



TESIS - SS235401

METODE REGULARISASI UNTUK *SPATIAL POINT PROCESS ON A LINEAR NETWORK* (STUDI KASUS: ANALISIS RISIKO KECELAKAAN LALU LINTAS DI KABUPATEN NGANJUK)

VANDA FITRIYANAH
NRP. 6003221019

Dosen Pembimbing
Dr. Achmad Choiruddin, S.Si., M.Sc.
Dr. Sutikno, S.Si., M.Si.

Program Studi Magister Statistika
Departemen Statistika
Fakultas Sains dan Analitika Data
Institut Teknologi Sepuluh Nopember
Surabaya
2024



TESIS - SSS235401

**METODE REGULARISASI UNTUK *SPATIAL POINT
PROCESS ON A LINEAR NETWORK* (STUDI KASUS:
ANALISIS RISIKO KECELAKAAN LALU LINTAS
DI KABUPATEN NGANJUK)**

**VANDA FITRIYANAH
NRP. 6003221019**

**Dosen Pembimbing
Dr. Achmad Choiruddin, S.Si., M.Sc.
Dr. Sutikno, S.Si., M.Si.**

**Program Studi Magister Statistika
Departemen Statistika
Fakultas Sains dan Analitika Data
Institut Teknologi Sepuluh Nopember
Surabaya
2024**



THESIS- SS235401

**REGULARIZATION METHOD FOR SPATIAL POINT
PROCESS ON A LINEAR NETWORK (CASE STUDY:
TRAFFIC ACCIDENCE RISK ANALYSIS IN NGANJUK
REGENCY)**

**VANDA FITRIYANAH
NRP. 6003221019**

Supervisors

Dr. Achmad Choiruddin, S.Si., M.Sc.

Dr. Sutikno, S.Si., M.Si.

**Master of Statistics Study Program
Department of Statistics
Faculty of Science and Data Analytics
Institut Teknologi Sepuluh Nopember
Surabaya
2024**

LEMBAR PENGESAHAN TESIS

Tesis disusun untuk memenuhi salah satu syarat memperoleh gelar

Magister Statistika (M.Stat)

di

Institut Teknologi Sepuluh Nopember

Oleh:

VANDA FITRIYANAH

NRP: 6003221019

Tanggal Ujian: 17 Juli 2024

Periode Wisuda: September 2024

Disetujui oleh:

Pembimbing:

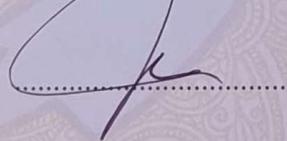
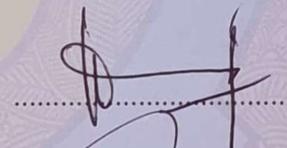
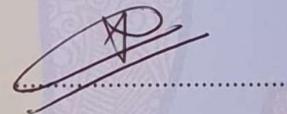
1. Dr. Achmad Choiruddin, S.Si., M.Sc.
NIP: 1991201911101

2. Dr. Sutikno, S.Si., M.Si.
NIP: 19710313 199702 1 001

Penguji:

1. Prof. Dr. Bambang Widjanarko Otok, M.Si.
NIP: 19810223 200812 1 003

2. Dr. Muhammad Ahsan, S.Si.
NIP: 1990202011001



Kepala Departemen Statistika
Fakultas Sains dan Analitika Data

Dr. Dra. Kartika Fithriasari, M.Si
NIP: 19691212 199303 2 002

(Halaman ini sengaja dikosongkan)

**METODE REGULARISASI UNTUK *SPATIAL POINT*
PROCESS ON A LINEAR NETWORK
(STUDI KASUS: ANALISIS RISIKO KECELAKAAN LALU
LINTAS DI KABUPATEN NGANJUK)**

Nama Mahasiswa : Vanda Fitriyanah
NRP : 6003221019
Pembimbing : Dr. Achmad Choiruddin, S.Si., M.Sc.
Co-Pembimbing : Dr. Sutikno, S.Si., M.Si.

ABSTRAK

Point process on a linear network digunakan untuk memodelkan persebaran lokasi objek atau kejadian yang terjadi di jaringan linier, seperti: peta jalur kereta api, sungai, kabel listrik, serabut saraf, rute penerbangan, saluran irigasi, patahan geologi, atau retakan tanah. Seiring dengan perkembangan teknologi pengambilan data, seringkali pengumpulan data lokasi objek atau kejadian disertai dengan data kovariat. Penelitian mengembangkan seleksi variabel untuk pemodelan intensitas *point process on a linear network* berdasarkan metode regularisasi dengan menggabungkan tiga metode, yaitu: Berman-Turner, *Iteratively Reweighted Least Square*, dan *Coordinate Descent Algorithm*. Bentuk *log-likelihood* model *inhomogeneous Poisson point process* dengan pendekatan regresi *Poisson* ditambahkan dengan fungsi regularisasi Lasso. Seleksi variabel dilakukan berdasarkan nilai *tuning parameter* yang berperan untuk mengoptimalkan hasil seleksi variabel menggunakan BIC_{γ} terendah. Hasil pemodelan *inhomogeneous Poisson point process* menggunakan regularisasi Lasso dengan nilai γ optimal dan nilai BIC_{γ} terendah menunjukkan bahwa variabel jumlah jalur, keberadaan kerb kanan, dan keberadaan kerb kiri terseleksi dari model risiko kecelakaan lalu lintas di Kabupaten Nganjuk.

Kata kunci: Kecelakaan Lalu Lintas, Keselamatan di Jalan Raya, *Linear Network*, Regularisasi Lasso, *Spatial Point Process*.

(Halaman ini sengaja dikosongkan)

**REGULARIZATION METHOD FOR SPATIAL POINT
PROCESS ON A LINEAR NETWORK
(CASE STUDY: TRAFFIC ACCIDENT RISK ANALYSIS IN
NGANJUK REGENCY)**

Student Name : Vanda Fitriyanah
Student Identity Number : 6003221019
Supervisor : Dr. Achmad Choiruddin, S.Si, M.Sc.
Co-Supervisor : Dr. Sutikno, S.Si., M.Si.

ABSTRACT

A point processes on a linear network is employed to model the distribution of object locations or events that occur on linear networks, such as: maps of railway lines, rivers, power lines, nerve fibres, flight routes, irrigation channels, geological faults, or ground cracks. The development of data capture technology has led to the collection of object or event location data being accompanied by covariate data. The research developed variable selection for intensity point process on a linear network modeling based on regularization method by combining three methods, namely: Berman-Turner, Iteratively Reweighted Least Square, and Coordinate Descent Algorithm. The log-likelihood form of the inhomogeneous Poisson point process model with a Poisson regression approach is supplemented with a Lasso regularization function. The variable selection based on the value of the tuning parameter (γ), which serves to optimize the variable selection results with the lowest $BIC\gamma$. The result of inhomogeneous Poisson point process using Lasso regularization with optimal γ value and the lowest $BIC\gamma$ value show that the variables number of lanes, presence of right kerb, and presence of left kerb are selected from the traffic accident risk model in Nganjuk Regency.

Keywords: Lasso Regularization, Linear Network, Road Safety, Spatial point Process, Traffic Accidents.

(Halaman ini sengaja dikosongkan)

KATA PENGANTAR

Puji dan syukur bagi Allah SWT yang senantiasa melimpahkan rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan tesis yang berjudul “**Metode Regularisasi untuk *Spatial Point Process on a Linear Network* (Studi Kasus: Analisis Risiko Kecelakaan Lalu Lintas di Kabupaten Nganjuk)**”. Penyusunan tesis ini tidak lepas dari dukungan dan bantuan yang penulis dapatkan dari berbagai pihak. Oleh karena itu, penulis haturkan ucapan terima kasih yang setulus-tulusnya kepada kedua orang tua tercinta, Ayahanda Moh. Syafi’i, Ibunda Alifah dan saudara-saudara yang telah menjadi inspirasi, mendidik, dan membesarkan penulis dengan bertabur cinta, kasih sayang, serta dengan ikhlas telah mengiringi setiap penulis dengan do’a dan restunya. Terima kasih telah menjadi penyemangat dan penghibur bagi penulis dalam menyusun tesis ini.

Penulis juga ingin menyampaikan penghargaan dan ucapan terima kasih yang penuh ketulusan kepada pihak-pihak yang telah membantu sebagai berikut.

1. Institut Teknologi Sepuluh Nopember yang telah memberikan penulis kesempatan untuk melanjutkan studi di jenjang Magister melalui beasiswa “*Fresh Graduate ITS*”.
2. Ibu Dr. Dra. Kartika Fithriasari, M.Si selaku Kepala Departemen Statistika ITS, Ibu Santi Wulan Purnami, M.Si., Ph.D selaku Sekretaris Departemen I (Bidang Akademik, Kemahasiswaan, Penelitian dan Pengabdian Kepada Masyarakat) dan Ibu Dr. Vita Ratnasari, M.Si selaku Sekretaris Departemen II (Bidang Sumber Daya Keuangan, Sumber Daya Manusia, dan Sarana Prasarana) yang telah memfasilitasi sarana dan prasarana untuk kegiatan belajar dan mengajar selama studi di ITS.
3. Bapak Dr.rer.pol. Dedy Dwi Prastyo, S.Si., M.Si selaku Kepala Program Studi Pascasarjana Departemen Statistika yang senantiasa memudahkan segala urusan akademik dan fasilitas yang menunjang di Departemen Statistika ini.
4. Bapak Dr. Achmad Choiruddin, S.Si., M.Sc dan Bapak Dr. Sutikno, S.Si., M.Si selaku dosen pembimbing tesis yang senantiasa meluangkan waktu untuk memberikan bimbingan, saran, masukan serta motivasi kepada penulis dalam menyelesaikan penyusunan tesis ini.

5. Bapak Prof. Dr. Bambang Widjanarko Otok, M.Si. dan Bapak Dr. Muhammad Ahsan, S.Si. selaku dosen penguji yang telah memberikan masukan dan saran untuk menjadikan tesis ini menjadi lebih baik.
6. Bapak/Ibu dosen pengajar dan karyawan di Departemen Statistika ITS yang berperan penting dalam menyelesaikan studi penulis.
7. Rekan-rekan seperjuangan Pasca Sarjana Statistika ITS dan rekan-rekan di pusat kajian PDPM ITS yang telah memberikan pengalaman dan kebersamaan yang sangat berharga.
8. Semua pihak yang tidak bisa disebutkan satu persatu. Penulis menyampaikan terima kasih atas semua dukungan dan bantuan yang diberikan selama menjalani studi.

Penulis meyakini sepenuhnya bahwa dalam penyusunan tesis ini masih jauh dari sempurna. Oleh karena itu, segala saran dan kritik yang membangun dari pembaca sangat penulis harapkan untuk perbaikan tesis ini. Akhir kata, penulis berharap semoga tesis ini dapat memberikan wawasan pengetahuan dan manfaat terhadap pembaca.

Surabaya, Juli 2024

Penulis

DAFTAR ISI

	Halaman
LEMBAR PENGESAHAN	i
ABSTRAK	iii
ABSTRACT	v
KATA PENGANTAR.....	vii
DAFTAR ISI.....	ix
DAFTAR TABEL	xi
DAFTAR GAMBAR.....	xiii
DAFTAR NOTASI.....	xv
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	5
1.3 Tujuan.....	5
1.4 Manfaat Penelitian.....	5
1.5 Batasan Masalah	6
BAB 2 TINJAUAN PUSTAKA.....	7
2.1 Bidang Spasial di \mathbb{R}^2 dan <i>Linear Network</i>	7
2.2 <i>Spatial Point Process</i> di \mathbb{R}^2 dan <i>Linear Network</i>	8
2.3 Karakteristik <i>Point Process on a Linear Network</i>	11
2.3.1 Intensitas	11
2.3.2 Uji Homogenitas	13
2.3.3 <i>Correlation Pixel Image</i>	14
2.3.4 Uji Independensi	15
2.4 <i>Poisson Point Process on a Linear Network</i>	18
2.5 Estimasi Parameter	18
2.5.1 Estimasi θ Berbasis Regresi Poisson.....	19
2.5.2 Seleksi Variabel dengan Regularisasi dan Metode Pemilihan <i>Tuning Parameter</i>	24
2.6 Kecelakaan Lalu Lintas dan Faktor yang Mempengaruhinya	26

BAB 3 METODE PENELITIAN	31
3.1 Sumber Data.....	31
3.2 Variabel Penelitian	32
3.3 Proses Pengambilan Data Kovariat Spasial dan <i>Pre-Processing</i> Data	36
3.4 Struktur Data	41
3.5 Langkah Penelitian.....	44
BAB 4 HASIL DAN PEMBAHASAN	49
4.1 Estimasi Parameter Model <i>Inhomogeneous Poisson Point Process</i>	49
4.1.1 Regresi Poisson	49
4.1.2 Penambahan Metode Lasso pada Pendekatan Regresi Poisson	53
4.2 Penerapan <i>Regularization Poisson Point Process on a Linear Network</i>	55
4.2.1 Peta Sebaran Data Kecelakaan Lalu Lintas di Kabupaten Nganjuk	55
4.2.2 Karakteristik Kovariat Spasial	57
4.2.3 Uji Homogenitas	59
4.2.4 Korelasi <i>Pixel Image</i> dan Uji Independensi	60
4.2.5 Pemodelan Risiko Kecelakaan Lalu Lintas di Kabupaten Nganjuk Menggunakan <i>Inhomogeneous Poisson Point Process</i> Lasso	62
4.2.6 Peta Prediksi Risiko Kecelakaan Lalu Lintas di Kabupaten Nganjuk	67
BAB 5 KESIMPULAN DAN SARAN	69
5.1 Kesimpulan	69
5.2 Saran.....	69
DAFTAR PUSTAKA	71
LAMPIRAN	75
BIODATA PENULIS	89

DAFTAR TABEL

	Halaman
Tabel 2.1 Ilustrasi interpretasi hasil estimasi parameter	12
Tabel 2.2 Tabel kontingensi b x c	15
Tabel 2.3 Tabel probabilitas b x c	16
Tabel 2.4 Algoritma CDA	26
Tabel 3.1 Variabel penelitian kecelakaan lalu lintas	32
Tabel 3.2 Struktur data kecelakaan lalu lintas di Kabupaten Nganjuk	41
Tabel 3.3 Struktur data koordinat lokasi lampu lalu lintas di Kabupaten Nganjuk	43
Tabel 3.4 Struktur data kovariat	43
Tabel 3.5 Struktur data pemodelan intensitas kecelakaan lalu lintas	44
Tabel 4.1 Algoritma <i>Iteratively Reweighted Least Square</i>	53
Tabel 4.2 Hasil uji <i>Cumulative Distribution Function</i> data kecelakaan di Kabupaten Nganjuk.....	60
Tabel 4.3 Hasil uji independensi antar kovariat numerik data kecelakaan di Kabupaten Nganjuk.....	61
Tabel 4.4 Hasil uji independensi antar kovariat kategorik data kecelakaan di Kabupaten Nganjuk.....	61
Tabel 4.4 Hasil estimasi parameter model <i>Inhomogeneous Poisson point process on a linear network</i> tanpa regularisasi	63
Tabel 4.5 Algoritma seleksi variabel dengan menggunakan regularisasi Lasso.....	64
Tabel 4.6 Hasil regularisasi Lasso model <i>Inhomogeneous Poisson point process</i>	65

(Halaman ini sengaja dikosongkan)

DAFTAR GAMBAR

	Halaman
Gambar 1.1 Peta prediksi risiko kecelakaan lalu lintas di jaringan jalan Kabupaten Nganjuk	2
Gambar 1.2 Peta alur penelitian.	5
Gambar 2.1 Bidang spasial pada (a) $W \subset \mathbb{R}^2$ dan (b) <i>linear network</i> $L \subset \mathbb{R}^2$...	7
Gambar 2.2 Ilustrasi <i>linear network</i> L dengan $M = 10$ segmen garis.	8
Gambar 2.3 <i>Oservation window</i> dan struktur data <i>spatial point pattern</i>	9
Gambar 2.4 Ilustrasi jarak pada <i>linear network</i>	10
Gambar 2.5 Ilustrasi <i>pixel image</i>	14
Gambar 2.6 Skema untuk menetapkan bobot ke titik kuadrat	21
Gambar 2.7 Kontur estimasi parameter dengan Lasso	24
Gambar 2.8 Kerangka konseptual penelitian	29
Gambar 3.1 Lokasi kecelakaan lalu lintas di Kabupaten Nganjuk tahun 2018-2020.....	31
Gambar 3.2 Ilustrasi jalur dan lajur di segmen jalan.....	34
Gambar 3.3 Jaringan Jalan Kabupaten Nganjuk	36
Gambar 3.4 Gambaran Ruas Jalan Setapak di Kabupaten Nganjuk	37
Gambar 3.5 Opsi <i>street view</i> pada <i>Google Maps</i>	37
Gambar 3.6 Pemilihan tahun pengambilan foto jalan pada <i>Google Maps</i>	37
Gambar 3.7 Kondisi jalan pada koordinat jalan 7,604810238° LS dan 112,0810013° BT.....	38
Gambar 3.8 Kondisi jalan pada koordinat jalan 7,602650166° LS dan 112,0899963° BT.....	39
Gambar 3.9 Kondisi jalan pada koordinat jalan 7,676750183° LS dan 111,9160004° BT.....	40
Gambar 3.10 Ilustrasi jarak terpendek suatu titik x_i ke lokasi lampu lalu lintas	41
Gambar 3.11 Ilustrasi nilai variabel pada setiap segmen di <i>linear network</i>	44
Gambar 3.12 Diagram analisis data.....	47
Gambar 4.1 Peta persebaran kecelakaan lalu lintas di Kabupaten Nganjuk.	55
Gambar 4.2 Peta persebaran kecelakaan lalu lintas di Kabupaten Nganjuk setiap tahun.....	56
Gambar 4.3 Peta kovariat spasial	57
Gambar 4.4 Ilustrasi <i>quadrature scheme</i> metode Berman-Turner.....	62
Gambar 4.5 Peta prediksi risiko kecelakaan lalu lintas di Kabupaten Nganjuk.	68

(Halaman ini sengaja dikosongkan)

DAFTAR NOTASI

\mathbf{u}	: Vektor titik koordinat lokasi objek atau kejadian
\mathbb{R}^2	: Ruang berdimensi dua
W	: <i>Observation window</i>
$ W $: Luas <i>observation window</i>
\mathbf{x}	: <i>Spatial point pattern</i>
\mathbf{X}	: <i>Spatial point process</i>
$N(\mathbf{X} \cap W)$: Banyaknya kejadian yang berada pada <i>observation window</i> W
L	: <i>Linear network</i>
$ L $: Total panjang <i>linear network</i>
ℓ_j	: Segmen garis ke- j pada <i>linear network</i>
M	: Banyaknya segmen garis pada <i>linear network</i>
$d_L(\mathbf{x}_i, \mathbf{x}_j)$: Jarak terpendek (<i>shortest-path distance</i>) antara titik lokasi \mathbf{x}_i dan \mathbf{x}_j pada <i>linear network</i> L
$\lambda(\boldsymbol{\theta}; \mathbf{u})$: Fungsi intensitas kecelakaan lalu lintas pada <i>linear network</i>
$\boldsymbol{\theta}$: Vektor parameter pada model
p	: Banyaknya kovariat pada model
$\mathbf{z}(\mathbf{u})$: Vektor kovariat yang mempengaruhi kecelakaan lalu lintas
$N(\mathbf{X} \cap B)$: Banyaknya kejadian yang berada pada segmen garis <i>linear network</i> L
$\mu(B)$: Nilai harapan variabel random $N(\mathbf{X} \cap B)$
B	: Segmen garis <i>linear network</i> $B \subset L$
n	: Banyaknya kejadian kecelakaan lalu lintas dalam <i>linear network</i>
d	: Banyaknya <i>dummy point</i>
y_j	: Variabel indikator pada pendekatan <i>quadrature</i>
w_j	: Bobot <i>quadrature</i>

(halaman ini sengaja dikosongkan)

BAB 1

PENDAHULUAN

Bagian ini membahas mengenai latar belakang penelitian, rumusan masalah, tujuan penelitian, manfaat yang diharapkan, batasan masalah, serta kontribusi penelitian. Pembahasan mengenai masing-masing sub bab adalah sebagai berikut.

1.1 Latar Belakang

Spatial point process merupakan model yang digunakan untuk analisis *spatial point pattern*, data *point pattern* umumnya merupakan kumpulan data yang berisi koordinat objek atau kejadian yang diamati pada suatu *observation window*. Contoh data *point pattern* adalah lokasi spesies pohon di hutan (Thurman, et al., 2015), titik pusat gempa bumi (Zhuang, et al., 2002; Choiruddin, et al., 2021), alamat tempat tinggal pasien yang menderita penyakit tertentu (Diggle, 2014; Choiruddin, et al., 2023), serta lokasi kecelakaan lalu lintas (Llagostera, et al., 2022; Hasanah, et al., 2022; Chaudhuri, et al., 2023). Tujuan utama analisis ini adalah untuk memperkirakan fungsi intensitas, seperti dijelaskan oleh Møller dan Waagepetersen (2004). Fungsi intensitas didefinisikan sebagai probabilitas munculnya suatu titik dalam radius yang sangat kecil di sekitar lokasi spasial, yang biasanya diwakili oleh koordinat titik dari suatu kejadian yang sedang diteliti (Baddeley, et al., 2015). Analisis fungsi intensitas bermanfaat untuk memahami distribusi spasial suatu objek atau kejadian.

Seiring dengan perkembangan teknologi dan kebutuhan untuk pengambilan keputusan dengan permasalahan yang kompleks, seringkali distribusi *spatial point pattern* dihubungkan dengan berbagai kovariat atau faktor eksternal. Faktor-faktor ini secara signifikan mempengaruhi kejadian dan intensitas peristiwa yang dimodelkan. Misalnya, dalam kasus gempa bumi model dikaitkan dengan faktor-faktor geologis, seperti zona subduksi, sesar, dan gunung berapi (Choiruddin, et al., 2021; Kartikasari & Choiruddin, 2022), penyebaran penyakit COVID-19 dihubungkan dengan faktor tingkat kepadatan populasi masyarakat di suatu area, tingkat vaksinasi, dan pusat kerumunan (Virania, et al., 2021; Choiruddin, et al.,

2023), persebaran suatu pohon yang dipengaruhi oleh faktor abiotik seperti ketinggian dan kemiringan (Daniel, et al., 2018; Choiruddin, et al., 2017) serta kecelakaan lalu lintas dikaitkan dengan faktor-faktor lingkungan seperti jarak ke lampu lalu lintas terdekat ke dalam model di *linear network* yang diilustrasikan pada Gambar 1.1 (Hasanah, et al., 2022). Analisis *point pattern on a linear network* memiliki aplikasi yang luas dengan *linear network* didefinisikan sebagai kumpulan segmen garis pada suatu bidang seperti peta jalur kereta api, sungai, kabel listrik, serabut saraf, rute penerbangan, saluran irigasi, patahan geologi, atau retakan tanah. Penerapan analisis ini dapat berupa lokasi kecelakaan lalu lintas, pencurian kendaraan atau kejahatan jalanan (Yamada & Thill, 2004) pepohonan di tepi jalan; kios pinggir jalan atau taman kota (Okabe & Kitamura, 1996) atau sarang serangga (Ang, et al., 2012), sedangkan titik-titik menunjukkan lokasi objek yang diamati di sepanjang garis.



Gambar 1.1 Peta prediksi risiko kecelakaan lalu lintas di jaringan jalan Kabupaten Nganjuk yang dikaitkan dengan lokasi lampu lalu lintas terdekat sebagai kovariat (Hasanah, et al., 2022).

Estimasi parameter pada *linear network* terutama yang melibatkan kovariat spasial dilakukan dengan metode *Maximum Likelihood Estimation* (MLE) dan menerapkan pendekatan Berman-Turner. Parameter dari *point process* pada *linear network* diestimasi menggunakan metode MLE dan mengimplementasikan pendekatan Berman-Turner. Perkembangan metode pada *linear network* meliputi

nonparametric kernel estimate (Okabe & Sugihara, 2012; McSwiggan, et al., 2016), definisi *spatial correlation* dan *K-function* (Ang, 2010), hingga pemodelan intensitas yang melibatkan kovariat pada *linear network* (Ang, 2010; Ang, et al., 2012).

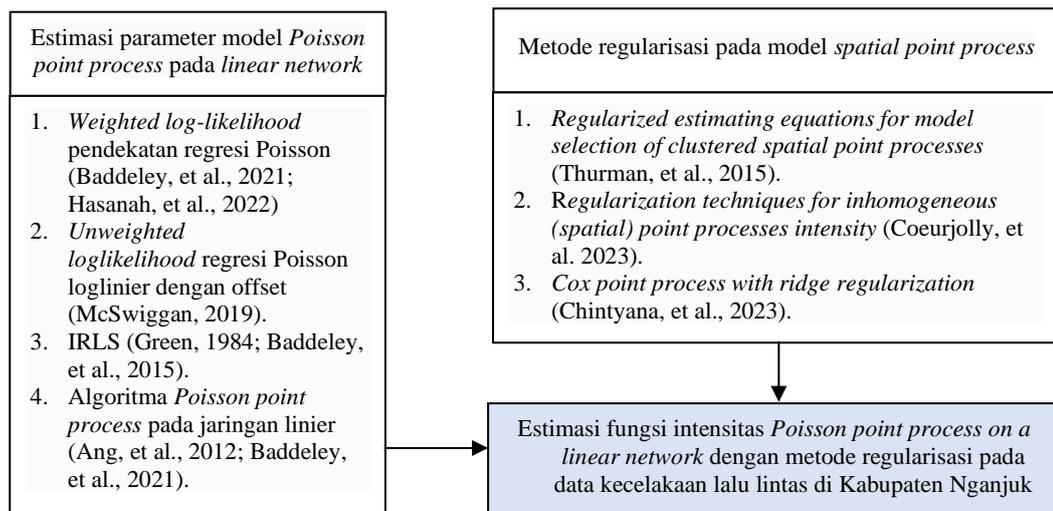
Penggunaan kovariat dalam analisis memiliki tantangan pada identifikasi variabel yang paling signifikan dari berbagai kovariat yang digunakan. Kebutuhan seleksi variabel menjadi krusial dalam proses pemodelan *spatial point process* untuk mendapatkan model yang *parsimonious*. Model yang dihasilkan melalui proses ini cenderung lebih sederhana namun tetap efisien dalam prediksi. Seleksi variabel pada *spatial point process* telah berkembang di beberapa aplikasi khususnya pada *point pattern* berdimensi dua seperti dilakukan oleh Thurman, et al. (2015), Husain (2022), Chintyana, et al. (2023), dan Coeurjolly, et al. (2023). Integrasi pengembangan ini telah dilakukan dalam fungsi model `glmnet` pada `spatstat.model` R *packages* melalui penambahan struktur objek model regularisasi `improve.type`. Akan tetapi, aplikasi ini masih terbatas pada ruang dimensi dua dan belum diterapkan pada *linear network*.

Aplikasi seleksi variabel dilakukan misalnya pada studi kasus kecelakaan lalu lintas di Kabupaten Nganjuk, analisis melibatkan faktor lingkungan dan kondisi jalan meliputi jarak ke lokasi lampu lalu lintas terdekat, jenis jalan, jumlah lajur, jumlah jalur, keberadaan kerb di sisi kanan dan kiri jalan. Data kecelakaan lalu lintas di Kabupaten Nganjuk ini sudah pernah dianalisis sebelumnya oleh Hasanah, et al. (2022) dengan menerapkan metode *point process on a linear network*. Penelitian ini bertujuan untuk menghindari kelemahan yang ada pada pendekatan *crash-frequency* serta sebagai penyesuaian metodologi dari *point process* pada ruang berdimensi dua ke *linear network*. Namun, studi tersebut terbatas hanya menggunakan dua variabel kovariat, meskipun sebenarnya terdapat variabel lain yang mempengaruhi kecelakaan lalu lintas yang tersedia. Jika semua variabel dimasukkan, model bisa jadi terlalu kompleks. Selain itu, pemilihan model dengan variabel yang terbatas namun relevan menghindari *overfitting* dan mempercepat komputasi tanpa mengorbankan kualitas hasil analisis yang diperoleh. Interaksi antara lokasi kejadian kecelakaan diasumsikan saling independen, dimana kejadian

antar kecelakaan tidak saling mempengaruhi satu sama lain sehingga model yang digunakan adalah model *Poisson*.

