

PEMODELAN *GEOGRAPHICALLY WEIGHTED NEGATIVE BINOMIAL REGRESSION* DAN *MIXED GEOGRAPHICALLY WEIGHTED NEGATIVE BINOMIAL REGRESSION* PADA JUMLAH KASUS TUBERKULOSIS DI SULAWESI SELATAN

MODELING *GEOGRAPHICALLY WEIGHTED NEGATIVE BINOMIAL REGRESSION* AND *MIXED GEOGRAPHICALLY WEIGHTED NEGATIVE BINOMIAL REGRESSION* ON THE NUMBER OF TUBERCULOSIS CASES IN SOUTH SULAWESI

A. AINUN NURFAJRIN. S



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

HALAMAN JUDUL
PEMODELAN *GEOGRAPHICALLY WEIGHTED NEGATIVE*
BINOMIAL REGRESSION* DAN *MIXED GEOGRAPHICALLY
***WEIGHTED NEGATIVE BINOMIAL REGRESSION* PADA JUMLAH**
KASUS TUBERKULOSIS DI SULAWESI SELATAN

Tesis

Sebagai salah ssatu syarat untuk mencapai gelar magister

Program Studi Magister Statistika

Disusun dan diajukan oleh

A.AINUN NURFAJRIN S

H062212002

Kepada

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

LEMBAR PENGESAHAN

TESIS

PEMODELAN GEOGRAPHICALLY WEIGHTED NEGATIVE BINOMIAL REGRESSION DAN MIXED GEOGRAPHICALLY WEIGHTED NEGATIVE BINOMIAL REGRESSION PADA JUMLAH KASUS TUBERKULOSIS DI SULAWESI SELATAN

A.AINUN NURFAJRIN S

H062212002

Telah dipertahankan di hadapan Panitia Ujian yang dibentuk dalam rangka Penyelesaian Program Studi Magister Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam

Universitas Hasanuddin

pada tanggal 06 Desember 2023

dan dinyatakan telah memenuhi syarat kelulusan

Menyetujui

Pembimbing Utama



Prof. Dr. Nurtit Sunusi, S.Si., M.Si

NIP. 19720117 199703 2 002

Pembimbing Pendamping



Dr. Erna Tri Herdiani, S.Si., M.Si

NIP. 19750429 200003 2 001

Ketua Program Studi
Magister Statistika



Dr. Erna Tri Herdiani, S.Si., M.Si

NIP. 19750429 200003 2 001

Dekan Fakultas Matematika dan
Ilmu Pengetahuan Alam
Universitas Hasanuddin



Dr. Eng. Amiruddin, M.Si

NIP. 19720515 199702 1 002

PERNYATAAN KEASLIAN TESIS DAN PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa tesis berjudul *Pemodelan Geographically Weighted Negative Binomial Regression dan Mixed Geographically Weighted Negative Binomial Regression* pada Jumlah Kasus Tuberkulosis di Sulawesi Selatan adalah karya saya dengan arahan dari komisi pembimbing (Prof. Dr. Nurititi Sunusi S.Si., M.Si dan Dr. Erna Tri Herdiani, S.Si., M.Si). karya ilmiah ini belum diajukan dan tidak sedang diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka tesis ini. Sebagian dari isi tesis ini telah dipublikasikan di Jurnal *International Communications in Mathematical Biology and Neuroscience* sebagai artikel dengan judul "*Modeling Mixed Geographically Weighted Negative Binomial Regression on the Number of Tuberculosis Cases in South Sulawesi*".

Dengan ini saya melimpahkan hak cipta dari karya tulis saya berupa tesis ini kepada Universitas Hasanuddin.

Makassar, 06 Agustus 2023

Yang menyatakan,



A. Ainun Nurfaejrin S
NIM. H06221200

UCAPAN TERIMA KASIH

Bismillahirrahmanirrahim

Assalamualaikum Warahmatullahi Wabarakatuh

Puji syukur kehadiran Tuhan Yang Maha Esa atas limpahan rahmat dan karunia-Nya sehingga penulis dapat menyusun dan menyelesaikan tesis ini. Penulis menyadari sepenuhnya bahwa apa yang dikemukakan dalam tesis ini masih jauh dari kesempurnaan yang merupakan sebagai akibat dari keterbatasan kemampuan serta berbagai kesulitan yang penulis hadapi dalam penyusunan tesis ini.

Penulis memanjatkan doa kepada Tuhan Yang Maha Esa agar memberikan rahmat-Nya kepada pihak yang banyak membantu dalam penyelesaian tesis ini. Penulis juga percaya tesis ini dapat selesai bukan hanya dengan kekuatan pikiran penulis semata akan tetapi karena bantuan dari berbagai pihak juga, baik selama proses perkuliahan bahkan sampai proses pengerjaan tesis di Program Magister Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Hasanuddin. Namun demikian, penulis dengan senang hati menerima kritik dan saran yang bersifat membangun dari pembaca karya tulis ini demi sempurnanya tesis ini.

Terima Kasih yang tak terhingga kepada kedua orang tuaku tercinta Ayah (**A. Muhammad Saerdi, S.Pd**) dan Ibu (**Andi Sukawati AL.**) yang selalu mendengarkan keluh kesahku, memberikan kasih sayang tak terhingga, materi, semangat, motivasi dan doa yang tak pernah putus, juga kepada saudara-saudaraku, kakak (**A.M.Alief S.S.H**) dan adikku (**A.Muh.Iqbal S.**) yang selalu memberikan dukungan, semangat dan dalam penyelesaian tesis ini. Ucapan rasa hormat dan juga terima kasih yang tulus kepada:

1. Yth. **Prof. Dr. Ir. Jamaluddin Jompa, M.Sc.** selaku Rektor Universitas Hasanuddin.
2. Yth. **Dr. Eng. Amiruddin, M.Si.** selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam beserta seluruh jajarannya.

3. Yth. **Dr. Anna Islamiyati, S.Si., M.Si.** selaku Ketua Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam dan salah satu tim penguji tesis yang memberikan masukan dan saran.
4. Yth. **Dr. Erna Tri Herdiani, S.Si., M.Si.** selaku Ketua Program Studi Magister Statistika sekaligus pembimbing pertama yang senantiasa meluangkan waktu, tenaga, pemikiran dalam membimbing dan mengarahkan penulis dalam menyelesaikan tesis.
5. Yth. **Prof. Dr. Nurtiti Sunusi, S.Si., M.Si.** selaku pembimbing utama yang senantiasa meluangkan waktunya untuk membimbing dengan penuh kesabaran, memberikan arahan, motivasi, serta memberikan saran dan masukkan dalam menyelesaikan tesis.
6. Yth. **Dr. Nirwan, M.Si** dan **Dr. Dr. Georgina Maria Tinungki, M.Si.** selaku penguji yang telah bersedia menguji serta memberikan masukan-masukan dan arahan dalam penyusunan tesis.
7. Teman-teman seperjuangan **Aliyah Az-zahra Ibrahim, S.Stat., Firawati, S.Stat., Nalto Batty Mangiri, S.Stat., Nurfauzul Akbar, S.Stat., Nurwan, S.Stat** dan **Thesya Atarezcha Pangruruk, S.Stat., M.Si.**
8. Semua pihak yang telah membantu penulis yang tidak bisa disebutkan satu per satu terima kasih atas doa serta dukungannya.

Semoga Tuhan Yang Maha Esa memberikan balasan yang berlipat ganda, kasih dan hikmat-Nya ata segala kebaikan yang telah diberikan kepada penulis. Penulis berharap penulisan tesis ini dapat memberikan manfaat khususnta bagi penulis dan umumnya bagi para pembaca.

Makassar, 06 Desember 2023

Yang menyatakan



A. Ainun Nurfajrin S
NIM. H062212002

ABSTRAK

A. AINUN NURFAJRIN S. *Pemodelan Geographically Weighted Negative Binomial Regression dan Mixed Geographically Weighted Negative Binomial Regression pada Jumlah Kasus Tuberkulosis di Sulawesi Selatan* (dibimbing oleh Nurtiti Sunusi dan Erna Tri Herdiani).

