis Kabupaten/Kota di Provinsi Sulawesi Selatan	54
Lampiran 3. Matriks <i>Euclidean</i>	54
Lampiran 4. Matriks Pembobot <i>Fixed Gaussian Kernel Bandwidth CV</i>	57
Lampiran 5. Hasil Pendugaan Parameter Model GWR <i>Bandwidth CV</i>	60
Lampiran 6. Model GWR <i>Bandwidth CV</i> untuk Setiap Kabupaten/Kota	61
Lampiran 7. Hasil Pengujian Parameter Model GWR <i>Bandwidth CV</i>	63
Lampiran 8. Matriks Pembobot <i>Fixed Gaussian Kernel Bandwidth AICc</i>	64
Lampiran 9. Hasil Pendugaan Parameter Model GWR <i>Bandwidth AICc</i>	67
Lampiran 10. Model GWR <i>Bandwidth AICc</i> untuk Setiap Kabupaten/Kota	68
Lampiran 11. Hasil Pengujian Parameter Model GWR <i>Bandwidth AICc</i>	70

BAB I PENDAHULUAN

1.1. Latar Belakang

Jumlah penduduk di setiap daerah sangat beragam dengan laju pertumbuhan yang beragam pula. Provinsi Sulawesi Selatan merupakan salah satu provinsi dengan jumlah penduduk terbanyak. Jumlah penduduk di Sulawesi Selatan mengalami peningkatan setiap tahunnya. Pada tahun 2015 jumlah penduduk Sulawesi Selatan sebanyak 8.520.304 jiwa, lalu tahun 2018 meningkat dengan laju pertumbuhan penduduk 0,94% menjadi 8.771.970 jiwa. Jumlah penduduk tertinggi di Provinsi Sulawesi Selatan berada di Kota Makassar, yaitu sebanyak 1.508.154 jiwa dan jumlah penduduk terendah berada di Kabupaten Kepulauan Selayar, yaitu sebanyak 134.280 jiwa. Laju pertumbuhan penduduk yang relatif tinggi akan membawa dampak terhadap pembangunan, termasuk dalam penentuan kebijakan kependudukan. Jumlah penduduk yang relatif tinggi menyebabkan penentuan kebijakan harus mempertimbangkan banyak hal, misalnya penyediaan sarana dan prasarana untuk menunjang implementasi kebijakan di bidang kependudukan tersebut (Badan Pusat Statistik, 2019).

Salah satu cara yang dapat dilakukan agar dapat mengetahui penentuan kebijakan kependudukan yang tepat berdasarkan jumlah penduduk adalah dengan memodelkan jumlah penduduk. Analisis statistik yang digunakan untuk mengetahui pengaruh suatu variabel terhadap variabel lain adalah analisis regresi linier. Variabel respon adalah variabel yang dipengaruhi dan variabel prediktor adalah variabel yang mempengaruhi. Analisis regresi linier mengandung asumsi bahwa variabel respon tidak dipengaruhi oleh unsur spasial. Namun, suatu wilayah yang berdekatan memiliki keterkaitan satu sama lain, seperti jumlah penduduk yakni jumlah penduduk suatu wilayah dipengaruhi oleh faktor pengaruh jumlah penduduk yang berada di wilayah sekitarnya. Hal ini sesuai dengan teori Tobler (1970) yang dikenal dengan hukum geografi pertama bahwa *“Everything is related to everything else, but near things are more related than distant things.”*

(Fotheringham & Charlton, 2009). Teori tersebut mengandung makna suatu hal pasti memiliki kaitan dengan hal lain, tapi hal-hal yang jaraknya berdekatan lebih erat kaitannya dibandingkan hal-hal yang jaraknya berjauhan. Hal inilah yang dinamakan efek spasial.

Analisis statistik yang memodelkan hubungan antar variabel dan mengandung unsur spasial adalah *Geographically Weighted Regression (GWR)*. GWR merupakan pengembangan dari analisis regresi linier yang bersifat lokal untuk setiap wilayah pengamatan. Pada analisis regresi linier, model yang diperoleh berlaku untuk seluruh lokasi pengamatan, sedangkan GWR menggunakan pembobotan berdasarkan lokasi setiap pengamatan sehingga model yang diperoleh berlaku hanya untuk lokasi tersebut (Saefuddin, dkk, 2011). Pendugaan parameter model GWR membutuhkan fungsi pembobot berdasarkan lokasi setiap pengamatan. Fungsi pembobot tersebut tergantung pada nilai *bandwidth*. *Bandwidth* merupakan lingkaran dengan radius (h) dari titik pusat lokasi pengamatan yang digunakan sebagai dasar penentuan pembobot setiap lokasi pengamatan. Oleh karena itu, metode pemilihan *bandwidth* penting dalam menentukan fungsi pembobot. Menurut Fotheringham, *et. al.* (2002), terdapat beberapa metode dalam pemilihan *bandwidth* optimum, di antaranya, metode *Cross Validation (CV)* dan *Akaike Information Criterion Corrected (AICc)*.