Fokus utama dalam penelitian ini adalah mengembangkan prosedur seleksi variabel pada estimasi model *point process on a linear network* menggunakan metode *quadrature* dengan regularisasi. Metode regularisasi yang dipertimbangkan dalam penelitian ini yaitu regularisasi Lasso. Metode diaplikasikan dengan memodifikasi *first-step estimation* dengan menambahkan fungsi penalti. Fungsi penalti diselesaikan dengan *Coordinate Descent Algorithm* (CDA) yang dapat memberikan solusi secara efisien dengan penurunan koordinat mengoptimalkan fungsi target secara parsial dan iteratif sampai kriteria konvergensi tercapai (Friedman, et al., 2007). *Tuning parameter* dalam konteks regularisasi memiliki peranan dalam membentuk model yang optimal. Parameter ini berfungsi sebagai pengontrol kekuatan dari penalti yang diterapkan terhadap koefisien model, memastikan bahwa koefisien yang tidak signifikan berpengaruh terhadap model dieliminasi sepenuhnya, sehingga menghasilkan model yang lebih sederhana. Pemilihan *tuning parameter* dilakukan dengan menggunakan nilai *Bayesian Information Criterion* (Choiruddin, et al., 2021). Penelitian ini menggunakan kovariat jarak ke lokasi lampu lalu lintas terdekat, jenis jalan, banyak lajur dan jalur, keberadaan kerb di sisi kiri dan kanan jalan yang diharapkan dapat memberikan gambaran kinerja regularisasi pada *point process on a linear network*. Penggunaan tujuh kovariat dalam pemodelan dianggap cukup untuk melakukan seleksi variabel karena keterbatasan jumlah data spasial terutama pada kasus kecelakaan lalu lintas di Indonesia.

Secara umum, peta alur penelitian berdasarkan ringkasan latar belakang yang menunjukkan posisi penelitian relatif terhadap penelitian sebelumnya ditunjukkan pada Gambar 1.2



Gambar 1.2 Peta alur penelitian.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, maka rumusan masalah pada penelitian ini adalah sebagai berikut.

1. Bagaimana estimasi parameter pada fungsi intensitas *point process on a linear network* dengan metode regularisasi?
2. Bagaimana model dan prediksi intensitas kecelakaan lalu lintas di Kabupaten Nganjuk dengan *regularization Poisson point process on a linear network*?

1.3 Tujuan

Berdasarkan rumusan masalah yang telah diuraikan, tujuan dari penelitian ini adalah sebagai berikut.

1. Mengembangkan proses estimasi parameter pada fungsi intensitas *point process on a linear network* dengan metode regularisasi.
2. Mendapatkan model dan prediksi intensitas kecelakaan lalu lintas di Kabupaten Nganjuk dengan *regularization Poisson point process on a linear network*.

1.4 Manfaat Penelitian

Manfaat penelitian terkait pemodelan kecelakaan lalu lintas di Kabupaten Nganjuk adalah sebagai berikut.

1. Manfaat Teoritis

Manfaat teoritis dalam penelitian ini memberikan sumbangan pengetahuan dan tinjauan statistik mengenai pemodelan *spatial point process on a linear network*, khususnya dalam pemodelan intensitas kecelakaan lalu lintas di Kabupaten Nganjuk dengan penambahan metode regularisasi.

2. Manfaat Praktis

Hasil penelitian ini dapat memberikan informasi yang berguna bagi masyarakat dan pemerintah mengenai lokasi-lokasi yang rawan terjadi kecelakaan lalu lintas dengan informasi tambahan faktor jalan dan lingkungan yang mempengaruhinya.

1.5 Batasan Masalah

Batasan masalah yang digunakan pada penelitian ini adalah sebagai berikut.

1. Jaringan jalan yang digunakan adalah jalan arteri, jalan kolektor, dan jalan lokal di Kabupaten Nganjuk.
2. Perbedaan tahun periode data kovariat (tahun 2020) dan data kecelakaan lalu lintas (tahun 2018-2020). Diasumsikan dengan kondisi jalan yang tetap konstan, termasuk jumlah lajur, jumlah jalur, tipe jalan, dan keberadaan kerb jalan yang tidak berubah.
3. Model *point process* yang digunakan adalah *Poisson point process* karena setiap kejadian kecelakaan lalu lintas diasumsikan saling independen.
4. Metode estimasi parameter menggunakan metode *maximum likelihood estimation* (MLE)
5. Metode regularisasi yang digunakan dalam penelitian ini adalah *Coordinate Descent Algorithm* (CDA).
6. Pemilihan *tuning parameter* γ menggunakan nilai *Bayesian Information Criterion* (BIC).
7. Fokus aplikasi regularisasi pada metode Lasso untuk seleksi variabel karena metode ini efektif dalam mengurangi *overfitting* sekaligus menyederhanakan model dengan secara otomatis memilih variabel yang paling relevan.

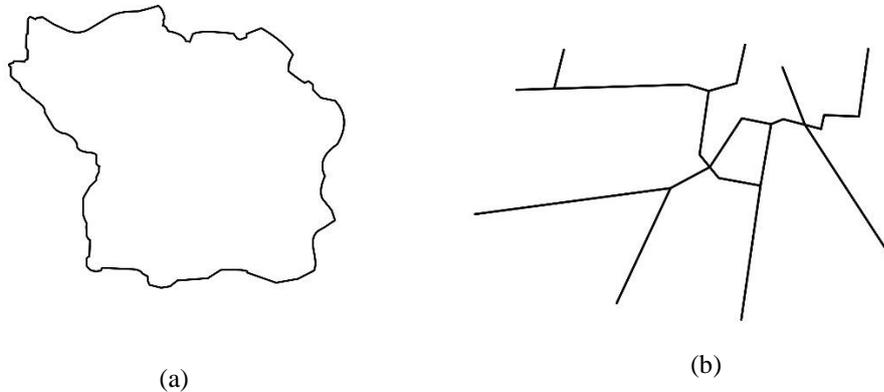
BAB 2

TINJAUAN PUSTAKA

Bagian ini menjelaskan mengenai konsep dan teori dari metode yang digunakan, yaitu *inhomogeneous Poisson point process on a linear network*. Selain itu, bagian ini juga akan membahas mengenai topik yang digunakan, yaitu kecelakaan lalu lintas. Uraian mengenai konsep dan teori terkait metode yang digunakan dan topik yang diangkat adalah sebagai berikut.

2.1 Bidang Spasial di \mathbb{R}^2 dan *Linear Network*

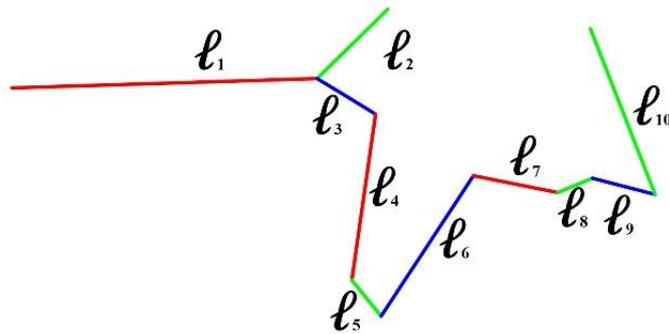
Secara umum, dalam analisis spasial berada pada \mathbb{R}^2 yang melam-bangkan wilayah suatu daerah seperti pada Gambar 2.1 (a). Namun, ada beberapa bidang yang ditentukan oleh *linear network* seperti peta jalur kereta api, sungai, kabel listrik, serabut saraf, rute penerbangan, saluran irigasi, patahan geologi, retakan tanah, dan jalan raya yang terdiri dari kumpulan segmen garis pada suatu bidang seperti pada Gambar 2.1 (b).



Gambar 2.1 Bidang spasial pada (a) $W \subset \mathbb{R}^2$ dan (b) *linear network* $L \subset \mathbb{R}^2$.

Linear network L merupakan gabungan dari M segmen garis $\ell_1, \ell_2, \ell_3, \dots, \ell_M$ sehingga $L = \bigcup_{j=1}^M \ell_j$, $M \leq \infty$. Misalnya $M=10$ segmen garis, ilustrasi linear network L digambarkan pada Gambar 2.2. Setiap segmen garis ℓ_j dinotasikan sebagai $\ell_j = [u, v] = \{w: w = tu + (1-t)v, 0 \leq t \leq 1\}$ dengan $u, v \in \mathbb{R}^2, u \neq v$ sebagai ujung dari ℓ_j . Total Panjang seluruh segmen garis pada L dinotasikan dengan $|L| = \sum_{j=1}^M |\ell_j|$ dimana $|\ell_j|$ adalah panjang setiap segmen garis.

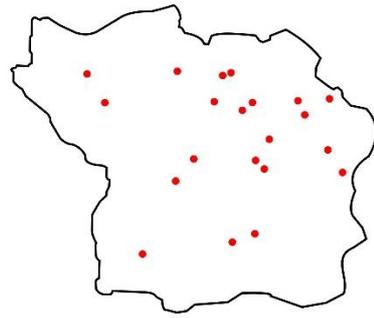
Perpotongan dari dua segmen garis ℓ_i dan ℓ_j untuk $i \neq j$ diasumsikan kosong atau merupakan titik akhir dari kedua segmen garis. Misalkan L adalah *linear network*. Sebuah titik $v \in L$ disebut sebagai titik simpul jika $\{v\} = \ell_i \cap \ell_j$ untuk $i, j, i \neq j$ (Ang, et al., 2012).



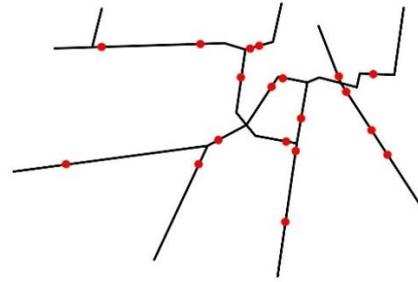
Gambar 2.2 Ilustrasi *linear network* L dengan $M = 10$ segmen garis.

2.2 *Spatial Point Process* di \mathbb{R}^2 dan *Linear Network*

Spatial point process merupakan suatu proses stokastik yang terbentuk dari *spatial point pattern*, *point pattern* merupakan himpunan titik yang acak dalam suatu *observation window* pada dimensi dua $W \subset \mathbb{R}^2$. Titik-titik tersebut merupakan lokasi dari suatu objek penelitian atau peristiwa yang diamati di suatu wilayah, seperti lokasi spesies pohon, pusat gempa, lokasi kecelakaan, kasus penyakit langka dan lain-lain (Baddeley, et al., 2015). *Spatial point process* pada ruang berdimensi dua, objek atau peristiwa yang diamati akan dibatasi pada wilayah pengamatan (*observation window*) tertentu, yaitu berupa luasan $|W|$. $\mathbf{X} \subset \mathbb{R}^2$ adalah subset acak dari \mathbb{R}^2 yang lokal finit, artinya dalam wilayah terbatas W , jumlah titik di W yaitu $N(\mathbf{X} \cap W)$. Realisasi \mathbf{X} di W merupakan *spatial point pattern* yaitu $\mathbf{x} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}, \mathbf{x}_i \in W, i = 1, 2, \dots, n$ dimana n adalah jumlah titik data yang diamati di *observation window* W dan \mathbf{x}_i adalah koordinat *longitude* dan *latitude* kejadian ke- i di *observation window* W . *Observation window* W pada dimensi dua tidak harus berbentuk segi empat, tetapi juga bisa berbentuk *polygonal*. Ilustrasi dari suatu *spatial point process* dan *spatial point pattern* dapat dilihat pada Gambar 2.3 (a).



(a)



(b)

<i>Point Pattern</i> (\mathbf{x}_i)	<i>Longitude</i> $(x_{long,i})$	<i>Latitude</i> $(x_{lat,i})$
\mathbf{x}_1	$x_{long,1}$	$x_{lat,1}$
\mathbf{x}_2	$x_{long,2}$	$x_{lat,2}$
\vdots	\vdots	\vdots
\mathbf{x}_n	$x_{long,n}$	$x_{lat,n}$

(c)

<i>Point Pattern</i> (\mathbf{x}_i)	<i>Longitude</i> $(x_{long,i})$	<i>Latitude</i> $(x_{lat,i})$
\mathbf{x}_1	$x_{long,1}$	$x_{lat,1}$
\mathbf{x}_2	$x_{long,2}$	$x_{lat,2}$
\vdots	\vdots	\vdots
\mathbf{x}_n	$x_{long,n}$	$x_{lat,n}$

(d)

Gambar 2.3 *Observation window* dan struktur data *spatial point pattern*, titik-titik merah (•) menunjukkan lokasi objek atau kejadian pada *observation window* (a) *spatial point pattern* pada \mathbb{R}^2 (b) *spatial point pattern* pada *linear network* (c) struktur data *point pattern* di \mathbb{R}^2 (d) struktur data *point pattern* di *linear network*. Struktur data *point pattern* di \mathbb{R}^2 dan *linear network* sama meskipun dengan *observation window* yang berbeda.

Penerapan pemodelan *point process* dalam kehidupan sehari-hari tidak hanya terjadi pada suatu *observation window* \mathbb{R}^2 , tetapi sering kali ditemukan pada objek pada *linear network*, misalnya kecelakaan lalu lintas di jalan raya, pencurian dan kejahatan di jalanan, kios di pinggir jalan, dan lain-lain. Perbedaan utama *point process on a linear network* dan *point process* pada dimensi dua terletak pada jenis *window* pengamatan meskipun memiliki struktur data *point pattern* yang sama seperti pada Gambar 2.3. Pada \mathbb{R}^2 , *window* pengamatan biasanya mencakup suatu area, sedangkan pada *linear network* pengamatan berorientasi pada panjang segmen garis. Perbedaan ini menyebabkan adanya justifikasi dalam metodologi. Salah satunya terkait dengan pengukuran jarak terdekat antar objek. Pada \mathbb{R}^2 jarak antara dua titik diukur dengan ukuran metrik seperti jarak *Euclidean*. Sedangkan, penggunaan ukuran jarak kejadian dalam *linear network* hanya bisa terjadi di sepanjang *linear network* sebagai contoh kejadian kecelakaan hanya terjadi di sepanjang jalan yang mengikuti bentuk *linear network* dengan konsep perhitungan

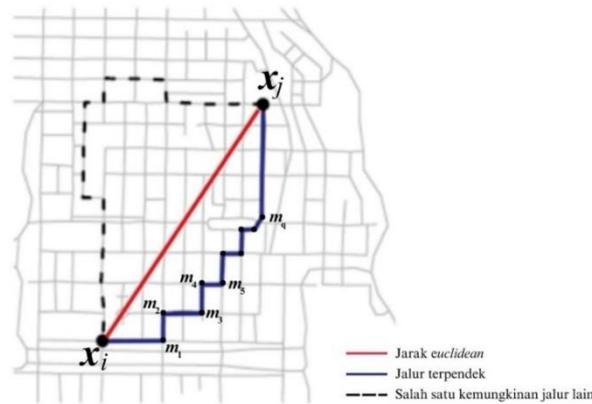
jarak antara dua titik menggunakan jarak jalur minimum (*shortest-path distance*). Panjang jalur minimum $d_L(\mathbf{x}_i, \mathbf{x}_j)$ antara titik \mathbf{x}_i dan \mathbf{x}_j di L adalah panjang minimum yang terbentuk dari \mathbf{x}_i ke \mathbf{x}_j yang dihitung menggunakan Persamaan (2.1).

$$d_L(\mathbf{x}_i, \mathbf{x}_j) = \|\mathbf{x}_i - \mathbf{m}_1\| + \sum_{k=1}^q \|\mathbf{m}_{k+1} - \mathbf{m}_k\| + \|\mathbf{m}_q - \mathbf{x}_j\| \quad (2.1)$$

dimana $\|\cdot\|$ merupakan jarak *Euclidean* yang dihitung dengan Persamaan (2.2) dan q merupakan banyak simpul pada *shortest path distance* dari \mathbf{x}_i ke \mathbf{x}_j (lihat Gambar 2.4).

$$\|\mathbf{m}_{i+1} - \mathbf{m}_i\| = \sqrt{(m_{lat,i+1} - m_{lat,i})^2 + (m_{long,i+1} - m_{long,i})^2} \quad (2.2)$$

Jika tidak terdapat jalur dari \mathbf{x}_i ke \mathbf{x}_j (menunjukkan bahwa jalur tidak terhubung) maka didefinisikan $d_L(\mathbf{x}_i, \mathbf{x}_j) = \infty$.



Gambar 2.4 Ilustrasi jarak pada *linear network*.

Ilustrasi jarak pada *linear network* digambarkan pada Gambar 2.4, garis putus-putus menunjukkan salah satu kemungkinan jalur lain yang terbentuk antara dua titik. Garis berwarna merah menunjukkan jarak *Euclidean* yang diterapkan pada \mathbb{R}^2 . Garis berwarna biru merupakan jalur terpendek dari kedua titik yang disebut *shortest-path distance* pada lintasan jalur *linear network*.

Point process X pada sebuah *linear network* L adalah suatu proses stokastik yang hasilnya berupa *point pattern*, sehingga jumlah titik pada segmen garis $L \subseteq \mathbb{R}^2$ adalah sebuah variabel random. *Point pattern* \mathbf{x} pada *linear network* L merupakan himpunan berhingga $\mathbf{x} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}, \mathbf{x}_i \in L, i = 1, 2, \dots, n$.

2.3 Karakteristik *Point Process on a Linear Network*

Point process on a linear network merupakan metode analisis spasial yang fokus pada distribusi *point pattern* di dalam sebuah *linear network*. Sub bab ini akan menjelaskan karakteristik *Point process on a linear network* dan eksplorasi data *point pattern* untuk memberikan informasi mengenai fungsi intensitas yang mempertimbangkan faktor eksternal berupa faktor lingkungan dan geometri jalan.

2.3.1 Intensitas

Langkah pertama dan terpenting dari analisis data spasial *point pattern* adalah dengan menginvestigasi intensitasnya. Fungsi intensitas $\lambda(\mathbf{u})$ dari *point process* \mathbf{X} adalah peluang mengobservasi sebuah titik di suatu lokasi yang sangat kecil yang berpusat di $\mathbf{u} \in \mathbb{R}^2$ dengan panjang segmen $d\mathbf{u}$ dengan $\mathbf{u} = [u_{long}, u_{lat}]^T$ merupakan sembarang titik dalam *longitude* dan *latitude* yang berada di $L \subset \mathbb{R}^2$ (Baddeley, et al., 2015). Setiap *point process* \mathbf{X} , ekspektasi jumlah titik yang diamati dalam *observation window* $L \subset \mathbb{R}^2$ dapat dituliskan seperti pada Persamaan (2.3),

$$\mu(L) = E[N(\mathbf{X} \cap L)] = \int_L \lambda(\mathbf{u}) d\mathbf{u}, \mathbf{u} \in L \subset \mathbb{R}^2 \quad (2.3)$$

dimana $\lambda(\cdot)$ merupakan fungsi intensitas dengan $|L|$ adalah total panjang *observation window* L . Apabila fungsi intensitas tidak konstan, maka untuk mengestimasi nilai λ adalah dengan mengasumsikannya sebagai fungsi parametrik dari parameter lain yang selanjutnya diestimasi menggunakan *maximum likelihood*.

Fungsi intensitas pada *point process* non stasioner bersifat tidak konstan di setiap segmen L dan diduga dipengaruhi oleh variabel prediktor dalam hal ini merupakan kovariat. Fungsi intensitas *point process* non stasioner dapat dituliskan sebagai $\lambda(\boldsymbol{\theta}; \mathbf{u})$. Menurut Choiruddin, et al. (2021) dan Baddeley, et al. (2015) fungsi intensitas dari *point process* non stasioner membentuk model *log linear* yang bergantung pada parameter $\boldsymbol{\theta}$ dan dipengaruhi oleh kovariat spasial $\mathbf{z}(\mathbf{u})$, dituliskan pada Persamaan (2.4) (Choiruddin, et al., 2021),

$$\lambda(\mathbf{u}) = \lambda(\boldsymbol{\theta}; \mathbf{u}) = \exp(\boldsymbol{\theta}^T \mathbf{z}(\mathbf{u})), \mathbf{u} \in L \subset \mathbb{R}^2 \quad (2.4)$$

dimana $\mathbf{z}(\mathbf{u}) = (1, z_1(\mathbf{u}), z_2(\mathbf{u}), \dots, z_p(\mathbf{u}))^T$ merupakan vektor kovariat dan $\boldsymbol{\theta} = (\theta_0, \theta_1, \theta_2, \dots, \theta_p)^T$ merupakan parameter regresi untuk setiap variabel prediktor dalam model. Keunggulan dari model ini adalah adanya fungsi eksponensial dalam persamaan (2.4) yang mencerminkan sifat dari fungsi intensitas yang bernilai lebih dari atau sama dengan nol, terlepas dari berapapun nilai $\boldsymbol{\theta}$ dan nilai kovariat. Selain itu, fungsi intensitas pada persamaan (2.4) dapat lebih mudah dalam estimasi parameter (Baddeley, et al., 2015). Interpretasi model pada kovariat bertipe kategorik, salah satu kategori dari setiap kovariat dianggap sebagai referensi untuk membandingkan risiko kecelakaan lalu lintas di lokasi \mathbf{u} terhadap kondisi kategori lainnya. Contohnya, untuk kovariat keberadaan kerb di sisi kiri jalan $z_5(\mathbf{u})$, terdapat dua kategori yaitu $z_{5(\text{tidak ada kerb kiri})}(\mathbf{u})$ dan $z_{5(\text{ada kerb kiri})}(\mathbf{u})$. Kategori “tidak ada marka” pada penelitian ini dijadikan sebagai basis untuk membandingkan risiko kecelakaan di jalan dengan kerb kiri terhadap jalan tanpa kerb kiri. Ilustrasi interpretasi parameter pada model *Poisson point process on a linear network* berdasarkan tanda dari $\hat{\boldsymbol{\theta}}$ dan jenis kovariat dijelaskan pada Tabel 2.1.

Tabel 2.1 Ilustrasi interpretasi hasil estimasi parameter

Jenis Kovariat	Tanda Estimasi Parameter ($\hat{\theta}_i$)	Interpretasi
Numerik	Positif (+)	Semakin besar nilai kovariat $z_i(\mathbf{u})$ di lokasi \mathbf{u} , semakin besar risiko kecelakaan yang terjadi di lokasi tersebut.
	Negatif (-)	Semakin besar nilai kovariat $z_i(\mathbf{u})$ di lokasi \mathbf{u} , semakin kecil risiko kecelakaan yang terjadi di lokasi tersebut.
Kategorik	Positif(+)	Risiko kecelakaan lalu lintas di lokasi \mathbf{u} dengan kondisi $z_i(\mathbf{u})$ adalah sebesar $\exp(\hat{\theta}_i)$ kali atau $(\exp(\hat{\theta}_i)-1) \times 100\%$ lebih besar dari risiko kecelakaan lalu lintas di lokasi \mathbf{u} dengan kondisi kovariat basis kategori $z_i(\mathbf{u})$.
	Negatif (-)	Risiko kecelakaan lalu lintas di lokasi \mathbf{u} dengan kondisi $z_i(\mathbf{u})$ adalah sebesar $1/\exp(\hat{\theta}_i)$ kali atau $(1-\exp(\hat{\theta}_i)) \times 100\%$ lebih kecil dari risiko kecelakaan lalu lintas di lokasi \mathbf{u} dengan kondisi kovariat basis kategori $z_i(\mathbf{u})$.

2.3.2 Uji Homogenitas

Uji homogenitas pada *point process on a linear network* digunakan untuk mengetahui apakah intensitas memiliki pola yang homogen atau tidak, sehingga pada saat melakukan estimasi parameter akan didapatkan model yang sesuai dengan karakteristik persebaran dari *point pattern* yang diteliti. Salah satu uji homogenitas yang dapat digunakan adalah *Cumulative Distribution Function (CDF) test* menggunakan *Kolmogorov-Smirnov* dengan hipotesis dari pengujian homogenitas dituliskan sebagai berikut:

$$H_0 : F_{0,j}(z) = \hat{F}_j(z) \quad (\text{intensitas homogen})$$

$$H_1 : F_{0,j}(z) \neq \hat{F}_j(z) \quad (\text{intensitas inhomogen})$$

Statistik uji yang digunakan adalah yaitu *Kolmogorov-Smirnov* pada Persamaan (2.5) (Baddeley, et al., 2015).

$$D = \max |\hat{F}_j(z) - F_{0,j}(z)| \quad (2.5)$$

dengan

$$\hat{F}_j(z) = \frac{1}{n} \sum_{i=1}^n I\{z_{ij} \leq z\} \quad (2.6)$$

dan

$$F_{0,j}(z) = \frac{|\{u \in L : z_j(u) \leq z\}|}{|L|} \quad (2.7)$$

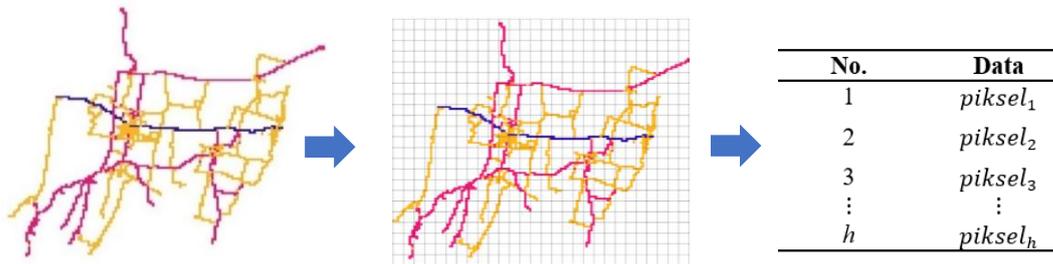
dimana

$\hat{F}_j(z)$: fungsi distribusi kumulatif empiris dari nilai-nilai kovariat z_j untuk menggambarkan variasi intensitas di setiap titik pengamatan yang diamati.

$F_{0,j}(z)$: fungsi distribusi kumulatif teoritis atau yang diharapkan dari kovariat z_j di bawah asumsi homogenitas dengan berdistribusi *uniform* pada seluruh *observation window*.

Tolak H_0 jika nilai statistik uji $D \geq$ nilai kritis D atau $p\text{-value} <$ nilai taraf signifikansi α , dengan $\alpha = 0,05$. Asumsi homogenitas dengan berdistribusi *uniform* yang dimaksud adalah memiliki nilai intensitas yang sama pada semua titik di *observation window*.

2.3.3 Correlation Pixel Image



Gambar 2.5 Ilustrasi pixel image

Correlation pixel image digunakan untuk mengetahui korelasi jarak antar kovariat pada format data *pixel image*. Metode *Pearson correlation* diadaptasi pada perhitungan korelasi data piksel. Perhitungan koefisien korelasi antar kovariat, dengan h adalah banyaknya piksel (lihat Gambar 2.5), dituliskan pada persamaan (2.8):

$$r_{z_a, z_b} = \frac{h \sum_{i=1}^h z_{ai} z_{bi} - \sum_{i=1}^h z_{ai} \sum_{i=1}^h z_{bi}}{\sqrt{h \sum_{i=1}^h z_{ai}^2 - \left(\sum_{i=1}^h z_{ai}\right)^2} \sqrt{h \sum_{i=1}^h z_{bi}^2 - \left(\sum_{i=1}^h z_{bi}\right)^2}} \quad (2.8)$$

Korelasi dihitung untuk semua kovariat z_a, z_b , $a, b = 1, 2, \dots, p$ dengan p merupakan banyaknya kovariat. Sama halnya dengan *Pearson correlation*, koefisien korelasi berada pada rentang $-1 \leq r \leq 1$. Jika koefisien korelasi mendekati -1 atau 1 maka hubungan keeratan dua variabel semakin kuat. Tanda positif dan negatif pada koefisien korelasi menunjukkan arah hubungan. Koefisien korelasi bertanda positif artinya hubungan berbanding lurus. Koefisien korelasi bertanda negatif memiliki arti hubungan berbanding terbalik.