Tuberkulosis adalah suatu penyakit infeksi yang disebabkan bakteri yang dikenal *Mycrobacterium Tuberculosis*, dimana kasus ini merupakan masalah diberbagai wilayah, salah satunya di Provinsi Sulawesi Selatan yang mengalami masalah tuberkulosis dalam beberapa tahun terakhir. Data Tuberkulosis di Sulawesi Selatan merupakan data diskrit yang berheterogenitas spasial antar wilayah dan mengalami overdispersi dalam data, yang kemungkinan disebabkan oleh letak geografis setiap wilayah berbeda-beda. Masalah overdispersi dalam data dapat diatasi dengan menggunakan model Binomial Negatif. Namun model ini hanya bersifat global sedangkan pada kasus tuberkulosis memiliki karakteristik lokasi yang berbeda-beda. Oleh karena itu diperlukan metode dengan yang mempertimbangkan efek heterogenitas spasial, yang datanya mengalami overdispersi. Geographically Weighted Negative Binomial Regression (GWNBR) dan Mixed Geographically Weighted Negative Binomial Regression (MGWNBR) adalah model yang digunakan untuk data diskrit yang berheterogenitas spasial yang dapat mengatasi overdispersi dalam data. Hasil penelitian menggunakan GWNBR dengan pembobot adaptive bisquare kernel menunjukkan bahwa variabel yang berpengaruh signifikan terhadap jumlah kasus tuberkulosis di seluruh lokasi amatan atau disebut juga dengan variabel global adalah jumlah sarana kesehatan dan tenaga medis, sedangkan variabel lokal yang berpengaruh signifikan terhadap jumlah kasus tuberkulosis di beberapa lokasi amatan adalah jumlah penduduk dan kepadatan penduduk. MGWNBR dengan pembobot adaptive bisquare kernel menunjukkan bahwa variabel yang berpengaruh signifikan terhadap jumlah kasus tuberkulosis di seluruh lokasi amatan atau disebut juga dengan variabel global adalah jumlah tenaga medis, sedangkan variabel lokal yang berpengaruh signifikan terhadap jumlah kasus tuberkulosis di beberapa lokasi amatan adalah jumlah sarana kesehatan, jumlah penduduk dan kepadatan penduduk. Model terbaik untuk memodelkan jumlah kasus tuberkulosis berdasarkan nilai AIC paling terkecil adalah model MGWNBR.

Kata Kunci: Tuberkulosis, Overdispersi, Binomial Negatif, GWNBR, MGWNBR, Adaptive Bisquare Kernel, AIC.

ABSTRACT

A. AINUN NURFAJRIN S. *Modeling Geographically Weighted Negative Binomial Regression and Mixed Geographically Weighted Negative Binomial Regression on the Number of Tuberculosis Cases in South Sulawesi* (dibimbing oleh Nurtiti Sunusi dan Erna Tri Herdiani).

Tuberculosis is an infectious disease caused by bacteria known as Mycobacterium Tuberculosis, where this case is a problem in various regions, one of which is in South Sulawesi Province which has experienced tuberculosis problems in recent years. Tuberculosis data in South Sulawesi is discrete data with spatial heterogeneity between regions and overdispersion in the data, which may be caused by the geographical location of each region. The problem of overdispersion in the data can be overcome by using the Negative Binomial model. However, this model is only global while tuberculosis cases have different location characteristics. Therefore, a method is needed that considers the effects of spatial heterogeneity, where the data is overdispersed. Geographically Weighted Negative Binomial Regression (GWNBR) and Mixed Geographically Weighted Negative Binomial Regression (MGWNBR) are models used for spatially heterogeneous discrete data that can overcome overdispersion in the data. The results of the study using GWNBR with adaptive bisquare kernel weights show that the variables that have a significant effect on the number of tuberculosis cases in all observed locations or also known as global variables are the number of health facilities and medical personnel, while the local variables that have a significant effect on the number of tuberculosis cases in some observed locations are population and population density. MGWNBR with adaptive bisquare kernel weights shows that the variable that has a significant effect on the number of tuberculosis cases in all observed locations or also known as the global variable is the number of medical personnel, while the local variables that have a significant effect on the number of tuberculosis cases in some observed locations are the number of health facilities, population and population density. The best model to model the number of tuberculosis cases based on the smallest AIC value is the MGWNBR model.

Keywords: Tuberculosis, Overdispersion, Negative Binomial, GWNBR, MGWNBR, Adaptive Bisquare Kernel, AIC.

DAFTAR ISI

HALAMAN JUDUL	i
LEMBAR PENGESAHAN	ii
PERNYATAAN KEASLIAN TESIS	iii
UCAPAN TERIMA KASIH.....	iv
ABSTRAK.....	vi
ABSTRACT.....	vii
DAFTAR ISI	viii
DAFTAR TABEL	xi
DAFTAR GAMBAR	xii
BAB I PENDAHULUAN	1
1.1. Latar Belakang.....	1
1.2. Rumusan Masalah	4
1.3. Batasan Masalah	4
1.4. Tujuan Penelitian	5
1.5. Manfaat Penelitian	5
BAB II TINJAUAN PUSTAKA	6
2.1. Multikolinearitas	6
2.2. Overdispersi	7
2.3. Regresi Binomial Negatif	7
2.3.1. Estimasi Parameter Binomial Negatif.....	8
2.3.2. Pengujian Parameter Binomial Negatif	8
2.4. Efek Spasial	9
2.5. Heterogenitas Spasial.....	9
2.6. Matriks Pembobot Spasial	10

2.6.1.	<i>Adaptive Gaussian Kernel</i>	10
2.6.2.	<i>Adiptive Bisquare Kernel</i>	11
2.7.	Penentuan <i>Bandwidth</i> Optimum.....	11
2.8.	<i>Geographically Weighted Negative Binomial Regression</i> (GWNBR)....	12
2.8.1.	Estimasi Parameter Model GWNBR	13
2.8.2.	Pengujian Parameter GWNBR	14
2.9.	<i>Mixed Geographically Weighted Negative Binomial Regression</i> (MGWNBR).....	15
2.9.1.	Pengujian Parameter MGWNBR.....	16
2.10.	Tuberkulosis.....	16
2.11.	Penelitian Terdahulu.....	17
2.12.	Kerangka Konseptual	19
BAB III METODOLOGI PENELITIAN.....		20
3.1.	Sumber Data	20
3.2.	Identifikasi Variabel.....	20
3.3.	Analisis Data	21
BAB IV HASIL DAN PEMBAHASAN.....		22
4.1.	Estimasi Parameter Model <i>Geographically Weighted Negative Binomial Regression</i> (GWNBR)	22
4.2.	Estimasi Parameter Model <i>Mixed Geographically Weighted Negative Binomial Regression</i> (MGWNBR)	25
4.3.	Karakteristik Jumlah Kasus Tuberkulosis di Sulawesi Selatan Tahun 2021-2022.....	29
4.4.	Rasio Sarana Kesehatan tiap Kabupaten/Kota di Sulawesi Selatan	30
4.5.	Rasio Tenaga Medis tiap Kabupaten/Kota di Sulawesi Selatan.....	31
4.6.	Rasio Jumlah Penduduk tiap Kabupaten/Kota di Sulawesi Selatan.....	32

4.7. Kepadatan Penduduk	32
4.8. Pendeteksian Multikolinieritas	33
4.9. Pemodelan Regresi Binomial Negatif.....	34
4.10. Heterogenitas Spasial.....	35
4.11. Pemodelan <i>Geographically Weighted Negative Binomial Regression</i> ..	35
4.11.1. Pengujian Serentak	36
4.11.2. Pengujian Parsial	36
4.12. Pemodelan <i>Mixed Geographically Weighted Negative Binomial Regression</i>	39
4.12.1. Pengujian Serentak	39
4.12.2. Pengujian Parsial	40
4.13. Pemilihan Model Terbaik	43
BAB V KESIMPULAN DAN SARAN	44
5.1. Kesimpulan	44
5.2. Saran.....	44
DAFTAR PUSTAKA.....	45
LAMPIRAN	48

DAFTAR TABEL

Tabel 2. 1 Penelitian Terdahulu	18
Tabel 4. 1 Nilai VIF dari Variabel Prediktor	34
Tabel 4. 2 Regresi Binomial Negatif.....	34
Tabel 4. 3 Uji <i>Breusch-Pagan</i> (BP)	35
Tabel 4. 4 Estimasi Parameter Model GWNBR di Kota Makassar.....	37
Tabel 4. 5 Pengelompokan Kabupaten/Kota berdasarkan variabel yang signifikan.....	38
Tabel 4. 6 Estimasi Parameter Model MGWNBR di Kota Makassar	40
Tabel 4. 7 Pengelompokan Kabupaten/Kota berdasarkan variabel yang signifikan	41
Tabel 4. 8 Nilai AIC	43

