

**MODEL REGRESI SPASIAL DATA KECELAKAAN LALU
LINTAS UNTUK PREDIKSI FREKUENSI KLAIM
SULAWESI SELATAN**

SKRIPSI



NUR ALIAH AMIN

H081191007

**PROGRAM STUDI ILMU AKTUARIA
DEPARTEMEN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS HASANUDDIN
2023**

**MODEL REGRESI SPASIAL DATA KECELAKAAN
LALU LINTAS UNTUK PREDIKSI FREKUENSI
KLAIM SULAWESI SELATAN**

SKRIPSI

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

NUR ALIAH AMIN

H081191007

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

2023

LEMBAR PERNYATAAN KEOTENTIKAN

Yang bertanda tangan di bawah ini:

Nama : Nur Aliah Amin

Nim : H081191007

Program Studi : Ilmu Aktuaria

Jenjang : S1

Menyatakan dengan ini bahwa karya tulisan saya yang berjudul

MODEL REGRESI SPASIAL DATA KECELAKAAN LALU LINTAS UNTUK PREDIKSI FREKUENSI KLAIM SULAWESI SELATAN

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

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

Makassar, 1 Agustus 2023

Yang menyatakan,



Nur Aliah Amin

Nur Aliah Amin

NIM: H081191007

**MODEL REGRESI SPASIAL DATA KECELAKAAN LALU
LINTAS UNTUK PREDIKSI FREKUENSI KLAIM
SULAWESI SELATAN**

Disusun dan diajukan oleh

NUR ALIAH AMIN

H081191007

UNIVERSITAS HASANUDDIN

Menyetujui,

Pembimbing Utama

Pembimbing Pertama


Dr. Amran, S.Si., M.Si.
NIP.197011011998021001


Edy Saputra Rusdi, S.Si., M.Si.
NIP.199104102020053001



Kepala Program Studi


Dr. Amran, S.Si., M.Si.
NIP.197011011998021001

HALAMAN PENGESAHAN

Skripsi ini diajukan oleh :

Nama : Nur Aliah Amin

NIM : H081191007

Program Studi : Ilmu Aktuaria

Judul Skripsi : Model Regresi Spasial Data Kecelakaan Lalu Lintas Untuk Prediksi Frekuensi Klaim Sulawesi Selatan.

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

DEWAN PENGUJI

Ketua : Dr. Amran, S.Si., M.Si

Sekretaris : Edy Saputra Rusdi, S.Si., M.Si

Anggota : Mauliddin, S.Si., M.Si.

Anggota : Ainun Mawaddah Abdal, S.Si., M.Si

Tanda Tangan



(.....)



(.....)

Ditetapkan di : Makassar

Tanggal : 1 Agustus 2023



KATA PENGANTAR

Assalamu'alaikum Warahmatullahi Wabarakatuh

Puji syukur penulis panjatkan kepada Allah SWT, berkat limpahan rahmat, hidayahnya, dan karunia-Nya. Sholawat serta salam semoga senantiasa tercurah kepada junjungan kita semua baginda Rasulullah *Shallallahu'alaihi Wasallam* yang mengantarkan manusia dari zaman kegelapan ke zaman yang terang berderang. Alhamdulillahirobbil'alamiin berkat rahmat dan hidayah-Nya penulis dapat menyelesaikan skripsi yang berjudul. “***Model Regresi Spasial Data Kecelakaan Lalu Lintas Untuk Prediksi Frekuensi Klaim Sulawesi Selatan***”. Penulisan skripsi ini dilakukan dalam rangka memenuhi syarat dalam menyelesaikan pendidikan starta satu (S1) Sarjana Sains di Departemen Matematika, Fakultas Matematika dan Ilmu Pengerahuan Alam, Universitas Hasanuddin.

Pada kesempatan ini penulis menyampaikan terima kasih yang sebesar-besarnya kepada kedua orang tua tercinta, Ayahanda **Muh Amin** dan Ibunda **Ramlah**, yang selalu memberikan doa, dukungan, dan kasih sayang yang tak pernah putus sehingga penulis dapat menyelesaikan skripsi ini.

1. Bapak **Prof. Dr. Ir. Jamaluddin Jompa, M.Sc.** selaku Rektor Universitas Hasanuddin.
2. Bapak **Dr. Eng. Amiruddin, M.Si.** selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam beserta jajarannya.
3. **Bapak Prof. Dr. Nurdin, S.Si., M.Si** selaku Ketua Departemen Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin serta Bapak dan Ibu **Dosen Prodi Ilmu Aktuaria** yang telah memberikan banyak ilmu dan pengetahuan kepada penulis selama menjadi mahasiswa di Program Studi Ilmu Aktuaria, serta para **Staf Departemen Matematika** yang telah membantu dan memudahkan penulis dalam berbagai hal administrasi.
4. Bapak **Dr. Amran, S.Si., M.Si.**, selaku Ketua Program Studi Ilmu Aktuaria sekaligus Penasehat Akademik dan Pembimbing Utama penulis yang senantiasa sabar, tulus, dan ikhlas meluangkan banyak

waktu di tengah kesibukan dan prioritasnya untuk membimbing dan memberi masukan serta motivasi dalam penulisan skripsi ini.