Setelah diperoleh nilai koefisien korelasi, tahap selanjutnya adalah melakukan pengujian terhadap koefisien korelasi dengan hipotesis sebagai berikut.
 $H_0 : \hat{\rho} = 0$ (keeratan hubungan linier yang tidak signifikan antara kovariat i dan j)
 $H_1 : \hat{\rho} \neq 0$ (keeratan hubungan linier yang signifikan antara kovariat i dan j)
 Statistik uji yang digunakan untuk pengujian hipotesis koefisien korelasi adalah statistik uji Z yang dapat diperoleh menggunakan Persamaan (2.9) (Draper & Smith, 2014).

$$Z_{hit} = \frac{\sqrt{h-3}}{2} \ln \left(\frac{1+r_{z_a z_b}}{1-r_{z_a z_b}} \right) \quad (2.9)$$

Tolak H_0 pada taraf signifikansi (α) jika nilai statistik uji $Z_{hit} \geq Z_{\alpha/2}$ atau $Z_{hit} \leq -Z_{\alpha/2}$ atau menggunakan nilai $p\text{-value} < \alpha$. Jika keputusannya gagal tolak H_0 , maka dapat disimpulkan terdapat keeratan hubungan yang signifikan secara statistik antara kedua kovariat.

2.3.4 Uji Independensi

Uji independensi digunakan untuk mengetahui adanya hubungan atau dependensi antara dua kovariat kategorik yang memiliki skala nominal. Dengan kata lain, uji ini membantu menentukan apakah perubahan pada satu kovariat berkaitan dengan perubahan pada kovariat lainnya. Uji independensi juga dapat digunakan untuk mendeteksi adanya kasus multikolinearitas. Langkah pertama dalam pengujian ini adalah membentuk tabel kontingensi. Tabel kontingensi merupakan sebuah tabel yang digunakan untuk menunjukkan frekuensi distribusi dari dua atau lebih kovariat kategori. Tabel ini menyajikan data dalam bentuk matriks, dimana baris dan kolomnya merepresentasikan kategori dari masing-masing kovariat, dan setiap sel dalam tabel menunjukkan jumlah observasi yang termasuk dalam kategori-kategori tersebut. Bentuk tabel kontingensi $b \times c$ ditunjukkan pada Tabel 2.2.

Tabel 2.2 Tabel kontingensi $b \times c$

		Kovariat B						Total
		B_1	B_2	...	B_j	...	B_c	
Kovariat A	A_1	n_{11}	n_{12}	...	n_{1j}	...	n_{1c}	n_{1+}
	A_2	n_{21}	n_{22}	...	n_{2j}	...	n_{2c}	n_{2+}
	\vdots	\vdots	\vdots	\ddots	\vdots	\ddots	\vdots	\vdots
	A_i	n_{i1}	n_{i2}	...	n_{ij}	...	n_{ic}	n_{i+}
	\vdots	\vdots	\vdots	\ddots	\vdots	\ddots	\vdots	\vdots
	A_b	n_{b1}	n_{b2}	...	n_{bj}	...	n_{bc}	n_{b+}
Total		n_{+1}	n_{+2}	...	n_{+j}	...	n_{+c}	n_{++}

Pembentukan tabel kontingensi $b \times c$ harus memenuhi syarat-syarat sebagai berikut:

1. Homogen: setiap kategori harus memiliki sifat yang serupa atau setara, artinya setiap kategori dalam variabel yang sama harus mengukur konsep yang sama dan relevan.
2. *Mutually Exhausting*: dekomposisi secara lengkap sampai pada unit terkecil. Setiap observasi harus termasuk dalam salah satu kategori yang ada. Hal ini berarti tidak ada observasi yang dikecualikan atau tidak terklasifikasi.
3. *Mutually Exclusive*: antar kategori saling asing. Setiap observasi hanya dapat dimasukkan ke dalam satu kategori saja. Artinya, tidak ada observasi yang dapat masuk ke lebih dari satu kategori pada saat yang bersamaan.
4. Setiap variabel memiliki skala nominal atau ordinal.

Uji independensi dapat dilakukan menggunakan uji *Chi-square* dengan mengasumsikan pengambilan sampel secara multinomial dengan probabilitas bersama tiap sel dalam tabel kontingensi adalah π_{ij} (lihat Tabel 2.3), dengan

$$\pi_{ij} = \frac{n_{ij}}{n_{++}}$$

Tabel 2.3 Tabel probabilitas b x c

		Kovariat B						Total
		B_1	B_2	...	B_j	...	B_c	
Kovariat A	A_1	π_{11}	π_{12}	...	π_{1j}	...	π_{1c}	π_{1+}
	A_2	π_{21}	π_{22}	...	π_{2j}	...	π_{2c}	π_{2+}
	\vdots	\vdots	\vdots	\ddots	\vdots	\ddots	\vdots	\vdots
	A_i	π_{i1}	π_{i2}	...	π_{ij}	...	π_{ic}	π_{i+}
	\vdots	\vdots	\vdots	\ddots	\vdots	\ddots	\vdots	\vdots
	A_b	π_{b1}	π_{b2}	...	π_{bj}	...	π_{bc}	π_{b+}
Total		π_{+1}	π_{+2}	...	π_{+j}	...	π_{+c}	π_{++}

Uji independensi menggunakan uji *Chi-square* dengan hipotesis sebagai berikut:

$H_0 : \pi_{ij} = \pi_{i+}\pi_{+j}$ (tidak terdapat dependensi antara dua variabel yang diamati)

$H_1 : \pi_{ij} \neq \pi_{i+}\pi_{+j}$ (terdapat dependensi antara dua variabel yang diamati)

Statistik uji yang digunakan adalah *Pearson's Chi-squared* dengan Persamaan sebagai berikut.

$$\chi^2 = \sum_{j=1}^b \sum_{i=1}^c \frac{(n_{ij} - e_{ij})^2}{e_{ij}} \quad (2.10)$$

dimana n_{ij} adalah jumlah observasi kovariat A kategori i dan variabel B kategori j dan $e_{ij} = \frac{n_{i+}n_{+j}}{n_{++}}$ adalah nilai ekspektasi frekuensi. Tolak H_0 jika $\chi^2 \geq \chi_{(b-1)(c-1);\alpha}^2$. Penggunaan statistik uji untuk pengujian *Chi-square* bergantung pada karakteristik dari tabel kontingensi dan nilai ekspektasi dalam tiap sel. Berikut rincian penggunaan rumus-rumusnya.

- Secara umum, statistik uji yang digunakan untuk tabel kontingensi $b \times c$ dengan $b > 2$ dan $c > 2$ adalah *Pearson's Chi-squared* seperti pada Persamaan (2.10)
- Apabila tabel kontingensi berbentuk 2×2 , maka statistik uji yang digunakan adalah koreksi Yates seperti pada Persamaan (2.11).

$$\chi^2 = \frac{n_{++}(n_{11}n_{22} - n_{12}n_{21})^2}{(n_{11} - n_{12})(n_{21} - n_{22})(n_{11} + n_{21})(n_{12} + n_{22})} \quad (2.11)$$

Tolak H_0 jika $\chi^2 \geq \chi_{1;\alpha}^2$.

- Apabila tabel kontingensi berbentuk 2×2 , dan terdapat sel dengan frekuensi harapan kurang dari 5, maka pengujian independensi dilakukan menggunakan uji *Fisher Exact*. Uji *Fisher* ini tidak melibatkan uji *Chi-Square*, akan tetapi langsung menggunakan peluang dan sama sekali tidak melibatkan nilai ekspektasi. Hipotesis dalam pengujian ini adalah sebagai berikut:

$H_0 : \psi = 1$ (tidak terdapat dependensi antara dua variabel yang diamati)

$H_1 : \psi \neq 1$ (terdapat dependensi antara dua variabel yang diamati)

Tahapan pertama dalam uji *Fisher* adalah mencari konfigurasi-konfigurasi tabel yang lebih ekstrim dari tabel yang diamati. Kemudian menghitung nilai P_i dari tabel-tabel tersebut, misalkan $P_0, P_1, \dots, P_i, \dots, P_k$ dengan Persamaan (2.12).

$$p_i = \frac{n_{1+}! n_{2+}! n_{+1}! n_{+2}!}{n_{11}! n_{12}! n_{21}! n_{22}! n_{++}!} \quad (2.12)$$

dilakukan penjumlahan terhadap nilai P_i dari masing-masing konfigurasi tabel sehingga $P = \sum_{i=0}^k P_i$. Bandingkan nilai P dengan taraf signifikansi (α) yang sudah ditentukan. Tolak H_0 jika $P < \alpha$ (Agresti, 2007).

2.4 *Poisson Point Process on a Linear Network*

Model *Poisson point process* merupakan suatu pendekatan pada *spatial point process* dengan asumsi bahwa setiap titik lokasi adalah saling independen. *Poisson point process* dibagi menjadi dua, yaitu *homogeneous* dan *inhomogeneous Poisson point process*. *Inhomogeneous Poisson point process* merupakan suatu *Poisson process* dengan fungsi intensitas $\lambda(\mathbf{u}): L \rightarrow [0, \infty)$, $\mathbf{u} \in B \subseteq L$ yang tidak konstan atau bervariasi secara spasial dan didefinisikan oleh karakteristik berikut:

1. Jumlah titik acak di setiap segmen garis $B \subseteq L \subset \mathbb{R}^2$ atau $N(\mathbf{X} \cap B) < \infty$ berdistribusi *Poisson* dengan rata-rata $\lambda(B) = \int_B \lambda(\mathbf{u}) d\mathbf{u}$;
2. Jika terdapat beberapa segmen garis $B_1, B_2, \dots, B \subseteq L$, dan setiap segmen tidak saling beririsan antara satu dengan yang lainnya, maka banyaknya titik pada setiap segmen $N(\mathbf{X} \cap B_1), N(\mathbf{X} \cap B_2), \dots$ merupakan variabel random yang saling independen;
3. Jika $N(\mathbf{X} \cap B) = n$ adalah banyaknya titik pada setiap segmen garis $B \subseteq L \subset \mathbb{R}^2$, maka titik-titik yang berada pada segmen garis B merupakan titik-titik acak yang independen.

Pada *inhomogeneous Poisson point process*, titik-titik yang berada dalam *observation window* B adalah independen dan berdistribusi identik, dengan fungsi intensitas *inhomogeneous Poisson point process* pada Persamaan (2.13) (Møller & Waagepetersen, 2004),

$$f(\mathbf{u}) = \frac{\lambda(\mathbf{u})}{\mu(\mathbf{u})}, \mathbf{u} \in B \subseteq L \quad (2.13)$$

2.5 **Estimasi Parameter**

Pada penelitian ini estimasi parameter dilakukan untuk penaksiran nilai parameter θ terkait dengan *trend* spasial yang dilakukan dengan regularisasi.

2.5.1 Estimasi θ Berbasis Regresi Poisson

Estimasi parameter pada model *point process* dilakukan dengan metode *maximum likelihood estimation (MLE)*. Misalkan fungsi intensitas X adalah λ dan fungsi intensitas X_1 adalah λ_1 sehingga $\mu(L) = \int_L \lambda(\theta; \mathbf{u}) d\mathbf{u}$ dan $\mu_1(L) = \int_L \lambda_1(\theta; \mathbf{u}) d\mathbf{u}$ bersifat *finite*. Jika X merupakan *Poisson point process*, maka fungsi distribusi probabilitas dari X terhadap X_1 dinyatakan sebagai $f(\mathbf{X})$ dimana $\mathbf{X} \subset L$ pada Persamaan (2.14) (Moller & Waagepetersen, 2004)

$$f(\mathbf{X}) = \exp(\mu_1(L) - \mu(L)) \prod_{\mathbf{u} \in \mathbf{X} \cap L} \frac{\lambda(\theta; \mathbf{u})}{\lambda_1(\theta; \mathbf{u})} \quad (2.14)$$

Asumsikan bahwa $\lambda_1(\theta; \mathbf{u}) = 1$ yang berarti bahwa proses Poisson memiliki intensitas dengan konstan sebesar 1 sehingga $\mu_1(L) = \int_L \lambda_1(\theta; \mathbf{u}) d\mathbf{u} = \int_L 1 d\mathbf{u} = |L|$. Berdasarkan asumsi tersebut, maka Persamaan (2.14) dapat ditulis menjadi Persamaan (2.15). Persamaan (2.15) juga disebut sebagai fungsi *likelihood*

$$L(\mathbf{X}) = \exp(|L| - \mu(L)) \prod_{\mathbf{u} \in \mathbf{X} \cap L} \frac{\lambda(\theta; \mathbf{u})}{1}$$

$$L(\mathbf{X}) = \exp(|L| - \mu(L)) \prod_{\mathbf{u} \in \mathbf{X} \cap L} \lambda(\theta; \mathbf{u}) \quad (2.15)$$

Fungsi *log-likelihood* berdasarkan Persamaan (2.15) adalah

$$\log L(\theta) = \sum_{\mathbf{u} \in \mathbf{X} \cap L} \log \lambda(\theta; \mathbf{u}) + |L| - \mu(L) \quad (2.16)$$

dalam hal ini, $\mu(L) = \int_L \lambda(\theta; \mathbf{u}) d\mathbf{u}$, maka Persamaan (2.16) dapat dituliskan sebagai berikut

$$\log L(\theta) = \sum_{\mathbf{u} \in \mathbf{X} \cap L} \log \lambda(\theta; \mathbf{u}) + |L| - \int_L \lambda(\theta; \mathbf{u}) d\mathbf{u} \quad (2.17)$$

Nilai $|L|$ pada persamaan (2.13) merupakan nilai konstan dan tidak berpengaruh terhadap parameter θ , sehingga Persamaan (2.17) diaproksimasi dengan Persamaan (2.18).

$$\log L(\theta) = \sum_{\mathbf{u} \in \mathbf{X} \cap L} \log \lambda(\theta; \mathbf{u}) - \int_L \lambda(\theta; \mathbf{u}) d\mathbf{u}$$

$$\ell(\theta) = \sum_{\mathbf{u} \in \mathbf{X} \cap L} \log \lambda(\theta; \mathbf{u}) - \int_L \lambda(\theta; \mathbf{u}) d\mathbf{u} \quad (2.18)$$

Pendugaan parameter θ dengan memaksimalkan Persamaan (2.18) memerlukan pendekatan *numerical quadrature* adalah salah satu strategi komputasi yang sederhana dan tetap efisien untuk integrasi numerik (Baddeley, et al., 2015). Terdapat dua strategi untuk memodelkan *spatial point pattern* agar dapat didekati dengan dua pendekatan, yaitu dengan pendekatan Berman-Turner atau biasa disebut sebagai regresi Poisson dan pendekatan regresi logistik. Kedua pendekatan ini dilakukan dengan membagi *observation window* menjadi beberapa subinterval panjang tertentu, kemudian akan membangkitkan *dummy point* yang digunakan untuk mengkonstruksi variabel respon dan variabel prediktor dalam *spatial point pattern*.

Pendekatan Berman-Turner akan dibangun *dummy point* yang tersebar secara seragam pada *linear network* dan semua titik dalam *quadrature* akan menerima bobot w_j . Langkah pertama pada penerapan Berman-Turner untuk *inhomogeneous Poisson point process* agar fungsi *likelihood* nya mendekati fungsi *likelihood* dari *Generalized Linear Model* (GLM) dengan mendiskritkan integral $\int_L f(\mathbf{u})d\mathbf{u}$ dengan penjumlahan bobot $\sum_j w_j f(u_j)$. Fungsi integral pada Persamaan (2.18) jika didekati dengan pemberian bobot pada regresi, akan menjadi persamaan berikut, (Baddeley, et al., 2015)

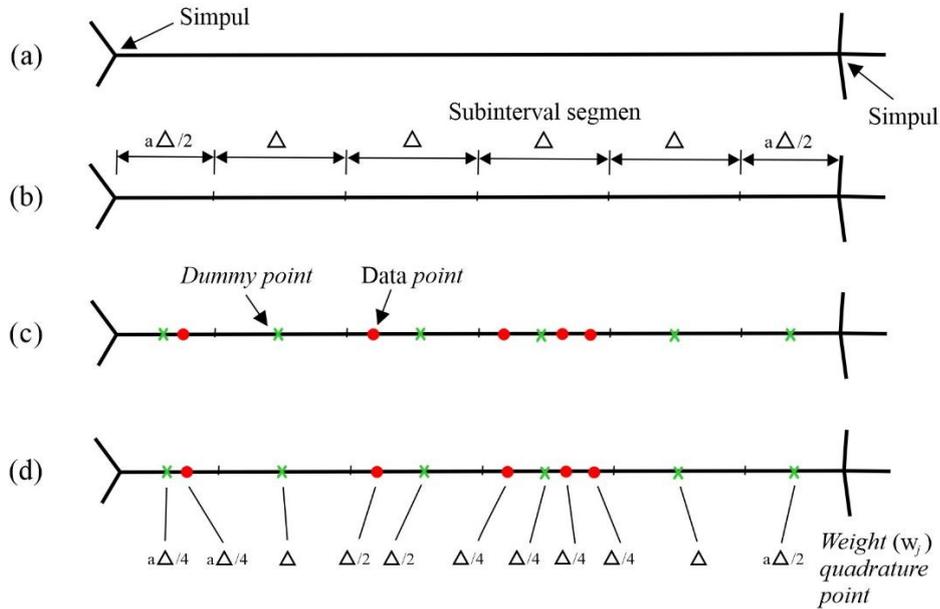
$$\int_L \lambda(\theta; \mathbf{u})d\mathbf{u} \approx \sum_{j=1}^{n+d} \lambda(\theta; \mathbf{u})w_j \quad (2.19)$$

dimana $u_j, j = 1, \dots, n + d$ merupakan titik dalam *observation window* L yang terdiri dari n data *point* dan d *dummy point*, dan $\sum_{j=1}^{n+d} w_j = |L|$. *Summy point* sebanyak d titik dalam pendekatan *Berman-Turner* digabungkan dengan titik data sebanyak n untuk membentuk $n + d$ *quadrature point*. Setiap segmen garis dengan panjang ℓ seperti pada Gambar 2.6 (a) dipartisi menjadi subinterval dengan panjang yang sama yaitu Δ , dengan dua subinterval yang lebih pendek di setiap ujung segmen (lihat Gambar 2.6 (b)). Subinterval di ujung segmen memiliki panjang yang sama yaitu $r/2$ di mana $r = \ell - \left\lfloor \frac{\ell}{\Delta} \right\rfloor \Delta$. Cara untuk menghindari masalah numerik ketika r sangat kecil adalah dengan mengurangi jumlah subinterval Δ sebanyak satu, sehingga $r/2$ bertambah sebesar $\Delta/2$. Sebuah *dummy point* ditempatkan di

tengah subinterval, sehingga setiap subinterval akan memiliki satu *dummy point* dengan atau tanpa titik data (lihat Gambar 2.6 (c)). Rumus penghitungan pembobot w_j dari setiap *quadrature point* yaitu:

$$w_j = \frac{\Delta}{d} \quad (2.20)$$

dengan Δ adalah panjang subinterval di mana poin itu berada dan d adalah jumlah *quadrature point* yang terdapat pada subinterval tersebut (lihat Gambar 2.6 (d)). Gambaran mengenai pendekatan *quadrature* pada *linear network* disajikan pada Gambar 2.6



Gambar 2.6 Skema untuk menetapkan bobot ke titik kuadrat. Setiap interval garis berisi satu titik *dummy* (lingkaran) dan mungkin berisi satu atau lebih titik data (kotak) (McSwiggan, 2019).

Dilakukan substitusi Persamaan (2.19) pada Persamaan (2.18) diperoleh fungsi *log-likelihood* dengan pendekatan Berman-Turner pada Persamaan (2.21)

$$\ell(\boldsymbol{\theta}) \approx \sum_{j=1}^n \log \lambda(\boldsymbol{\theta}; \mathbf{u}_j) - \sum_{j=1}^{n+d} \lambda(\boldsymbol{\theta}; \mathbf{u}_j) w_j \quad (2.21)$$

dan didefinisikan I_j , yaitu:

$$I_j = \begin{cases} 1, & \text{jika } \mathbf{u}_j \text{ adalah data point} \\ 0, & \text{jika } \mathbf{u}_j \text{ adalah dummy point} \end{cases}$$

maka *log-likelihood* pada Persamaan (2.21) dapat dituliskan menjadi

$$\begin{aligned}
\ell(\boldsymbol{\theta}) &\approx \sum_{j=1}^n I_j \log \lambda(\boldsymbol{\theta}; \mathbf{u}_j) + \sum_{j=1}^d I_j \log \lambda(\boldsymbol{\theta}; \mathbf{u}_j) - \sum_{j=1}^{n+d} \lambda(\boldsymbol{\theta}; \mathbf{u}_j) w_j \\
&= \sum_{j=1}^{n+d} I_j \log \lambda(\boldsymbol{\theta}; \mathbf{u}_j) - \sum_{j=1}^{n+d} \lambda(\boldsymbol{\theta}; \mathbf{u}_j) w_j \\
&= \sum_{j=1}^{n+d} [I_j \log \lambda(\boldsymbol{\theta}; \mathbf{u}_j) - \lambda(\boldsymbol{\theta}; \mathbf{u}_j) w_j]
\end{aligned} \tag{2.22}$$

Persamaan (2.22) dapat dituliskan sebagai berikut

$$\ell(\boldsymbol{\theta}) \approx \sum_{j=1}^{n+d} [y_j \log \lambda(\boldsymbol{\theta}; \mathbf{u}_j) - \lambda(\boldsymbol{\theta}; \mathbf{u}_j)] w_j \tag{2.23}$$

dengan $y_j = \frac{I_j}{w_j}$. Penerapan skema Berman-Turner pada Persamaan (2.23) ekuivalen dengan *likelihood* dari model regresi Poisson terboboti dengan variabel respon y_j dan bobot w_j yang diimplementasikan pada `spatstat.linnet` R *packages* dalam fungsi `lppm` (Baddeley, et al., 2021).

Fungsi *likelihood* yang dihasilkan pada Persamaan (2.23) tidak *closed form*, maka dilakukan dengan metode iterasi numerik yaitu *Iteratively Reweighted Least Square* (IRLS). Jika estimasi parameter saat ini dalam proses iteratif adalah $\tilde{\boldsymbol{\theta}}$, dimana nilai $\tilde{\boldsymbol{\theta}}$ merupakan nilai parameter di titik baru yang mendekati nilai $\boldsymbol{\theta}$. Selanjutnya dilakukan pembaruan iteratif dari parameter model dalam mendapatkan solusi optimal melalui pendekatan *quadrature* untuk *log-likelihood* dengan menerapkan ekspansi Taylor:

$$\begin{aligned}
f(x) &= \sum_{h=0}^{\infty} \frac{1}{h!} f^{(h)}(g)(x - g)^h \\
\ell_T(\boldsymbol{\theta}) &= \ell(\tilde{\boldsymbol{\theta}}) + (\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}})^T \ell'(\tilde{\boldsymbol{\theta}}) + \frac{(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}})^T}{2!} \ell''(\tilde{\boldsymbol{\theta}})(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}) \\
&\quad + \frac{(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}})^T}{3!} \ell'''(\tilde{\boldsymbol{\theta}})(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}})^T (\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}) + \dots
\end{aligned}$$

Iterasi numerik IRLS dilakukan dengan ekspansi Taylor orde dua karena efisiensi dan kestabilan antara kecepatan konvergensi dan kompleksitas komputasi. Meskipun secara teoritis penggunaan ekspansi Taylor dengan orde yang lebih tinggi

dapat meningkatkan akurasi dan mempercepat konvergensi, tetapi disisi lain juga meningkatkan kompleksitas komputasi dan perhitungan derivatif pada orde tinggi. Pendekatan ekspansi Taylor orde dua ditulis pada Persamaan (2.24)

$$\ell(\boldsymbol{\theta}) \approx \ell(\tilde{\boldsymbol{\theta}}) + (\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}})^T \ell'(\tilde{\boldsymbol{\theta}}) + \frac{(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}})^T}{2!} \ell''(\tilde{\boldsymbol{\theta}})(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}) \quad (2.24)$$

dimana:

- $\tilde{\boldsymbol{\theta}}$: nilai parameter di titik baru yang mendekati nilai $\boldsymbol{\theta}$
- $\ell(\tilde{\boldsymbol{\theta}})$: fungsi *likelihood* di suatu titik baru yang mendekati titik asli dari $\ell(\boldsymbol{\theta})$
- $\ell'(\tilde{\boldsymbol{\theta}})$: vektor gradien fungsi *likelihood* di suatu titik berukuran $[(p+1) \times 1]$
- $\ell''(\tilde{\boldsymbol{\theta}})$: matriks Hessian fungsi *likelihood* di suatu titik baru berdimensi $[(p+1) \times (p+1)]$.

Setelah dilakukan estimasi parameter, maka dilakukan uji signifikansi untuk mengetahui kovariat mana saja yang signifikan berpengaruh terhadap model. Pengujian signifikansi parameter secara parsial dilakukan menggunakan uji Z terhadap parameter $\boldsymbol{\theta}$ dengan hipotesis sebagai berikut:

$H_0 : \theta_i = 0; i = 0,1,2, \dots, p$ (parameter tidak signifikan)

$H_1 : \theta_i \neq 0; i = 0,1,2, \dots, p$ (parameter signifikan)

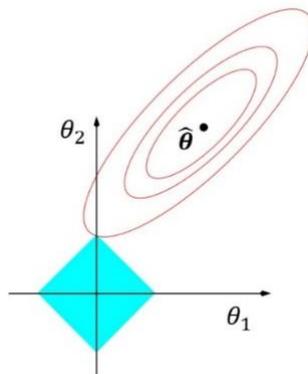
Satastistik uji yang digunakan adalah uji Z, dengan rumus sebagai berikut (Baddeley, et al., 2015)

$$Z = \frac{\hat{\theta}_i}{\widehat{se}(\hat{\theta}_i)} \sim N(0,1) \quad (2.25)$$

$\widehat{se}(\hat{\theta}_i)$ adalah standar error dari $\hat{\theta}_i$ dengan $se(\hat{\theta}_i) = \sqrt{\hat{v}_{ii}}$ yaitu akar kuadrat dari elemen diagonal matriks $\mathbf{I}^{-1}(\hat{\boldsymbol{\theta}}_i)$ pada baris ke- i dan kolom ke- i dimana $\mathbf{I}(\boldsymbol{\theta})$ merupakan matriks Hessian atau turunan kedua dari fungsi *likelihood* $\ell(\boldsymbol{\theta})$. Daerah penolakan uji Z yaitu tolak H_0 jika $|Z| > Z_{\alpha/2}$ atau $p\text{-value} < \alpha$ dengan α merupakan taraf signifikansi yang ditentukan, maka dapat disimpulkan bahwa kovariat berpengaruh signifikan terhadap intensitas *point process*.

2.5.2 Seleksi Variabel dengan Regularisasi dan Metode Pemilihan *Tuning Parameter*

Pemodelan intensitas dapat dilakukan dengan menggunakan Persamaan (2.23), namun persamaan tersebut belum dapat digunakan untuk seleksi variabel. Seleksi variabel pada penelitian ini dilakukan dengan menggunakan regularisasi dengan $\ell(\boldsymbol{\theta})$ pada Persamaan (2.23). Tujuan utama dalam melakukan seleksi variabel adalah mencari kovariat yang relevan dengan intensitas dari model *point process* \mathbf{X} dan mengestimasi koefisien parameter pada setiap variabel prediktor yang terseleksi di *point process* \mathbf{X} . Regularisasi merupakan metode alternatif dalam seleksi variabel dan estimasi parameter secara bertahap dengan menambahkan fungsi penalti untuk mengontrol kompleksitas model terutama ketika data berdimensi tinggi yang memungkinkan mengalami *overfitting* dan ketidakstabilan seleksi variabel. Regularisasi juga dapat menangani prediktor yang berkorelasi dengan lebih baik daripada metode *stepwise*, karena cenderung memilih hanya satu dari sekelompok prediktor yang berkorelasi (Kumar, et al., 2019).