DAFTAR GAMBAR

Gambar 2. 1 Kerangka Konseptual.....	19
Gambar 4. 1 Penyebaran Jumlah Kasus TB di Provinsi Sulawesi Selatan Tahun 2021-2022	29
Gambar 4. 2 Persebaran Sarana Kesehatan di Provinsi Sulawesi Selatan	30
Gambar 4. 3 Persebaran Tenaga Medis di Provinsi Sulawesi Selatan	31
Gambar 4. 4 Persebaran Jumlah Penduduk di Provinsi Sulawesi Selatan Tahun 2022.....	32
Gambar 4. 5 Persebaran Kepadatan Penduduk di Provinsi Sulawesi Selatan ..	33
Gambar 4. 6 Peta Persebaran Pengelompokan Kabupaten/Kota berdasarkan variabel yang signifikan.....	38
Gambar 4. 7 Peta Persebaran Pengelompokan Kabupaten/Kota berdasarkan variabel yang signifikan.....	42

BAB I

PENDAHULUAN

1.1. Latar Belakang

Regresi adalah suatu metode yang berfungsi untuk memprediksi pengaruh dari dua atau lebih variabel. Regresi pertama kali diperkenalkan oleh Francis Galton. Menurut Galton, analisis regresi adalah metode untuk menelaah hubungan dua variabel atau lebih. Dalam analisis regresi terdapat dua jenis variabel yaitu variabel bebas (variabel prediktor) dan variabel terikat (variabel respon). Penerapan analisis regresi digunakan untuk menganalisis data variabel respon yang berupa kontinu, namun sering diperoleh data variabel respon yang berupa data diskrit (*count*). Salah satu model regresi yang dapat digunakan untuk menjelaskan hubungan antara variabel respon Y yang berupa data diskrit (*count*) dengan variabel prediktor X berupa data diskrit, kontinu, kategori atau campuran adalah model regresi Poisson (Casella & Berger, 1990). Regresi Poisson merupakan metode yang sering digunakan untuk menganalisis data *count* (Cameron & Trivedi, 1998). Data *count* adalah data yang menyatakan banyaknya kejadian dalam interval waktu, ruang, atau volume tertentu. Regresi Poisson merupakan model regresi nonlinear yang digunakan untuk menganalisis data diskrit (*count*). Dalam regresi Poisson terdapat asumsi yang harus dipenuhi yaitu variabel respon (Y) berupa data diskrit dan equidispersi. Equidispersi yaitu nilai rata-rata sama dengan nilai varian. Seringkali hitungan data memiliki variansi yang lebih besar dari meannya atau biasa disebut istilah overdispersi (Yasin et al., 2021). Pelanggaran dari asumsi ini menyebabkan parameter yang dihasilkan dari regresi Poisson menjadi kurang akurat, sedangkan jika nilai variansi lebih kecil dari nilai mean disebut underdispersi (McCullagh & Nelder, 1989). Model Negatif Binomial merupakan model yang dapat mengatasi overdispersi dan lebih fleksibel daripada model regresi Poisson karena asumsi mean dan variansinya tidak harus sama. (Yasin et al., 2021).

Distribusi Binomial Negatif tidak mengharuskan nilai variansi sama dengan meannya serta memiliki parameter dispersi yang membuat variansi dapat bervariasi menjadi lebih besar dari rata-rata. Apabila nilai parameter dispersi pada regresi Binomial Negatif sama dengan nol, maka akan menghasilkan nilai variansi sama dengan nilai rata-rata sesuai dengan asumsi pada regresi Poisson (Yuli & Indriani,

2015). Penelitian ini tidak hanya sampai pada penanganan overdispersi, namun model regresi ini memperhatikan faktor heterogenitas spasial karena pada kenyataannya kondisi geografis, sosial budaya dan ekonomi pada suatu wilayah tentunya akan berbeda-beda atau disebut heterogenitas spasial antar wilayah. Pengembangan model regresi yang memperhatikan faktor heterogenitas spasial dengan pembobotan geografi yaitu Geographically Weighted Regression (GWR) (Fotheringham, 2002). Sebelumnya penelitian dilakukan oleh Brunsdon (1996) yang menunjukkan bahwa model regresi Poisson tidak dapat menjelaskan hubungan antara beberapa set peubah jika terdapat faktor geografis, akan tetapi GWR dapat menjelaskan hubungan yang berbeda pada setiap titik lokasi pengamatan, namun GWR hanya digunakan ketika data atau variabel respon yang kontinu. Sehingga untuk mengatasi variabel respon yang tidak kontinu atau diskrit, Nakaya et al. (2005) telah mengembangkan model *Geographically Weighted Poisson Regression* (GWPR). Model GWPR cukup baik digunakan apabila asumsi ekuidispersi terpenuhi, namun bila terjadi overdispersi, terdapat *model Geographically Weighted Negative Binomial Regression* (GWNBR) yang diperkenalkan pertama kali oleh Ricardo & Carvalho (2014) dalam penelitiannya menyatakan bahwa metode GWNBR baik digunakan untuk memodelkan data diskrit, khususnya ketika data tersebut terjadi overdispersi. Metode GWNBR ini lebih baik dibandingkan dengan regresi Poisson, GWPR, dan regresi Binomial Negatif berdasarkan nilai AIC.

Geographically Weighted Negative Binomial Regression (GWNBR) merupakan salah satu metode untuk memodelkan data cacah yang berheterogenitas spasial dan overdispersi dan GWNBR juga salah satu solusi yang tepat untuk membentuk analisis regresi yang bersifat lokal untuk setiap lokasi pengamatan. Namun seringkali tidak semua variabel prediktor dalam model GWNBR berpengaruh secara lokal. Terkadang, beberapa variabel prediktor berpengaruh secara global, sedangkan prediktor lainnya dapat mempertahankan pengaruh lokal atau spasialnya. Oleh karena itu, Model GWNBR dikembangkan menjadi model *Mixed Geographically Weighted Negative Binomial Regression* (MGWNBR). Model MGWNBR adalah gabungan dari model regresi Binomial Negatif (NB) dengan model GWNBR, sehingga pada model MGWNBR akan dihasilkan penaksir parameter yang sebagian bersifat global dan sebagian lainnya bersifat lokal sesuai dengan pengamatan data. Model MGWNBR dalam pengamatannya hanya memiliki satu variabel respon yang tergantung pada lokasi

pengamatan. Penelitian sebelumnya menggunakan model GWNBR dan model MGWNBR telah dilakukan oleh peneliti, Juniardi & Salamah (2015) dengan melakukan analisis faktor-faktor yang mempengaruhi jumlah kasus Kusta di Jawa Timur menggunakan *Geographically Weighted Negative Binomial Regression* (GWNBR). Fitriyah (2017) tentang estimasi parameter model GWNBR dengan pembobot *Adaptive Gaussian Kernel* yang diaplikasikan terhadap pemetaan kemiskinan dan menghasilkan 15 kelompok wilayah berdasarkan variabel yang signifikan. Yasin et al. (2021) Grafik model GWNBR menggunakan R-*Shiny* yang menghasilkan model GWNBR dengan pembobot Adaptive Boxcar model terbaik berdasarkan nilai AIC. Putera et al. (2021) tentang Pemodelan Spasial Kasus Terkonfirmasi Covid-19 di Kalimantan, Indonesia menggunakan GWNBR dan MGWNBR yang menghasilkan bahwa nilai MGWNBR lebih baik dari model lainnya dalam hal AIC. Model GWNBR dan MGWNBR sama-sama digunakan untuk mengatasi kasus overdispersi dalam data diskrit (*count*) yang berheterogenitas spasial dan memperhatikan pembobot berupa lintang dan bujur dari titik-titik pengamatan yang diamati. Perbedaan dari kedua model tersebut terletak pada variabel prediktornya yang dipartisi menjadi dua bagian yaitu koefisien lokal dan koefisien global.