Metode CV melibatkan konsep dua kelompok data yaitu, data validasi dan data keseluruhan. Kelompok data validasi diambil dengan cara menghilangkan nilai titik data ke- i untuk dibandingkan dengan kelompok data keseluruhan yang dilakukan berulang sampai diperoleh nilai CV minimum. Sedangkan, metode AICc merupakan pengembangan dari *Akaike Information Criterion (AIC)* yang mengoreksi bias pada AIC dan baik digunakan pada sampel yang ukurannya kecil.

Guo, *et.al.* (2008) melakukan penelitian mengenai model GWR dengan *bandwidth* metode CV dan AIC pada pembobot *fixed* dan *adaptive Gaussian kernel* serta *fixed* dan *adaptive bisquare kernel* yang menghasilkan kesimpulan bahwa penggunaan pembobot *fixed kernel* menghasilkan model yang lebih baik dibandingkan dengan penggunaan pembobot *adaptive kernel*. Rosa (2015) dalam penelitian mengenai model GWR dengan *bandwidth* metode CV pada pembobot

fixed Gaussian kernel dan *fixed bisquare kernel* menjelaskan bahwa pembobot *fixed Gaussian kernel* memberikan hasil yang lebih baik dibandingkan pembobot *fixed bisquare kernel*. Ariyanto (2017) melakukan perbandingan *bandwidth* dengan metode CV dan AIC pada fungsi pembobot *adaptive Gaussian kernel* dan menghasilkan kesimpulan bahwa model GWR dengan *bandwidth CV* memberikan model yang lebih baik dibandingkan model GWR dengan *bandwidth AIC*.

Penelitian – penelitian sebelumnya menggunakan metode AIC ataupun CV dalam menentukan nilai *bandwidth*, sehingga dalam penelitian ini penulis tertarik untuk menggunakan metode AICc dan CV dalam menentukan nilai *bandwidth*. Berdasarkan uraian tersebut, maka penulis akan melakukan penelitian tugas akhir mengenai performa *bandwidth CV* dan AICc dalam model GWR pada data jumlah penduduk kabupaten/kota di Provinsi Sulawesi Selatan tahun 2018.

1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan, maka rumusan masalah dalam penelitian adalah sebagai berikut:

1. Bagaimana model GWR dengan *bandwidth CV* dan AICc pada data jumlah penduduk kabupaten/kota di Provinsi Sulawesi Selatan tahun 2018?
2. Bagaimana perbandingan model GWR yang dihasilkan dengan *bandwidth CV* dan AICc pada data jumlah penduduk kabupaten/kota di Provinsi Sulawesi Selatan tahun 2018?

1.3. Batasan Masalah

Penelitian ini difokuskan pada model GWR dengan *bandwidth CV* dan AICc. Penelitian ini menggunakan data jumlah penduduk kabupaten/kota di Provinsi Sulawesi Selatan tahun 2018. Adapun fungsi pembobot yang digunakan dalam penelitian ini adalah fungsi pembobot *fixed Gaussian kernel*.

1.4. Tujuan Penelitian

Berdasarkan rumusan masalah yang telah ditentukan, maka tujuan penelitian adalah sebagai berikut:

1. Memperoleh model GWR dengan *bandwidth* CV dan AICc pada data jumlah penduduk kabupaten/kota di Provinsi Sulawesi Selatan tahun 2018.
2. Mendapatkan hasil perbandingan antara model GWR dengan *bandwidth* CV dan AICc pada data jumlah penduduk kabupaten/kota di Provinsi Sulawesi Selatan tahun 2018.

1.5. Manfaat Penelitian

Adapun manfaat yang dapat diambil dari penelitian ini yaitu dapat memberikan gambaran tentang model GWR dengan *bandwidth* CV dan AICc. Selain itu, dapat memberikan manfaat berupa tambahan kepustakaan bagi pengguna ilmu statistika.

BAB II

TINJAUAN PUSTAKA

2.1. Model Regresi Linier Berganda

Model regresi adalah aplikasi dari model linier dengan variabel respon (y) diidentifikasi dari nilai numerik oleh satu atau lebih variabel kuantitatif yang disebut variabel prediktor (x). Regresi linier berganda merupakan pengembangan dari regresi linier sederhana dengan variabel prediktor (x) lebih dari satu. Analisis regresi linier berganda digunakan untuk menganalisis hubungan antara variabel respon (y) dan variabel prediktor (x). Bentuk umum dari model regresi linier berganda adalah sebagai berikut:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \cdots + \beta_p x_{ip} + \varepsilon_i, \quad (2.1)$$

dengan

$i = 1, 2, \dots, n$ (jumlah pengamatan)