5. Bapak **Edy Saputra Rusdi, S.Si., M.Si** selaku Pembimbing Pertama penulis yang senantiasa sabar, tulus, dan ikhlas meluangkan banyak waktu di tengah kesibukan dan prioritasnya untuk membimbing dan memberi masukan serta motivasi dalam penulisan skripsi ini.
6. Dosen Penguji, Bapak **Mauliddin, S.Si., M.Si** dan Ibu **Ainun Mawaddah Abdal, S.Si., M.Si.**, yang telah meluangkan waktunya sejak seminar proposal hingga sidang skripsi untuk memberikan saran dan masukan dalam proses penulisan skripsi penulis.
7. **Diah Eka Damayanti** sebagai sahabat tercinta yang telah kebersamai penulis dari SMP sampai bangku perkuliahan. Terima kasih selalu menemani meluangkan waktunya, mendukung ataupun menghibur dalam sedih dan memberikan semangat untuk terus maju tanpa kenal kata menyerah dalam segala hal untuk meraih apa yang menjadi impian saya.
8. Sahabat “**GengBell**” yaitu, Elsa, Ekky, Ammi, Firda, Fadillah, Kiki yang selalu ada dalam suka maupun duka, memberikan dukungan doa dan semangat satu sama lain sejak SMP, perkuliahan hingga selesainya penulisan skripsi.
9. **Nena, Marhama, Neno, Muhajir dan Iyal**, yang telah menjadi teman dari bangku SMA sampai bangku perkuliahan saling menguatkan dan menyemangati satu sama lain.
10. Terima kasih kepada **Syarifah Fatimiyah** telah menjadi teman dibangku perkuliahan yang senantiasa saling menyemangati dan kebersamai dari proses perkuliahan, bimbingan skripsi sampai dengan selesainya skripsi ini.
11. Terima kasih kepada **Rahmi, Tasya dan Putri** telah menjadi teman yang saling menguatkan dan menyemangati satu sama lain .
12. Teman-teman seperjuangan **Ilmu Aktuaria 2019** kepada **Ayu, Ica, Yuyu, Meli, Uyuun, Ona, Adrian, Wahyu, Firja, Lintar, Ageng,**

Lija, Chika. Terima kasih atas kebersamaannya dan bantuan yang diberikan kepada penulis selama ini dalam menjalani rutinitas perkuliahan dan mendukung dalam menyelesaikan skripsi.

13. Teman-teman seperjuangan KKN UNHAS GEL.108 Bowong Cindea yang telah menjadi teman serta keluarga baru dan semoga kedepannya silaturahmi yang telah dibangun bersama tetap terjalin dengan baik.
14. Teman-teman POL190N dan Indira yang telah memberikan pengalaman berharga selama masa perkuliahan.
15. Terakhir, Saya ingin berterima kasih kepada diri saya sendiri karena telah mampu berusaha keras dan berjuang sejauh ini. Mampu mengendalikan diri dari berbagai tekanan diluar keadaan dan tak pernah memutuskan menyerah sesulit apapun proses penyusunan skripsi ini hingga selesainya skripsi .

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna dikarenakan terbatasnya pengalaman dan pengetahuan yang dimiliki penulis. Oleh karena itu, penulis mengharapkan segala bentuk saran serta masukan kritik yang membangun dari berbagai pihak. Semoga skripsi ini dapat bermanfaat bagi yang membacanya, terutama untuk pengembangan ilmu pengetahuan.

Makassar, 01 Agustus 2023



Nur Aliah Amin

**PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK
KEPENTINGAN AKADEMIS**

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

Nama : Nur Aliah Amin

Nim : H0811191007

Program Studi : Ilmu Aktuaria

Departemen : Matematika

Fakultas : Matematika dan Ilmu Pengetahuan Alam

Jenis Karya : Skripsi

demikian pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Hasanuddin **Hak Bebas Royalti Noneksklusif (*Non-exclusive Royalty- Free Right*)** atas karya ilmiah saya yang berjudul:

**Model Regresi Spasial Data Kecelakaan Lalu Lintas Untuk Prediksi
Frekuensi Klaim Sulawesi Selatan**

Beserta perangkat yang ada (jika diperlukan). Terkait dengan hal di atas, maka pihak universitas berhak menyimpan, mengalih-media/format-kan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan memublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di Makassar Pada Tanggal 01 Agustus 2023

Yang menyatakan


(Nur Aliah Amin)

ABSTRAK

Provinsi Sulawesi Selatan memiliki angka kecelakaan lalu lintas yang cukup tinggi. Salah satu upaya untuk mengurangi tingkat kecelakaan itu ialah dengan mencari penyebab utama terjadinya kecelakaan. Penelitian ini membahas tentang pemodelan dan prediksi frekuensi klaim kecelakaan di Provinsi Sulawesi Selatan. Model yang digunakan ialah Model *Geographically Weighted Regression* (GWR) dan variabel prediktor ialah data jumlah penduduk, rasio jenis kelamin, dan jumlah usia anak. Hasil penelitian ini menunjukkan terdapat keberagaman nilai parameter di setiap kabupaten/kota. Variabel prediktor memiliki pengaruh yang signifikan terhadap rata-rata frekuensi klaim ialah Jumlah penduduk dan jumlah penduduk usia lebih dari 15 tahun. Hasil prediksi frekuensi klaim kecelakaan lalu lintas di Provinsi Sulawesi Selatan berdasarkan hasil model GWR memiliki nilai rata-rata residual sebesar 0.21 dan Nilai koefisien determinasi sebesar 87%. Prediksi frekuensi klaim di Sulawesi Selatan yang tertinggi berada di Kota Makassar sebesar 1062 klaim dan nilai prediksi terendah berada di Kepulauan Selayar sebesar 97 klaim.