Jumlah kasus tuberkulosis merupakan salah satu contoh data diskrit (*count*) yang terdapat heterogenitas spasial antar wilayah didalamnya dan sering terjadi overdispersi pada data. Tuberkulosis merupakan salah satu penyakit saluran pernapasan bawah yang menular yang disebabkan bakteri *Mycrobacterium Tuberkulosis* (Lestari, 2014) Kasus tuberkulosis merupakan masalah diberbagai wilayah. Kondisi geografis wilayah menyebabkan adanya perbedaan jumlah kasus tuberkulosis antara satu wilayah dengan wilayah lainnya. Kepadatan penduduk menjadi salah satu faktor yang memicu cepat berkembangnya penyakit tuberkulosis. Faktor lain yang mempengaruhi banyaknya kasus tuberkulosis yaitu jumlah rumah sakit, jumlah puskesmas, dan perilaku hidup bersih dan sehat. Kejadian tuberkulosis dapat berpindah dari suatu wilayah ke wilayah lain. Provinsi Sulawesi Selatan mengalami masalah tuberkulosis dalam beberapa tahun terakhir. Berdasarkan data dari Dinas Kesehatan Provinsi Sulawesi Selatan, pada tahun 2018 tercatat sebanyak 18 ribu kasus, tahun 2019 naik hampir 19 ribu kasus, sedangkan di tahun 2020 mengalami penurunan menjadi 12 ribu kasus, kemudian di tahun 2021 terjadi kenaikan kasus sebanyak 15 ribu kasus. Angka itu disebut

masih cukup tinggi, apabila tidak segera ditangani maka dapat menyebabkan setiap penderita aktif yang menginfeksi.

Geographically Weighted Negative Binomial Regression digunakan untuk data yang memiliki heterogenitas spasial untuk untuk data diskrit (*count*) yang memiliki overdispersi. Model GWNBR juga akan menghasilkan parameter lokal dengan masing-masing lokasi dan memiliki parameter yang berbeda-beda. Sama halnya dengan *Mixed Geographically Weighted Negative Binomial Regression* digunakan untuk data diskrit yang berheterogenitas spasial dan terjadi overdispersi dalam data tetapi yang membedakan dari model GWNBR yaitu terdapat pada variabel prediktornya, dimana pada model MGWNBR variabel prediktornya dipartisi menjadi dua bagian yaitu variabel prediktor lokal dan variabel prediktor global.

Berdasarkan uraian tersebut, penelitian ini mengambil judul "**Pemodelan *Geographically Weighted Negative Binomial Regression* dan *Mixed Geographically Weighted Negative Binomial Regression* pada Jumlah Kasus Tuberkulosis di Sulawesi Selatan**"

1.2. Rumusan Masalah

Berdasarkan latar belakang penelitian yang telah diuraikan, maka rumusan masalah pada penelitian ini adalah:

1. Bagaimana estimasi parameter model *Geographically Weighted Negative Binomial Regression* (GWNBR) ?
2. Bagaimana estimasi parameter model *Mixed Geographically Weighted Negative Binomial Regression* (MGWNBR) ?
3. Bagaimana pemodelan *Geographically Weighted Negative Binomial Regression* (GWNBR) pada kasus tuberkulosis di Sulawesi Selatan?
4. Bagaimana pemodelan *Mixed Geographically Weighted Negative Binomial Regression* (MGWNBR) pada kasus tuberkulosis di Sulawesi Selatan?
5. Bagaimana menentukan model terbaik pada kasus Tuberkulosis di Sulawesi Selatan dengan nilai AIC terkecil pada model GWNBR dan MGWNBR ?

1.3. Batasan Masalah

1. Data yang digunakan adalah banyaknya kasus tuberkulosis tiap Kabupaten/Kota di Sulawesi Selatan tahun 2020-2021.
2. Pembobot yang di gunakan adalah *Adaptive Bisquare Kernel*.

3. Metode estimasi parameter adalah *Maximum Likelihood Estimation* (MLE).

1.4. Tujuan Penelitian

Berdasarkan rumusan masalah yang telah diuraikan. Maka tujuan penulisan penelitian ini adalah:

1. Memperoleh estimasi parameter model *Geographically Weighted Negative Binomial Regression* (GWNBR).
2. Memperoleh estimasi *Mixed Geographically Weighted Negative Binomial Regression* (MGWNBR).
3. Memperoleh hasil pemodelan *Geographically Weighted Negative Binomial Regression* (GWNBR).
4. Memperoleh hasil pemodelan *Mixed Geographically Weighted Negative Binomial Regression* (MGWNBR).
5. Memperoleh model terbaik pada kasus Tuberkulosis di Sulawesi Selatan dengan nilai AIC terkecil pada model GWNBR dan MGWNBR

1.5. Manfaat Penelitian

Hasil penelitian ini diharapkan dapat memberi tambahan ilmu tentang penerapan ilmu statistika dalam kehidupan sehari-hari, memberi pemahaman tentang model *Geographically Weighted Negative Binomial Regression* (GWNBR) dan *Mixed Geographically Weighted Negative Binomial Regression* (MGWNBR), memberikan informasi efektifitas mengenai tuberkulosis, memberi informasi tambahan pada pemerintah mengenai penanganan tuberkulosis pada masing-masing wilayah karna memiliki pengaruh kondisi lokal sehingga setiap wilayah dapat di tangani dengan cara yang berbeda-beda.

BAB II

TINJAUAN PUSTAKA

2.1. Multikolinearitas

Salah satu asumsi yang harus dipenuhi pada analisis regresi dengan beberapa peubah prediktor adalah tidak adanya korelasi antara satu peubah prediktor dengan peubah prediktor yang lain. Sembiring (1995) menyatakan bahwa jika dalam suatu model regresi terdapat kolinieritas maka akan mempengaruhi hubungan antara peubah prediktor dengan peubah respon. Penduga koefisien dari peubah prediktor menjadi tidak tunggal, melainkan tak terhingga banyaknya sehingga sulit untuk melakukan pendugaan. Selain itu, multikolinieritas dalam regresi dapat mengakibatkan besarnya simpangan baku dari penduga koefisien regresi. Kejadian multikolinieritas akan menyebabkan kesalahan yang besar pada penduga parameter regresi. Sehingga menyebabkan hasil penduga parameter bias dan tidak baik untuk dianalisis lebih lanjut. Oleh sebab itu, perlu dilakukan pemeriksaan multikolinieritas agar hasil penduga parameter tersebut bisa dianalisis lebih lanjut. Pemeriksaan multikolinieritas dapat dilakukan melalui perhitungan nilai koefisien korelasi *Pearson (ril)* antar variabel prediktor. Jika nilai-nilai koefisien korelasi *Pearson (ril)* antar variabel prediktor lebih besar dari 0,95 maka terjadi multikolinieritas dalam model tersebut. Selain dengan nilai koefisien korelasi *Pearson (ril)*, pemeriksaan multikolinieritas dapat dilakukan melalui nilai *Variance Inflation Faktors (VIF)*.

Pendeteksian kasus multikolinearitas menurut Hocking (1996) dapat dilihat dengan beberapa cara sebagai berikut :

- a. Jika koefisien korelasi *Pearson (r_{ij})* antar variabel prediktor lebih besar dari 0,95 ($r_{ij} > 0,95$) maka terdapat korelasi antar variabel tersebut.
- b. Nilai VIF lebih besar dari 10 ($VIF > 10$) menunjukkan bahwa adanya multikolinearitas antar variabel prediktor atau terdapat pelanggaran asumsi non-multikolinieritas. Nilai VIF dinyatakan sebagai berikut:

$$VIF_j = \frac{1}{1-R_j^2} \quad ; j = 1, 2, \dots, p \quad (2.1)$$

Dengan j adalah jumlah variabel predictor dan R_j^2 adalah koefisien determinasi antara satu variabel prediktor dengan variabel prediktor lainnya. Solusi untuk

mengatasi adanya kasus multikolinieritas yaitu dengan mengeluarkan variabel prediktor satu per satu mulai dari yang memiliki nilai VIF paling besar, melakukan transformasi variabel, penambahan variabel ataupun melakukan metode *Principal Component Analysis* (PCA). Pelanggaran asumsi non-multikolinieritas dapat menyebabkan nilai taksiran parameter regresi yang didapatkan memiliki standar *error* besar sehingga asumsi non-multikolinieritas menjadi salah satu syarat utama yang harus dipenuhi dalam pemodelan regresi.