$k = 1, 2, \dots, p$ (jumlah variabel prediktor),

serta

y_i : nilai variabel respon pada pengamatan ke- i

x_{ik} : nilai variabel prediktor ke- k pada pengamatan ke- i

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

β_k : parameter regresi ke- k

ε_i : sisaan pada pengamatan ke- i

Jika didefinisikan matriks \mathbf{Y} , \mathbf{X} , $\boldsymbol{\beta}$, dan $\boldsymbol{\varepsilon}$ sebagai berikut:

$$\mathbf{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \mathbf{X} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}, \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}, \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}.$$

Maka Persamaan (2.1) dapat dituliskan menjadi persamaan sebagai berikut (Freund, *et. al.*, 2006):

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (2.2)$$

2.2. Pendugaan Parameter Model Regresi Linier Berganda

Pendugaan parameter dalam model regresi linier berganda adalah metode *Ordinary Least Square* (OLS). Tujuan dari metode OLS adalah untuk meminimumkan fungsi jumlah kuadrat residual (Weisberg, 2005). Berdasarkan Persamaan (2.2), maka didapatkan persamaan sebagai berikut:

$$\boldsymbol{\varepsilon} = \mathbf{Y} - \mathbf{X}\boldsymbol{\beta} . \quad (2.3)$$

Karena tujuan dari metode OLS adalah untuk meminimumkan jumlah kuadrat residual, maka berdasarkan Persamaan (2.3):

$$\begin{aligned} \boldsymbol{\varepsilon}^2 &= \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \\ &= \mathbf{Y}'\mathbf{Y} - 2\boldsymbol{\beta}'\mathbf{X}'\mathbf{Y} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta} . \end{aligned}$$

Kemudian $\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}$ diturunkan terhadap $\boldsymbol{\beta}$ dan disamakan dengan 0, maka:

$$\begin{aligned} \left. \frac{\partial(\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon})}{\partial\boldsymbol{\beta}} \right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} &= 0 \\ \frac{\partial(\mathbf{Y}'\mathbf{Y} - 2\boldsymbol{\beta}'\mathbf{X}'\mathbf{Y} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta})}{\partial\boldsymbol{\beta}} &= 0 \\ -2\mathbf{X}'\mathbf{Y} + 2\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} &= \mathbf{0} \\ \mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} &= \mathbf{X}'\mathbf{Y} . \end{aligned}$$

Sehingga diperoleh penduga dengan metode OLS sebagai berikut:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} . \quad (2.4)$$

2.3. Heterogenitas Spasial

Heterogenitas spasial adalah suatu keadaan ketika satu variabel prediktor memberikan respon yang berbeda pada lokasi yang berbeda dalam suatu wilayah penelitian. Karakteristik data spasial dengan pendekatan titik adalah adanya heterogenitas spasial atau keragaman variansi yang terdapat di setiap lokasi. Pengujian heterogenitas spasial dilakukan menggunakan statistik uji *Breusch-Pagan* (BP) dengan hipotesis sebagai berikut:

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

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

Statistik uji:

$$BP = \left(\frac{1}{2}\right) \mathbf{f}' \mathbf{Z} (\mathbf{Z}' \mathbf{Z})^{-1} \mathbf{Z}' \mathbf{f}, \quad (2.5)$$

dan elemen vektor \mathbf{f} :

$$f_i = \left(\frac{e_i^2}{\sigma^2} - 1\right), \quad (2.6)$$

dengan

\mathbf{f} : matriks berukuran $n \times 1$

e_i^2 : kuadrat sisaan untuk pengamatan ke- i

σ^2 : varians sisaan e_i

\mathbf{Z} : matriks berukuran $n \times (p + 1)$ berisi vektor standar (\mathbf{z}) untuk setiap lokasi

Kriteria keputusan:

H_0 ditolak jika $BP \geq \chi_{(\alpha;p)}^2$ artinya terdapat heterogenitas spasial dan H_0 diterima jika $BP < \chi_{(\alpha;p)}^2$ artinya tidak terdapat heterogenitas spasial (Anselin, 2009).