Kata Kunci: Kecelakaan lalu lintas, Model *Geographically Weighted Regression*, Prediksi Frekuensi Klaim, Rata-rata residual, Koefisien Determinasi.

ABSTRACT

South Sulawesi province has a fairly high number of traffic accidents. One effort to reduce the number of accidents is to find the main causes of accidents. This study discusses modeling and predicting the frequency of accident claims in South Sulawesi Province. The model used is the Geographically Weighted Regression (GWR) Model and the predictor variables are population data, sex ratio, and number of children's ages. The results of this study indicate that there are variations in parameter values in each district/city. The predictor variable has a significant effect on the average frequency of claims, namely the number of residents and the total population aged over 15 years. The prediction results for the frequency of traffic accident claims in South Sulawesi Province based on the results of the GWR model have an average residual value of 0.21 and a coefficient of determination of 87%. The highest claim prediction frequency in South Sulawesi is in Makassar City with 1062 claims and the lowest predicted value is in the Selayar Islands with 97 claims.

Keywords: *Number of Accidents, Geographically Weighted Regression Model, Claim Frequency, Average residual, Coefficient of Determination.*

DAFTAR ISI

HALAMAN JUDUL.....	ii
LEMBAR PERNYATAAN KEOTENTIKAN.....	iii
HALAMAN PENGESAHAN.....	v
KATA PENGANTAR	vi
PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS	ix
ABSTRACT.....	xi
DAFTAR ISI.....	xii
DAFTAR TABEL.....	xiv
DAFTAR GAMBAR	xv
DAFTAR LAMPIRAN	xvi
DAFTAR NOTASI	xvii
BAB I	1
PENDAHULUAN	1
I.1 Latar Belakang.....	1
I.2 Rumusan Masalah.....	3
I.3 Tujuan Penelitian	3
I.4 Batasan Masalah.....	3
I.5 Manfaat Penelitian	3
I.6 Sistematika Penulisan	4
BAB II.....	5
TINJAUAN PUSTAKA	5
II.1 Statistik Deskriptif.....	5
II.2 Uji Efek Spasial.....	5
II.1.2 Uji Dependensi Spasial.....	6
II.2.2 Heterogenitas Spasial	6
II.3 Model <i>Geographically Weighted Regression</i> (GWR).....	7
II.4 <i>Bandwidth</i>	7
II.5 Penaksiran Parameter Model <i>Geographically Weighted Regression</i>	8
II.6 Pembobot Model <i>Geographically Weighted Regression</i>	8
II.7 Pengujian Model <i>Geographically Weighted Regression</i>	10
II.7.1 Uji Parsial Model <i>Geographically Weighted Regression</i>	10
II.8 Model Frekuensi Klaim.....	11

BAB III	13
METODE PENELITIAN.....	13
III.1 Jenis dan Data Penelitian	13
III.2 Metode Pengumpulan data	13
III.3 Metode Analisis Data	13
IV.1 Karakteristik Data	16
IV.1.1 Jumlah Kecelakaan	17
IV.1.2 Jumlah Penduduk	19
IV.1.3 Rasio Jenis Kelamin.....	20
IV.1.4 Jumlah Penduduk Usia 15 Tahun ke Atas (SMA).....	21
IV.2 Uji Efek Spasial	22
IV.2.1 Uji Dependensi Spasial	22
IV.2.2 Heterogenitas	23
IV.3 Model <i>Geographically Weighted Regression</i> (GWR)	23
IV.3.4 Uji Parsial Model <i>Geographically Weighted Regression</i>	25
IV.3 Model Frekuensi Klaim	26
BAB V.....	31
KESIMPULAN DAN SARAN.....	31
V.1 Kesimpulan	31
DAFTAR PUSTAKA	32
LAMPIRAN.....	34

DAFTAR TABEL

Tabel 4.1 Statistik deskriptif Data Kecelakaan lalu lintas Direktorat Kepolisian Daerah Sulawesi Selatan , Jumlah penduduk, Rasio Jenis Kelamin dan Jumlah Penduduk Usia 15 Tahun ke atas (SMA) Badan Pusat Statistik Sulawesi Selatan	16
Tabel 4.2 Nilai Morans'I	22
Tabel 4.3 Nilai Breush-Pangan (BP)	23
Tabel 4.4 Bandwidth untuk setiap lokasi	24
Tabel 4.4 Bandwidth untuk setiap lokasi(Lanjutan).....	25
Tabel 4.5 Nilai t_{hitung} di setiap kabupaten/kota	26
Tabel 4.6 Model GWR di setiap kabupaten/kota	27
Tabel 4.7 Prediksi Frekuensi Klaim	28

DAFTAR GAMBAR

Gambar 1. <i>Flowchart</i>	15
Gambar 4.1 Jumlah Kecelakaan Lalu Lintas Sulawesi Selatan 2021.....	18
Gambar 4.2 Jumlah Penduduk Sulawesi Selatan 2021.....	20
Gambar 4.3 Rasio Jenis Kelamin Sulawesi Selatan 2021	24
Gambar 4.4 Jumlah Penduduk Usia 15 tahun ke atas (SMA) 2021	25