2.2. Overdispersi

Regresi Poisson dikatakan overdispersi apabila nilai variansinya lebih besar dari nilai rata-ratanya ($Var(Y) > E(Y)$). Jika pada data diskrit terjadi overdispersi dan tetap menggunakan regresi Poisson sebagai metode penyelesaiannya, maka nilai *standard error* menjadi *underestimate* sehingga kesimpulan yang akan ditarik menjadi tidak valid. Overdispersi merupakan nilai dispersi *pearson chisquare* dan *deviance* yang dibagi dengan derajat bebasnya, diperoleh nilai lebih besar dari 1. Selain itu, θ juga merupakan parameter dispersi dengan kriteria (Famoye, et al, 2004) sebagai berikut:

- a. jika $\theta > 0$ artinya terjadi overdispersi pada regresi Poisson.
- b. jika $\theta < 0$ artinya terjadi underdispersi.
- c. jika $\theta = 0$ menunjukkan bahwa tidak terjadi kasus *over/under* dispersi (*equidispersi*).

2.3. Regresi Binomial Negatif

Regresi Binomial Negatif, variabel responnya diasumsikan memiliki distribusi Binomial Negatif yang dihasilkan dari distribusi campuran Poisson-Gamma. (Yasin dkk, 2021). Binomial Negatif merupakan salah satu solusi untuk mengatasi adanya kasus overdispersi. Model Binomial Negatif didasarkan pada distribusi *mixture* Poisson-Gamma (Hardin dan Hilbe 2018). Sebaran binomial negatif memiliki rata-rata $E(Y) = \mu$ dan ragam $Var(Y) = \mu + \mu^2\theta$ (Hilbe 2011). Fungsi peluang distribusi binomial negative adalah (Greene 2007):

$$f(y, \mu, \theta) = \frac{\Gamma(y + \theta^{-1})}{\Gamma(\theta^{-1})\Gamma(y + 1)} \left(\frac{1}{1 + \theta\mu} \right)^{\theta^{-1}} \left(\frac{\theta\mu}{1 + \theta\mu} \right)^y \quad (2.2)$$

Keterangan :

$f(y, \mu, \theta)$ = peluang sebaran binomial negatif

μ = nilai harapan peubah respon Y

y = nilai dari peubah respon Y

θ = parameter *disperse*

Binomial Negatif memiliki bentuk model regresi yang sama seperti model regresi Poisson pada persamaan (2.2) karena memiliki fungsi penghubung yang sama, yaitu fungsi penghubung *log*. Model regresi binomial negatif dituliskan sebagai berikut:

$E(Y) = \mu$ dan $Var(Y) = \mu + a\mu^2$ dengan θ adalah parameter dispersi.

Kemudian fungsi probabilitas Binomial Negatif dapat dinyatakan sebagai berikut:

$$\ln(\mu_i) = e \mathbf{x}_i^T \boldsymbol{\beta} = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki} \quad (2.3)$$

Hilbe (2011) menyatakan bahwa model binomial negatif pada umumnya menggunakan fungsi penghubung logaritma yaitu:

$$\eta_i = \ln(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta} \quad (2.4)$$

2.3.1. Estimasi Parameter Binomial Negatif

Estimasi parameter model Binomial Negatif dapat dilakukan dengan menggunakan metode *Maximum Likelihood Estimation* (MLE). Pada estimasi regresi Binomial Negatif, untuk memaksimumkan fungsi *likelihood* maka digunakan metode iterasi *newton raphson*. Fungsi *likelihood* dari regresi binomial negatif dinyatakan sebagai berikut:

$$L(\boldsymbol{\beta}, \theta | y_i) = \prod_{i=1}^n \left\{ \frac{\Gamma(y_i + \frac{1}{\theta_i})}{\Gamma(\frac{1}{\theta_i}) y_i!} \left(\frac{1}{1 + \theta_i \mu_i} \right)^{\frac{1}{\theta_i}} \left(\frac{\theta_i \mu_i}{1 + \theta_i \mu_i} \right)^{y_i} \right\} \quad (2.5)$$

Berikut adalah fungsi *log-likelihood* dari regresi binomial negatif:

$$\ln L(\boldsymbol{\beta}, \theta | y_i) = \sum_i \left\{ \ln \left(\frac{\Gamma(y_i + \frac{1}{\theta_i})}{\Gamma(\frac{1}{\theta_i}) y_i!} \right) + y_i \ln(\theta_i \mu_i) - \left(y_i + \frac{1}{\theta_i} \right) \ln(1 + \theta_i \mu_i) \right\} \quad (2.6)$$

2.3.2. Pengujian Parameter Binomial Negatif

Pengujian parameter model regresi Binomial Negatif dilakukan untuk mengetahui apakah parameter model mempengaruhi variabel respon. Pengujian parameter regresi Binomial Negatif terdiri dari uji serentak dan uji parsial. Hipotesis uji signifikansi secara serentak yaitu sebagai berikut :

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0$$

$$H_1 : \text{paling sedikit ada satu } \beta_k \neq 0, k = 1, 2, \dots, p$$

Statistik uji yang digunakan adalah sebagai berikut :

$$D(\hat{\beta}) = -2 \ln \Lambda = -2 \ln \left(\frac{L(\hat{\omega})}{L(\hat{\Omega})} \right) = 2 \left(\ln L(\hat{\Omega}) - L(\hat{\omega}) \right) \quad (2.7)$$

Kriteria penolakan : tolak H_0 jika nilai $D(\hat{\beta})$ lebih besar dari $X_{(p,a)}^2$ yang artinya minimal terdapat satu parameter dalam regresi Binomial Negatif yang berpengaruh signifikan terhadap variabel respon.

Setelah dilakukan pengujian serentak, dilakukan pengujian secara parsial untuk mengetahui parameter mana saja yang memberikan pengaruh signifikan terhadap variabel respon dengan hipotesis sebagai berikut :

$$H_0 : \beta_k = 0, k = 1, 2, \dots, p$$

$$H_1 : \beta_k \neq 0$$

Statistik uji yang digunakan sebagai berikut :

$$Z_{hit} = \frac{\beta_k}{se(\hat{\beta}_k)} \quad (2.8)$$

Daerah penolakan adalah tolak H_0 jika nilai dari $|Z_{hit}|$ lebih besar dari $Z_{\alpha/2}$ yang artinya parameter tersebut berpengaruh signifikan terhadap variabel respon dalam model regresi Binomial Negatif.

2.4. Efek Spasial

Efek spasial ditunjukkan adanya dependensi/autokorelasi spasial dan heterogenitas spasial. Autokorelasi spasial adalah penilaian korelasi antar pengamatan yang berdekatan pada suatu variabel. Jika pengamatan X_1, X_2, \dots, X_n menunjukkan saling ketergantungan terhadap ruang, maka data tersebut dikatakan terdapat autokorelasi spasial. Metode pengujian untuk mengetahui apakah terdeteksi autokorelasi spasial adalah dapat menggunakan indikator *Moran's I*, *Rasio Geary's*, dan *Local Indicator of Spatial Autocorrelation (LISA)*.

2.5. Heterogenitas Spasial

Heterogenitas spasial dapat terjadi apabila satu variabel prediktor yang sama memberikan respon yang berbeda pada lokasi lain dalam satu wilayah penelitian. Menurut Anselin (1988) pengujian heterogenitas spasial dilakukan untuk

melihat perbedaan karakteristik antara satu titik pengamatan dengan titik pengamatan lainnya, sehingga parameter regresi yang dihasilkan berbeda-beda secara spasial. Pengujian heterogenitas spasial dilakukan menggunakan statistik uji *Breusch-Pagan* (BP) dengan hipotesis sebagai berikut:

$$H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_p^2 = \sigma^2 \text{ (varians antarlokasi sama)}$$

$$H_1 : \text{minimal ada satu } \sigma_i^2 \neq \sigma^2 \text{ (varians antarlokasi berbeda)}$$

Dimana p adalah banyaknya variabel prediktor, σ^2 adalah varians e_i , dan σ_i^2 adalah varians dari variabel prediktor ke- i .