2.4. Model Geographically Weighted Regression

Model *Geographically Weighted Regression* (GWR) adalah model yang dikembangkan dari model regresi linier menjadi model regresi terboboti geografis. GWR menghasilkan pendugaan parameter yang hanya dapat digunakan untuk memprediksi setiap titik atau lokasi data tersebut diamati dan dikumpulkan. Model GWR merupakan pengembangan regresi linier dengan memperhitungkan lokasi data penelitian yang diamati. Secara matematis, model GWR dapat dituliskan sebagai berikut (Fotheringham, *et.al.*, 2002):

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^p \beta_k(u_i, v_i) x_{ik} + \varepsilon_i ; i = 1, 2, \dots, n, \quad (2.7)$$

dengan

y_i : nilai variabel respon pada pengamatan ke- i

x_{ik} : nilai variabel prediktor ke- k pada pengamatan ke- i

$\beta_0(u_i, v_i)$: nilai intersep model GWR untuk setiap lokasi ke- i

$\beta_k(u_i, v_i)$: parameter regresi ke- k untuk setiap lokasi ke- i

(u_i, v_i) : titik koordinat (lintang, bujur) pada lokasi ke- i

ε_i : sisaan pada pengamatan ke- i

2.5. Fungsi Pembobot

Fungsi pembobot dalam GWR digunakan untuk memberikan hasil penaksiran parameter yang berbeda untuk tiap lokasi pengamatan. Secara umum, diasumsikan bahwa penaksiran parameter titik spasial (u_i, v_i) dipengaruhi oleh titik-titik yang saling berdekatan lokasinya dibanding lokasi yang jauh dari titik tersebut (Chasco, *et.al.*, 2007).

Fungsi pembobot yang digunakan dalam penelitian ini adalah fungsi *fixed Gaussian kernel*, yaitu sebagai berikut:

$$w_{ij} = \exp \left[-\frac{1}{2} \left(\frac{d_{ij}}{h} \right)^2 \right], \quad (2.8)$$

dengan h adalah *bandwidth* dan d_{ij} adalah jarak *euclidean* antara lokasi (u_i, v_i) ke lokasi (u_j, v_j) yang diperoleh dari persamaan berikut:

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

2.6. Bandwidth

Definisi *bandwidth* secara teoritis adalah lingkaran dengan radius (h) dari titik pusat lokasi pengamatan yang digunakan sebagai dasar penentuan pembobot untuk setiap lokasi pengamatan. Pemilihan *bandwidth* memiliki dampak yang besar dalam GWR. Titik-titik lokasi pengamatan yang berada dalam lingkaran dengan radius (h) pada titik lokasi pengamatan ke- i memiliki berpengaruh dalam membentuk parameter di titik lokasi pengamatan ke- i . *Bandwidth* dianggap sebagai parameter penghalus (*smoothing*). Nilai *bandwidth* yang kecil akan mengakibatkan variansi yang dihasilkan semakin besar sehingga model yang diperoleh *undersmoothed*. Nilai *bandwidth* yang besar mengakibatkan bias semakin besar sehingga model yang diperoleh *oversmoothed*. Model yang *undersmoothed* akan menghasilkan parameter dengan banyak variasi lokal sehingga sulit untuk melihat pola yang terbentuk. Sedangkan, model yang *oversmoothed* akan menghasilkan parameter yang nilainya hampir sama di seluruh lokasi pengamatan. Sehingga untuk menghindari kedua hal tersebut diperlukan metode untuk menghasilkan nilai *bandwidth* optimum. Metode yang digunakan

untuk menentukan *bandwidth* optimum adalah *Cross Validation* (CV) dan *Akaike Information Criterion Corrected* (AICc).

2.6.1. Golden Section Search

Proses untuk mendapatkan *bandwidth* optimum dilakukan dengan menggunakan teknik *Golden section search*. Teknik ini dilakukan dengan mengevaluasi fungsi dengan 3 nilai berbeda. Misalkan nilai a , b , dan c , dengan $a < b < c$. Nilai fungsi pada ketiga titik ini adalah $f(a)$, $f(b)$, dan $f(c)$ yang disebut sebagai triplet. Fungsi ini kemudian dievaluasi lagi pada nilai baru, d , yang dapat dipilih di antara a dan b atau b dan c sehingga menghasilkan nilai fungsi baru, yaitu $f(d)$ dengan aturan sebagai berikut:

Jika $f(b) < f(d)$, maka triplet baru adalah $a < b < d$

Jika $f(b) > f(d)$, maka triplet baru adalah $b < d < c$

Nilai a diperoleh dari nilai minimum d_{ij} , nilai c diperoleh dari nilai maksimum d_{ij} , dan nilai b dipilih sedemikian rupa sehingga nilainya 0.382 kali jarak antara a dan c . Proses iterasi pemilihan nilai d dilakukan hingga dua nilai $f(d)$ yang dihasilkan mendekati sama atau memiliki selisih yang sangat kecil (Fotheringham, *et. al.*, 2002).