DAFTAR LAMPIRAN

Lampiran 1. Data Jumlah Kecelakaan Lalu Lintas Sulawesi Selatan tahun 2021, Jumlah penduduk Badan Pusat Statistik Sulawesi, Rasio Jenis Kelamin Badan Pusat Statistik Sulawesi dan Jumlah Penduduk Usia 15 Tahun ke atas (SMA) Badan Pusat Statistik Sulawesi Selatan.....	34
Lampiran 2. Statistik Deskriptif Variabel Respon dan Prediktor.....	35
Lampiran 3. Hasil Pengujian Morans I dengan Software R.....	36
Lampiran 4. Hasil Pengujian <i>Breusch – pagan</i> dengan Software R.....	37
Lampiran 5. Outpot nilai Bandwidth untuk setiap lokasi pengamatan.....	38
Lampiran 6. Jarak <i>Euclidean</i> antar Kabupaten/Kota.....	39
Lampiran 6. Jarak <i>Euclidean</i> antar Kabupaten/Kota (lanjutan).....	40
Lampiran 6. Jarak <i>Euclidean</i> antar Kabupaten/Kota (lanjutan).....	41
Lampiran 7. Model GWR dengan Software R.....	42
Lampiran 7. Model GWR dengan Software R (lanjutan).....	43
Lampiran 7. Model GWR dengan Software R (lanjutan).....	44
Lampiran 8. Perhitungan Model GWR disetiap kabupaten/kota.....	45
Lampiran 8. Perhitungan Model GWR disetiap kabupaten/kota (Lanjuta).....	46
Lampiran 9. Estimasi Parameter Model GWR.....	47
Lampiran 9. Syntax Menggunakan Software R.....	48
Lampiran 9. Syntax Menggunakan Software R (lanjutan).....	49
Lampiran 9. Syntax Menggunakan Software R (lanjutan).....	50
Lampiran 9. Syntax Menggunakan Software R (lanjutan).....	51

DAFTAR NOTASI

Notasi	Keterangan
s^2	<i>Sample variance.</i>
x_i	Banyaknya data.
e_i	<i>Error</i> untuk pengamatan ke- i .
σ^2	<i>Variance error.</i>
Z	Standarisasi variabel x .
y_i	Variabel respon untuk lokasi ke- i ($i = 1, 2, \dots, n$).
x_{ik}	Variabel prediktor ke- k pada lokasi ke- i ($i = 1, 2, \dots, n$).
(u_i, v_i)	Koordinat lintang dan bujur dari titik ke- i .
$\beta_0(u_i, v_i)$	Nilai <i>intercept</i> model regresi GWR.
ε_i	<i>Error</i> ke- i .
$W(u_i, v_i)$	Matriks pembobot.
$\hat{\beta}_k(u_i, v_i)$	Koefisien regresi koordinat letak geografis (bujur, lintang).
$SE\hat{\beta}_k(u_i, v_i)$	Standar <i>error</i> pada penduga parameter β_k .
x_{ij}	Nilai observasi variabel prediktor ke- j pada pengamatan lokasi (u_i, v_i) .
μ_i	Nilai tengah model Poisson pada lokasi ke- i .
d_{ij}	Jarak antara titik lokasi pengamatan ke- i dengan titik lokasi pengamatan ke- j .
b	<i>Bandwith</i>
$b_{i(q)}$	<i>Bandwith adaptif</i> yang menentukan q sebagai jarak tetangga terdekat dari titik lokasi ke- i .
$\beta_0(u_i, v_i)$	Nilai <i>intercept</i> model regresi GWR.

BAB I

PENDAHULUAN

I.1 Latar Belakang

Sulawesi Selatan merupakan salah satu provinsi dengan jumlah wilayah yang padat. Seiring bertambahnya jumlah penduduk setiap tahun menyebabkan semakin meningkatnya transportasi. Transportasi merupakan sektor yang sangat penting bagi masyarakat. Namun, ketergantungan masyarakat Provinsi Sulawesi Selatan terhadap sektor transportasi berdampak meningkatnya kecelakaan lalu lintas. Menurut Saputra (2017) kecelakaan lalu lintas didefinisikan sebagai peristiwa di jalan yang tidak diduga dan tidak disengaja melibatkan kendaraan dengan tanpa pengguna jalan lain mengakibatkan korban manusia dan kerugian harta benda. Kepolisian Republik Indonesia (POLRI) mencatat 98.419 kecelakaan pada tahun 2017. Secara keseluruhan *World Health Organization* (WHO) mencatat hampir 3.400 orang meninggal dunia akibat kecelakaan lalu lintas dan mengalami peningkatan setiap tahun (Fisu, 2019).

Melihat pertumbuhan jumlah kejadian dan korban kecelakaan lalu lintas yang meningkat tiap tahunnya, Pemerintah Indonesia berupaya menyediakan beberapa peraturan sebagai jaminan perlindungan terhadap korban kecelakaan lalu lintas di Indonesia, yang diatur dalam Undang-Undang (UU) Nomor 33 Tahun 1964 tentang Dana Pertanggungjawaban Wajib Kecelakaan Penumpang dan Nomor 34 Tahun 1964 tentang Dana Kecelakaan Lalu Lintas Jalan. Kemudian pada undang-undang nomor 40 Tahun 2004 mengatur tentang sistem jaminan sosial nasional untuk memberikan bantuan dalam bentuk jaminan sosial yang akan diberikan pada mereka yang mengalami kecelakaan lalu lintas jalan.