Statistik Uji :

$$BP = \left(\frac{1}{2}\right) f^T Z (Z^T Z)^{-1} Z^T f \sim X_{(p)}^2 \quad (2.9)$$

Dimana $f = (f_1, f_2, \dots, f_n)^T$ dengan $f_i = \frac{e_i^2}{\sigma^2} - 1$, dan e_i adalah error untuk observasi ke- i , Z adalah matriks berukuran $n \times (p+1)$ yang berisi vector konstan.

Daerah penolakan adalah tolak H_0 jika nilai $BP > X_{(p,\alpha)}^2$ atau $p\text{-value} < \alpha(0,05)$ yang berarti varians antarlokasi berbeda.

2.6. Matriks Pembobot Spasial

Matriks pembobot spasial menggunakan kedekatan antar wilayah atau jarak antar wilayah satu sama lain. Beberapa fungsi yang digunakan dalam pembentukan matriks pembobot spasial adalah sebagai berikut (Fotheringham et al., 2002):

2.6.1. Adaptive Gaussian Kernel

Menurut Leung, Mei dan Zhang (2000) dalam analisis spasial, suatu pengamatan yang berdekatan dengan lokasi i diasumsikan lebih berpengaruh dalam pendugaan parameter daripada pengamatan yang lebih jauh. Oleh karena itu ketika parameter di lokasi ke- i diduga, harus ditempatkan pada pengamatan yang dekat dengan lokasi ke- i . Berdasarkan pokok, salah satu fungsi pembobot yang paling umum digunakan adalah fungsi Gaussian seperti berikut:

$$w_{ij} = \exp \left[-\frac{1}{2} \left(\frac{d_{ij}}{b_i} \right)^2 \right] \quad (2.10)$$

Keterangan:

$i, j = 1, 2, \dots, n ; i \neq j$

d_{ij} = Jarak Euclidean antara lokasi amatan i dengan lokasi amatan j

b_i = lebar *bandwidth* pada lokasi pengamatan ke- i

Jika i dan j sama, pembobot dari titik akan tunggal sedangkan pembobotan dari data lain akan menurun berdasarkan pada kurva *Gaussian* seiring dengan meningkatnya jarak diantara i dan j .

2.6.2. Adaptive Bisquare Kernel

Menurut Leung, dkk (2000) fungsi pembobotan yang lain adalah untuk mengatur bobot bernilai nol jika berada di luar bandwidth dan menurun secara monoton ke nol jika berada dalam bandwidth seiring dengan peningkatan jarak dari kedua lokasi pengamatan.

$$w_{ij} = \begin{cases} \left[-\frac{1}{2} \left(\frac{d_{ij}}{b_i} \right)^2 \right]^2, & \text{jika } d_{ij} < b_i, \\ 0 & \text{lainnya} \end{cases} \quad (2.11)$$

Keterangan:

$i, j = 1, 2, \dots, n ; i \neq j$

d_{ij} = Jarak Euclidean antara lokasi amatan i dengan lokasi amatan j

b_i = lebar *bandwidth* pada lokasi pengamatan ke- i

Dengan $d_{ij} = \sqrt{(u_i - u_j)^2 + (v_i - v_j)^2}$ adalah jarak *euclidean* antara titik lokasi pengamatan ke- i dan titik lokasi pengamatan ke- j .

2.7. Penentuan Bandwidth Optimum

Secara teoritis *bandwidth* merupakan luasan dengan radius b dari titik pusat lokasi yang digunakan sebagai dasar menentukan bobot setiap pengamatan terhadap model regresi pada lokasi tersebut. Pengamatan-pengamatan yang terletak di dalam radius b masih dianggap berpengaruh terhadap model pada lokasi tersebut sehingga akan diberi bobot tergantung pada fungsi yang digunakan. Nilai *bandwidth* yang sangat kecil menyebabkan varians semakin besar. Hal ini dikarenakan jika nilai *bandwidth* sangat kecil maka akan semakin sedikit pengamatan yang berada dalam radius b , sehingga model yang diperoleh akan sangat kasar (*under smoothing*). Sebaliknya nilai *bandwidth* yang besar akan menimbulkan bias yang semakin besar karena semakin banyak pengamatan yang berada dalam radius b , sehingga model yang diperoleh akan terlampau halus (*over smoothing*).

Pemilihan *bandwidth* optimum menjadi sangat penting karena akan mempengaruhi ketepatan model terhadap data, yaitu mengatur varians dan bias

dari model. Penentuan bandwidth optimum dilakukan menggunakan metode *Cross Validation* (CV) sebagai berikut:

$$CV(b) = \sum_{i=1}^n (y_i - \hat{y}_{\neq i}(b))^2 \quad (2.12)$$

$\hat{y}_{\neq i}(b)$ merupakan nilai penaksiran untuk pengamatan ke- i yang dimana saat pemodelannya pengamatan ke- i dihilangkan dari proses penaksiran. *Bandwidth* optimum tercapai apabila diperoleh nilai CV paling minimum.

Pemilihan *bandwidth* optimum salah satunya adalah *Akaike Information Criterion* (AIC). Menurut Akaike (1974) fungsi AIC yaitu:

$$AIC = 2n \log_e(\delta) + n \log_e(2\pi) + n + tr(S) \quad (2.13)$$

Proses pemilihan lebar bandwidth optimum menggunakan teknik *Golden Section Search*. Teknik ini dilakukan secara iterasi dengan mengevaluasi nilai AIC terkecil pada interval jarak minimum dan maksimum lokasi pengamatan sehingga diperoleh nilai AIC minimum.

2.8. Geographically Weighted Negative Binomial Regression (GWNBR)

Model *Geographically Weighted Negative Binomial Regression* (GWNBR) adalah salah satu metode yang cukup efektif untuk menduga data yang memiliki heterogenitas spasial untuk data *count* yang memiliki overdispersi. *Geographically Weighted Negative Binomial Regression* (GWNBR) merupakan salah satu metode yang cukup efektif untuk menduga parameter lokal dengan unit amatan berupa wilayah pada data *count* yang memiliki heterogenitas spasial dan mengalami kasus overdispersi. GWNBR juga merupakan metode hasil pengembangan dari model regresi Binomial Negatif. Model GWNBR akan menghasilkan parameter lokal yang masing-masing memiliki parameter yang berbeda-beda. Model GWNBR dapat dirumuskan sebagai berikut (Ricardo dan Carvalho 2014):

$$y_i \sim NB \left[\exp \left(\sum_k \beta_k(u_i, v_i) x_{ki} \right), \theta(u_i, v_i) \right] \quad i = 1, 2, \dots, n$$

$$y_i \sim NB [\exp(\mathbf{x}'_i \boldsymbol{\beta}(u_i, v_i), \theta(u_i, v_i))] \quad (2.14)$$

Keterangan :

i = 1, 2, 3, ... , n

y_i = nilai observasi respon pada lokasi ke- i

x_{ki} = variabel prediktor lokasi (u_i, v_i)
 $\beta_k(u_i, v_i)$ = koefisien regresi di lokasi (u_i, v_i)
 $\theta(u_i, v_i)$ = parameter dispersi pada lokasi (u_i, v_i)

Fungsi sebaran binomial negatif untuk setiap lokasi dapat ditulis sebagai berikut (Yasin et al., 2021):

$$f(y_i | x_{ki} \beta_k(u_i, v_i), \theta(u_i, v_i)) = \frac{\Gamma(y_i + \frac{1}{\theta_i})}{\Gamma(\frac{1}{\theta_i}) \Gamma(y_i + 1)} \left(\frac{1}{1 + \theta_i \mu_i} \right)^{\frac{1}{\theta_i}} \left(\frac{\theta_i \mu_i}{1 + \theta_i \mu_i} \right)^{y_i} \quad (2.15)$$

Model GWNBR dapat ditulis sebagai berikut:

$$y_i = \exp(\beta_0 + \beta_1(u_i, v_i)x_1 + \beta_2(u_i, v_i)x_2 + \dots + \beta_k(u_i, v_i)x_k) \quad (2.16)$$