2.6.2. Cross Validation

Proses *cross validation* (CV) sendiri melibatkan konsep dua kelompok data, yaitu kelompok data validasi (*test data*) dan kelompok data rill (*training data*) yang diambil dari populasi yang sama. CV jika diterapkan ke dalam konsep *bandwidth*, maka kelompok data validasi (*test data*) diambil dengan cara menghilangkan nilai titik data ke- i untuk dibandingkan dengan kelompok data keseluruhan (*training data*) dengan cara iterasi sampai diperoleh nilai CV minimum. Nilai *bandwidth* dikatakan optimum ketika nilai CV yang dihasilkan minimum. Secara matematis, CV dapat dituliskan sebagai berikut (Fotheringham, *et. al.*, 2002):

$$CV = \sum_{i=1}^n (y_i - \hat{y}_{\neq i}(h))^2, \quad (2.10)$$

dengan

- $\hat{y}_{\neq i}(h)$: penduga y_i pada lokasi pengamatan (u_i, v_i) dihilangkan dari proses Pendugaan
 y_i : nilai variabel respon pada pengamatan ke- i
 n : jumlah pengamatan

2.6.3. Akaike Information Criterion Corrected

Akaike Information Criterion (AIC) merupakan metode untuk membandingkan berbagai model yang mungkin dan menentukan model yang paling cocok untuk data. Metode *Akaike Information Criterion Corrected* (AICc) merupakan pengembangan dari *Akaike Information Criterion* (AIC). Metode AICc adalah penyesuaian dari AIC yang digunakan ketika sampel berukuran kecil dengan ketentuan $k > \frac{n}{40}$ (Portet, 2020). Metode pemilihan *bandwidth* dengan AICc dilakukan secara iterasi dengan mengevaluasi nilai AICc terkecil pada interval jarak minimum dan maksimum lokasi pengamatan sehingga diperoleh nilai AICc minimum. Nilai *bandwidth* dikatakan optimum ketika nilai AICc yang dihasilkan minimum. Secara matematis, AICc dapat dituliskan sebagai berikut (Fotheringham, *et. al.*, 2002):

$$AIC_c = 2n \ln(\hat{\sigma}) + n \ln(2\pi) + n \left\{ \frac{n + \text{tr}(\mathbf{L})}{n - 2 - \text{tr}(\mathbf{L})} \right\}, \quad (2.11)$$

dengan

$$\hat{\sigma} = \sqrt{\frac{\mathbf{y}'(\mathbf{I}-\mathbf{L})'(\mathbf{I}-\mathbf{L})\mathbf{y}}{n}}$$

$$\pi = 3.14$$

$$\mathbf{L}_{(n \times n)} = \begin{pmatrix} \mathbf{x}'_1[\mathbf{X}'\mathbf{W}(u_1, v_1)\mathbf{X}]^{-1}\mathbf{X}'\mathbf{W}(u_1, v_1) \\ \mathbf{x}'_2[\mathbf{X}'\mathbf{W}(u_2, v_2)\mathbf{X}]^{-1}\mathbf{X}'\mathbf{W}(u_2, v_2) \\ \vdots \\ \mathbf{x}'_n[\mathbf{X}'\mathbf{W}(u_n, v_n)\mathbf{X}]^{-1}\mathbf{X}'\mathbf{W}(u_n, v_n) \end{pmatrix}$$

serta

- \mathbf{I} : matriks identitas
 n : jumlah pengamatan

2.7. Pendugaan Parameter Model *Geographically Weighted Regression*

Pendugaan parameter dalam model GWR menggunakan metode *Weighted Least Square* (WLS). Pada model GWR pembobot di setiap lokasi pengamatan (u_i, v_i) mewakili jarak dari lokasi pengamatan (u_i, v_i) ke lokasi pengamatan lain. Misalkan pembobot pada lokasi pengamatan (u_i, v_i) adalah $w_i(u_i, v_i)$; $i = 1, 2, \dots, n$. Maka parameter pada lokasi pengamatan (u_i, v_i) diduga dengan meminimumkan:

$$\sum_{i=1}^n W_{ij}(u_i, v_i) \varepsilon_i^2 = \sum_{i=1}^n W_{ij}(u_i, v_i) \left[y_i - \beta_0(u_i, v_i) - \sum_{k=1}^p \beta_k(u_i, v_i) x_{ik} \right]^2 \quad (2.12)$$

dan dalam bentuk matriks jumlah kuadrat residual adalah sebagai berikut:

$$\begin{aligned} \varepsilon' \mathbf{W}(u_i, v_i) \varepsilon &= (\mathbf{y} - \boldsymbol{\beta}(u_i, v_i) \mathbf{X})' \mathbf{W}(u_i, v_i) (\mathbf{y} - \boldsymbol{\beta}(u_i, v_i) \mathbf{X}) \\ &= (\mathbf{y}' - \boldsymbol{\beta}'(u_i, v_i) \mathbf{X}') \mathbf{W}(u_i, v_i) (\mathbf{y} - \boldsymbol{\beta}(u_i, v_i) \mathbf{X}) \\ &= \mathbf{y}' \mathbf{W}(u_i, v_i) \mathbf{y} - \mathbf{y}' \mathbf{W}(u_i, v_i) \boldsymbol{\beta}(u_i, v_i) \mathbf{X} - \boldsymbol{\beta}'(u_i, v_i) \mathbf{X}' \mathbf{W}(u_i, v_i) \mathbf{y} \\ &\quad + \boldsymbol{\beta}'(u_i, v_i) \mathbf{X}' \mathbf{W}(u_i, v_i) \mathbf{X} \boldsymbol{\beta}(u_i, v_i) \\ &= \mathbf{y}' \mathbf{W}(u_i, v_i) \mathbf{y} - (\mathbf{y}' \mathbf{W}(u_i, v_i) \boldsymbol{\beta}(u_i, v_i) \mathbf{X})' \\ &\quad - \boldsymbol{\beta}'(u_i, v_i) \mathbf{X}' \mathbf{W}(u_i, v_i) \mathbf{y} + \boldsymbol{\beta}'(u_i, v_i) \mathbf{X}' \mathbf{W}(u_i, v_i) \mathbf{X} \boldsymbol{\beta}(u_i, v_i) \\ &= \mathbf{y}' \mathbf{W}(u_i, v_i) \mathbf{y} - \boldsymbol{\beta}'(u_i, v_i) \mathbf{X}' \mathbf{W}(u_i, v_i) \mathbf{y} \\ &\quad - \boldsymbol{\beta}'(u_i, v_i) \mathbf{X}' \mathbf{W}(u_i, v_i) \mathbf{y} + \boldsymbol{\beta}'(u_i, v_i) \mathbf{X}' \mathbf{W}(u_i, v_i) \mathbf{X} \boldsymbol{\beta}(u_i, v_i) \\ &= \mathbf{y}' \mathbf{W}(u_i, v_i) \mathbf{y} - 2 \boldsymbol{\beta}'(u_i, v_i) \mathbf{X}' \mathbf{W}(u_i, v_i) \mathbf{y} \\ &\quad + \boldsymbol{\beta}'(u_i, v_i) \mathbf{X}' \mathbf{W}(u_i, v_i) \mathbf{X} \boldsymbol{\beta}(u_i, v_i) \end{aligned}$$

Kemudian diturunkan terhadap $\boldsymbol{\beta}(u_i, v_i)$ dan disamakan dengan nol, seperti berikut:

$$\begin{aligned} \frac{\partial \varepsilon' \mathbf{W} \varepsilon}{\partial \boldsymbol{\beta}(u_i, v_i)} &= 0 \\ -2 \mathbf{X}' \mathbf{W}(u_i, v_i) \mathbf{y} + 2 \mathbf{X}' \mathbf{W}(u_i, v_i) \mathbf{X} \boldsymbol{\beta}(u_i, v_i) &= 0 \\ \mathbf{X}' \mathbf{W}(u_i, v_i) \mathbf{X} \boldsymbol{\beta}(u_i, v_i) &= \mathbf{X}' \mathbf{W}(u_i, v_i) \mathbf{y} \\ \boldsymbol{\beta}(u_i, v_i) &= [\mathbf{X}' \mathbf{W}(u_i, v_i) \mathbf{X}]^{-1} \mathbf{X}' \mathbf{W}(u_i, v_i) \mathbf{y} \end{aligned}$$

Sehingga pendugaan parameter untuk model GWR adalah sebagai berikut:

$$\boldsymbol{\beta}(u_i, v_i) = [\mathbf{X}' \mathbf{W}(u_i, v_i) \mathbf{X}]^{-1} \mathbf{X}' \mathbf{W}(u_i, v_i) \mathbf{y} \quad (2.13)$$

dengan

$$\mathbf{X} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}, \mathbf{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix},$$

dan $\mathbf{W}(u_i, v_i) = \text{Diag}[w_1(u_i, v_i), w_2(u_i, v_i), \dots, w_n(u_i, v_i)]$.

Misalkan $\mathbf{x}'_i = (1 \ x_{i1} \ x_{i2} \ \cdots \ x_{ip})$ adalah elemen baris ke- i dari \mathbf{X} . Maka nilai dugaan untuk y pada lokasi pengamatan (u_i, v_i) diperoleh dari persamaan berikut:

$$\hat{y}_i = \mathbf{x}'_i \hat{\beta}(u_i, v_i) = \mathbf{x}'_i [\mathbf{X}' \mathbf{W}(u_i, v_i) \mathbf{X}]^{-1} \mathbf{X}' \mathbf{W}(u_i, v_i) \mathbf{Y} \quad (2.14)$$

dengan $\hat{\mathbf{Y}} = (\hat{y}_1 \ \hat{y}_2 \ \cdots \ \hat{y}_n)'$ dan $\hat{\boldsymbol{\varepsilon}} = (\hat{\varepsilon}_1 \ \hat{\varepsilon}_2 \ \cdots \ \hat{\varepsilon}_n)'$ adalah vektor dari nilai dugaan dari y dan vektor dari residual pada n lokasi pengamatan $(u_i, v_i) = 1, 2, \dots, n$ (Mei, 2005). Sehingga:

$$\begin{cases} \hat{\mathbf{Y}} = \mathbf{L} \mathbf{Y} \\ \hat{\boldsymbol{\varepsilon}} = \mathbf{Y} - \hat{\mathbf{Y}} = (\mathbf{I} - \mathbf{L}) \mathbf{Y} \end{cases} \quad (2.15)$$

dengan

$$\mathbf{L}_{(n \times n)} = \begin{pmatrix} \mathbf{x}'_1 [\mathbf{X}' \mathbf{W}(u_1, v_1) \mathbf{X}]^{-1} \mathbf{X}' \mathbf{W}(u_1, v_1) \\ \mathbf{x}'_2 [\mathbf{X}' \mathbf{W}(u_2, v_2) \mathbf{X}]^{-1} \mathbf{X}' \mathbf{W}(u_2, v_2) \\ \vdots \\ \mathbf{x}'_n [\mathbf{X}' \mathbf{W}(u_n, v_n) \mathbf{X}]^{-1} \mathbf{X}' \mathbf{W}(u_n, v_n) \end{pmatrix}.$$

2.8. Pengujian Kesesuaian Model *Geographically Weighted Regression*

Pengujian kesesuaian model GWR dilakukan untuk menguji pengaruh fungsi pembobot $W(u_i, v_i)$ atau dengan kata lain untuk menguji pengaruh faktor geografis pada model GWR. Bentuk hipotesis pengujian sebagai berikut:

$$H_0: \beta_1(u_i, v_i) = \beta_2(u_i, v_i) = \cdots = \beta_p(u_i, v_i) = \beta_k; i = 1, 2, \dots, n$$

(Tidak ada pengaruh faktor geografis pada model)

$$H_1: \text{Ada } \beta_k(u_i, v_i) \neq \beta_k; k = 1, 2, \dots, p \text{ dan } i = 1, 2, \dots, n$$

(Ada pengaruh faktor geografis pada model)

Statistik uji:

$$F_{hitung} = \frac{SSE(H_0) - SSE(H_1) / \tau_1}{SSE(H_0) / df_2}, \quad (2.16)$$

dengan

$$SSE(H_1) = \mathbf{Y}' (\mathbf{I} - \mathbf{L})' (\mathbf{I} - \mathbf{L}) \mathbf{Y}$$

$$SSE(H_0) = Y'(I - H)Y$$

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

$$\tau_i = tr([(I - H) - (I - L)'(I - L)]^i), i = 1, 2$$

$$df_1 = \frac{\tau_1^2}{\tau_2}$$

$$df_2 = n - p - 1$$

$$L_{(n \times n)} = \begin{pmatrix} x'_1 [X'W(u_1, v_1)X]^{-1} X'W(u_1, v_1) \\ x'_2 [X'W(u_2, v_2)X]^{-1} X'W(u_2, v_2) \\ \vdots \\ x'_n [X'W(u_n, v_n)X]^{-1} X'W(u_n, v_n) \end{pmatrix}$$

Kriteria keputusan:

H_0 ditolak jika $F_{hitung} \geq F_{tabel(\alpha; df_1; df_2)}$ artinya terdapat pengaruh faktor geografis pada model GWR dan H_0 diterima jika $F_{hitung} < F_{tabel(\alpha; df_1; df_2)}$ artinya tidak terdapat pengaruh faktor geografis pada model GWR (Leung, *et. al.*, 2000).

2.9. Pengujian Parameter Model Geographically Weighted Regression

Pengujian parameter model GWR bertujuan untuk mengetahui parameter yang secara parsial signifikan terhadap model GWR. Hipotesis pengujian parameter model adalah sebagai berikut:

$$H_0 : \beta_k(u_i, v_i) = 0; k = 0, 1, 2, \dots, p; i = 1, 2, \dots, n$$

(parameter $\beta_k(u_i, v_i)$ tidak signifikan terhadap model GWR)

$$H_1 : \beta_k(u_i, v_i) \neq 0$$

(parameter $\beta_k(u_i, v_i)$ signifikan terhadap model GWR)

Statistik uji:

$$t_{hitung} = \frac{\hat{\beta}_k(u_i, v_i)}{\hat{\sigma} \sqrt{c_{kk}}}, \quad (2.17)$$

dengan

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

$$\hat{\sigma} = \sqrt{\frac{SSE(H_1)}{\delta_1}}$$

$$\delta_i = tr([(I - L)'(I - L)]^i), i = 1,2$$

$$df = \frac{\delta_1^2}{\delta_2}$$

c_{kk} : elemen diagonal ke- k dari matriks CC'

Kriteria keputusan:

H_0 ditolak jika $|t_{hitung}| \geq t_{(\frac{\alpha}{2}, df)}$ artinya parameter $\beta_k(u_i, v_i)$ berpengaruh terhadap model GWR dan H_0 diterima jika $|t_{hitung}| < t_{(\frac{\alpha}{2}, df)}$ artinya parameter $\beta_k(u_i, v_i)$ tidak berpengaruh terhadap model GWR (Purhadi dan Yasin, 2012).