Gambar 2.7 Kontur estimasi parameter dengan Lasso (Tibshirani, 1996).

Least Absolute Shrinkage and Selection Operator (Lasso) menggunakan regularisasi L_1 yang mengimplementasikan fungsi penalti berdasarkan nilai absolut dari koefisien dikalikan dengan parameter regularisasi. Ketika divisualisasikan dalam ruang dua dimensi, batasan yang dihasilkan berbentuk diamond seperti Gambar 2.7. Efek dari pendekatan ini adalah kemampuan Lasso untuk menekan koefisien menjadi tepat nol, memungkinkan sebagai alat dalam pemilihan variabel yang signifikan. Hal ini dikarenakan solusi optimal dengan batasan L_1 cenderung ditemukan pada kontur fungsi menyentuh batasan diamond. Karena

bentuk diamond dengan titik-titik di sumbu diamond, ada kecenderungan bahwa kontur akan menyentuh batasan di salah satu sudutnya yang menandakan koefisien tepat menjadi nol (Tibshirani, 1996).

Misalkan $\ell(\boldsymbol{\theta})$ merupakan pendekatan *likelihood Poisson point process on a linear network* pada Persamaan (2.20) dengan vektor koefisien regresi $\boldsymbol{\theta} = (\theta_0, \theta_1, \theta_2, \dots, \theta_p)$. Fungsi *Penalized likelihood* Lasso dituliskan pada Persamaan (2.26) (Choiruddin, et al., 2018)

$$\rho^{Pen}(\boldsymbol{\theta}) = \ell(\boldsymbol{\theta}) - \gamma \sum_{k=1}^p |\theta_k| \quad (2.26)$$

dimana $\ell(\boldsymbol{\theta})$ merupakan fungsi *log-likelihood* model *likelihood Poisson* (2.23) $k = 1, 2, \dots, p$, p merupakan jumlah parameter kovariat yang digunakan, $\gamma \geq 0$ merupakan *tuning parameter* dari metode regularisasi yang digunakan. Selanjutnya untuk memperoleh nilai estimasi parameter θ_k dengan regularisasi, digunakan CDA. CDA diterapkan untuk menyelesaikan *penalized problem*.

Misalkan terdapat estimator $\tilde{\theta}_1, \tilde{\theta}_2, \dots, \tilde{\theta}_p$, maka metode *coordinate descent* terdiri dari pengoptimalan secara parsial (2.27) terhadap θ_k :

$$\arg \min_{\theta_k \in \mathbb{R}} \rho^{Pen}(\tilde{\theta}_1, \dots, \tilde{\theta}_{k-1}, \theta_k, \tilde{\theta}_{k+1}, \dots, \tilde{\theta}_p) \quad (2.27)$$

dan Persamaan (2.27) berjalan untuk $k = 1, 2, \dots, p$ secara iteratif. Penentuan *tuning parameter* γ menjadi hal penting karena digunakan untuk mengontrol regularisasi yang diterapkan pada model. Untuk mendapatkan nilai γ yang optimal, terlebih dahulu ditentukan barisan menurun $\gamma \in |\gamma_{min}, \gamma_{max}|$ dengan γ_{max} sampai dengan $\gamma_{min} = 10^{-3}\gamma_{max}$ dimana γ_{max} adalah nilai terkecil dari γ yang menghasilkan semua koefisien dalam model menjadi nol dihitung berdasarkan Persamaan (2.28) (Friedman, et al., 2010).

$$\gamma_{max} = \frac{\max |X^T y|}{N} \quad (2.28)$$

Pemilihan model terbaik dapat menggunakan perhitungan nilai *Bayesian Information Criterion* (BIC). Pada penelitian ini nilai BIC digunakan untuk pemilihan model terbaik dengan nilai $\gamma \geq 0$ optimal. Perhitungan BIC dirumuskan dalam persamaan (2.29) berikut (Choiruddin, et al., 2021)

$$BIC(\gamma) = -2\ell(\hat{\theta}(\gamma)) + h(\gamma)\log(n) \quad (2.29)$$

dimana $\ell(\hat{\theta}(\gamma))$ merupakan *likelihood given* γ , $h(\gamma)$ adalah jumlah koefisien $\hat{\theta}(\gamma)$ bukan nol dalam model dengan *tuning parameter* γ dan n adalah banyaknya titik pada *observation window* $L \subset \mathbb{R}^2$.

Secara umum, algoritma estimasi parameter menggunakan CDA dapat dilihat pada Tabel 2.4.

Tabel 2.4 Algoritma CDA

Langkah 1:	Melakukan pendekatan fungsi <i>likelihood</i> dengan ekspansi Taylor.
Langkah 2:	Membangun fungsi <i>penalized likelihood</i> .
Langkah 3:	Mengidentifikasi barisan menurun dari <i>tuning parameter</i> $\gamma \in \gamma_{min}, \gamma_{max} $. Untuk setiap nilai γ , <i>outer loop</i> diulang sebanyak elemen barisan γ untuk menghitung $\ell(\theta)$ pada $\hat{\theta}$.
Langkah 4:	Menerapkan <i>Coordinate Descent Algorithm</i> untuk menyelesaikan <i>penalized problem</i> dengan mengoptimasi <i>partially</i> (tiap k) dan mengevaluasi nilai estimasi parameter.
Langkah 5:	Mendapatkan nilai γ yang optimal menggunakan $BIC(\gamma)$ sehingga nilai $\hat{\theta}$ diperoleh menyesuaikan dengan nilai γ terpilih.
Langkah 6:	Melakukan iterasi untuk mendapatkan penaksir parameter baru.
Langkah 7:	Memeriksa konvergensi parameter baru hasil iterasi, apabila sudah konvergen maka iterasi dapat dihentikan.

2.6 Kecelakaan Lalu Lintas dan Faktor yang Mempengaruhinya

Undang-Undang Republik Indonesia Nomor 22 Tahun 2009 mendefinisikan kecelakaan lalu lintas sebagai kejadian tak terduga dan tak disengaja di jalan yang melibatkan kendaraan, dengan atau tanpa pengguna jalan lain, menyebabkan cedera manusia dan/atau kerusakan harta benda. Penelitian terkait kecelakaan lalu lintas, Liu, et al. (2018) meneliti hubungan antara frekuensi kecelakaan di China dengan beberapa variabel independen melalui regresi binomial negatif. Variabel tersebut mencakup faktor manusia (seperti pengemudi profesional, pengaruh alkohol atau obat-obatan, dan kelelahan), faktor kendaraan (jenis kendaraan, kelebihan beban, dan masalah rem), serta faktor lingkungan dan jalan (cuaca, jenis jalan, medan, dan lokasi). Penelitian ini menemukan bahwa setiap faktor memiliki setidaknya satu variabel yang signifikan dalam

mempengaruhi kecelakaan. Sementara itu, Manan, et al. (2018) menggunakan metode multinomial logit dan mixed logit untuk mengkaji berbagai karakteristik jalan dan lingkungan yang diduga berkontribusi pada tiga jenis kasus kecelakaan sepeda motor di Malaysia, menemukan bahwa karakteristik tertentu meningkatkan risiko kecelakaan sepeda motor. Rakshit, et al. (2021) mengidentifikasi faktor-faktor signifikan dengan melibatkan pemilihan variabel untuk model *spatial point process* pada *linear network* mencakup efek kovariat spasial, efek *mark*, dan interaksi *mark*-kovariat pada data kecelakaan lalu lintas di Australia Barat. Kovariat sebagai faktor eksternal yang digunakan diantaranya batas kecepatan legal, radius kurva jalan, lebar jalan, lebar sisi jalan, jumlah lajur, keberadaan kerb (pelengkap jalan) di samping kanan dan kiri bahu jalan, keberadaan bantalan di bahu jalan, serta keberadaan jembatan di atas jalan.

Berdasarkan modul Pengenalan Rekayasa Keselamatan Jalan yang dipublikasikan oleh Kementerian Pekerjaan Umum dan Perumahan Rakyat (2016), berikut ini merupakan penjelasan dari masing-masing faktor penyebab kecelakaan lalu lintas.

1. Faktor Manusia

Faktor manusia merupakan faktor penyebab dominan atas terjadinya kecelakaan. Terdapat dua pihak yang menggunakan jalan, yaitu pejalan kaki dan pengemudi kendaraan. Hampir semua kejadian kecelakaan akibat faktor manusia didahului dengan terjadinya pelanggaran rambu-rambu atau peraturan lalu lintas, baik disengaja maupun tidak disengaja. Sebagai contoh yaitu pelanggaran rambu batas kecepatan. Banyak pengemudi yang mengemudikan kendaraannya melebihi batas kecepatan yang telah ditentukan sesuai dengan kategori fungsi jalan. Kecelakaan di jalan raya dapat disebabkan oleh kondisi fisik, mental, dan sikap pengemudi yang buruk, keterampilan mengemudi yang tidak memadai, serta mengemudi di bawah pengaruh alkohol atau obat-obatan. Meskipun keterampilan berkendara diuji sebagai persyaratan untuk mendapatkan surat ijin mengemudi, seorang pengemudi masih dapat memiliki risiko yang tinggi mengalami kecelakaan karena perasaan percaya diri yang berlebihan.

2. Faktor Kendaraan

Kendaraan bermotor sebagai hasil produksi suatu pabrik telah dirancang dengan memperhatikan faktor keamanan untuk menjamin keselamatan bagi pengendaranya. Kendaraan dapat menjadi faktor penyebab kecelakaan apabila tidak dapat dikendalikan sebagaimana mestinya sebagai akibat dari kondisi teknis yang tidak layak jalan atau penggunaannya tidak sesuai ketentuan. Ada beberapa hal yang perlu diperhatikan pada kendaraan untuk mengurangi risiko kecelakaan lalu lintas yaitu

a. Perawatan.

Kendaraan harus dirawat dengan baik untuk menghindari kerusakan pada bagian kendaraan seperti rem yang tidak berfungsi, kerusakan mesin, ban pecah, kemudi tidak baik, lampu yang mati, dan lain lain.

b. Muatan.

Pengendara harus membawa muatan secukupnya atau tidak melebihi kapasitas dan juga harus menempatkan muatan dengan benar dan seimbang.

c. Perlengkapan Keamanan.

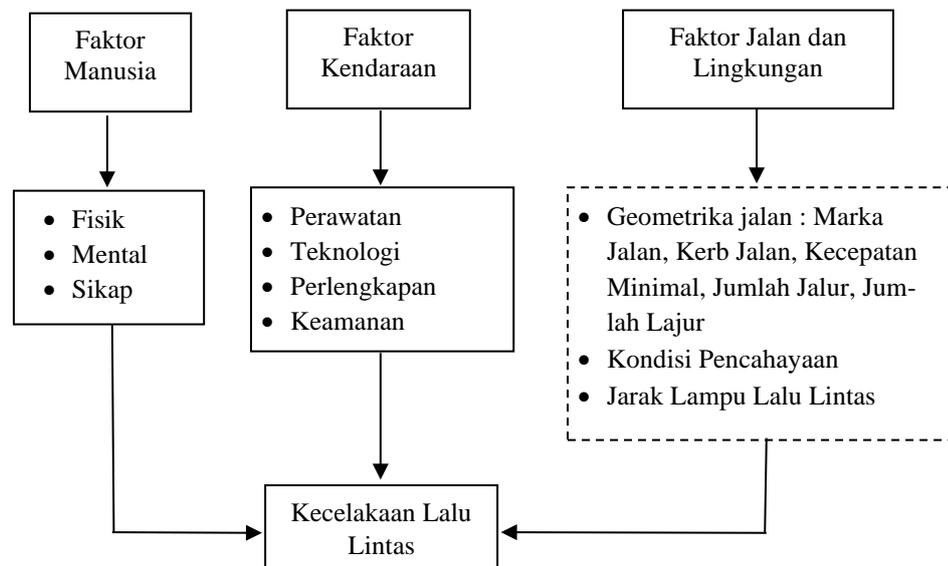
Sebelum berkendara, pengendara juga perlu memastikan perlengkapan keamanan dapat berfungsi dengan baik, sebagai contoh sabuk pengaman, air bag atau kantong udara, helm bagi pengendara kendaraan beroda dua, dan lain-lain.

3. Faktor Jalan dan Lingkungan

Faktor jalan dan lingkungan juga merupakan salah satu penyebab kecelakaan lalu lintas. Hal-hal yang berkaitan dengan faktor jalan antara lain geometrik jalan (lurus, menikung, tanjakan, turunan), batas kecepatan yang telah ditetapkan, pagar pengaman di daerah pegunungan, ada tidaknya marka jalan dan rambu-rambu lalu lintas, jumlah jalur, jenis jalan, dan lain-lain. Sedangkan faktor lingkungan yang dapat menjadi penyebab kecelakaan antara lain yaitu perubahan cuaca, pergantian siang dan malam, bencana alam, keamanan (daerah rawan tindak kriminal atau daerah konflik), dan lain-lain. Faktor lingkungan tersebut dapat mengakibatkan

kondisi jalanan berubah, sebagai contoh adalah jalanan yang licin dan jarak pandang yang berkurang akibat hujan.

Kerangka konseptual dari penelitian kecelakaan lalu lintas dapat dilihat pada Gambar 2.8



Gambar 2.8 Kerangka konseptual penelitian (KPUPR, 2016; Austroads, 2002).

Keterangan :

 Variabel yang diteliti

(halaman ini sengaja dikosongkan)

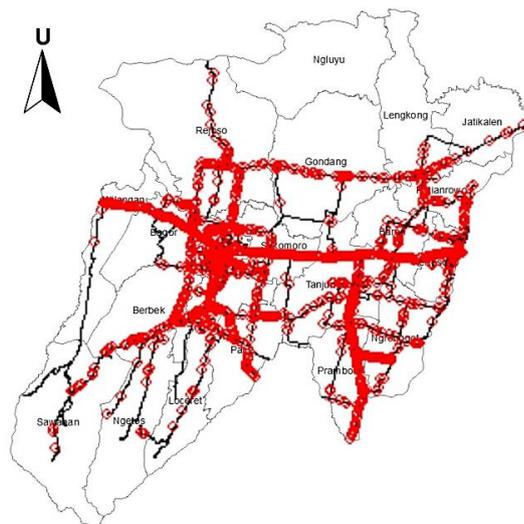
BAB 3

METODE PENELITIAN

3.1 Sumber Data

Penelitian ini menggunakan data sekunder dengan rincian sebagai berikut.

1. Data koordinat lokasi kecelakaan lalu lintas pada periode 1 Januari 2018 hingga 31 Desember 2020 sebanyak 1.944 kejadian kecelakaan lalu lintas serta koordinat lokasi lampu lalu lintas di Kabupaten Nganjuk sebanyak 100 lokasi lampu lalu lintas diperoleh dari penelitian sebelumnya yang dilakukan oleh Hasanah, et al. (2022) terlampir pada Lampiran 1 dan Lampiran 2. Koordinat lokasi kecelakaan lalu lintas di Kabupaten Nganjuk tahun 2018-2020 pada *linear network* yang digunakan dalam penelitian ini ditampilkan pada Gambar 3.1.



Gambar 3.1 Lokasi kecelakaan lalu lintas di Kabupaten Nganjuk tahun 2018-2020.

2. Data faktor-faktor yang diduga mempengaruhinya meliputi jenis jalan, jumlah lajur, jumlah jalur, keberadaan kerb (pelengkap jalan) di sisi kanan dan kiri jalan diperoleh dari observasi melalui *google maps* dengan nilai pada setiap ruas jalan di seluruh *observation window L* di Kabupaten Nganjuk terlampir pada Lampiran 3.
3. Data jenis jalan yang di *treatment* sebagai kovariat diperoleh dari dari Shapefile RBI Provinsi Jawa Timur tahun 2020 yang diakses melalui website <https://www.indonesia-geospasial.com>.

Data kovariat termasuk faktor lingkungan dan faktor geometri jalan dalam penelitian merupakan data tahun 2020, sementara data kecelakaan lalu lintas yang dianalisis mencakup periode dari tahun 2018 hingga 2020. Asumsi dasar yang mendasari penggunaan data ini adalah bahwa kondisi jalan tetap konstan selama periode analisis, tanpa adanya perluasan yang berarti bahwa jumlah lajur dan jalur cenderung tidak berubah. Selain itu, mengingat tipe jalan yang tidak berubah selama periode ini, kecepatan minimum di jalan diasumsikan konsisten dengan karakteristik desain jalan yang berlaku. Ketiadaan kerb kanan dan kiri jalan konsisten dengan apa yang diamati pada tahun-tahun sebelumnya, karena sebagian besar jalan tidak memiliki trotoar. Oleh karena itu, perbedaan tahun antara data kecelakaan dan data kovariat tidak menimbulkan masalah dalam analisis ini, karena kondisi dan karakteristik jalan diasumsikan tetap konsisten sepanjang periode yang diteliti.

3.2 Variabel Penelitian

Variabel yang digunakan pada penelitian ini adalah berupa titik koordinat *latitude* dan *longitude* lokasi kecelakaan lalu lintas serta kovariat spasial faktor lingkungan yang mempengaruhi kecelakaan lalu lintas pada seluruh segmen garis *observation window L* di Kabupaten Nganjuk.

Tabel 3.1 Variabel penelitian kecelakaan lalu lintas

Variabel	Keterangan	Skala Pengukuran	Keterangan
x	Himpunan titik koordinat lintang dan bujur lokasi kecelakaan lalu lintas di Kabupaten Nganjuk	Interval	
$z_1(\mathbf{u})$	Jarak ke lokasi lampu lalu lintas terdekat	Rasio	
$z_2(\mathbf{u})$	Jenis jalan	Nominal	0: Jalan lokal 1: Jalan kolektor 2: Jalan arteri
$z_3(\mathbf{u})$	Jumlah lajur	Rasio	
$z_4(\mathbf{u})$	Jumlah jalur	Rasio	
$z_5(\mathbf{u})$	Keberadaan kerb di sisi kiri jalan	Nominal	0: Tidak ada 1: Ada
$z_6(\mathbf{u})$	Keberadaan kerb di sisi kanan jalan	Nominal	0: Tidak ada 1: Ada

Variabel penelitian ditunjukkan pada Tabel 3.1, dan definisi operasional dari setiap variabel yang digunakan pada penelitian ini adalah sebagai berikut:

1. Lokasi kecelakaan lalu lintas

Titik koordinat geografi yang terdiri dari *latitude* dan *longitude* lokasi kecelakaan lalu lintas pada seluruh segmen garis *observation window* *L* di Kabupaten Nganjuk.

2. Jarak ke lokasi lampu lalu lintas terdekat

Kovariat ini menunjukkan jarak suatu titik dengan lokasi lampu lalu lintas terdekat yang dihitung menggunakan *shortest-path distance* pada Persamaan (2.1). Lampu lalu lintas yang dimaksud yaitu lampu merah yang terdapat di persimpangan jalan atau disebut sebagai *traffic light* dan lampu kuning yang biasa terdapat di zona penyeberangan pejalan kaki (*zebra cross*) atau disebut dengan *flashlight*. Lampu lalu lintas secara umum terletak di persimpangan yang sibuk, dimana terdapat peningkatan volume lalu lintas, baik kendaraan maupun pejalan kaki seperti di sekitar sekolah, pasar, persimpangan jalan, kawasan industri, atau fasilitas umum lainnya. Peningkatan volume ini dapat meningkatkan kecelakaan. Penggunaan kovariat jarak ke lokasi lampu lalu lintas terdekat di sepanjang segmen jalan sebagai indikator jarak suatu titik ke lokasi keramaian pada *observation window* dalam bentuk *linear network*. Hipotesis pengaruh kovariat jarak ke lokasi lampu lalu lintas terhadap risiko kecelakaan lalu lintas yaitu: jarak suatu titik ke lokasi lampu lalu lintas terdekat memberikan pengaruh negatif terhadap risiko kecelakaan lalu lintas di Kabupaten Nganjuk. Artinya, semakin jauh jarak suatu titik ke lokasi lalu lintas, maka semakin kecil risiko terjadinya kecelakaan lalu lintas.

3. Jenis jalan

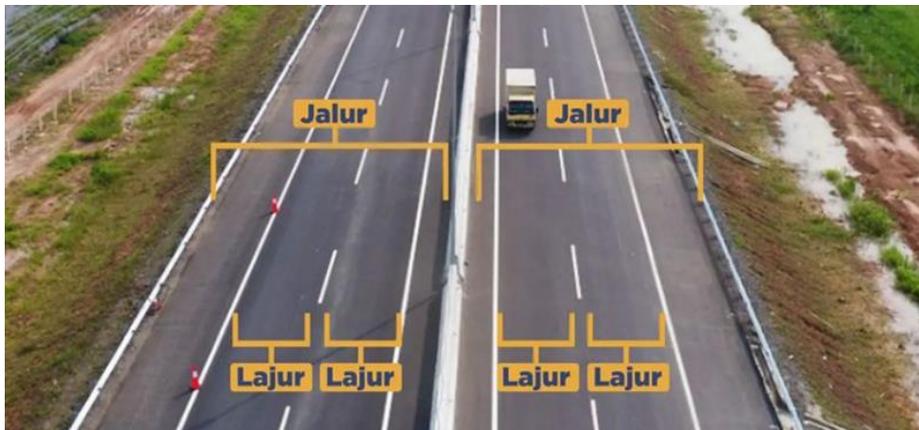
Tiga jenis jalan yang digunakan sebagai dalam penelitian ini adalah jalan arteri, jalan kolektor, dan jalan lokal. Pengertian dari ketiga jenis jalan tersebut tercantum dalam Undang-Undang Republik Indonesia No. 2 tentang Jalan, yang dijelaskan sebagai berikut. Definisi masing-masing kategori jalan dideskripsikan sebagai berikut:

- a. Jalan arteri merupakan jalan umum yang berfungsi melayani angkutan utama dengan ciri perjalanan jarak jauh, kecepatan rata-rata tinggi, dan jumlah jalan masuk dibatasi secara berdaya guna. Jalan arteri didesain dengan kecepatan rencana paling rendah 60 km/jam dengan lebar badan jalan paling sedikit 11 meter.
- b. Jalan kolektor merupakan jalan umum yang berfungsi melayani angkutan pengumpul atau pembagi dengan ciri perjalanan jarak sedang, kecepatan rata-rata sedang, dan jumlah jalan masuk dibatasi. Jalan arteri didesain dengan kecepatan rencana paling rendah 40 km/jam dengan lebar badan jalan paling sedikit 9 meter.
- c. Jalan lokal merupakan jalan umum yang berfungsi melayani angkutan setempat dengan ciri perjalanan jarak dekat, kecepatan rata-rata rendah, dan jumlah jalan masuk tidak dibatasi. Jalan arteri didesain dengan kecepatan rencana paling rendah 20 km/jam dengan lebar badan jalan paling sedikit 7,5 meter.

Hipotesis pengaruh kovariat jenis jalan terhadap risiko kecelakaan lalu lintas yaitu: risiko kecelakaan lalu lintas yang terjadi di jalan lokal lebih rendah dibandingkan dengan risiko kecelakaan lalu lintas yang terjadi di jalan arteri dan jalan kolektor.

4. Jumlah lajur

Lajur adalah bagian jalur yang memanjang, dengan atau tanpa marka jalan, yang memiliki lebar cukup untuk satu kendaraan bermotor sedang berjalan, selain sepeda motor seperti pada ilustrasi Gambar 3.2.



Gambar 3.2 Ilustrasi jalur dan lajur di segmen jalan.

Jumlah lajur dalam satu arah bisa bervariasi, mulai dari satu lajur (untuk jalan-jalan kecil atau di daerah terpencil) hingga empat lajur atau lebih (untuk jalan bebas hambatan atau jalan arteri di daerah perkotaan). Hipotesis pengaruh kovariat jumlah lajur terhadap risiko kecelakaan lalu lintas yaitu: jumlah lajur memberikan pengaruh positif terhadap risiko kecelakaan lalu lintas di Kabupaten Nganjuk. Artinya, semakin banyak jumlah lajur maka risiko kecelakaan lalu lintas akan semakin meningkat.

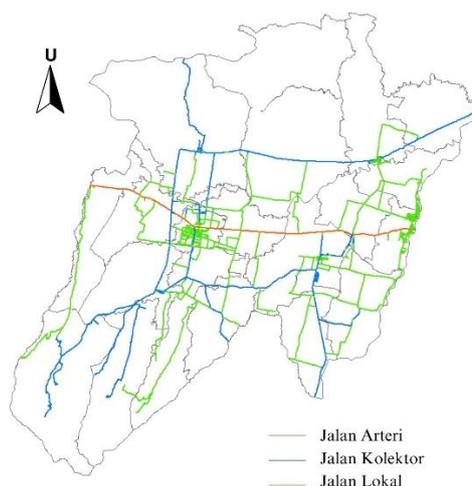
5. Jumlah jalur

Jalur adalah bagian jalan yang dipergunakan untuk lalu lintas kendaraan seperti ilustrasi pada Gambar 3.2. Jumlah jalur pada suatu segmen jalan merujuk pada jumlah jalur yang tersedia untuk kendaraan. Jalur ini biasanya dipisahkan oleh garis-garis marka jalan. Jumlah jalur dalam satu arah dapat bervariasi, tergantung pada desain dan kapasitas jalan tersebut. Hipotesis pengaruh kovariat jumlah jalur terhadap risiko kecelakaan lalu lintas yaitu: jumlah jalur memberikan pengaruh positif terhadap risiko kecelakaan lalu lintas di Kabupaten Nganjuk. Artinya, semakin banyak jumlah jalur yang diterapkan, maka risiko kecelakaan lalu lintas akan semakin meningkat.

6. Keberadaan kerb di sisi kanan dan kiri jalan

Keberadaan kerb jalan menunjukkan ada atau tidaknya kerb di sisi kanan dan kiri jalan. Menurut SNI 03-2442-1991 Kerb adalah bagian dari jalan sebagai pelengkap jalan untuk membantu keamanan serta kenyamanan para pemakai jalan. Beberapa fungsi kerb, antara lain untuk menghalangi atau mencegah kendaraan keluar dari jalur lalu-lintas; untuk membentuk sistem drainase perkerasan jalan; sebagai proteksi terhadap pejalan kaki; untuk mempertegas batas jalur lalu-lintas kendaraan dengan jalur-jalur lainnya; dan untuk menambah estetika. Sesuai fungsi aslinya, hipotesis pengaruh kovariat kerb jalan terhadap risiko kecelakaan lalu lintas yaitu: risiko kecelakaan lalu lintas yang terjadi di jalan yang memiliki kerb di sisi kanan/kiri jalan lebih rendah dibandingkan dengan risiko kecelakaan lalu lintas yang terjadi di jalan tanpa kerb di sisi kanan/kiri jalan.

Observation window yang digunakan berupa jaringan jalan di Kabupaten Nganjuk yang dikategorikan menjadi jalan arteri, kolektor, lokal, dan lingkungan menurut fungsinya sesuai dengan Undang-Undang Republik Indonesia No. 2 tentang Jalan. Pada penelitian ini dibatasi pada jalan arteri, kolektor, dan lokal (seperti yang ditunjukkan pada Gambar 3.3).



Gambar 3.3 Jaringan Jalan Kabupaten Nganjuk (Shapefile RBI Provinsi Jawa Timur).