Tiap daerah memiliki karakteristik geografis yang berbeda antara daerah satu dengan daerah lainnya. Geografi adalah salah satu ilmu yang mempelajari interaksi dan keterkaitan antara objek dalam suatu wilayah, hal ini memungkinkan permasalahan kecelakaan lalu lintas dapat dikaji dari perspektif geografi (Wijayanti, 2014). Hal tersebut menyebabkan faktor-faktor yang mempengaruhi kecelakaan antara daerah satu dan daerah lainnya diduga dipengaruhi oleh aspek spasial. Metode regresi yang menggunakan pertimbangan geografis

adalah metode *Geographically Weighted Regression* (GWR). Menurut Fotheringham dkk (2002) sebuah metode untuk menganalisis data spasial dengan mempertimbangkan faktor geografis yang kemudian dikenal dengan nama *Geographically Weighted Regression* (GWR). Model GWR merupakan salah satu metode pendekatan titik yaitu pendekatan berdasarkan posisi koordinat garis lintang (*latitude*) dan garis bujur (*longitude*). Pada model GWR, nilai – nilai taksiran parameter di setiap lokasi berbeda – beda.

Model *Geographically Weighted Regression* (GWR) ini belum pernah digunakan untuk menganalisis hubungan regresi data kecelakaan lalu lintas dengan klaim asuransi kecelakaan. Namun beberapa penelitian tentang *Geographically Weighted Regression* (GWR) pada kasus kecelakaan lalu lintas telah dilakukan oleh Utari dkk (2019) menyatakan bahwa pemodelan jumlah kecelakaan lalu lintas di Provinsi Bali pembobot *adaptive kernel gaussian* lebih akurat dibandingkan menggunakan *fixed kernel gaussian*. Penelitian juga dilakukan oleh Dewi dan Zain (2016) dengan judul “Pemodelan faktor penyebab kecelakaan lalu lintas berdasarkan metode *Geographically Weighted Regression* di Jawa Timur”. Penelitian ini dihasilkan bahwa metode *Geographically Weighted Regression* (GWR) lebih akurat dibandingkan menggunakan metode *Ordinary Least Square* (OLS). Selanjutnya Penelitian yang dilakukan oleh Agustianto dkk (2018) dengan judul “Pemodelan faktor penyebab kecelakaan lalu lintas di Kalimantan Barat dengan metode *Geographically Weighted Regression* (GWR)”. Hasil dari penelitian ini bahwa metode GWR dengan menggunakan model *Fixed Gaussian* lebih akurat dibandingkan menggunakan regresi linear. Selanjutnya penelitian yang dilakukan oleh Karim dan Mutaqin (2020) berjudul “Pemodelan frekuensi klaim asuransi kendaraan bermotor Indonesia menggunakan using *Generalized Poission - Lindley Linear Model*”. Hasil dari penelitian ini bahwa data frekuensi klaim asuransi kendaraan bermotor kategori 4 wilayah untuk polis tahun 2013 di Indonesia.

Berdasarkan data yang diperoleh dari Direktorat lalu lintas Kepolisian Daerah Sulawesi Selatan yaitu jumlah kendaraan sangat mempengaruhi tingkat kecelakaan lalu lintas. Oleh karena itu penulis tertarik untuk melakukan penelitian terhadap tingkat kecelakaan lalu lintas di Provinsi Sulawesi Selatan dengan

memakai variabel jumlah kecelakaan, kepadatan penduduk, rasio jenis kelamin dan jumlah usia yang digunakan untuk mengetahui frekuensi klaim asuransi di Sulawesi selatan. Masalah yang ingin diselesaikan dalam penelitian ini adalah pemodelan data kecelakaan lalu lintas menggunakan *Geographically Weighted Regression* (GWR) dan prediksi frekuensi klaim di Sulawesi Selatan. Adapun Judul Penelitian ini adalah **“Model Regresi Spasial Data Kecelakaan Lalu Lintas Untuk Prediksi Frekuensi Klaim Sulawesi Selatan”**

I.2 Rumusan Masalah

Berdasarkan uraian latar belakang tersebut di atas, maka penulis mengangkat permasalahan sebagai berikut:

1. Bagaimana pemodelan data kecelakaan lalu lintas di Provinsi Sulawesi Selatan dengan menggunakan metode *Geographically Weighted Regression* (GWR)?
2. Bagaimana prediksi frekuensi klaim di Provinsi Sulawesi Selatan?

I.3 Tujuan Penelitian

Berdasarkan rumusan masalah, maka tujuan dari peneliti adalah sebagai berikut:

1. Memodelkan data kecelakaan lalu lintas di Provinsi Sulawesi Selatan dengan menggunakan metode *Geographically Weighted Regression*.
2. Memperdiksi frekuensi klaim di Provinsi Sulawesi Selatan.