2.10. Pemilihan Model Terbaik

Kualitas suatu model regresi dapat dilihat dengan menggunakan nilai koefisien determinasi (R^2). Nilai R^2 menunjukkan kemampuan model dalam menjelaskan variabilitas data. Nilai R^2 yang semakin tinggi menyatakan model tersebut memiliki kualitas yang semakin baik. Secara sistematis nilai R^2 dituliskan sebagai berikut (Gujarati, 1993):

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad (2.18)$$

dengan:

n : jumlah pengamatan

\hat{y}_i : penduga y_i

y_i : nilai variabel respon pada pengamatan ke- i

\bar{y} : rata-rata pengamatan ke- i

Namun, setiap penambahan variabel respon akan menyebabkan terjadinya kenaikan nilai R^2 yang berlebihan sehingga untuk menghindari hal tersebut digunakan bentuk modifikasi R^2 yang disebut nilai koefisien determinasi disesuaikan atau R^2 adjusted (R_a^2) yang secara sistematis dituliskan sebagai berikut (Harel, 2009):

$$R_a^2 = 1 - \left[(1 - R^2) \left(\frac{n - 1}{n - p - 1} \right) \right] \quad (2.19)$$

dengan:

- R^2 : koefisien determinasi
 n : jumlah pengamatan
 p : jumlah prediktor

Selain dari nilai R_a^2 ukuran kebaikan hasil pendugaan model regresi dapat dilihat dari nilai *Root Mean Square Error* (RMSE) (Eubank, 1999). Nilai RMSE yang semakin kecil menunjukkan model regresi tersebut semakin baik. Perhitungan nilai RMSE dituliskan sebagai berikut:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}, \quad (2.20)$$

dengan

- n : jumlah pengamatan
 \hat{y}_i : penduga y_i
 y_i : nilai variabel respon pada pengamatan ke- i

2.11. Jumlah Penduduk

Jumlah dan pertumbuhan penduduk akan sangat berpengaruh pada proses pembangunan. Kondisi kependudukan Indonesia dan Provinsi Sulawesi Selatan yang tinggi secara tidak langsung berpengaruh pada pembangunan. Jumlah penduduk Provinsi Sulawesi Selatan pada tahun 2018 mencapai 8.771.970 jiwa dengan laju pertumbuhan sebesar 0,94%. Penduduk Sulawesi Selatan yang terus bertambah dan jumlahnya relatif besar akan memberikan dampak positif dan negatif. Jumlah penduduk yang besar merupakan indikator tersedianya tenaga kerja yang cukup memadai dan juga berdampak pada penyediaan bahan makanan serta berbagai fasilitas hidup layak. Jumlah penduduk yang terus bertambah, mengakibatkan pemerintah juga harus terus menambah jumlah fasilitas hidup layak bagi masyarakatnya. Dua fasilitas yang paling mendasar adalah fasilitas pendidikan dan kesehatan (Badan Pusat Statistik, 2019).

Berdasarkan kelompok umur, penduduk terbanyak di Provinsi Sulawesi Selatan berada di kelompok umur 0 – 4 tahun. Tahun 2018, rasio jenis kelamin di Provinsi Sulawesi Selatan berjumlah 95,58% dengan jumlah laki-laki 4.286.893

jiwa dan perempuan 4.485.077 jiwa. Faktor-faktor yang dapat mempengaruhi besarnya jumlah penduduk adalah jumlah fasilitas dan tenaga kesehatan. Jumlah fasilitas dan tenaga kesehatan dikatakan berpengaruh terhadap banyaknya jumlah penduduk karena semakin banyak fasilitas dan tenaga kesehatan maka dapat meningkatkan kualitas kesehatan penduduk, sehingga kualitas hidup dan umur penduduk diharapkan semakin tinggi yang akan berdampak pada penurunan tingkat kematian dan bertambahnya jumlah penduduk. Jumlah balita gizi buruk dan kasus penyakit juga diduga dapat mempengaruhi jumlah penduduk karena semakin banyak jumlah balita gizi buruk dan kasus penyakit, maka secara tidak langsung dapat meningkatkan tingkat kematian yang berdampak pada jumlah penduduk. Selain itu, salah satu faktor yang juga diduga dapat mempengaruhi jumlah penduduk adalah jumlah klinik keluarga berencana (KB) dan pos pelayanan keluarga berencana (KB) desa. Semakin banyak jumlah klinik KB dan pos pelayanan KB desa diduga dapat menekan jumlah penduduk.