2.8.1. Estimasi Parameter Model GWNBR

Pendugaan parameter koefisien GWNBR dapat dilakukan dengan metode Maximum Likelihood Estimation (MLE) dan metode iterasi Newton-Raphson melalui matriks Hessian. Faktor pembobot yang digunakan yaitu letak geografis berdasarkan matriks pembobot spasial yang diperoleh dari nilai bandwidth optimum sehingga mempunyai nilai yang tidak sama untuk setiap wilayah dan menggambarkan sifat khusus atau lokal pada model. Estimasi parameter model GWNBR dengan metode MLE dengan fungsi *likelihood* dengan $L(\boldsymbol{\beta}(u_i, v_i), \theta_i | y_i, x_i)$ yaitu,

$$L(\boldsymbol{\beta}(u_i, v_i), \theta_i | y_i, x_i) = \prod_{i=1}^n \left[\frac{\Gamma(y_i + \theta(u_i, v_i)^{-1})}{\Gamma(\theta(u_i, v_i)^{-1}) y_i!} \left(\frac{1}{1 + \theta(u_i, v_i) \exp(x_i^T \boldsymbol{\beta}(u_i, v_i))} \right)^{\frac{1}{\theta_i}} \left(\frac{\theta(u_i, v_i) \exp(x_i^T \boldsymbol{\beta}(u_i, v_i))}{1 + \theta(u_i, v_i) \exp(x_i^T \boldsymbol{\beta}(u_i, v_i))} \right)^{y_i} \right] \quad (2.17)$$

Faktor letak geografis merupakan faktor pembobot pada model GWNBR. Faktor ini memiliki nilai yang berbeda untuk setiap daerah yang menunjukkan sifat lokal pada model GWNBR. Oleh karena itu pembobot kernel adaptif dapat disesuaikan dengan kondisi titik-titik pengamatan. Jika titik-titik lokasi pengamatan tersebar secara padat di sekitar lokasi pengamatan ke- i maka *bandwidth* yang diperoleh semakin sempit. Sebaliknya, jika titik-titik lokasi pengamatan memiliki jarak yang relatif jauh dari titik lokasi pengamatan ke- i , maka *bandwidth* yang diperoleh akan semakin luas. Dengan pembobot untuk setiap lokasi (u_i, v_i) yang diberikan, maka diperoleh fungsi *in likelihood* sebagai berikut (Fitriyah, 2017):

$$\ln\{L(\boldsymbol{\beta}(u_i, v_i), \theta_i | y_i, x_i)\} = \sum_{i=1}^n w_j(u_i, v_i) [a + b - c] \quad (2.18)$$

2.8.2. Pengujian Parameter GWNBR

Pengujian signifikansi parameter model GWNBR terdiri dari uji serentak dan uji parsial. Menentukan nilai statistik uji, terlebih dahulu ditentukan dua buah fungsi *likelihood* yang berhubungan dengan model regresi yang diperoleh. Fungsi-fungsi *likelihood* yang dimaksud adalah $L(\hat{\Omega})$ yaitu nilai *maximum likelihood* untuk model yang lebih lengkap dengan melibatkan variabel prediktor dan $L(\hat{\omega})$ yaitu nilai *maximum likelihood* untuk model sederhana tanpa melibatkan prediktor. Uji signifikansi secara serentak dapat menggunakan *Maximum Likelihood Ratio Test* (MLRT). Dengan hipotesis sebagai berikut :

$$H_0 : \beta_1(u_i, v_i) = \beta_2(u_i, v_i) = \dots = \beta_p(u_i, v_i) = 0$$

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

Statistik uji :

$$D(\hat{\beta}) = -2 \ln \left(\frac{L(\hat{\omega})}{L(\hat{\Omega})} \right) = 2 (\ln L(\hat{\Omega}) - L(\hat{\omega})) \quad (2.19)$$

Kriteria penolakan adalah tolak H_0 jika nilai $D(\hat{\beta})$ lebih besar dari $\chi^2_{(p,a)}$ yang artinya minimal terdapat satu parameter dalam model GWNBR yang berpengaruh signifikan terhadap variabel respon. Setelah dilakukan pengujian serentak, dilakukan pengujian secara parsial untuk mengetahui parameter mana saja yang memberikan pengaruh signifikan terhadap variabel respon pada tiap – tiap lokasi dengan hipotesis sebagai berikut :

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

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

Statistik Uji :

$$Z_{hit} = \frac{\hat{\beta}_j(u_i, v_i)}{se(\hat{\beta}_j(u_i, v_i))} \quad (2.20)$$

Dimana, $\hat{\beta}_j$ adalah koefisien model variabel prediktor ke-j, $se(\hat{\beta}_j)$ adalah *standard error* dan estimasi *maximum likelihood*. Daerah penolakan adalah tolak H_0 jika nilai dari $|Z_{hit}|$ lebih besar dari $Z_{\alpha/2}$ yang artinya parameter tersebut berpengaruh signifikan terhadap variabel respon pada tiap – tiap lokasi dalam model GWNBR.

2.9. Mixed Geographically Weighted Negative Binomial Regression (MGWNBR)

MGWNBR merupakan metode yang dikembangkan dari model Negatif Binomial dan GWNBR yang mengkombinasikan antara parameter yang bersifat lokal dan parameter yang bersifat global. Pada model MGWNBR peubah respon (y) diduga dengan variabel prediktor (x) yang masing-masing koefisien regresinya $\beta_k(u_i, v_i)$ bergantung pada lokasi geografis dan γ_p bersifat konstan. Model MGWNBR dengan $Y_i \sim$ binomial negatif dapat dinyatakan sebagai berikut:

$$y_i = \exp \left(\sum_{i=0}^{q^*} \beta_k(u_i, v_i) x_{ik} + \sum_{p=q^*+1}^q \gamma_p x_p + \varepsilon_i \right) \quad i = 1, 2, \dots, q^*; p = q^* + 1, \dots, q \quad (2.21)$$

Dalam model Binomial Negatif y_i adalah variabel yang berupa data cacah sehingga y_i merupakan bilangan bulat non-negatif, maka nilai ekspektasi dari y_i juga tidak mungkin negatif. Perbedaan dari model GWNBR dan model MGWNBR terletak pada variabel prediktornya yaitu pada model MGWNBR variabel prediktornya X dipartisi menjadi dua bagian yaitu variabel prediktor dengan koefisien lokal dengan variabel prediktor dengan koefisien global. Berikut partisi vektor \mathbf{x}

$$\mathbf{x} = \begin{bmatrix} 1 \\ x_1 \\ x_2 \\ \vdots \\ x_q \end{bmatrix} = \begin{bmatrix} 1 \\ x_1 \\ \vdots \\ x_{q^*} \\ x_{q^*+1} \\ x_{q^*+2} \\ \vdots \\ x_q \end{bmatrix} = \begin{bmatrix} 1 \\ x_1 \\ \vdots \\ x_{q^*} \\ x_p \\ x_{p+1} \\ \vdots \\ x_q \end{bmatrix} \quad (2.22)$$

Sehingga partisi dari vektor \mathbf{x}_i

$$\mathbf{x} = \begin{bmatrix} 1 \\ x_1 \\ x_2 \\ \vdots \\ x_{q^*} \end{bmatrix}_{(q^*+1) \times 1} \quad \text{dan} \quad \mathbf{x}_i^* = \begin{bmatrix} x_p \\ x_{p+1} \\ \vdots \\ x_q \end{bmatrix}_{(q^*-q) \times 1} \quad (2.23)$$

2.9.1. Pengujian Parameter MGWNBR

Pengujian signifikansi parameter model MGWNBR sama dengan GWNBR yaitu dengan uji serentak dan uji parsial. Uji signifikansi secara serentak dapat menggunakan *Maximum Likelihood Ratio Test* (MLRT). Dengan hipotesis sebagai berikut :

$$H_0 : \beta_1(u_i, v_i) = \beta_2(u_i, v_i) = \dots = \beta_p(u_i, v_i) = 0$$

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

Statistik uji :

$$D(\hat{\beta}) = -2 \ln \left(\frac{L(\hat{\omega})}{L(\hat{\Omega})} \right) = 2 (\ln L(\hat{\Omega}) - L(\hat{\omega})) \quad (2.24)$$

Kriteria penolakan adalah tolak H_0 jika nilai $D(\hat{\beta})$ lebih besar dari $X_{(p,a)}^2$ yang artinya minimal terdapat satu parameter dalam model GWNBR yang berpengaruh signifikan terhadap variabel respon. Setelah dilakukan pengujian serentak, dilakukan pengujian secara parsial untuk mengetahui parameter mana saja yang memberikan pengaruh signifikan terhadap variabel respon pada tiap – tiap lokasi dengan hipotesis sebagai berikut :

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

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

Statistik Uji :

$$Z_{hit} = \frac{\hat{\beta}_j(u_i, v_i)}{se(\hat{\beta}_j(u_i, v_i))} \quad (2.25)$$

Dimana, $\hat{\beta}_j$ adalah koefisien model variabel prediktor ke-j, $se(\hat{\beta}_j)$ adalah *standard error* dan estimasi maximum *likelihood*. Daerah penolakan adalah tolak H_0 jika nilai dari $|Z_{hit}|$ lebih besar dari $Z_{\alpha/2}$ yang artinya parameter tersebut berpengaruh signifikan terhadap variabel respon pada tiap – tiap lokasi dalam model GWNBR.