3.3 Proses Pengambilan Data Kovariat Spasial dan *Pre-Processing* Data

Jaringan jalan Kabupaten Nganjuk yang digunakan sebagai *observation window* diperoleh dari *Shapefile* RBI Provinsi Jawa Timur tahun 2020 yang diakses melalui website <https://www.indonesia-geospasial.com>. *Shapefile* berisi kumpulan data ruas-ruas jalan yang dilengkapi informasi mengenai titik koordinat longitude dan latitude pada bagian awal dan akhir setiap ruas jalan serta informasi mengenai jenis jalan meliputi jalan arteri, kolektor, lokal, jalan setapak, dan jalan lainnya. *Pre-processing* mencakup verifikasi menggunakan *Google Maps* untuk mengevaluasi jenis jalan. Jalur seperti jalan setapak dan jalur lain yang tidak umum untuk lalu lintas kendaraan bermotor, seperti jalan di tengah hutan atau di antara sawah yang biasanya hanya dapat dilewati oleh pejalan kaki atau sepeda, tidak termasuk dalam kategori jalan umum yang dapat dilalui kendaraan, seperti yang ditunjukkan pada Gambar 3.4. Berdasarkan hal ini, dilakukan filter ruas jalan sebagai batasan dalam penelitian ini yang hanya terdiri dari tiga jenis jalan saja yaitu jalan arteri, jalan kolektor, dan jalan lokal untuk membentuk jaringan jalan di Kabupaten Nganjuk sebagai *observation window* yang terdiri dari 11.736 ruas jalan.



Gambar 3.4 Gambaran Ruas Jalan Setapak di Kabupaten Nganjuk (*Google Maps*). Lokasi di Jl. Wilis I, Pacar, Sono Patik, Kec. Berbek, Kabupaten Nganjuk, Jawa Timur 64473.

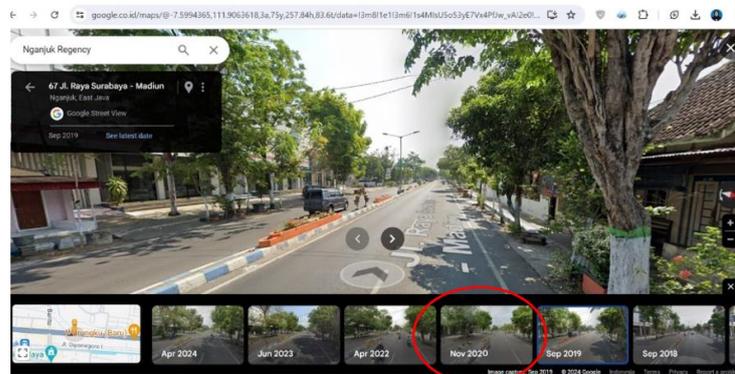
Data kovariat spasial berupa geometrika jalan pada penelitian ini didapatkan melalui observasi *google maps* dengan langkah-langkah sebagai berikut.

1. Menelusuri satu persatu koordinat (longitude dan latitude) ruas jalan pada *Shapefile* ruas jalan Kabupaten Nganjuk yang sudah difilter sesuai jenis jalan arteri, kolektor dan lokal.
2. Memilih opsi *street view* untuk melihat kondisi geometrika jalan pada setiap ruas jalan seperti pada Gambar 3.5



Gambar 3.5 Opsi *street view* pada *Google Maps*.

3. Menyesuaikan kondisi jalan pada tahun 2020 dengan memilih opsi tahun pengambilan foto ruas jalan pada *Google Maps* seperti Gambar 3.6



Gambar 3.6 Pemilihan tahun pengambilan foto jalan pada *Google Maps*. Lokasi gambar di Jl. Raya Surabaya - Madiun No.67, Mangundikaran, Mangun Dikaran, Kec. Nganjuk, Kabupaten Nganjuk, Jawa Timur 64419.

Selanjutnya dilakukan identifikasi terhadap kondisi jalan pada tiap ruas jalan di *google maps* pada tahun 2018, 2019, dan 2020 untuk memastikan tidak terdapat perubahan struktur jalan yang signifikan selama periode data kecelakaan lalu lintas yang digunakan yaitu 2018-2020. Berikut merupakan kondisi jalan pada ketiga tahun tersebut dari tiga titik koordinat ruas jalan.



Gambar 3.7 Kondisi jalan pada koordinat jalan $7,604810238^{\circ}$ LS dan $112,0810013^{\circ}$ BT (a) tahun 2018 (b) tahun 2019 (c) tahun 2020.

Berdasarkan hasil analisis visual yang ditampilkan pada Gambar 3.7 dapat disimpulkan bahwa kondisi jalan pada tahun 2018, 2019, dan 2020 tidak

menunjukkan perbedaan yang signifikan. Hal ini mencakup aspek-aspek komponen jalan seperti kondisi marka jalan yang tetap konsisten, jumlah lajur yang tidak mengalami perubahan, jumlah jalur yang tetap sama, serta kondisi kerb kanan dan kerb kiri yang tidak menunjukkan perubahan yang berarti. Maka, penggunaan data kondisi jalan dari tahun 2020 untuk analisis dan evaluasi ini dapat dianggap valid dan tidak menimbulkan masalah, karena kondisi jalan yang digunakan sebagai acuan tetap representatif untuk keseluruhan periode yang dianalisis. Kondisi jalan pada dua titik lainnya ditunjukkan pada Gambar 3.8 dan Gambar 3.9



Gambar 3.8 Kondisi jalan pada koordinat jalan $7,602650166^{\circ}$ LS dan $112,0899963^{\circ}$ BT (a) tahun 2018 (b) tahun 2019 (c) tahun 2020.



Gambar 3.9 Kondisi jalan pada koordinat jalan $7,676750183^{\circ}$ LS dan $111,9160004^{\circ}$ BT
 (a) tahun 2018 (b) tahun 2019 (c) tahun 2020.

4. Mengumpulkan informasi mengenai kovariat spasial yang dibutuhkan dalam penelitian berupa:
 - a. Jumlah lajur
 - b. Jumlah jalur
 - c. Keberadaan kerb kanan jalan
 - d. Keberadaan kerb kiri jalan
5. Memberikan kode pada kovariat spasial sesuai Tabel 3.1

6. Penyesuaian shapefile jaringan jalan Kabupaten Nganjuk yang sudah difilter berdasarkan jenis jalan arteri, kolektor, dan lokal dengan tambahan kovariat spasial dengan bantuan *software* ArcGIS.

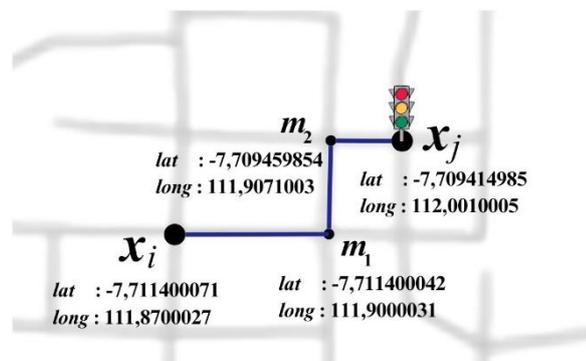
3.4 Struktur Data

Tabel 3.1 menguraikan sembilan variabel yang relevan untuk penelitian ini, dengan rincian setiap struktur data variabel tersebut dijelaskan secara mendetail dalam Tabel 3.2 hingga Tabel 3.5. Struktur data variabel himpunan titik koordinat lintang dan bujur lokasi kecelakaan lalu lintas pada *observation window* L di Kabupaten Nganjuk (x_i) ditampilkan dalam Tabel 3.2.

Tabel 3.2 Struktur data kecelakaan lalu lintas di Kabupaten Nganjuk

Titik lokasi kecelakaan lalu lintas di Kabupaten Nganjuk	Longitude	Latitude
x_1	$x_{long,1}$	$x_{lat,1}$
x_2	$x_{long,2}$	$x_{lat,2}$
x_3	$x_{long,3}$	$x_{lat,3}$
\vdots	\vdots	\vdots
$x_{1.944}$	$x_{long,1944}$	$x_{lat,1944}$

Variabel kovariat himpunan titik koordinat lintang dan bujur lokasi lampu lalu lintas dalam analisis ini dilakukan transformasi terhadap *dataset* titik koordinat lintang dan bujur setiap lampu lalu lintas menggunakan perhitungan *shortest-path distance* pada Persamaan (2.1) untuk mengetahui jarak suatu titik ke lokasi lampu lalu lintas terdekat ($z_1(\mathbf{u})$). Sebagai ilustrasi misalkan dihitung jarak suatu titik x_i ke lokasi lampu lalu lintas terdekat di titik x_j dengan lintasan jalur terpendek seperti Gambar 3.10 dengan perhitungan *shortest-path distance* sebagai berikut.



Gambar 3.10 Ilustrasi jarak terpendek suatu titik x_i ke lokasi lampu lalu lintas terdekat di titik x_j .

- a. Menghitung jarak euclidean antar titik pada Gambar 3.10 menggunakan Persamaan (2.2).

$$\begin{aligned}\|x_i - m_1\| &= \sqrt{(m_{lat,1} - x_{lat,i})^2 + (m_{long,1} - x_{long,i})^2} \\ &= \sqrt{(-7,711400042 - (-7,711400071))^2 + (111,9000031 - 111,8700027)^2} \\ &= 0,03000^\circ\end{aligned}$$

$$\begin{aligned}\|m_1 - m_2\| &= \sqrt{(m_{lat,2} - m_{lat,1})^2 + (m_{long,2} - m_{long,1})^2} \\ &= \sqrt{(-7,709459854 - (-7,711400042))^2 + (111,9071003 - 111,9000031)^2} \\ &= 0,007358^\circ\end{aligned}$$

$$\begin{aligned}\|m_2 - x_j\| &= \sqrt{(x_{lat,j} - m_{lat,2})^2 + (x_{long,j} - m_{long,2})^2} \\ &= \sqrt{(-7,709414985 - (-7,709459854))^2 + (112,0010005 - 111,9071003)^2} \\ &= 0,093900^\circ\end{aligned}$$

- b. Menghitung *shortest-path distance* pada Persamaan (2.1) dengan menjumlahkan seluruh jarak *euclidean* antar titik yang dilakukan pada tahap (a).

$$\begin{aligned}d_L(x_i, x_j) &= \|x_i - m_1\| + \|m_1 - m_2\| + \|m_2 - x_j\| \\ &= 0,03000 + 0,007358 + 0,093900 \\ &= 0,131258^\circ \\ &= 0,131258^\circ \times \frac{10000}{90} \text{ km}\end{aligned}$$

$$d_L(x_i, x_j) = 1458,4222 \text{ km} = 145842,22 \text{ dam}$$

Sehingga jarak terdekat suatu titik x_i ke lokasi lampu lalu lintas terdekat di titik x_j adalah 145842,22 dam. Adapun struktur data awal variabel kovariat sebelum dilakukan transformasi untuk koordinat lampu lalu lintas ditampilkan pada Tabel 3.3.

Tabel 3.3 Struktur data koordinat lokasi lampu lalu lintas di Kabupaten Nganjuk

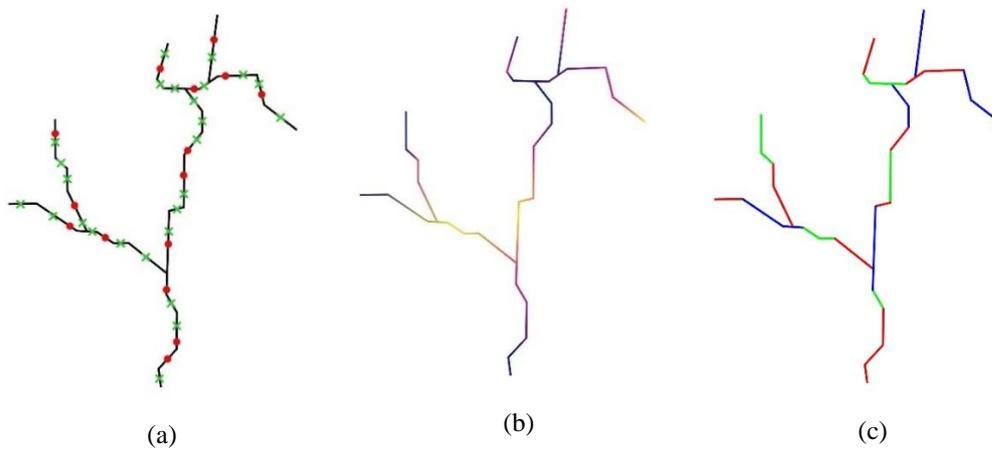
Titik Lampu Lalu Lintas	Longitude	Latitude
tl_1	$tl_{long,1}$	$tl_{lat,1}$
tl_2	$tl_{long,2}$	$tl_{lat,2}$
tl_3	$tl_{long,3}$	$tl_{lat,3}$
\vdots	\vdots	\vdots
tl_{100}	$tl_{long,100}$	$tl_{lat,100}$

Kovariat lainnya dalam penelitian ini tidak perlu dilakukan transformasi karena data yang tersedia sudah ada pada setiap segmen jalan sebanyak 11.736 segmen dengan struktur data ditampilkan pada Tabel 3.4.

Tabel 3.4 Struktur data kovariat

Segmen jalan	$z_2(u_j)$	$z_3(u_j)$...	$z_6(u_j)$
Segmen $_1$	$z_2(u_1)$	$z_3(u_1)$...	$z_6(u_1)$
Segmen $_2$	$z_2(u_2)$	$z_3(u_2)$...	$z_6(u_2)$
Segmen $_3$	$z_2(u_3)$	$z_3(u_3)$	$z_6(u_3)$
\vdots	\vdots	\vdots	\vdots	\vdots
Segmen $_{11.736}$	$z_2(u_{11.736})$	$z_3(u_{11.736})$...	$z_6(u_{11.736})$

Selanjutnya, ketiga struktur data tersebut dijadikan dalam satu tabel untuk pemodelan *Poisson point process* dengan pendekatan *quadrature scheme* yaitu dengan membagi setiap segmen garis dengan panjang ℓ menjadi subinterval dengan panjang yang sama untuk mendapatkan bobot *quadrature* serta dengan menambahkan *dummy point* sebanyak d lalu dilakukan evaluasi nilai kovariat untuk setiap *data point* dan *dummy point* yang terpilih sehingga indeks pada struktur data pemodelan menjadi $j = 1, 2, \dots, n + d$ dengan n jumlah *data point* dan d jumlah *dummy point*. Ilustrasi struktur data pada segmen jalan digambarkan pada Gambar 3.11. Struktur data pemodelan intensitas kecelakaan lalu lintas memiliki indeks j , $z_1(u_j), z_2(u_j), \dots, z_6(u_j)$ yang mana masing-masing adalah nilai kovariat pada titik data/*dummy* ke- j , I_j adalah variabel yang menyatakan indikator apakah titik ke- j adalah titik data atau titik *dummy*, w_j adalah *weight* dari titik data/ *dummy* ke- j . Secara rinci, metode untuk membangkitkan titik *dummy* dijelaskan pada Sub bab 2.5.1.



Gambar 3.11 Ilustrasi nilai variabel pada setiap segmen di *linear network* (a) data point (•) dan *dummy point* (x) (b) kovariat jarak ke lokasi lampu lalu lintas terdekat bernilai kontinu pada setiap segmen (c) kovariat $z_2(\mathbf{u})$ hingga $z_6(\mathbf{u})$ bernilai konstan pada setiap segmen.

Jumlah data kecelakaan di Kabupaten Nganjuk tahun 2018-2020 yang digunakan sebanyak 1.944 titik dengan jumlah *dummy point* yang dibangkitkan yaitu sebanyak 48.890, sehingga jumlah *quadrature point* yang dihasilkan sebanyak $n + d = 1.944 + 23.473 = 25.417$ titik. Struktur data dalam pemodelan fungsi intensitas kasus kecelakaan lalu lintas di Kabupaten Nganjuk disajikan pada Tabel 3.5.

Tabel 3.5 Struktur data pemodelan intensitas kecelakaan lalu lintas

j	$z_1(u_j)$	$z_2(u_j)$...	$z_6(u_j)$	I_j	w_j	$y_j = I_j/w_j$
1	$z_1(u_1)$	$z_2(u_1)$...	$z_6(u_1)$	1	w_1	$1/w_1$
2	$z_1(u_2)$	$z_2(u_2)$...	$z_6(u_2)$	1	w_2	$1/w_2$
⋮	⋮	⋮	...	⋮	⋮	⋮	⋮
$n=1.944$	$z_1(u_n)$	$z_2(u_n)$...	$z_6(u_n)$	1	w_n	$1/w_n$
$n+1$	$z_1(u_{n+1})$	$z_2(u_{n+1})$...	$z_6(u_{n+1})$	0	w_{n+1}	0
⋮	⋮	⋮	...	⋮	⋮	⋮	⋮
$n+d-1$	$z_1(u_{n+d-1})$	$z_2(u_{n+d-1})$...	$z_6(u_{n+d-1})$	0	w_{n+d-1}	0
$n+d=25.417$	$z_1(u_{n+d})$	$z_2(u_{n+d})$...	$z_6(u_{n+d})$	0	w_{n+d}	0

3.5 Langkah Penelitian

Langkah-langkah yang dilakukan dalam melakukan penelitian ini adalah untuk menjawab tujuan penelitian dan diuraikan sebagai berikut.

A. Tahapan estimasi parameter model intensitas *point process on a linear network* (Tujuan 1).

1. Menentukan model intensitas *point process* seperti pada Persamaan (2.4).

$$\lambda(\mathbf{u}) = \lambda(\boldsymbol{\theta}; \mathbf{u}) = \exp(\boldsymbol{\theta}^T \mathbf{z}(\mathbf{u}))$$

dimana $\mathbf{z}(\mathbf{u}) = (1, z_1(\mathbf{u}), z_2(\mathbf{u}), \dots, z_p(\mathbf{u}))^T$ dan $\boldsymbol{\theta} = (\theta_0, \theta_1, \theta_2, \dots, \theta_p)^T$

2. Mendefinisikan fungsi *log-likelihood point process* seperti pada persamaan (2.18).

$$\ell(\boldsymbol{\theta}) = \sum_{\mathbf{u} \in \mathcal{X}} \log \lambda(\boldsymbol{\theta}; \mathbf{u}) - \int_L \lambda(\boldsymbol{\theta}; \mathbf{u}) du$$

3. Melakukan diskritisasi integral pada fungsi *log likelihood* dengan skema Berman-Turner pada persamaan (2.19).

$$\int_L \lambda(\boldsymbol{\theta}; \mathbf{u}) du \approx \sum_{j=1}^{n+d} \lambda(\boldsymbol{\theta}; \mathbf{u}_j) w_j$$

4. Mendapatkan fungsi *likelihood* yang ekuivalen dengan *weighted log-likelihood* dari model regresi poisson terboboti pada persamaan (2.23).

$$\ell(\boldsymbol{\theta}) \approx \sum_{j=1}^{n+d} [y_j \log \lambda(\boldsymbol{\theta}; \mathbf{u}_j) - \lambda(\boldsymbol{\theta}; \mathbf{u}_j)] w_j$$

5. Melakukan pendekatan fungsi *likelihood* berbasis regresi Poisson pada Persamaan (2.23) menggunakan ekspansi Taylor orde 2 pada Persamaan (2.24).

$$\ell(\boldsymbol{\theta}) \approx \ell(\tilde{\boldsymbol{\theta}}) + (\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}})^T \ell'(\tilde{\boldsymbol{\theta}}) + \frac{(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}})^T}{2!} \ell''(\tilde{\boldsymbol{\theta}}) (\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}})$$

dimana $\ell(\boldsymbol{\theta})$: fungsi *likelihood* setelah didekati dengan dengan ekspansi Taylor orde 2.

6. Menggabungkan fungsi *penalized likelihood* Lasso pada Persamaan (2.26).

$$\ell^{Pen}(\boldsymbol{\theta}) = \ell(\boldsymbol{\theta}) - \gamma \sum_{k=1}^p |\theta_k|$$

dimana $\ell(\boldsymbol{\theta})$ merupakan pendekatan *likelihood Poisson point process on a linear network* pada Persamaan (2.23) dengan vektor koefisien regresi $\boldsymbol{\theta} = (\theta_0, \theta_1, \theta_2, \dots, \theta_p)$ dan $k = 1, 2, \dots, p$.

7. Melakukan *tuning parameter*

- a. Mengidentifikasi barisan menurun dari *tuning parameter* $\gamma \in |\gamma_{min}, \gamma_{max}|$ dengan γ_{max} sampai dengan $\gamma_{min} = 10^{-3} \gamma_{max}$. Setiap nilai γ , *outer loop* diulang sebanyak elemen barisan γ untuk menghitung $\ell(\boldsymbol{\theta})$ pada $\hat{\boldsymbol{\theta}}$;
- b. Menerapkan *Coordinate Descent Algorithm* untuk menyelesaikan *penalized problem* dengan mengoptimasi secara parsial (tiap k) pada Persamaan (2.27)

$$\min_{\tilde{\theta}_k \in \mathbb{R}} \ell_{lasso}(\tilde{\theta}_1, \dots, \tilde{\theta}_{k-1}, \theta_k, \tilde{\theta}_{k+1}, \dots, \tilde{\theta}_p)$$

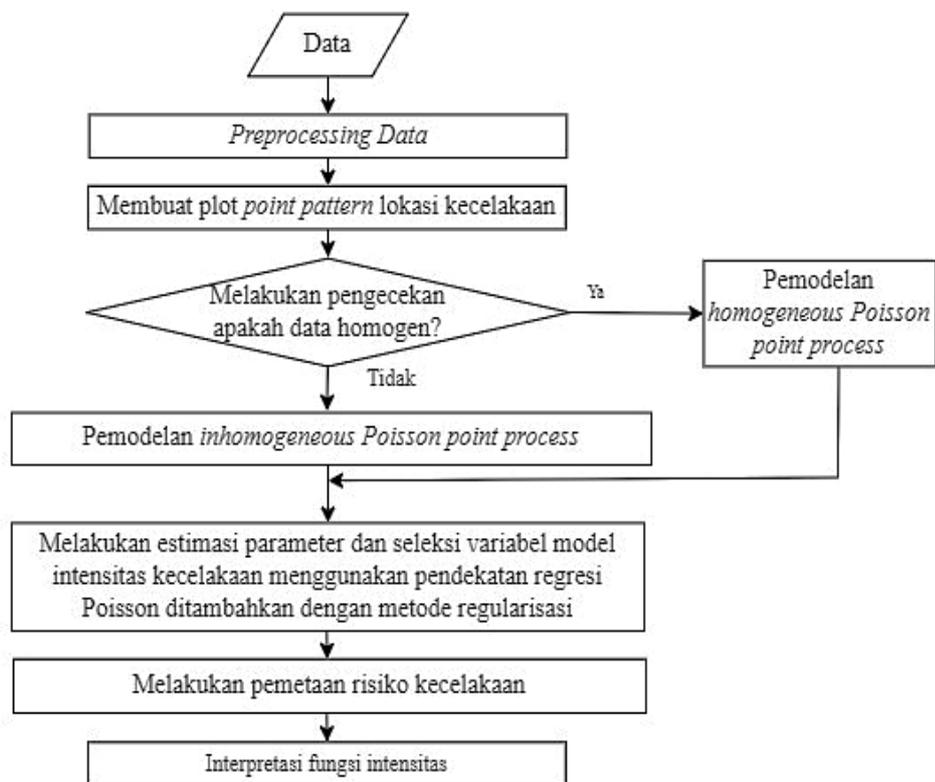
- c. Memilih γ yang optimal dengan menggunakan nilai BIC (γ) terkecil berdasarkan Persamaan (2.29).
- d. Mengevaluasi nilai estimasi parameter sampai tercapai konvergensi (*inner loop*) jika kriteria $|\tilde{\boldsymbol{\theta}}^{(q+1)} - \tilde{\boldsymbol{\theta}}^{(q)}| \leq \varepsilon$ tercapai dengan ε adalah suatu nilai positif yang kecil, dimana pada fungsi *glmnet software R* nilai $\varepsilon = 10^{-7}$.

B. Tahapan pemodelan intensitas kecelakaan lalu lintas di Kabupaten Nganjuk dengan menggunakan metode regularisasi (Tujuan 2).

1. Mendefinisikan *observation window* jaringan jalan di Kabupaten Nganjuk dengan fungsi ``as.linnet.SpatialLines`` pada *software R*.
2. Melakukan *preprocessing* data *point pattern* kecelakaan lalu lintas serta kovariat yang digunakan. *Preprocessing* yang dilakukan antara lain sebagai berikut:
 - a. Melakukan perubahan skala *observation window* dari satuan derajat ke dekameter (10 meter) dengan fungsi ``rescale`` pada *software R*.
 - b. Melakukan pengecekan masing-masing koordinat lokasi kecelakaan lalu lintas sesuai dengan *observation window*.
 - c. Memberikan kode kovariat pada file *shp* jaringan jalan Kabupaten Nganjuk sesuai hasil informasi di *google maps*.
3. Membuat plot *point pattern* untuk data titik lokasi kecelakaan lalu lintas pada *observation window* berupa *linear network*.

4. Melakukan eksplorasi data kecelakaan lalu lintas dengan uji homogenitas menggunakan *CDF test* dengan statistik uji pada Persamaan (2.7), *correlation pixel images* pada sub bab 2.3.3, dan uji independensi pada sub bab 2.3.4.
5. Melakukan estimasi parameter dan seleksi variabel model intensitas kecelakaan lalu lintas di Kabupaten Nganjuk menggunakan pendekatan regresi Poisson dengan ditambahkan metode regularisasi.
6. Memvisualisasi pemetaan risiko terjadinya kecelakaan lalu lintas di Kabupaten Nganjuk berdasarkan model terbaik.
7. Menarik kesimpulan dan saran.

Tahapan analisis data pada tujuan kedua ditunjukkan pada Gambar 3.12.



Gambar 3.12 Diagram analisis data.

(halaman ini sengaja dikosongkan)

BAB 4

HASIL DAN PEMBAHASAN

Analisis dan pembahasan dalam penelitian ini menjawab permasalahan dan mencapai tujuan penelitian. Estimasi parameter model *Inhomogeneous Poisson point process* dilakukan dengan pendekatan regresi Poisson. Selanjutnya pada fungsi *log-likelihood* ditambahkan fungsi penalti Lasso dengan proses *tuning parameter* untuk mendapatkan nilai γ optimum. Pemilihan nilai γ optimum menggunakan nilai *Bayesian Information Criteria* (BIC) terendah.