1.4 Batasan Masalah

Batasan masalah penelitian ini diantaranya sebagai berikut:

1. Data yang digunakan merupakan data kecelakaan Direktorat Lalu Lintas Kepolisian Daerah Sulawesi Selatan dan Badan Pusat Statistik Sulawesi Selatan.
2. Data jumlah kecelakaan terdiri dari 24 kabupaten/kota pada Provinsi Sulawesi Selatan.
3. Metode *Geographically Weighted Regression* (GWR)

I.5 Manfaat Penelitian

Adapun beberapa manfaat yang diharapkan dari peneliti diantaranya adalah sebagai berikut:

1. Bagi peneliti sendiri yaitu sebagai sarana untuk mengaplikasi ilmu yang diperoleh pada masa perkuliahan tentang persamaan regresi.

2. Sebagai referensi bagi peneliti yang lain ingin melakukan penelitian yang berkaitan dengan model regresi spasial.

I.6 Sistematika Penulisan

Adapun sistematika penulisan yang dibagi menjadi lima bab, sebagai berikut:

1. BAB I PENDAHULUAN

Pada bab ini. berisikan latar belakang penelitian, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian dan sistematika penulisan.

2. BAB II KAJIAN PUSTAKA

Berisi mengenai teori-teori yang berhubungan dengan pembahasan penelitian yaitu statistik deskriptif, uji efek spasial, model (GWR), Pengujian model *Geographically Weighted Regression*, model frekuensi klaim.

3. BAB III METODE PENELITIAN

Bab ini berisikan tentang pendekatan penelitian, jenis dan sumber data dan metode analisis.

4. BAB IV ANALISIS DATA DAN PEMBAHASAN

Pada bab ini menjelaskan pembahasan mengenai model *Geographically Weighted Regression*, model frekuensi klaim.

5. BAB V KESIMPULAN DAN SARAN

Pada bab terakhir penelitian berisi kesimpulan mengenai hasil dari pembahasan dan saran untuk penelitian selanjutnya.

BAB II

TINJAUAN PUSTAKA

II.1 Statistik Deskriptif

Menurut Solikhah (2016) menyatakan bahwa statistik deskriptif merupakan bagian dari ilmu statistik yang pengerjaannya untuk mengatur, mengelola data dan memberikan keterangan gambar yang jelas mengenai suatu kejadian. Statistik deskriptif dalam penelitian ini digunakan untuk mendeskripsikan karakteristik data kecelakaan Direktorat lalu lintas Kepolisian Daerah Sulawesi Selatan yaitu koefisien varians, rata-rata, nilai maksimum dan nilai minimum. Rumus dari rata-rata sebagai berikut:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

Keterangan:

x_i = data ke- i .

n = banyaknya data.

Selain rata-rata, *variens* juga digunakan untuk menjelaskan keragaman data.

Berikut rumus dari *variens*.

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$$

Keterangan:

s^2 = *sample variens*.

$\bar{x}$ = nilai rata-rata.

x_i = banyaknya data ke- i .

Sedangkan nilai maksimum dan nilai minimum menyatakan nilai tertinggi dan terendah dari suatu variabel.

II.2 Uji Efek Spasial

Menurut Rajabidfard dan Williamson (2000) menyatakan bahwa data spasial merupakan data yang membentuk objek yang berhubungan dengan ruang bumi termasuk permukaan bumi, dibawa permukaan bumi, atmosfer, perairan, permukaan bumi dan kelautan. Uji efek spasial terbagi menjadi dua yaitu :

II.1.2 Uji Dependensi Spasial

Menurut Jurniardi dan Salamah (2015), uji dependensi spasial untuk melihat pengamatan di suatu lokasi berpengaruh terhadap lokasi lainnya. Pengujian dependensi spasial dilakukan menggunakan statistik uji *Morans' I* adalah sebagai berikut:

Hipotesis:

$H_0: I = 0$ (tidak terdapat dependensi spasial)

$H_1: I \neq 0$ (ada dependensi spasial)

Statistik uji *Morans' I* adalah:

$$Z_I = \frac{\hat{I} - E(\hat{I})}{\sqrt{\text{var}(\hat{I})}}$$

Keterangan:

$\hat{I}$ = nilai indeks *Moran's I*

Z_I = nilai statistik uji indeks *Moran's I*

$\text{var}(\hat{I})$ = nilai varians dari indeks *Moran's I*

$E(\hat{I})$ = nilai statistik uji indeks *Moran's I*

Kriteria penolakannya adalah tolak H_0 jika nilai $p - \text{value} < \alpha$ yang artinya terdapat dependensi spasial.

II.2.2 Heterogenitas Spasial

Heterogenitas Spasial terjadi akibat adanya perbedaan antara satu wilayah dengan wilayah lainnya. Pengujian heterogenitas spasial menggunakan uji *Breusch-pagan* dalam suatu wilayah penelitian yang pada dasarnya tidak homogen. Adanya heterogenitas spasial dapat menunjukkan bahwa data dapat dimodelkan dengan GWR. Anselin (1998) menyatakan bahwa hipotesis yang mendasari pengujian heterogenitas spasial menggunakan uji *Breusch-pagan* adalah sebagai berikut:

Hipotesis:

$H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_i^2$ (tidak terdapat heterogenitas spasial)

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

Statistik uji *breusch-pagan* (BP) adalah:

$$BP = \frac{1}{2} F^T Z (Z^T Z)^{-1} Z^T f$$

dengan unsur-unsur vektor f adalah:

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

Keterangan:

e_i = error untuk pengamatan ke- i .

σ^2 = variance error.

Z = standarisasi variabel x .