2.10. Tuberkulosis

Tuberkulosis di Indonesia menjadi salah satu penyakit yang angka kasusnya cukup tinggi. Bila dibandingkan dengan Negara lain, Indonesia termasuk Negara yang memiliki banyak penderita tuberkulosis. Tuberkulosis adalah suatu penyakit infeksi yang disebabkan bakteri yang dikenal *Mycrobacterium Tuberculosis*.

Penularan penyakit ini disebabkan oleh penderita tuberculosis melalui udara yaitu dengan batuk, bersin, berbicara, tertawa atau meludah. butiran-butiran air ludah tersebut berterbangan diudara dan terhirup oleh orang sehat dan masuk kedalam paru-paru yang kemudian menyebabkan penyakit tuberculosis. Tuberculosis dibedakan menjadi dua klasifikasi, yaitu tuberculosis pada paru dan tuberculosis ekstra paru. Tuberculosis paru adalah penyakit yang menyerang jaringan paru, sedangkan tuberculosis ekstra paru merupakan penyakit yang menyerang organ tubuh selain paru, diantaranya organ selaput otak, selaput jantung, kelenjar getah bening, tulang, limfa, persendian, kulit, usus, ginjal, saluran kencing dan lain- lain.

Faktor risiko yang berperan dalam kejadian tuberculosis adalah faktor karakteristik individu dan faktor karakteristik lingkungan (sunarni,2009)

1. Faktor karakteristik individu

Faktor karakteristik individu yang menjadi faktor risiko terhadap kejadian tuberculosis adalah faktor umur, jenis kelamin, tingkat pendidikan, pekerjaan, kebiasaan merokok, status gizi, kondisi sosial ekonomi, dan perilaku.

2. Faktor risiko lingkungan

Beberapa faktor lingkungan yang menjadi faktor risiko terhadap kejadian tuberculosis adalah kepadatan penduduk, pencahayaan, ventilasi, kondisi rumah, jumlah sarana kesehatan, kelembapan udara, suhu, dan ketinggian wilayah.

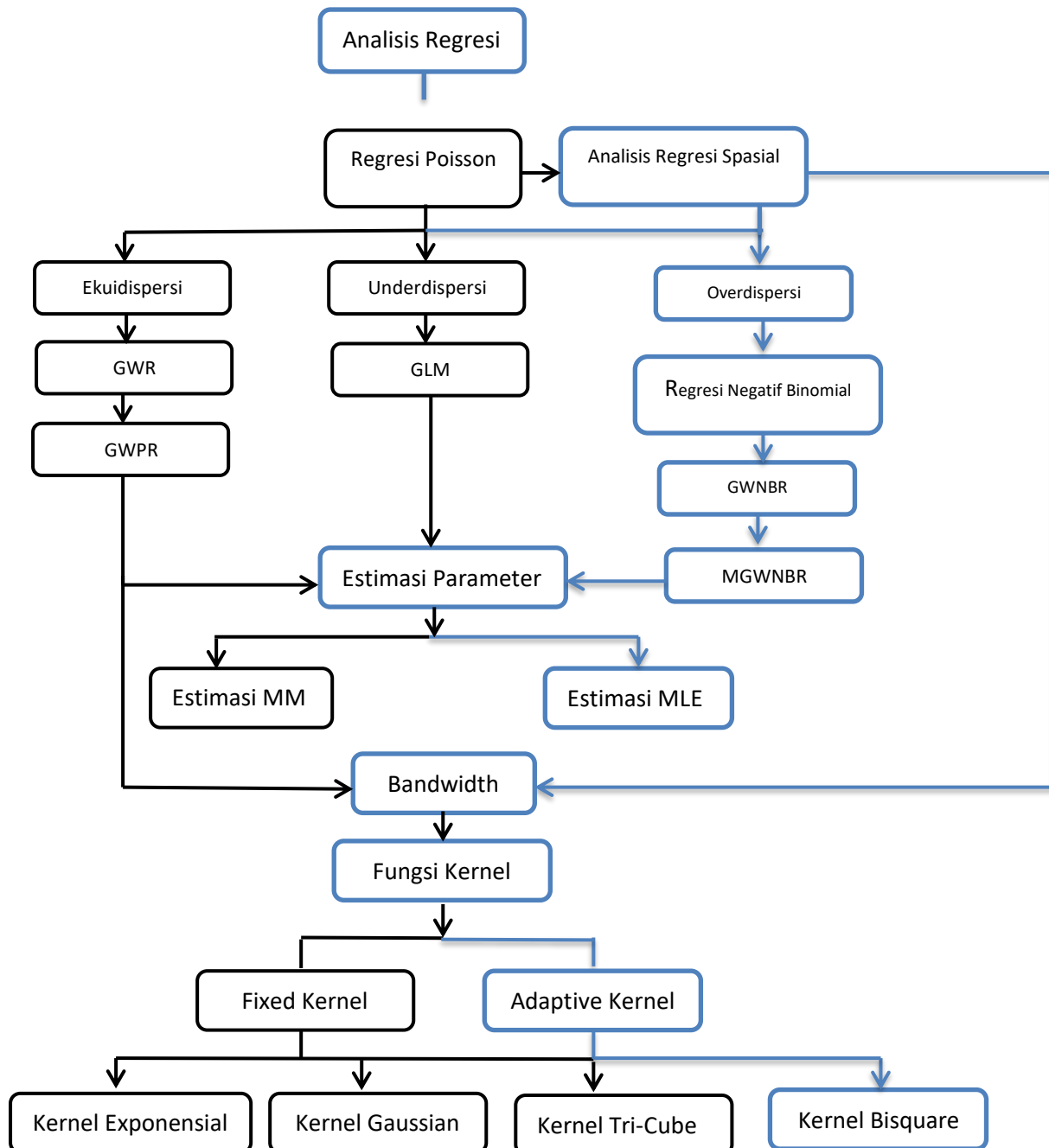
2.11. Penelitian Terdahulu

Nama Peneliti	Tahun	Judul
Brunsdon C, Fotheringham A. S, dan Charlton M.E	1996	<i>Geographically Weighted Regression: A Method for Exploring Spatial Nonstationary. Geographical Analysis</i>
Nakaya T, Fotheringham A.S, Brunsdon C, dan Charlton M.	2005	<i>Geographically Weighted Poisson Regression for Disease Associating Mapping</i>
Ricardo, A., & Carvalho, T.	2014	<i>Geographically Weighted Negative Binomial Regression-Incorporating Overdispersion.</i>
Juniardi L. C dan Salamah M.	2015	<i>Analisis Faktor-Faktor yang mempengaruhi Jumlah Kasus Kusta di Jawa Timur pada Tahun 2013 Menggunakan Geographically Weighted Negative Binomial Regression (GWNBR).</i>
Fitriyah I. I.	2017	Estimasi Parameter Model Geographically Weighted Negative Binomial Regression (GWNBR) dengan

		Pembobot Adaptive Gaussian Kernel.
Yasin H, Suriyani I, Kartikasari P.	2021	<i>Graphical Interface of Geographically Weighted Negative Binomial Regression (GWNBR) model using R-Shiny.</i>
Putera M. L. S, Wahyunita L, dan Yusup F.	2021	<i>Spatial Modelling of Covid-19 Confirmed Cases in Kalimantan, Indonesia: How Neighborhood Matters?</i>

Tabel 2. 1 Penelitian Terdahulu

2.12. Kerangka Konseptual



Gambar 2. 1 Kerangka Konseptual