4.1 Estimasi Parameter Model *Inhomogeneous Poisson Point Process*

4.1.1 Regresi Poisson

Penentuan model *Inhomogeneous Poisson point process* dari Persamaan (2.18) menggunakan metode *maximum likelihood*. Bentuk likelihood pada Persamaan (2.18) bersifat kompleks sehingga didekati dengan pendekatan Berman-Turner pada Persamaan (2.23) sebagai berikut.

$$\ell(\boldsymbol{\theta}) \approx \sum_{j=1}^{n+d} [y_j \log \lambda(\boldsymbol{\theta}; \mathbf{u}_j) - \lambda(\boldsymbol{\theta}; \mathbf{u}_j)] w_j$$

Persamaan (2.23) ekuivalen dengan regresi Poisson terboboti dengan variabel respon y_j dan bobot w_j . Persamaan (2.23) dikonstruksi kedalam bentuk *Iteratively Reweighted Least Square* (IRLS) menggunakan ekspansi Taylor orde dua. Jika estimasi parameter saat ini adalah $\tilde{\boldsymbol{\theta}}$, dimana nilai $\tilde{\boldsymbol{\theta}}$ merupakan nilai parameter di titik baru yang mendekati nilai $\boldsymbol{\theta}$, maka pendekatan *quadrature scheme* untuk *log-likelihood* diberikan sebagai berikut:

$$\ell_{IPP}(\boldsymbol{\theta}) \approx \ell(\tilde{\boldsymbol{\theta}}) + (\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}})^T \ell'(\tilde{\boldsymbol{\theta}}) + \frac{(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}})^T}{2!} \ell''(\tilde{\boldsymbol{\theta}}) (\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}) \quad (4.1)$$

dengan

$$\begin{aligned} \ell'(\tilde{\boldsymbol{\theta}}) &= \frac{d}{d\tilde{\boldsymbol{\theta}}^T} \left(\sum_{j=1}^{n+d} [y_j \log \lambda(\tilde{\boldsymbol{\theta}}; \mathbf{u}_j) - \lambda(\tilde{\boldsymbol{\theta}}; \mathbf{u}_j)] w_j \right) \\ &= \sum_{j=1}^{n+d} w_j \frac{\partial}{\partial \tilde{\boldsymbol{\theta}}^T} [y_j \log \lambda(\tilde{\boldsymbol{\theta}}; \mathbf{u}_j) - \lambda(\tilde{\boldsymbol{\theta}}; \mathbf{u}_j)] \end{aligned} \quad (4.2)$$

Kemudian dilakukan imputasi Persamaan (2.4) $\lambda(\tilde{\boldsymbol{\theta}}; \mathbf{u}_j) = \exp(\tilde{\boldsymbol{\theta}}^T \mathbf{z}(\mathbf{u}_j))$ dimana $\mathbf{z}(\mathbf{u}) = (1, z_1(\mathbf{u}), z_2(\mathbf{u}), \dots, z_p(\mathbf{u}))^T$ dan $\boldsymbol{\theta} = (\theta_0, \theta_1, \theta_2, \dots, \theta_p)^T$ pada persamaan (4.2) sebagai berikut

$$\begin{aligned} \ell'(\tilde{\boldsymbol{\theta}}) &= \sum_{j=1}^{n+d} w_j \frac{d}{d\tilde{\boldsymbol{\theta}}^T} [y_j (\tilde{\boldsymbol{\theta}}^T \mathbf{z}(\mathbf{u}_j)) - \exp(\tilde{\boldsymbol{\theta}}^T \mathbf{z}(\mathbf{u}_j))] \\ &= \sum_{j=1}^{n+d} w_j [y_j \mathbf{z}(\mathbf{u}_j) - \mathbf{z}(\mathbf{u}_j) \exp(\tilde{\boldsymbol{\theta}}^T \mathbf{z}(\mathbf{u}_j))] \\ &= \sum_{j=1}^{n+d} w_j \mathbf{z}(\mathbf{u}_j) [y_j - \exp(\tilde{\boldsymbol{\theta}}^T \mathbf{z}(\mathbf{u}_j))] \end{aligned} \quad (4.3)$$

$$\begin{aligned} \ell''(\tilde{\boldsymbol{\theta}}) &= \frac{d\ell'(\tilde{\boldsymbol{\theta}})}{d\tilde{\boldsymbol{\theta}}^T} \left(\sum_{j=1}^{n+d} w_j \mathbf{z}(\mathbf{u}_j) [y_j - \exp(\tilde{\boldsymbol{\theta}}^T \mathbf{z}(\mathbf{u}_j))] \right) \\ &= \sum_{j=1}^{n+d} w_j \mathbf{z}(\mathbf{u}_j) \frac{d}{d\tilde{\boldsymbol{\theta}}^T} [y_j - \exp(\tilde{\boldsymbol{\theta}}^T \mathbf{z}(\mathbf{u}_j))] \\ &= \sum_{j=1}^{n+d} w_j \mathbf{z}(\mathbf{u}_j) \mathbf{z}(\mathbf{u}_j)^T [-\exp(\tilde{\boldsymbol{\theta}}^T \mathbf{z}(\mathbf{u}_j))] \end{aligned} \quad (4.4)$$

Dari hasil Persamaan (4.3) sebagai gradien dan Persamaan (4.4) sebagai *Hessian*, selanjutnya disubstitusikan ke Persamaan (4.1)

$$\begin{aligned} \ell(\boldsymbol{\theta}) &\approx \sum_{j=1}^{n+d} [y_j \log \lambda(\tilde{\boldsymbol{\theta}}; \mathbf{u}_j) - \lambda(\tilde{\boldsymbol{\theta}}; \mathbf{u}_j)] w_j \\ &\quad + \sum_{j=1}^{n+d} w_j \mathbf{z}(\mathbf{u}_j) (\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}})^T [y_j - \exp(\tilde{\boldsymbol{\theta}}^T \mathbf{z}(\mathbf{u}_j))] \\ &\quad - \frac{1}{2} \sum_{j=1}^{n+d} w_j (\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}})^T \mathbf{z}(\mathbf{u}_j) \mathbf{z}(\mathbf{u}_j)^T (\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}) [-\exp(\tilde{\boldsymbol{\theta}}^T \mathbf{z}(\mathbf{u}_j))] \\ \ell(\boldsymbol{\theta}) &\approx -\frac{1}{2} \sum_{j=1}^{n+d} w_j (\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}})^T \mathbf{z}(\mathbf{u}_j) \mathbf{z}(\mathbf{u}_j)^T (\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}) [-\exp(\tilde{\boldsymbol{\theta}}^T \mathbf{z}(\mathbf{u}_j))] \\ &\quad + \sum_{j=1}^{n+d} w_j \mathbf{z}(\mathbf{u}_j) (\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}})^T [y_j - \lambda(\tilde{\boldsymbol{\theta}}; \mathbf{u}_j)] + c(\tilde{\boldsymbol{\theta}}; w_j) \end{aligned}$$

dimana $\lambda(\tilde{\boldsymbol{\theta}}; \mathbf{u}_j) = \exp(\tilde{\boldsymbol{\theta}}^T \mathbf{z}(\mathbf{u}_j))$ dan $c(\tilde{\boldsymbol{\theta}}; w_j) = \sum_{j=1}^{n+d} [y_j \log \lambda(\tilde{\boldsymbol{\theta}}; \mathbf{u}_j) - \lambda(\tilde{\boldsymbol{\theta}}; \mathbf{u}_j)] w_j$ merupakan suatu konstanta yang tidak mengandung $\tilde{\boldsymbol{\theta}}$

$$\begin{aligned}
\ell(\boldsymbol{\theta}) &\approx \sum_{j=1}^{n+d} \frac{w_j}{2} \left[(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}})^T \mathbf{z}(\mathbf{u}_j) \mathbf{z}(\mathbf{u}_j)^T (\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}) \lambda(\tilde{\boldsymbol{\theta}}; \mathbf{u}_j) \right. \\
&\quad \left. - 2(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}})^T \mathbf{z}(\mathbf{u}_j) (y_j - \lambda(\tilde{\boldsymbol{\theta}}; \mathbf{u}_j)) \right] \\
&\approx \sum_{j=1}^{n+d} -\frac{w_j}{2\lambda(\tilde{\boldsymbol{\theta}}; \mathbf{u}_j)} \left[(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}})^T \mathbf{z}(\mathbf{u}_j) \lambda(\tilde{\boldsymbol{\theta}}; \mathbf{u}_j) - (y_j - \lambda(\tilde{\boldsymbol{\theta}}; \mathbf{u}_j)) \right]^2 \\
&= \sum_{j=1}^{n+d} -\frac{w_j \lambda(\tilde{\boldsymbol{\theta}}; \mathbf{u}_j)}{2} \left[\frac{(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}})^T \mathbf{z}(\mathbf{u}_j) \lambda(\tilde{\boldsymbol{\theta}}; \mathbf{u}_j) - (y_j - \lambda(\tilde{\boldsymbol{\theta}}; \mathbf{u}_j))}{\lambda(\tilde{\boldsymbol{\theta}}; \mathbf{u}_j)} \right]^2 \\
&= \sum_{j=1}^{n+d} -\frac{w_j \lambda(\tilde{\boldsymbol{\theta}}; \mathbf{u}_j)}{2} \left[(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}})^T \mathbf{z}(\mathbf{u}_j) - \frac{(y_j - \lambda(\tilde{\boldsymbol{\theta}}; \mathbf{u}_j))}{\lambda(\tilde{\boldsymbol{\theta}}; \mathbf{u}_j)} \right]^2 \\
&= \sum_{j=1}^{n+d} -\frac{w_j \lambda(\tilde{\boldsymbol{\theta}}; \mathbf{u}_j)}{2} \left[-\boldsymbol{\theta}^T \mathbf{z}(\mathbf{u}_j) + \tilde{\boldsymbol{\theta}}^T \mathbf{z}(\mathbf{u}_j) - \frac{(y_j - \lambda(\tilde{\boldsymbol{\theta}}; \mathbf{u}_j))}{\lambda(\tilde{\boldsymbol{\theta}}; \mathbf{u}_j)} \right]^2 \\
&= \sum_{j=1}^{n+d} \frac{1}{2} w_j \lambda(\tilde{\boldsymbol{\theta}}; \mathbf{u}_j) \left[\tilde{\boldsymbol{\theta}}^T \mathbf{z}(\mathbf{u}_j) + \frac{(y_j - \lambda(\tilde{\boldsymbol{\theta}}; \mathbf{u}_j))}{\lambda(\tilde{\boldsymbol{\theta}}; \mathbf{u}_j)} - \boldsymbol{\theta}^T \mathbf{z}(\mathbf{u}_j) \right]^2 \\
&= \sum_{j=1}^{n+d} \frac{1}{2} w_j \lambda(\tilde{\boldsymbol{\theta}}; \mathbf{u}_j) \left[\log \lambda(\tilde{\boldsymbol{\theta}}; \mathbf{u}_j) + \frac{(y_j - \lambda(\tilde{\boldsymbol{\theta}}; \mathbf{u}_j))}{\lambda(\tilde{\boldsymbol{\theta}}; \mathbf{u}_j)} - \boldsymbol{\theta}^T \mathbf{z}(\mathbf{u}_j) \right]^2 \tag{4.5}
\end{aligned}$$

Sehingga didapatkan:

$$\ell(\boldsymbol{\theta}) \approx \frac{1}{2} \sum_{j=1}^{n+d} v_j [y_j^* - \boldsymbol{\theta}^T \mathbf{z}(\mathbf{u}_j)]^2 \tag{4.6}$$

dengan y_j^* adalah *working response* dan v_j adalah bobot.

$$y_j^* = \log \lambda(\tilde{\boldsymbol{\theta}}; \mathbf{u}_j) + \frac{(y_j - \lambda(\tilde{\boldsymbol{\theta}}; \mathbf{u}_j))}{\lambda(\tilde{\boldsymbol{\theta}}; \mathbf{u}_j)} \text{ dan } v_j = w_j \lambda(\tilde{\boldsymbol{\theta}}; \mathbf{u}_j)$$

Untuk mendapatkan solusi nilai $\boldsymbol{\theta}$ yang memaksimalkan fungsi *log-likelihood* pada Persamaan (4.6), terlebih dahulu dihitung turunan parsial dari $\ell(\boldsymbol{\theta})$ terhadap nilai $\boldsymbol{\theta}$.

$$\begin{aligned}
\frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} &= \frac{\partial}{\partial \boldsymbol{\theta}} \left(\frac{1}{2} \sum_{j=1}^{n+d} v_j [y_j^* - \boldsymbol{\theta}^T \mathbf{z}(\mathbf{u}_j)]^2 \right) \\
&= \frac{1}{2} \sum_{j=1}^{n+d} v_j 2 [y_j^* - \boldsymbol{\theta}^T \mathbf{z}(\mathbf{u}_j)] (-\mathbf{z}^T(\mathbf{u}_j)) \\
&= - \sum_{j=1}^{n+d} v_j [y_j^* - \boldsymbol{\theta}^T \mathbf{z}(\mathbf{u}_j)] \mathbf{z}^T(\mathbf{u}_j) \tag{4.7}
\end{aligned}$$

Kemudian menyamakan turunan Persamaan (4.7) dengan nol dan menyelesaikan persamaan yang dihasilkan sebagai berikut

$$\begin{aligned}
0 &= \sum_{j=1}^{n+d} v_j [y_j^* - \boldsymbol{\theta}^T \mathbf{z}(\mathbf{u}_j)] (-\mathbf{z}^T(\mathbf{u}_j)) \\
0 &= \sum_{j=1}^{n+d} v_j \mathbf{z}^T(\mathbf{u}_j) [y_j^* - \boldsymbol{\theta}^T \mathbf{z}(\mathbf{u}_j)] \\
\sum_{j=1}^{n+d} v_j y_j^* \mathbf{z}^T(\mathbf{u}_j) &= \sum_{j=1}^{n+d} v_j \mathbf{z}^T(\mathbf{u}_j) \mathbf{z}(\mathbf{u}_j) \boldsymbol{\theta} \\
\left(\sum_{j=1}^{n+d} v_j \mathbf{z}(\mathbf{u}_j)^T \mathbf{z}(\mathbf{u}_j) \right) \boldsymbol{\theta} &= \sum_{j=1}^{n+d} v_j y_j^* \mathbf{z}^T(\mathbf{u}_j) \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \\
\hat{\boldsymbol{\theta}} &= \frac{\sum_{j=1}^{n+d} v_j y_j^* \mathbf{z}^T(\mathbf{u}_j)}{\sum_{j=1}^{n+d} v_j \mathbf{z}(\mathbf{u}_j)^T \mathbf{z}(\mathbf{u}_j)} \tag{4.8}
\end{aligned}$$

Persamaan (4.8) jika direpresentasikan dalam bentuk vektor dan matriks, maka dapat tertulis pada Persamaan (4.9)

$$\hat{\boldsymbol{\theta}} = (\mathbf{Z}^T \mathbf{V} \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{V} \mathbf{y}^* \tag{4.9}$$

dimana \mathbf{Z} adalah matriks yang berukuran $(n+d) \times (p+1)$ dimana setiap elemen berisi $z_l(\mathbf{u}_j)$ dengan $l = 1, 2, \dots, p$ dan $j = 1, 2, \dots, n+d$. \mathbf{V} adalah matriks diagonal dengan elemen v_j , dan \mathbf{y}^* adalah vektor dengan elemen y_j^* . Algoritma proses *Iteratively Rewieghted Least Square* yang digunakan untuk mengestimasi nilai $\hat{\boldsymbol{\theta}}$ tanpa regularisasi ditunjukkan pada Tabel 4.1.

Tabel 4.1 Algoritma *Iteratively Reweighted Least Square*

Langkah 1: Menentukan *initial value* $\tilde{\boldsymbol{\theta}}^{(0)}$, dimulai dengan $\tilde{\boldsymbol{\theta}}^{(0)}=0$.

Langkah 2: Menghitung *working response* y_j^* dan bobot v_j untuk setiap $q = 0,1,2, \dots$
initial value $\tilde{\boldsymbol{\theta}}^{(0)}$ adalah $\tilde{\boldsymbol{\theta}}^{(q)}$ dengan $q=0$.

Langkah 3: Mendapatkan estimator $\tilde{\boldsymbol{\theta}}^{(q+1)}$ dengan menggunakan Persamaan (4.6) dimulai dengan $q=0$, mensubstitusikan hasil *working response* dan pembobot $\tilde{\boldsymbol{\theta}}^{(q=0)}$ dari langkah ke-2 ke dalam pembobot dan *working response* dengan solusi

$$\tilde{\boldsymbol{\theta}}^{(q+1)T} = \left(\mathbf{z}(\mathbf{u}_j)^T v_j(\tilde{\boldsymbol{\theta}}^{(q)}) \mathbf{z}(\mathbf{u}_j) \right)^{-1} \mathbf{z}(\mathbf{u}_j)^T v_j(\tilde{\boldsymbol{\theta}}^{(q)}) y_j^*(\tilde{\boldsymbol{\theta}}^{(q)})^*$$

Langkah 4: Memeriksa konvergensi dari $\tilde{\boldsymbol{\theta}}^{(q+1)}$ dengan kriteria $\|\tilde{\boldsymbol{\theta}}^{(q+1)} - \tilde{\boldsymbol{\theta}}^{(q)}\| \leq \varepsilon$ dengan ε adalah suatu nilai positif yang sangat kecil. Jika terpenuhi maka proses iterasi selesai, apabila belum terpenuhi maka mengulangi langkah ke-3 hingga konvergen.

4.1.2 Penambahan Metode Lasso pada Pendekatan Regresi Poisson

Langkah selanjutnya adalah *Coordinate Descent Algorithm* (CDA) dengan mendefinisikan *penalized likelihood function* untuk regularisasi Lasso. Misalkan $\ell(\boldsymbol{\theta})$ merupakan pendekatan *likelihood Poisson point process on a linear network* pada Persamaan (2.23) dengan vektor koefisien regresi $\boldsymbol{\theta} = (\theta_0, \theta_1, \theta_2, \dots, \theta_p)$. Fungsi *Penalized likelihood* Lasso dituliskan pada Persamaan (4.10).

$$\begin{aligned} \ell^{Pen}(\boldsymbol{\theta}) &= \ell(\boldsymbol{\theta}) - \gamma \sum_{k=1}^p |\theta_k| \\ \ell^{Pen}(\boldsymbol{\theta}) &\approx \frac{1}{2} \sum_{j=1}^{n+d} v_j \left[y_j^* - \sum_{\substack{l=1 \\ l \neq k}}^p z_l(\mathbf{u}_j) \tilde{\theta}_l - z_k(\mathbf{u}_j) \tilde{\theta}_k \right]^2 - \gamma \sum_{k=1}^p |\theta_k| \\ &= -\frac{1}{2} \sum_{j=1}^{n+d} v_j \left[y_j^* - \sum_{\substack{l=1 \\ l \neq k}}^p z_l(\mathbf{u}_j) \tilde{\theta}_l - z_k(\mathbf{u}_j) \tilde{\theta}_k \right]^2 + \gamma \sum_{k=1}^p |\theta_k| \end{aligned} \quad (4.10)$$

dimana y_j^* adalah *working response* dan v_j adalah bobot. Misalkan terdapat estimator $\tilde{\theta}_l$, $l = 1,2, \dots, p$ dimana $l \neq k$ maka metode *coordinate descent* terdiri dari pengoptimalan secara parsial (2.27) terhadap θ_k untuk setiap $k = 1,2, \dots, p$ secara iteratif

$$\arg \min_{\theta_k \in l} \ell^{pen}(\tilde{\theta}_1, \dots, \tilde{\theta}_{k-1}, \theta_k, \tilde{\theta}_{k+1}, \dots, \tilde{\theta}_p)$$

$$\frac{\partial \ell^{pen}(\tilde{\theta}_1, \dots, \tilde{\theta}_{k-1}, \theta_k, \tilde{\theta}_{k+1}, \dots, \tilde{\theta}_p)}{\partial \theta_k} = 0$$

$$\frac{d \left(-\frac{1}{2} \sum_{j=1}^{n+d} v_j \left[y_j^* - \sum_{\substack{l=1 \\ l \neq k}}^p z_l(\mathbf{u}_j) \tilde{\theta}_l - z_k(\mathbf{u}_j) \tilde{\theta}_k \right]^2 + \gamma \sum_{k=1}^p |\theta_k| \right)}{\partial \theta_k} = 0$$

$$-\frac{1}{2} \left(-2 \sum_{j=1}^{n+d} v_j z_k(\mathbf{u}_j) \left[y_j^* - \sum_{\substack{l=1 \\ l \neq k}}^p z_l(\mathbf{u}_j) \tilde{\theta}_l - z_k(\mathbf{u}_j) \tilde{\theta}_k \right] \right) + \gamma \text{sign}|\theta|$$

$$0 = \sum_{j=1}^{n+d} v_j z_k(\mathbf{u}_j) \left[y_j^* - \sum_{\substack{l=1 \\ l \neq k}}^p z_l(\mathbf{u}_j) \tilde{\theta}_l - z_k(\mathbf{u}_j) \tilde{\theta}_k \right] + \gamma \text{sign}|\tilde{\theta}_k|$$

$$0 = \sum_{j=1}^{n+d} v_j z_k(\mathbf{u}_j) [y_j^* - y_{j_k}^* - z_k(\mathbf{u}_j) \tilde{\theta}_k] + \gamma \text{sign}|\tilde{\theta}_k|$$

$$0 = \sum_{j=1}^{n+d} (v_j z_k(\mathbf{u}_j) (y_j^* - y_{j_k}^*) - v_j z_k(\mathbf{u}_j) z_k(\mathbf{u}_j) \tilde{\theta}_k) + \gamma \text{sign}|\tilde{\theta}_k|$$

Sehingga evaluasi nilai parameter menggunakan Persamaan (4.11)

$$\tilde{\theta}_k = \frac{S \left(\sum_{j=1}^{n+d} v_j z_k(\mathbf{u}_j) [y_j^* - y_{j_k}^*], \gamma \right)}{\sum_{j=1}^{n+d} v_j (z_k(\mathbf{u}_j))^2} \quad (4.11)$$

dengan $S(Z, \gamma)$ adalah operator *soft-threshold* dimana *Soft-threshold* $S(a, b)$ didefinisikan sebagai:

$$S(a, b) = \text{sign}(a)(|a| - b)_+ = \begin{cases} a - b & \text{if } a > 0 \text{ and } b < |a| \\ a + b & \text{if } a < 0 \text{ and } b < |a| \\ 0 & \text{if } b \geq |a|. \end{cases}$$

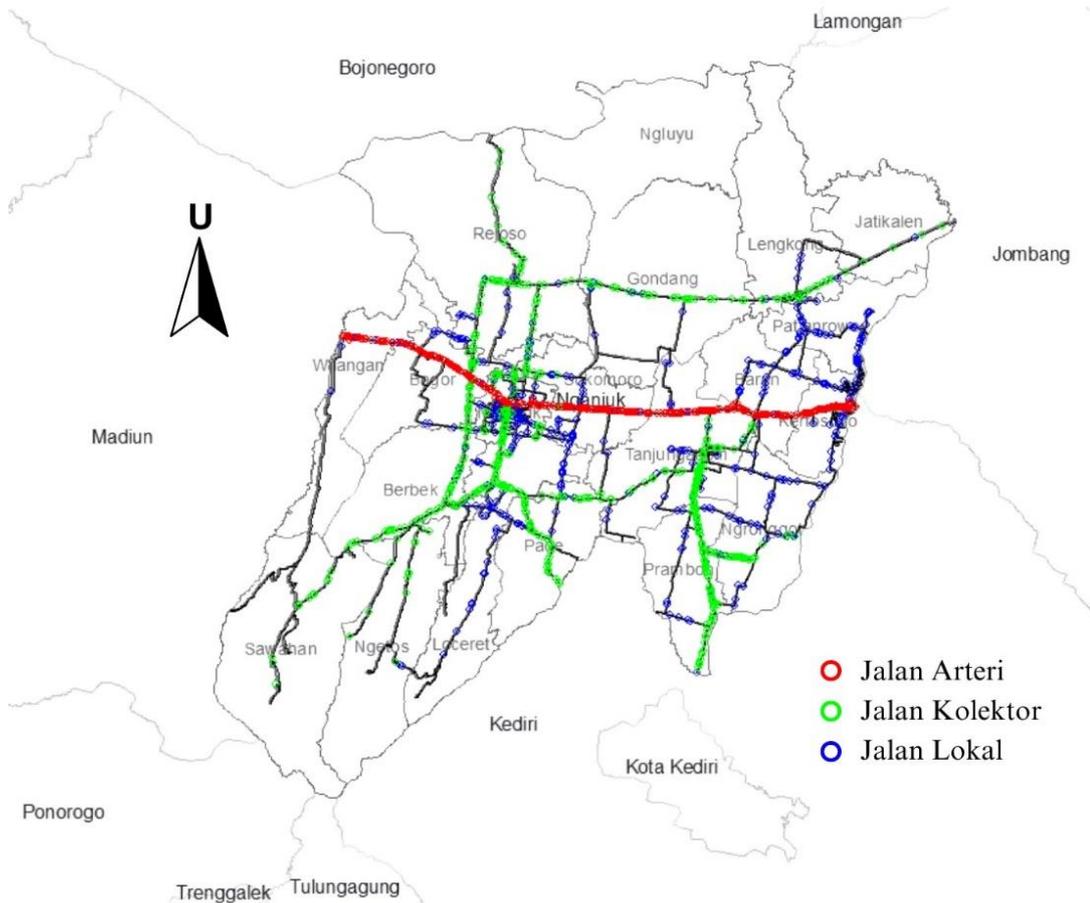
Jika diperoleh kondisi $\hat{\theta}_k \neq 0$ pada kondisi *outer loop* maka dilanjutkan pada proses estimasi tahap dua pada *inner loop*. Proses tersebut dilakukan untuk menyeleksi variabel pada setiap anggota variabel prediktor. Jika $\hat{\theta}_k \neq 0$ maka kondisi *subgradien* pada $\theta_k = 0$ jika

$$\sum_{j=1}^{n+d} v_j z_k(u_j) [y_j^* - y_{j_k}^*] \leq \gamma \quad (4.12)$$

Proses regularisasi dengan menggunakan metode Lasso selanjutnya dikonstruksi menggunakan *software* R Studio.

4.2 Penerapan *Regularization Poisson Point Process on a Linear Network*

Pemodelan *Poisson point process* pada *linear network* menggunakan data koordinat lokasi kecelakaan lalu lintas tahun 2018-2020 pada jaringan jalan di Kabupaten Nganjuk yang terbagi menjadi jalan arteri, jalan kolektor, dan jalan lokal. Persebaran kecelakaan lalu lintas di Kabupaten Nganjuk tahun 2018-2020 ditunjukkan pada Gambar 4.1.

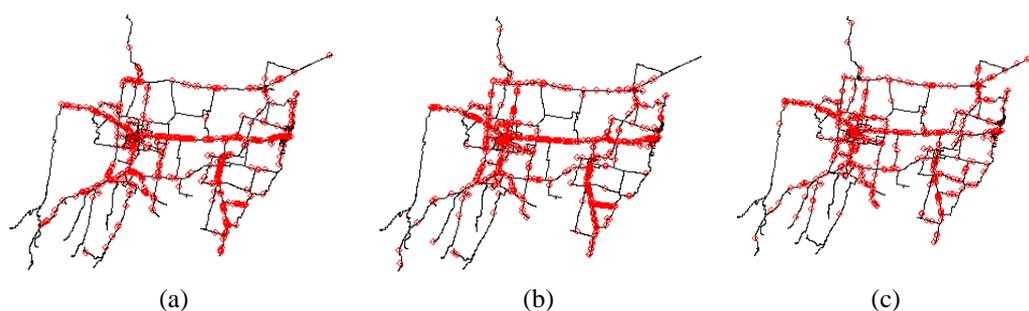


Gambar 4.1 Peta persebaran kecelakaan lalu lintas di Kabupaten Nganjuk.

4.2.1 Peta Sebaran Data Kecelakaan Lalu Lintas di Kabupaten Nganjuk

Pola persebaran kecelakaan lalu lintas di Kabupaten Nganjuk pada Gambar 4.1 terlihat tidak homogen. Sebagian besar kecelakaan terjadi di jalan

arteri, yaitu jalur utama yang menghubungkan Kabupaten Nganjuk dengan Kabupaten Madiun serta perbatasan Kabupaten Kediri dan Kabupaten Jombang. Selain berfungsi sebagai penghubung antar kabupaten, jalan arteri ini juga melewati beberapa kecamatan seperti Kecamatan Wilangan, Kecamatan Bagor, Nganjuk, Sukomoro, dan Kertosono. Tingginya angka kecelakaan di jalan arteri ditandai dengan banyaknya titik kecelakaan yang padat dan saling tumpang tindih, dibandingkan dengan titik kecelakaan di jalan kolektor dan lokal. Selain itu, jumlah kecelakaan yang tinggi juga terjadi di jalan lokal, terutama di Kecamatan Tanjunganom dan Kecamatan Prambon, serta menyebar ke jalan-jalan sekitarnya.