Kriteria penolakannya adalah tolak H_0 jika $BP > \chi_{(p+1)(\alpha)}^2$ dimana $\chi_{(p+1)(\alpha)}^2$ merupakan titik kritis uji χ^2 dengan taraf α dan derajat bebas $(p + 1)$, dengan menolak H_0 dapat dilakukan bahwa terdapat heterogenitas spasial (Rosa, 2015).

II.3 Model Geographically Weighted Regression (GWR)

Model GWR adalah model regresi sederhana yang diubah menjadi model regresi yang berbobot. Setiap nilai parameter akan dihitung pada setiap lokasi geografis sehingga setiap titik lokasi geografis mempunyai nilai parameter regresi yang berbeda – beda. Hal ini akan memberikan variansi pada nilai parameter regresi di suatu kumpulan wilayah geografis. Jika nilai parameter regresi konstan pada tiap-tiap wilayah geografis, maka model GWR adalah model global. Artinya tiap – tiap wilayah geografis mempunyai model yang sama.

Model umum untuk model GWR adalah (Fotheringham dkk 2002):

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

Keterangan:

y_i = variabel respon untuk lokasi ke- i .

x_{ik} = variabel prediktor ke- k pada lokasi ke- i .

(u_i, v_i) = koordinat lintang dan bujur dari titik ke- i pada suatu lokasi geografis.

$\beta_0(u_i, v_i)$ = nilai *intercept* model regresi GWR.

$\beta_k(u_i, v_i)$ = Koefisien regresi variabel prediktor ke- k untuk ke- i .

ε_i = error ke- i

II.4 Bandwidth

Bandwidth adalah ukuran jarak fungsi pembobot dan sejauh mana pengaruh lokasi terhadap lokasi lain. Secara teoritis *bandwidth* merupakan lingkaran dengan

radius b dari titik pusat lokasi, dimana digunakan sebagai dasar menentukan bobot setiap pengamatan yang terletak dekat lokasi i maka akan lebih berpengaruh dalam membentuk parameter model lokasi i .

Bandwidth optimal diperoleh dengan menggunakan *cross validation* (CV). Jika nilai CV semakin kecil, maka didapatkan bandwidth yang optimum (Fotheringham dkk 2002) dengan rumus sebagai berikut:

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

Keterangan:

i = lokasi ke- i .

b = bandwidth.

$\hat{y}_{\neq i}(b)$ = nilai taksiran dari model regresi tanpa pengamatan ke- i .

II.5 Penaksiran Parameter Model *Geographically Weighted Regression*

Model GWR merupakan model regresi yang menggunakan pendekatan titik sehingga penaksiran parameter yang diperoleh bersifat lokal. Penaksiran parameter model GWR dilakukan dengan metode kuadrat terkecil terboboti *Weighted Least Square* (WLS) dengan memberikan matriks pembobot yang berbeda untuk setiap lokasi pengamatan (Fotheringham dkk 2002).

Bentuk penaksiran parameter pada lokasi (u_i, v_i) dengan GWR adalah sebagai berikut:

$$\hat{\beta}(u_i, v_i) = (X^T W(u_i, v_i) X)^{-1} X^T W(u_i, v_i) Y$$

Keterangan:

X = matriks data dari variabel prediktor.

Y = vektor variabel respon.

$W(u_i, v_i)$ = matriks pembobot.

II.6 Pembobot Model *Geographically Weighted Regression*

Fotheringham dkk (2002), menyatakan bahwa Pembobot dalam GWR merupakan aspek penting pembobot tersebut bergantung pada jarak antara titik lokasi

pengamatan. Fungsi dari matriks pembobot adalah untuk menentukan parameter yang berbeda pada setiap titik lokasi pengamatan.

Matriks pembobot pada GWR merupakan matriks pembobot berbasis pada lokasi titik pengamatan ke- i dengan titik lokasi pengamatan lainnya. Lokasi pengamatan ke- i umumnya diasumsikan sebagai pengaruh paling besar terhadap penaksiran parameter di titik lokasi pengamatan ke- i . Oleh karena itu, matriks pembobot dari $W(u_i, v_i)$ akan semakin besar seperti jarak yang semakin dekat.

Terdapat beberapa cara yang dapat digunakan untuk menentukan nilai pembobot. Salah satu cara yang paling sederhana adalah dengan memberikan bobot sebesar 1 untuk setiap titik lokasi pengamatan i dan j sebagai berikut:

$$w_{ij} = 1, \forall i \text{ dan } j$$

Sehingga, model yang dihasilkan apabila menggunakan fungsi pembobot ini adalah model regresi linear klasik.

Pembobot dalam GWR juga didapatkan dengan menggunakan fungsi invers jarak sebagai berikut:

$$w_i(u_i, v_i) = \begin{cases} 1 & \text{jika } d_{ij} \leq b \\ 0 & \text{jika } d_{ij} > b \end{cases}$$

d_{ij} adalah jarak antara titik lokasi pengamatan ke- i dengan titik lokasi pengamatan ke- j (Fotheringham dkk (2002) dalam (Maulani, 2013).

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

Dimana b adalah *bandwith* atau lebar jendela yang persamaannya sebagai radius (b) suatu lingkaran, sehingga sebuah titik lokasi pengamatan yang berada dalam radius lingkaran masih dianggap berpengaruh dalam membentuk parameter di titik lokasi ke- i . Fungsi invers jarak akan memberikan bobot sama dengan 1, jika titik lokasi ke- j berada di dalam radius b . Sedangkan jika titik lokasi ke- j berada di luar radius b dari luar titik ke- i , maka fungsi invers jarak akan memiliki bobot sama dengan 0.

Selain itu, matriks pembobot $W(u_i, v_i)$ dapat ditentukan dengan menggunakan fungsi karnel. Fungsi karnel memberikan pembobot sesuai *bandwith* optimum yang nilainya bergantung pada kondisi data.

Fungsi kernel *adaptive* memiliki *bandwidth* yang berbeda – beda pada setiap lokasi pengamatan, dengan fungsi pembobot sebagai berikut:

1. *Fixed kernel gaussian* (Fotheringham dkk 2002)

$$w_i = \frac{1}{\sum_{j=1}^n \exp\left(-\frac{d_{ij}^2}{2h^2}\right)}$$

Keterangan:

w_i = bobot spasial untuk pengamatan ke- i .

d_{ij} = jarak Euclidean antara pengamatan ke i dan ke j .

h = *bandwidth* atau panjang *kernel* yang telah ditentukan sebelumnya.

II.7 Pengujian Model *Geographically Weighted Regression*

II.7.1 Uji Parsial Model *Geographically Weighted Regression*

Pengujian parameter model GWR dilakukan untuk mengetahui parameter mana yang berpengaruh signifikan terhadap variabel respon secara parsial.

Hipotesis (Fotheringham dkk 2002)

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

$$H_0: \text{Minimal ada satu } \beta_k(u_i, v_i) \neq 0$$

Statistik Uji:

$$t_{hitung} = \frac{\hat{\beta}_k(u_i, v_i)}{SE\hat{\beta}_k(u_i, v_i)}$$

Keterangan:

$\hat{\beta}_k(u_i, v_i)$ = koefisien regresi pada koordinat letak geografis (bujur, lintang).

$SE\hat{\beta}_k(u_i, v_i)$ = Standar *error* pada penduga parameter β_k .

t_{tabel} dengan distribusi t menggunakan derajat bebas $df = \frac{\delta_2^2}{\delta_2}$.

Daerah penolakan:

Tolak H_0 jika $t_{hitung} > t_{tabel}$ dengan taraf signifikan α 0.05. Hal ini berarti ada pengaruh signifikan antara variabel prediktor dengan variabel respon.

II.8 Model Frekuensi Klaim

Model frekuensi klaim di dapatkan dari model GWR. Model GWR yaitu setiap parameter dihitung pada setiap lokasi pengamatan, sehingga setiap lokasi pengamatan mempunyai nilai pengamatan regresi yang berbeda-beda. Model GWR dapat ditulis sebagai berikut (Fotheringham dkk 2002):

$$y_i = \beta_0(u_i, v_i) + \beta_1(u_i, v_i)x_{i1} + \beta_2(u_i, v_i)x_{i2} + \cdots + \beta_k(u_i, v_i)x_{ik} + \varepsilon_i \quad (2.2)$$

$$\hat{y}_i = \hat{\beta}_0(u_i, v_i) + \hat{\beta}_1(u_i, v_i)x_{i1} + \hat{\beta}_2(u_i, v_i)x_{i2} + \cdots + \hat{\beta}_k(u_i, v_i)x_{ik}$$

Keterangan:

y_i	= variabel respon untuk lokasi ke- i .
x_{ik}	= variabel prediktor ke- k pada lokasi ke- i .
(u_i, v_i)	= koordinat lintang dan bujur dari titik ke- i pada suatu lokasi geografis.
$\beta_0(u_i, v_i)$	= nilai <i>intercept</i> model regresi GWR.
$\beta_k(u_i, v_i)$	= Koefisien regresi variabel prediktor ke- k untuk ke- i .
ε_i	= <i>error</i> ke- i

Model GWR menghasilkan nilai residual untuk observasi dalam dataset yang dinyatakan sebagai selisih antara nilai observasi dari variabel prediktor dan nilai prediksi yang diestimasikan setiap lokasi dalam data, rumus residual untuk model GWR(Fotheringham dkk 2002):

$$e_i = y_i - \hat{y}_i$$

e_i = Residual untuk observasi ke - i

y_i = Nilai variabel respon untuk observasi ke- i

Jika e_i mendekati angka nol, berarti nilai $\hat{y}_i$ hampir sama dengan nilai variabel respon.

Koefisien determinasi sebagai ukuran sejauh mana variasi dalam variabel respon yang dijelaskan oleh model GWR masing-masing lokasi dalam dataset, rumus untuk R^2 dalam model GWR Fotheringham dkk 2002):

$$SST = \sum_{i=1}^n (y_i - \bar{y})^2$$

Keterangan:

n = Jumlah observasi dalam dataset

y_i = Nilai variabel respon untuk observasi ke- i

$\bar{y}_i$ = Rata-rata nilai seluruh nilai y_i dalam dataset.

Jika nilai R^2 berkisaran antara 0 hingga 1. Jika R^2 mendekati 0 artinya model hampir tidak dapat menjelaskan variasi dalam data dan dapat dikatakan model cocok untuk data yang diamati. Jika R^2 mendekati angka 1 menunjukkan bahwa model sangat baik dalam menjelaskan variasi dalam data dan model memberikan prediksi yang sangat baik dan akurat.