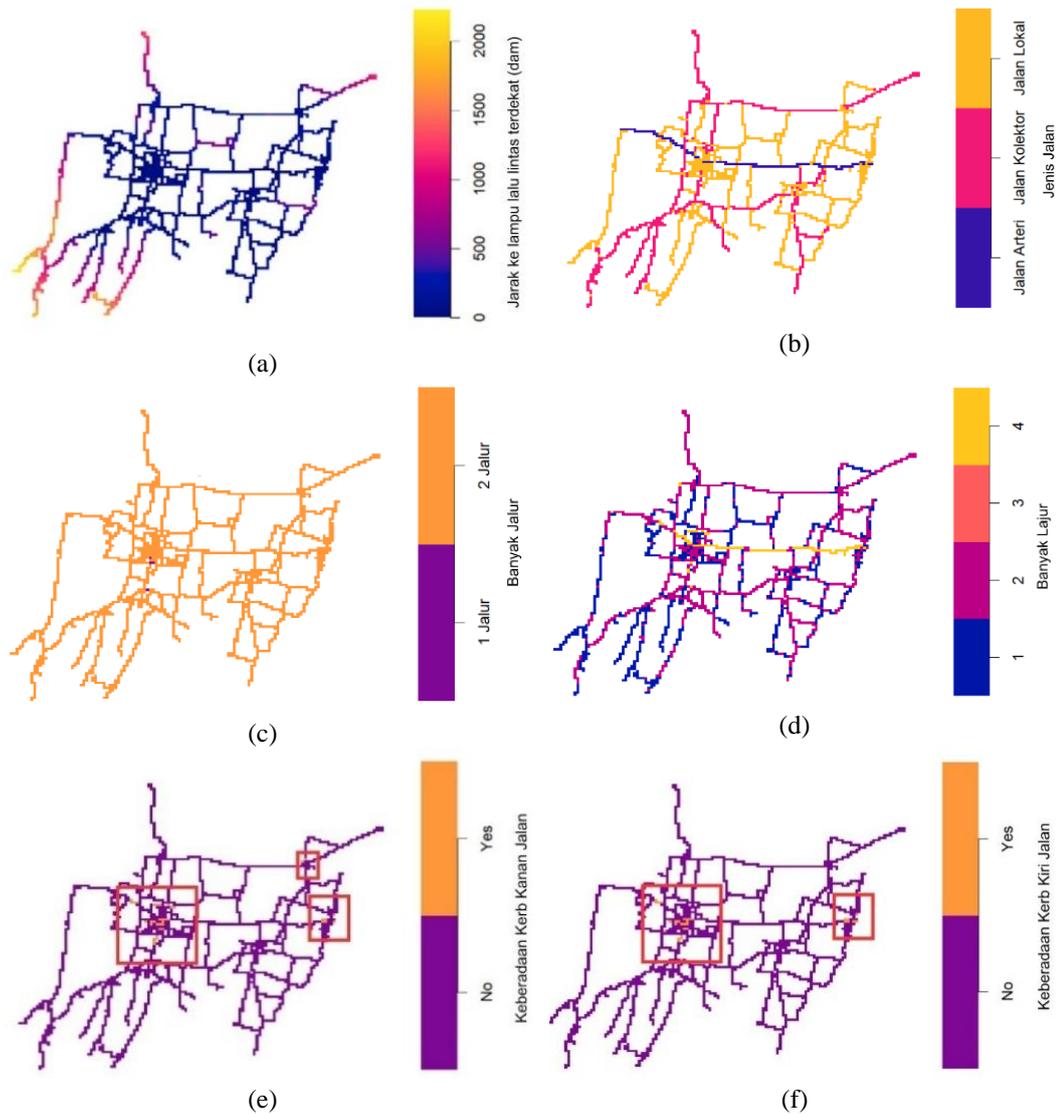


Gambar 4.2 Peta persebaran kecelakaan lalu lintas di Kabupaten Nganjuk setiap tahun (a) tahun 2018 (b) tahun 2019 (c) tahun 2020.

Sebaran kecelakaan lalu lintas per tahun di Kabupaten Nganjuk selama 2018-2020 pada Gambar 4.2 menunjukkan pola yang serupa untuk tahun 2018, 2019, dan 2020, dengan jumlah kecelakaan masing-masing sebanyak 671, 678, dan 596 kecelakaan. Pada tahun 2020, jumlah kecelakaan lalu lintas lebih rendah dibandingkan dua tahun sebelumnya, dengan sebaran titik kecelakaan yang cenderung lebih jarang dan tidak padat, terutama di jalan arteri. Hal ini dikarenakan adanya kebijakan pemerintah dalam pembatasan kegiatan di luar rumah untuk mengurangi kasus COVID-19 pada periode tahun 2020. Seperti pola keseluruhan pada Gambar 4.1, titik kecelakaan tahun 2018, 2019, dan 2020 paling padat terjadi di jalan arteri. Pada tahun 2018 dan 2019, titik kecelakaan juga cukup banyak tersebar di jalan lokal yang melalui Kecamatan Tanjunganom.

4.2.2 Karakteristik Kovariat Spasial

Pada penelitian ini terdapat tujuh kovariat berupa faktor lingkungan dan geometri jalan meliputi jarak ke lampu lalu lintas terdekat, jenis jalan, jumlah jalur, jumlah lajur, keberadaan kerb di kiri dan kanan jalan.



Gambar 4.3 Peta kovariat spasial (a) jarak ke lampu lalu lintas terdekat (dalam satuan dam) (b) jenis jalan (c) banyak jalur (d) banyak lajur (e) keberadaan kerb kanan jalan (f) keberadaan kerb kiri jalan.

Ruas jalan berwarna biru pada Gambar 4.3 (a) menunjukkan kedekatan dengan lampu lalu lintas. Ruas jalan ini terletak di pusat kota atau di daerah dengan kepadatan persimpangan yang tinggi dan memiliki lampu lalu lintas. Sebaliknya, ruas jalan dengan warna yang semakin terang (ke arah kuning) menunjukkan jarak

yang lebih jauh dari lampu lalu lintas. Ruas jalan ini berada di daerah pinggiran atau di jalan dengan volume jalan yang lebih rendah. Jika dibandingkan dengan kepadatan kecelakaan lalu lintas pada Gambar 4.1, kecelakaan cenderung lebih banyak pada ruas jalan dengan jarak yang dekat dengan lokasi lampu lalu lintas. Hal ini mungkin disebabkan oleh tingginya volume lalu lintas di area tersebut, yang meningkatkan peluang terjadinya kecelakaan, serta adanya pengaturan lalu lintas yang kompleks.

Gambar 4.3 (b) menunjukkan jenis jalan dengan tiga kategori jalan yang diwarnai berbeda untuk menandai tingkat kepentingannya: jalan lokal (kuning) yang paling banyak dan tersebar luas, mengindikasikan infrastruktur yang mendukung aksesibilitas tinggi pada tingkat lokal dan komunitas; jalan kolektor (orange) yang lebih sedikit dan tampak menghubungkan jalan-jalan lokal dengan jalan-jalan utama, berfungsi sebagai pengumpul lalu lintas dari jalan-jalan kecil menuju rute utama; serta jalan arteri (biru), yang paling sedikit tetapi terbesar dan paling dominan, berfungsi sebagai arteri utama untuk transportasi cepat dan efisien antar wilayah dan kabupaten, memfasilitasi mobilitas regional dan ekonomi lokal. Kovariat jenis jalan jika dibandingkan dengan distribusi persebaran kecelakaan lalu lintas pada Gambar 4.1 kecelakaan lalu lintas paling banyak terjadi pada jalan arteri, diikuti oleh jalan kolektor, sedangkan kecelakaan lalu lintas pada jalan lokal lebih sedikit jika dibandingkan dengan jalan arteri dan jalan kolektor. Hal ini menunjukkan bahwa jalan arteri berpotensi memiliki risiko kecelakaan yang lebih tinggi dibandingkan dengan jalan kolektor dan jalan lokal.

Pada Gambar 4.3 (c) sebagian besar ruas jalan memiliki dua jalur yang digambarkan dengan warna orange di hampir seluruh ruas jalan Kabupaten Nganjuk. Hal ini mengindikasikan hampir seluruh ruas jalan di Kabupaten Nganjuk menerapkan sistem dua arah. Namun, ketika membandingkan dengan Gambar 4.1 yang menampilkan distribusi lokasi kecelakaan lalu lintas dengan intensitas yang bervariasi sepanjang ruas jalan di Kabupaten Nganjuk, jumlah jalur tidak berdampak pada distribusi kepadatan kecelakaan. Meskipun jalan dua jalur mendominasi dalam peta, kecelakaan tidak terpusat hanya pada jalan tersebut. Sehingga, bisa dihipotesiskan bahwa variabel jumlah jalur tidak secara signifikan mempengaruhi risiko kecelakaan lalu lintas di daerah tersebut. Selanjutnya akan

dilakukan penambahan fungsi penalti Lasso pada model untuk mengkonfirmasi apakah jumlah jalur signifikan berpengaruh atau tidak terhadap risiko kecelakaan lalu lintas di Kabupaten Nganjuk.

Jumlah lajur di jalan arteri dan kolektor bervariasi dari satu hingga tiga, tergantung pada lebar jalan seperti terlihat pada Gambar 4.3 (d). Sementara itu, jalan arteri memiliki rata-rata empat lajur karena pada umumnya jalan arteri memiliki lebar yang lebih besar dibandingkan dengan jalan lainnya dan sering dilalui oleh volume kendaraan yang lebih tinggi

Sebagian besar jalan yang tidak memiliki kerb di sisi kanan atau kiri sebagai pembatas jalan baik di jalan arteri, kolektor, dan jalan lokal sesuai pada Gambar 4.3 (e) dan Gambar 4.3 (f). Hal ini mengindikasikan kovariat keberadaan kerb baik di sisi kanan atau kiri jalan diduga tidak berpengaruh terhadap risiko kecelakaan lalu lintas di Kabupaten Nganjuk seperti halnya kovariat jumlah jalur karena persebaran kecelakaan lalu lintas cenderung bervariasi tersebar di sepanjang ruas jalan dan tidak secara khusus terkonsentrasi di jalan yang memiliki kerb. Disisi lain, ruas jalan yang terdapat kerb di sisi kanan atau kirinya yang ditandai dengan kotak berwarna merah pada pada Gambar 4.3 (e) dan Gambar 4.3 (f) merupakan ruas jalan yang berada di area dengan kepadatan persimpangan yang tinggi, dimana lokasi kecelakaan lalu lintas yang terjadi pada area ruas jalan ini juga terlihat sangat padat. Oleh karena itu, keberadaan kerb baik di sisi kanan atau kiri jalan juga diduga dapat mempengaruhi risiko kecelakaan lalu lintas di Kabupaten Nganjuk, walaupun pengaruhnya sangat kecil.

4.2.3 Uji Homogenitas

Uji homogenitas dilakukan untuk mengetahui intensitas kecelakaan lalu lintas yang diteliti apakah pola distribusinya mengikuti pola yang homogen (tidak bergantung pada kovariat) atau inhomogen (bergantung pada kovariat). Pada penelitian ini, uji homogenitas dilakukan menggunakan *Cumulative Distribution Function (CDF) test*. Konsepnya adalah dengan membandingkan distribusi nilai kovariat yang diamati pada data *point* dengan nilai kovariat di semua lokasi spasial dalam *observation window* . Hasil *CDF test* ditampilkan pada Tabel 4.2.

Tabel 4.2 Hasil uji *Cumulative Distribution Function* data kecelakaan di Kabupaten Nganjuk

Deskripsi	D	P-Value	Kesimpulan
Jarak ke lokasi lampu lalu lintas terdekat	0,22682	$<2,2 \times 10^{-16}$	Inhomogen
Jenis jalan	0,32461	$<2,2 \times 10^{-16}$	Inhomogen
Jumlah lajur	0,27739	$<2,2 \times 10^{-16}$	Inhomogen
Jumlah jalur	0,01932	0,5360	Homogen
Keberadaan kerb di sisi kiri jalan	0,04959	0,0002	Inhomogen
Keberadaan kerb di sisi kanan jalan	0,02571	0,1560	Homogen

Hasil pengujian pada Tabel 4.2 diperoleh keputusan Tolak H_0 pada kovariat jarak ke lokasi lampu lalu lintas terdekat, jenis jalan, jumlah lajur, dan keberadaan kerb di sisi kiri jalan dengan menggunakan taraf kesalahan $\alpha=0,05$, karena nilai *p-value* hasil CDF test kurang dari taraf kesalahan $\alpha=0,05$. Hal ini mengindikasikan bahwa jumlah titik lokasi kecelakaan lalu lintas di Kabupaten Nganjuk tidak homogen (inhomogen). Artinya, terdapat kemungkinan faktor-faktor yang dapat mempengaruhi perbedaan intensitas kecelakaan lalu lintas di berbagai daerah di Kabupaten Nganjuk. Sedangkan, kovariat jumlah jalur dan keberadaan kerb di sisi kiri jalan dengan menggunakan taraf kesalahan $\alpha=0,05$, karena nilai *p-value* hasil CDF test kurang dari taraf kesalahan $\alpha=0,05$. Hal ini mengindikasikan bahwa jumlah titik lokasi kecelakaan lalu lintas di Kabupaten Nganjuk tidak homogen (inhomogen).

4.2.4 Korelasi *Pixel Image* dan Uji Independensi

Korelasi *pixel image* merupakan nilai yang menunjukkan kekuatan hubungan antara jarak kovariat dalam bentuk *pixel image* pada kovariat numerik. Metode yang digunakan adalah *Pearson correlation*. Hasil perhitungan korelasi *pixel image* antara kovariat numerik jarak ke lokasi lampu lalu lintas terdekat, jumlah jalur, dan jumlah lajur dituliskan pada Tabel 4.3.

Tabel 4.3 Hasil uji independensi antar kovariat numerik data kecelakaan di Kabupaten Nganjuk

No.	Kovariat 1	Kovariat 2	Nilai Koefisien Korelasi	P-Value	Kesimpulan
1.	Jarak ke lampu lalu lintas terdekat	Jumlah jalur	0,036	0,506	Independen
2.	Jarak ke lampu lalu lintas terdekat	Jumlah lajur	-0,266	0,516	Independen
3.	Jumlah jalur	Jumlah lajur	-0,020	0,392	Independen

Hasil pengujian pada Tabel 4.3 menunjukkan bahwa terdapat korelasi yang sangat lemah antar ketiga kovariat, yaitu jarak lampu lalu lintas terdekat, jumlah jalur dan jumlah lajur. Hal ini juga dibuktikan oleh nilai *p-value* yang lebih besar dari taraf signifikansi (α) = 0,05 sehingga diperoleh keputusan gagal tolak H_0 yang berarti tidak terdapat korelasi antara ketiga kovariat tersebut.

Uji independensi pada kovariat kategorik dilakukan menggunakan uji *Chi-square* atau *Fisher's exact* tergantung pada nilai ekpektasi frekuensi sel di setiap tabel kontingensi yang dijelaskan pada sub bab 2.3.4. Tabel 4.4 menunjukkan hasil uji independensi antar kovariat kategorik berdasarkan jenis pengujian yang digunakan.

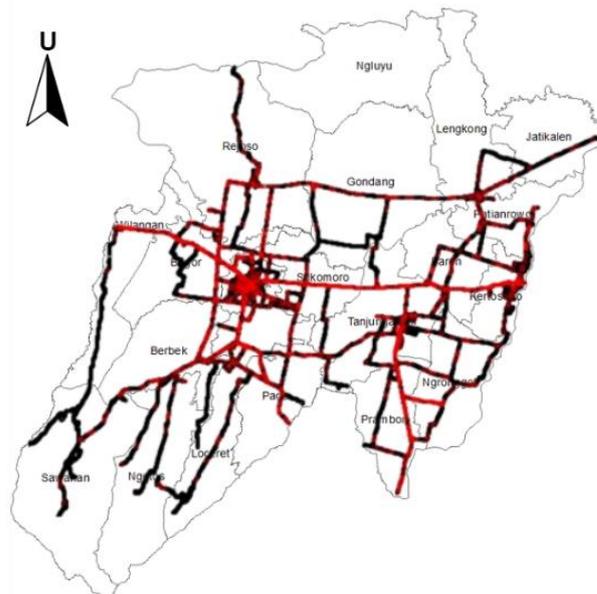
Tabel 4.4 Hasil uji independensi antar kovariat kategorik data kecelakaan di Kabupaten Nganjuk

No.	Kovariat 1	Kovariat 2	Jenis Pengujian	P-Value	Kesimpulan
1.	Kerb kanan	Kerb kiri	<i>Fisher's Exact</i>	$2,20 \times 10^{-16}$	Dependen
2.	Kerb kanan	Jenis jalan	<i>Chi-Square</i>	$1,46 \times 10^{-10}$	Dependen
3.	Kerb kiri	Jenis jalan	<i>Chi-Square</i>	$2,48 \times 10^{-16}$	Dependen

Hasil pengujian pada Tabel 4.4 menunjukkan bahwa kovariat jenis jalan, keberadaan kerb kanan dan keberadaan kerb kiri jalan memiliki dependensi antara satu kovariat dengan kovariat lainnya. Hal ini ditunjukkan oleh *p-value* < $\alpha=0,05$ pada hasil uji independensi antar kovariat kategorik. Adanya dependensi antar kovariat memang sangat mungkin terjadi, karena karakteristik geometrika jalan cenderung saling berhubungan atau memiliki keterkaitan. Adanya dependensi antar kovariat dapat menyebabkan terjadinya multikolinearitas, sehingga pada penelitian ini untuk mengatasi masalah tersebut dipertimbangkan regularisasi Lasso.

4.2.5 Pemodelan Risiko Kecelakaan Lalu Lintas di Kabupaten Nganjuk Menggunakan *Inhomogeneous Poisson Point Process Lasso*

Risiko kecelakaan lalu lintas di Kabupaten Nganjuk dimodelkan menggunakan *Inhomogeneous Poisson point process on a linear network* karena intensitas kecelakaan lalu lintas di Kabupaten Nganjuk tersebar secara inhomogen dan antar kejadian kecelakaan lalu lintas diasumsikan saling independen. Hasil pemodelan Skema Berman-Turner menjadi konstruksi awal dalam pendekatan reresi Poisson. Pada skema ini *dummy point* dibangkitkan dengan membagi *observation window* ke beberapa segmen tertentu. Kemudian pada setiap segmennya dibangkitkan *dummy point*. Konstruksi Berman-Turner pada *observation window* berupa *linear network* menggunakan fungsi ``linequad`` pada *packages spatstat.linnet* di R Studio.



Gambar 4.4 Ilustrasi *quadrature scheme* metode Berman-Turner (titik merah: *data point*, titik hitam: *dummy point*)

Berdasarkan Gambar 4.4, *data point* diilustrasikan sebagai titik berwarna merah dan *dummy point* sebagai titik berwarna hitam. Seluruh titik direpresentasikan sebagai pasangan *longitude* dan *latitude*. Selanjutnya setiap titik tersebut disesuaikan dengan nilai variabel prediktor yang digunakan. Diperoleh konstruksi data dari variabel prediktor dan variabel respon yang dapat dianalisis menggunakan *software* R Studio. Variabel respon pada skema Berman-Turner

dikonstruksi berdasarkan $y_j = I_j/w_j$, dimana I_j adalah fungsi indikator yang bernilai 1 jika I_j merupakan *data point* dan bernilai 0 jika *dummy point*. Nilai dari w_j diperoleh dengan pendekatan $\sum_{j=1}^{n+d} w_j = |L|$. Nilai dari variabel respon bergantung dari satuan atau panjang *observasi window* yang digunakan. Total panjang segmen observasi ($|L|$) yang digunakan sebesar 260,2781 Km. Total *dummy point* yang diperoleh sebesar 23.473 *point* yang terdapat pada setiap panjang segmen 59,49 dam dan data point sebanyak 1.944 *point*, sehingga terdapat 25.417 baris data yang dianalisis. Selanjutnya dilakukan konstruksi dalam bentuk *Iteratively Reweighted Least Square* (IRLS) menggunakan ekspansi Taylor orde dua untuk mendapatkan *working response* y_j^* dan pembobot v_j sesuai persamaan (4.5) menggunakan fungsi *ppm* dengan penambahan hasil *quadrature scheme* dari fungsi *linequad*.

Tabel 4.5 Hasil estimasi parameter model *Inhomogeneous Poisson point process on a linear network* tanpa regularisasi

Variabel	$\hat{\theta}$	$\widehat{se}(\hat{\theta})$	Z_{val}	Signifikansi
Intersep	-4,008971	0,609259	-6,580079	***
Jarak ke lokasi lampu lalu lintas terdekat	-0,001975	0,000137	-14,456065	***
Jenis jalan (kolektor)	0,980289	0,055972	17,513982	***
Jenis jalan (arteri)	1,565581	0,087115	17,971444	***
Jumlah lajur	0,275298	0,033346	8,255697	***
Jumlah jalur	-0,105069	0,303200	-0,346533	
Keberadaan kerb di sisi kiri jalan (terdapat kerb kiri)	0,522135	0,195142	2,675668	**
Keberadaan kerb di sisi kanan jalan (terdapat kerb kanan)	-0,618626	0,203024	-3,047105	**

Catatan: Tanda * merepresentasikan level signifikansi: * (0.1); ** (0.05); *** (0.01), sedangkan SE untuk standard error $\hat{\theta}$

Pemodelan awal dilakukan dengan menggunakan seluruh kovariat tanpa menggunakan regularisasi Lasso dan hasil estimasi parameter beserta signifikansi parameter model ditunjukkan pada Tabel 4.5. Model tanpa menggunakan regularisasi menunjukkan koefisien tidak konsisten, seperti variabel keberadaan kerb di sisi kanan jalan yang bertanda negatif, yang mungkin terjadi karena

multikolinearitas. Penambahan fungsi penalti Lasso digunakan untuk mengatasi multikolinearitas dengan menyeleksi variabel yang tidak signifikan, menyederhanakan model dan membuatnya lebih mudah diinterpretasikan. Selain itu, koefisien yang lebih kecil dan stabil dari model Lasso memberikan interpretasi yang lebih jelas tentang faktor-faktor utama yang berkontribusi terhadap risiko kecelakaan.

Selanjutnya dilakukan proses seleksi variabel dengan menambahkan fungsi penalti Lasso untuk mendapatkan informasi faktor lingkungan dan geometri jalan apa saja yang memiliki kecenderungan berpengaruh secara signifikan terhadap kecelakaan lalu lintas yang diamati. Algoritma yang digunakan dalam proses seleksi variabel untuk studi kasus data kecelakaan lalu lintas di Kabupaten Nganjuk dengan regularisasi Lasso ditunjukkan pada Tabel 4.6.

Tabel 4.6 Algoritma seleksi variabel dengan menggunakan regularisasi Lasso

Langkah 1:	Mengontruksi model <i>Inhomogeneous Poisson point process</i> dengan menambahkan metode Lasso seperti pada persamaan (4.10).
Langkah 2:	Proses regularisasi menggunakan Lasso <ol style="list-style-type: none"> 1. Langkah <i>outer loop</i> pada setiap variabel ke-k diperoleh $\hat{\theta}^k \neq 0$ jika memenuhi Persamaan (4.12) 2. Lakukan <i>inner loop</i> jika pada langkah satu diperoleh $\hat{\theta}^k \neq 0$ Lakukan proses 1 dan 2 untuk $k = 1, 2, \dots, p$
Langkah 3:	Proses <i>tuning parameter</i> γ <ol style="list-style-type: none"> 1. Menentukan barisan menurun dari <i>tuning parameter</i> $\gamma \in \gamma_{min}, \gamma_{max}$ dengan γ_{max} sampai dengan $\gamma_{min} = 10^{-3} \gamma_{max}$ 2. Melihat nilai BIC γ setiap berdasarkan Persamaan (2.29), gunakan hasil regularisasi terbaik dengan nilai kriteria BIC terendah.

Nilai γ_{max} dalam regularisasi Lasso adalah nilai γ yang paling kecil di mana semua koefisien, kecuali intersep, adalah nol (penalti terbesar). Semakin besar nilai γ maka semakin banyak kovariat yang tidak masuk dalam model. Penentuan γ_{max} ini sangat penting karena menentukan batas atas untuk rentang nilai γ yang akan dieksplorasi dalam proses fitting model. Penelitian ini menggunakan nilai $\gamma_{max} =$

0,032119 yang secara *default* dihasilkan oleh fungsi `cv.glmnet` dari paket R. Fungsi ini secara otomatis menetapkan γ_{max} sebagai nilai di mana semua koefisien kecuali intersep adalah nol, yang memberikan penalti terbesar dan γ_{min} yang secara *default* diatur ke 10^{-3} dari γ_{max} pada kasus di mana jumlah observasi lebih besar dari jumlah prediktor. Pilihan ini memungkinkan untuk mengeksplorasi berbagai model dari yang paling sederhana hingga yang paling kompleks, tanpa perlu secara manual mengatur nilai γ , sehingga mempermudah proses pemodelan dan meningkatkan efisiensi waktu. Berdasarkan nilai BIC setiap *tuning parameter* terlampir pada Lampiran 5, diperoleh bahwa nilai *tuning parameter* ke-36 ($\gamma = 0,001238$) memiliki nilai BIC 134526 dengan hasil seleksi variabel pada Tabel 4.7.

Tabel 4.7 Hasil regularisasi Lasso model *Inhomogeneous Poisson point process*

Variabel	$\hat{\theta}$	Exp ($\hat{\theta}$)	1/Exp ($\hat{\theta}$)
Intersep	-4,205171		
Jarak ke lokasi lampu lalu lintas terdekat	-0,001493		1,00149
Jenis jalan (kolektor)	0,844578	2,32700	
Jenis jalan (arteri)	1,481854	4,40110	
Jumlah lajur	0,274809	1,31628	
Jumlah jalur	0,000000	1,00000	
Keberadaan kerb di sisi kiri jalan (terdapat kerb kiri)	0,000000	1,00000	
Keberadaan kerb di sisi kanan jalan (terdapat kerb kanan)	0,000000	1,00000	

Berdasarkan hasil proses regularisasi dengan nilai *tuning parameter* ke-42 dengan nilai $\gamma = 0,001238$, menunjukkan bahwa terdapat tiga variabel lingkungan yang terseleksi ($\tilde{\theta} = \mathbf{0}$). Variabel tersebut adalah jumlah lajur, keberadaan kerb di sisi kanan, dan keberadaan kerb di sisi kiri jalan. Temuan ini dapat dijelaskan oleh konsistensi data untuk ketiga variabel tersebut; secara spesifik, hampir semua data menunjukkan bahwa sebagian besar jalan memiliki dua lajur, dan pada variabel keberadaan kerb di sisi kanan dan kiri, sebagian besar jalan tercatat tidak memiliki kerb. Ini menandakan bahwa ketiga variabel tersebut memiliki varian yang sangat rendah dalam sampel yang diobservasi, yang mengakibatkan pengaruhnya terhadap

model menjadi tidak signifikan dan oleh karena itu tidak dipilih oleh metode regularisasi yang diterapkan.

Berdasarkan Tabel 4.7, dapat diinterpretasikan pengaruh tiap kovariat terhadap risiko kecelakaan lalu lintas di Kabupaten Nganjuk adalah sebagai berikut.

- a) Kovariat jarak lampu lalu lintas terdekat: hasil estimasi parameter dari jarak suatu lokasi ke lampu lalu lintas terdekat bernilai negatif, yang berarti bahwa model memprediksi bahwa risiko kecelakaan akan lebih besar pada lokasi yang lebih dekat dengan lampu lalu lintas. Jika jarak ke lampu lalu lintas bertambah 10 meter atau 1 dekameter, risiko kecelakaan lalu lintas akan berkurang sebesar 0,00149 kali. Namun, hal ini secara langsung tidak dapat ditarik kesimpulan bahwa keberadaan lampu lalu lintas menjadi penyebab lebih meningkatnya risiko terjadinya kecelakaan. Perlu diketahui bahwa lampu lalu lintas secara umum diletakkan di lokasi dengan volume lalu lintas dan aktivitas masyarakat yang tinggi, seperti di persimpangan jalan, daerah kawasan pabrik, wilayah sekitar sekolah, pasar, atau fasilitas umum lainnya, sehingga secara tersirat menunjukkan bahwa risiko kecelakaan cenderung lebih tinggi di ruas jalan yang berdekatan dengan lampu lalu lintas karena volume lalu lintas yang tinggi.
- b) Kovariat jenis jalan: hasil prediksi menunjukkan bahwa risiko kecelakaan lalu lintas di jalan arteri 4,40110 kali lebih berisiko dibandingkan di jalan lokal. Sedangkan risiko kecelakaan lalu lintas di jalan kolektor 2.32700 kali lebih berisiko dibandingkan di jalan lokal. Berdasarkan fungsi yang dimiliki, jalan arteri dan kolektor sering digunakan untuk perjalanan dengan jarak menengah hingga jauh dan biasanya memiliki kecepatan rata-rata yang lebih tinggi dibandingkan jalan lokal. Kondisi ini menyebabkan jalan arteri dan kolektor memiliki risiko kecelakaan yang lebih tinggi daripada jalan lokal. Selain itu, kedua jenis jalan ini juga umumnya mengalami volume lalu lintas yang lebih besar dan lebih padat, meningkatkan kemungkinan interaksi antar kendaraan yang bisa menambah risiko kecelakaan lebih lanjut dibandingkan dengan di jalan lokal.
- c) Kovariat jumlah lajur: penambahan jumlah lajur di jalan raya diprediksi juga dapat meningkatkan risiko kecelakaan lalu lintas sebesar 32%. Hal ini

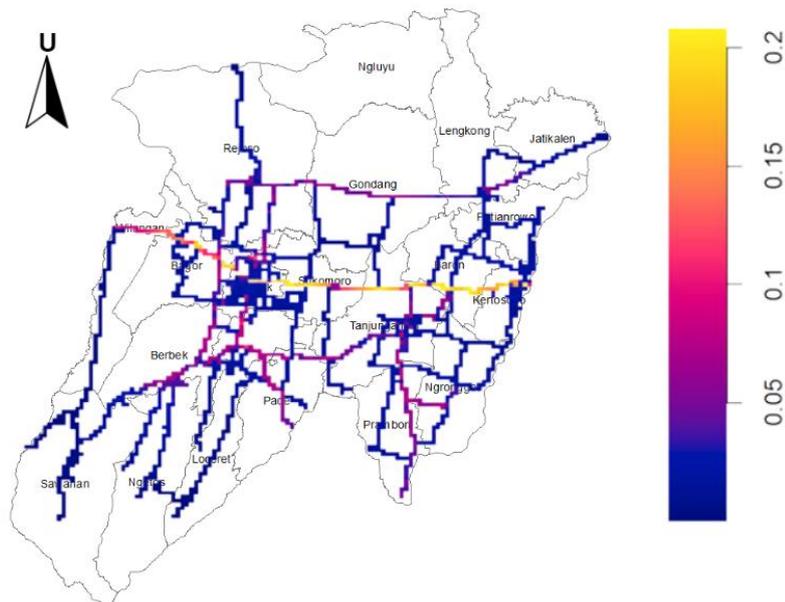
dikarenakan semakin banyak lajur yang diterapkan pada ruas jalan memungkinkan semakin lebar ruas jalannya, sehingga volume lalu lintas juga akan semakin besar dan potensi interaksi antar kendaraan juga meningkat yang dapat meningkatkan risiko kecelakaan lalu lintas. Namun, penting untuk dicatat bahwa mengurangi jumlah lajur bukan berarti mengurangi risiko kecelakaan. Sebaliknya, pengurangan lajur dapat menyebabkan masalah lain seperti kemacetan dan peningkatan waktu perjalanan. Oleh karena itu, penambahan atau pengurangan lajur harus dilakukan dengan pertimbangan yang matang terhadap berbagai faktor seperti volume lalu lintas, desain jalan, dan keselamatan pengguna jalan. Pernyataan ini berfungsi sebagai peringatan (*warning*) bahwa perubahan dengan menambah jumlah lajur dapat memiliki dampak signifikan terhadap keselamatan lalu lintas dan harus dikelola dengan baik untuk menghindari peningkatan risiko kecelakaan. Langkah-langkah seperti penegakan hukum lalu lintas, edukasi pengemudi, dan desain jalan yang baik juga diperlukan untuk memastikan keselamatan jalan raya secara keseluruhan.

4.2.6 Peta Prediksi Risiko Kecelakaan Lalu Lintas di Kabupaten Nganjuk

Hasil pemodelan fungsi intensitas risiko kecelakaan lalu lintas di Kabupaten Nganjuk dapat dilakukan prediksi sebagai *warning* dan mitigasi terjadinya kecelakaan lalu lintas di Kabupaten Nganjuk. Peta hasil prediksi risiko kecelakaan lalu lintas di Kabupaten Nganjuk berdasarkan model dengan kovariat yang signifikan disajikan pada Gambar 4.5. Nilai intensitas (tingkat kecelakaan per 10 meter) direpresentasikan oleh gradien warna di setiap ruas jalan. Semakin terang warna ruas jalan, maka semakin tinggi risiko kecelakaan lalu lintas di ruas jalan tersebut. Hasil prediksi menunjukkan, sebagian besar ruas jalan arteri memiliki risiko kecelakaan yang sangat tinggi, karena hampir sepanjang ruas jalan tersebut memiliki warna yang paling terang (kuning).

Risiko kecelakaan lalu lintas yang terjadi di sebagian besar jalan kolektor (direpresentasikan oleh warna merah muda) diprediksi tergolong cukup tinggi namun lebih rendah dibanding dengan risiko kecelakaan lalu lintas yang terjadi di jalan arteri, dan risiko kecelakaan lalu lintas paling rendah terjadi di jalan lokal

(warna biru). Hal ini sesuai dengan hasil estimasi parameter, dimana risiko kecelakaan lalu lintas yang terjadi di jalan lokal jauh lebih kecil dibandingkan risiko kecelakaan di jalan arteri dan kolektor. Peta prediksi juga menunjukkan bahwa risiko kecelakaan lalu lintas yang tergolong tinggi, terjadi pada ruas-ruas jalan yang berdekatan dengan lampu lalu lintas. Hal ini dibuktikan dengan tingginya risiko kecelakaan pada hasil prediksi yang terjadi di area pusat kabupaten dengan kepadatan lampu lalu lintas yang tinggi dan sebagian besar terjadi di jalan arteri dan kolektor.



Gambar 4.5 Peta prediksi risiko kecelakaan lalu lintas di Kabupaten Nganjuk.

BAB 5

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil analisis yang dilakukan dalam penelitian ini didapatkan kesimpulan sebagai berikut.

1. Proses estimasi parameter model *inhomogeneous Poisson point process* dengan pendekatan regresi Poisson dilakukan dengan tahapan:
 - a. Memaksimumkan bentuk *log-likelihood-nya* menggunakan *ekspansi Taylor* sehingga diperoleh *log-likelihood* dan secara natural diperoleh *working response* dan pembobot.
 - b. Mendapatkan hasil estimasi parameter menggunakan metode *Iteratively Reweighted Least Squared*.
 - c. Bentuk *log-likelihood* model *Inhomogeneous Poisson point process* dengan pendekatan regresi Poisson ditambahkan dengan fungsi regularisasi Lasso.
 - d. Seleksi variabel dilakukan dengan tahapan *inner loop* berdasarkan nilai *tuning parameter* (γ) yang berperan untuk mengoptimalkan hasil seleksi variabel menggunakan BIC_{γ} .
2. Hasil penambahan metode regularisasi didapatkan nilai γ optimal sebesar 0,001238 dengan nilai BIC 134526. Pada nilai γ optimal kovariat jumlah jalur, keberadaan kerb kanan, dan keberadaan kerb kiri terseleksi dari model risiko kecelakaan lalu lintas di Kabupaten Nganjuk.

5.2 Saran

Saran yang dapat diberikan dalam penelitian selanjutnya adalah sebagai berikut:

1. Data pada penelitian ini menggunakan data kecelakaan lalu lintas tahun 2018-2020 dan data kovariat tahun 2020. Penelitian selanjutnya diharapkan dapat menggunakan data terbaru dengan periode yang sama sehingga dapat memberikan informasi yang lebih aktual.

2. Pengujian dengan *K-function* untuk mendeteksi interaksi antar kecelakaan meskipun sudah dicoba untuk dijalankan pada penelitian ini tetapi membutuhkan komputasi yang tinggi. Oleh karena itu, kedepan perlu adanya penelitian untuk mereduksi waktu komputasi dalam pengujian ini.
3. Pemodelan dengan pendekatan berbasis logistik untuk membandingkan model yang diperoleh dengan pendekatan skema Berman-Turner berbasis Poisson. Pemodelan dengan pendekatan berbasis Poisson memerlukan waktu yang lama ketika datanya besar karena pada prinsipnya dilakukan *generate dummy point* dengan jumlah yang lebih besar dibandingkan dengan pendekatan logistik.
4. Penggunaan model *spatio-temporal* guna menganalisis interaksi antar kecelakaan secara lebih komprehensif. Pendekatan ini memungkinkan identifikasi pola spasial dan temporal dari kejadian kecelakaan, sehingga dapat memberikan wawasan yang lebih mendalam mengenai faktor-faktor penyebab dan lokasi rawan kecelakaan.

DAFTAR PUSTAKA

- Agresti, A., 2007. *An Introduction to Categorical Data Analysis*. 2nd. New York: Wiley.
- Ang, Q., 2010. *Statistical methodology for spatial point patterns on a linear network*. s.l.:MSc thesis, University of Western Australia.
- Ang, Q., Baddeley, A. & Nair, G., 2012. *Geometrically corrected second order analysis of events on a linear network, with applications to ecology and criminology*. *Scandinavian Journal of Statistics*, Volume 39, p. 591–617.
- Austroroads, 2002. *Road Safety Audit*, Sydney: Austroroads Publication.
- Baddeley, A. et al., 2021. *Analysing Point Patterns on Networks-a Review*. *Spatial Statistics*, Volume 42, p. 100435.
- Baddeley, A., Rubbak, E. & Turner, R., 2015. *Spatial Point Patterns Methodology and Applications with R*. New York: CRC Press.
- Chaudhuri, S., Juan, P. & Mateu, J., 2023. *Spatio-temporal modeling of traffic accidents incidence on urban road networks based on an explicit network triangulation*. *Journal of Applied Statistics*, 50(16), pp. 3229-3250 .
- Chintyana, A., Choiruddin, A. & Sutikno, 2023. *Cox Point Process with Ridge Regularization: A Better Approach for Statistical Modeling of Earthquake Occurrences*. Singapore, Springer Singapore.
- Choiruddin, A., Coeurjolly, J. F. & Letué, F., 2018. *Convex and non-convex regularization methods for spatial point processes intensity estimation*. *Electronic journal of statistics*, 12(1), pp. 1210-1255.
- Choiruddin, A., Coeurjolly, J. F. & Waagepetersen, R., 2021. *Information Criteria for Inhomogeneous Spatial Point Processes*. *Australian & New Zealand Journal of Statistics*, 63(1), pp. 119-143.
- Choiruddin, A., Coeurjolly, J.-F. & Latue, F., 2017. *Spatial point processes intensity estimation with a diverging number of covariates*. *arXiv:1712.09562v1*, pp. 1-30.
- Choiruddin, A., Hannanu, F., Mateu, J. & Fitriyanah, V., 2023. *COVID-19 transmission risk in Surabaya and Sidoarjo: an inhomogeneous marked Poisson point process approach*. *Stoch Environ Res Risk Assess*, 37(6), pp. 2271-2282.

- Choiruddin, A., Susanto, T. Y. & Metrikasari, R., 2021. *Two-Step Estimation for Modeling the Earthquake Occurrences in Sumatra by Neyman–Scott Cox Point Processes*. Singapore, Springer Singapore.
- Coeurjolly, J., Ba, I. & Choiruddin, A., 2023. *Regularization techniques for inhomogeneous (spatial) point processes intensity and conditional intensity estimation*. *Xiv preprint arXiv:2305.13470*.
- Daniel, J., Horrocks, J. & Umphrey, G., 2018. *Penalized composite likelihoods for inhomogeneous Gibbs point process models*. *Computational Statistics and Data Analysis*, pp. 104-116.
- Diggle, P. J., 2014. *Statistical Analysis of Spatial and Spatio Temporal Point Patterns*. 3rd penyunt. England: CRC Press.
- Draper, N. R. & Smith, H., 2014. *Applied Regression Analysis*. 3rd s.l.:Wiley-Interscience.
- Friedman, J., Hastie, T., Höfling, H. & Tibshirani, R., 2007. *Pathwise Coordinate Optimization*. *The Annals of Applied Statistics*, 1(2), pp. 302-332.
- Friedman, J., Hastie, T. & Tibshirani, R., 2010. *Regularization paths for generalized linear models via coordinate descent*. *Journal of statistical software*, 33(1), p. 1.
- Hasanah, A., Choiruddin, A. & Prastyo, D., 2022. *On the modeling of traffic accident risk in Nganjuk Regency by Poisson point process on a linear network*. s.l., AIP Publishing.
- Husain, A., 2022. *Sparse Group Lasso Untuk Estimasi Fungsi Intensitas Multivariate Poisson Point Process (Studi Kasus: Pemodelan Sebaran Pohon Di Pulau Barro Colorado)*. s.l.:Tesis Master, Institut Teknologi Sepuluh Nopember.
- Kartikasari, Y. M. & Choiruddin, A., 2022. *Analisis risiko gempa bumi di Sumatera dengan Cauchy cluster process*. *Inferensi*, 5(2), pp. 123-134.
- KPUPR, 2016. *Pengenalan Rekayasa Keselamatan Jalan*, s.l.: Kementerian Pekerjaan Umum dan Perumahan Rakyat.
- Kumar, S., Attri, S. D. & Singh., K. K., 2019. *Comparison of Lasso and stepwise regression technique for wheat yield prediction*. *Journal of Agrometeorology* 21, Volume 2, pp. 188-192.
- Liu, G. et al., 2018. *Risk factors for extremely serious road accidents: results from national road accident statistical annual report of China*. *PLoS one*, 13(8), p. e0201587.

- Llagostera, P., Comas, C. & López, N., 2022. *Modeling road traffic safety based on point patterns of wildlife-vehicle collisions. Science of the total environment*, Volume 846, p. 157237.
- Manan, M. M. A., Várhelyi, A., Çelik, A. K. & Hashim, H. H., 2018. *Road characteristics and environment factors associated with motorcycle fatal crashes in Malaysia. IATSS research*, 42(4), pp. 207-220.
- McSwiggan, G., 2019. *Spatial point process methods for linear networks with applications to road accident analysis*. s.l.:The University of Western Australi.
- McSwiggan, G., Baddeley, A. & Nair, G., 2016. Kernel density estimation on a linear network. *Scandinavian Journal of Statistics*, Volume 44, p. 324–345.
- Møller, J. & Waagepetersen, R. P., 2004. *Statistical Inference and Simulation for Spatial Point Processes*. Florida: Chapman & Hall.
- Okabe, A. & Kitamura, M., 1996. *A computational method for market area analysis on a network. Geographically Analysis*, 28(4), pp. 330-349.
- Okabe, A. & Sugihara, K., 2012. *Spatial analysis along networks*. s.l.:Wiley.
- Rakshit, S., McSwiggan, G., Nair, G. & Baddeley, A., 2021. *Variable selection using penalised likelihoods for point patterns on a linear network. Australian and New Zealand Journal of Statistics*, 63(3), pp. 417-454.
- Thurman, A., Fu, R., Guan, Y. & Zhu, J., 2015. *Regularized Estimating Equations for Model Selection of Clustered Spatial Point Processes. Statistical Sinica*, Volume 25, pp. 173-188.
- Tibshirani, R., 1996. *Regression shrinkage and selection via the lasso. Journal of The Royal Statistical Society Series B Methodological*, 58(1), pp. 267-288.
- Virania, T. A., Choiruddin, A. & Ratnasari, V., 2021. Analisis Resiko Penyebaran Kasus COVID-19 di Surabaya Raya menggunakan Thomas Cluster Process. *Inferensi*, 4(1), pp. 57-67.
- Yamada, I. & Thill, J. C., 2004. *Comparison of planar and network K-functions in traffic accident analysis. Journal of Transport Geography*, pp. 12(2), 149-158.
- Zhuang, J., Ogata, Y. & dan Vere-Jones, D., 2002. *Stochastic Declustering of Space-Time Earthquake Occurrences. Journal of The American Statistical Association*, 97(458), pp. 369-380.

(Halaman ini sengaja dikosongkan)

LAMPIRAN

Lampiran 1. Data Lokasi Kecelakaan Lalu Lintas di Kabupaten Nganjuk Tahun 2018-2020

No	Waktu	Jalan	Lat	Long	Tahun
1	Malam	Lokal	-7,5906	112,1009	2018
2	Malam	Kolektorprimer	-7,7311	112,0230	2018
3	Malam	Kolektorprimer	-7,6459	112,0143	2018
4	Malam	Arteriprimer	-7,5773	111,8683	2018
5	Siang	Arteriprimer	-7,5929	111,8926	2018
6	Siang	Kolektorprimer	-7,6149	111,8987	2018
7	Siang	Arteriprimer	-7,6068	112,0677	2018
8	Malam	Arteriprimer	-7,5837	111,8785	2018
9	Malam	Arteriprimer	-7,5905	111,8906	2018
10	Siang	Kolektorprimer	-7,6460	112,0142	2018
11	Siang	Kolektorprimer	-7,6987	112,0200	2018
12	Siang	Lokal	-7,6099	111,9313	2018
13	Siang	Kolektorsekunder	-7,6758	111,8332	2018
14	Siang	Arteriprimer	-7,6012	112,0957	2018
15	Siang	Arteriprimer	-7,5911	111,8899	2018
16	Siang	Lokal	-7,5815	112,0727	2018
17	Malam	Kolektorprimer	-7,6049	111,8995	2018
18	Malam	Kolektorsekunder	-7,6763	111,8307	2018
19	Siang	Kolektorsekunder	-7,5956	111,9046	2018
20	Siang	Lokal	-7,5970	111,9062	2018
21	Siang	Arteriprimer	-7,6012	112,104	2018
⋮	⋮	⋮	⋮	⋮	⋮
1943	Siang	Lokal	-7,6334	112,0947	2020
1944	Malam	Lokal	-7,6327	111,8836	2020
1945	Malam	Kolektorsekunder	-7,5356	112,0001	2020

Lampiran 2. Data Lokasi Lampu Lalu Lintas di Kabupaten Nganjuk Tahun 2020

No	Lat	Long
1	-7,5837645	111,8784862
2	-7,5993862	111,9044976
3	-7,6083844	111,9066522
4	-7,5823360	111,9047413
5	-7,6173932	111,8981693
6	-7,6160238	111,8863722
7	-7,6060808	111,9002504
8	-7,6147657	111,8768405
9	-7,6011257	112,1039751
10	-7,5928124	112,1073375
11	-7,6492869	111,8991812
12	-7,6381767	111,8945827
13	-7,5962945	111,8966980
14	-7,7197934	112,0242753
15	-7,6065116	112,0495876
16	-7,6290602	112,0185195
17	-7,5979320	111,8775057
18	-7,6013226	111,9313699
19	-7,5918319	111,8911822
20	-7,6019582	111,8887187
21	-7,6603859	111,8628324
22	-7,6002220	111,9144435
23	-7,5346884	112,0752017
24	-7,6108552	111,8994185
25	-7,6022058	111,9438398
26	-7,6818695	111,9263518
⋮	⋮	⋮
97	-7,6765766	112,0557946
98	-7,6285794	111,8962957
99	-7,5932827	111,8932056
100	-7,5908776	111,8913081

Lampiran 3. Data Kovariat Jenis Jalan, Jumlah Lajur, Jumlah Jalur, Keberadaan Kerb Di Sisi Kanan Dan Kiri Tahun 2020

No	n_lajur	n_jalur	Kerb_Kanan	Kerb_Kiri	Jenis_Jalan
1	1	2	0	0	Jalan Kolektor
2	1	2	0	0	Jalan Kolektor
3	1	2	0	0	Jalan Kolektor
4	1	2	0	0	Jalan Kolektor
5	1	2	0	0	Jalan Kolektor
6	1	2	0	0	Jalan Kolektor
7	1	2	0	0	Jalan Kolektor
8	1	2	0	0	Jalan Kolektor
9	1	2	0	0	Jalan Kolektor
10	1	2	0	0	Jalan Kolektor
11	1	2	0	0	Jalan Kolektor
12	1	2	0	0	Jalan Kolektor
13	1	2	0	0	Jalan Kolektor
14	1	2	0	0	Jalan Kolektor
15	1	2	0	0	Jalan Kolektor
16	1	2	0	0	Jalan Kolektor
17	1	2	0	0	Jalan Kolektor
18	1	2	0	0	Jalan Kolektor
19	1	2	0	0	Jalan Kolektor
20	2	2	0	0	Jalan Kolektor
21	2	2	0	0	Jalan Kolektor
22	2	2	0	0	Jalan Kolektor
23	1	2	0	0	Jalan Lokal
24	1	2	0	0	Jalan Lokal
25	1	2	0	0	Jalan Lokal
26	1	2	0	0	Jalan Kolektor
⋮	⋮	⋮	⋮	⋮	⋮
11732	4	2	0	0	Jalan Arteri
11733	4	2	0	0	Jalan Arteri
11734	4	2	0	0	Jalan Arteri
11735	4	2	0	0	Jalan Arteri
11736	4	2	0	0	Jalan Arteri

Lampiran 4. Data Diskritisasi Matriks GLM Pemodelan Kecelakaan Lalu Lintas di Kabupaten Nganjuk Berbasis Regresi Poisson

No	.mpl.W	.mpl.Y	Jenis_Jalan	n_Lajur	n_Jalur	Kerb_R	Kerb_L	f	.mpl.SUBSET
1	2,347	0,426	Jalan Lokal	2	2	No	No	61,390	TRUE
2	0,868	1,152	Jalan Kolektor	2	2	No	No	129,935	TRUE
3	1,827	0,547	Jalan Kolektor	2	2	No	No	70,963	TRUE
4	1,062	0,941	Jalan Arteri	2	2	Yes	Yes	0,449	TRUE
5	2,589	0,386	Jalan Arteri	4	2	No	No	0,261	TRUE
6	4,083	0,245	Jalan Kolektor	2	2	No	No	27,694	TRUE
7	1,125	0,889	Jalan Arteri	4	2	No	No	28,103	TRUE
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1942	0,914	1,095	Jalan Lokal	2	2	No	No	374,074	TRUE
1943	0,610	1,641	Jalan Lokal	2	2	No	No	2,961	TRUE
1944	2,405	0,416	Jalan Kolektor	2	2	No	No	74,998	TRUE
1945	0,007	0,000	Jalan Kolektor	1	2	No	No	56,205	TRUE
1946	0,007	0,000	Jalan Kolektor	1	2	No	No	56,213	TRUE
1947	2,727	0,000	Jalan Kolektor	1	2	No	No	57,580	TRUE
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
25415	3,249	0,000	Jalan Arteri	4	2	No	No	3,895	TRUE
25416	0,950	0,000	Jalan Arteri	4	2	No	No	1,320	TRUE
25417	0,237	0,000	Jalan Arteri	4	2	No	No	0,579	TRUE

Lampiran 5. Hasil tuning parameter γ

Tuning ke -	γ	BIC	AIC	Loglikelihood
1	0,032119	136252,0320	136257,8888	-68120,9444
2	0,029266	135985,3012	135981,0149	-67982,5074
3	0,026666	135803,8061	135799,5197	-67891,7599
4	0,024297	135676,9866	135672,7002	-67828,3501
5	0,022139	135552,7698	135538,3403	-67761,1701
6	0,020172	135448,4276	135433,9980	-67708,9990
7	0,018380	135368,3966	135353,9671	-67668,9836
8	0,016747	135306,3949	135291,9654	-67637,9827
9	0,015259	135258,0289	135243,5994	-67613,7997
10	0,013904	135220,1240	135205,6945	-67594,8472
11	0,012669	135190,3375	135175,9080	-67579,9540
12	0,011543	135166,9519	135152,5224	-67568,2612
13	0,010518	135148,5070	135134,0775	-67559,0387
14	0,009583	135118,2704	135093,6977	-67538,8488
15	0,008732	135047,2306	135012,5148	-67498,2574
16	0,007956	134969,3563	134934,6404	-67459,3202
17	0,007249	134902,4565	134867,7406	-67425,8703
18	0,006605	134844,9474	134810,2315	-67397,1157
19	0,006019	134795,5041	134760,7882	-67372,3941
20	0,005484	134753,0106	134718,2947	-67351,1474
21	0,004997	134716,5185	134681,8026	-67332,9013
22	0,004553	134685,2321	134650,5162	-67317,2581
23	0,004148	134658,7641	134624,0482	-67304,0241
24	0,003780	134635,8010	134601,0851	-67292,5425
25	0,003444	134616,1905	134581,4746	-67282,7373
26	0,003138	134599,4893	134564,7734	-67274,3867
27	0,002859	134585,3259	134550,6100	-67267,3050
28	0,002605	134573,3031	134538,5872	-67261,2936
29	0,002374	134563,1429	134528,4270	-67256,2135
30	0,002163	134554,5852	134519,8693	-67251,9346
31	0,001971	134547,4009	134512,6850	-67248,3425
32	0,001796	134541,3908	134506,6750	-67245,3375
33	0,001636	134536,3819	134501,6661	-67242,8330
34	0,001491	134532,2243	134497,5085	-67240,7542
35	0,001358	134528,7885	134494,0727	-67239,0363

Tuning ke -	γ	BIC	AIC	<i>Loglikelihood</i>
36	0,001238	134525,9629	134491,2470	-67237,6235
37	0,001128	134533,6744	134488,8154	-67236,4077
38	0,001028	134531,4181	134486,5590	-67235,2795
39	0,000936	134529,5886	134484,7296	-67234,3648
40	0,000853	134528,1142	134483,2552	-67233,6276
41	0,000777	134526,9349	134482,0759	-67233,0380
42	0,000708	134536,0158	134481,0136	-67232,5068
43	0,000645	134534,2218	134479,2196	-67231,6098
44	0,000588	134532,7209	134477,7187	-67230,8593
45	0,000536	134531,5447	134476,5424	-67230,2712
46	0,000488	134530,5989	134475,5967	-67229,7984
47	0,000445	134529,8602	134474,8580	-67229,4290
48	0,000405	134529,3073	134474,3051	-67229,1526
49	0,000369	134528,8760	134473,8737	-67228,9369
50	0,000336	134528,5549	134473,5527	-67228,7764
51	0,000307	134538,4611	134473,3157	-67228,6579
52	0,000279	134538,2854	134473,1400	-67228,5700
53	0,000255	134538,1687	134473,0233	-67228,5117
54	0,000232	134538,1072	134472,9618	-67228,4809
55	0,000211	134538,0743	134472,9289	-67228,4644
56	0,000193	134538,0692	134472,9238	-67228,4619
57	0,000175	134538,0861	134472,9407	-67228,4703
58	0,000160	134538,1231	134472,9777	-67228,4889
59	0,000146	134538,1677	134473,0223	-67228,5111
60	0,000133	134538,2204	134473,0750	-67228,5375
61	0,000121	134538,2788	134473,1334	-67228,5667
62	0,000110	134538,3407	134473,1953	-67228,5976
63	0,000100	134538,3929	134473,2476	-67228,6238
64	0,000091	134538,4672	134473,3218	-67228,6609
65	0,000083	134538,5143	134473,3689	-67228,6844
66	0,000076	134538,5909	134473,4455	-67228,7228
67	0,000069	134538,6330	134473,4876	-67228,7438

Lampiran 6. Syntax analisis data pendekatan regresi Poisson

a. Persiapan Data

```
#Package
library(spatstat)
library(spatstat.linnet)
library(sf)
library(sp)
library(maptools)
library(dplyr)
library(rgdal)
library(spNetwork)
library(glmnet)
library(stringr)
library(doParallel)
library(plotly)
library(MASS)

### Read data
#####
datacov<-read.csv("D:/Tesis/Data/Linnet_marks.csv", header=TRUE,
sep=";")
dataall<-read.csv("D:/Tesis/Data/dataall.csv",header=TRUE, sep=";")
datatl<-read.csv("D:/Tesis/Data/trafficlights.csv",header=TRUE,
sep=";")

### Observation Window Preparation
#####
peta<-st_read("D:/Tesis/Data/Nganjuk/Nganjuk_filter.shp") #SHP Jaringan
jalan Nganjuk (arteri, kolektor, lokal)
plot(peta[, "geometry"])
nganjuk<-st_transform(peta[, "geometry"], "+proj=utm +zone=49
ellps=WGS84")
nganjuk_sp<-as(nganjuk%>%st_zm(), "Spatial")
nganjuk_linnet<-as.linnet.SpatialLines(nganjuk_sp)
plot(nganjuk_linnet, col="black")
nganjuk_linnet_marks <- marks(nganjuk_linnet$lines) # Menambahkan marks
pada observation window
colnames(datacov) <- c("Min_Kec", "n_Lajur", "n_Jalur", "Marka",
"Kerb_R", "Kerb_L", "Jenis_Jalan")
marks(nganjuk_linnet$lines) <- datacov
nganjuk_linnet_marks <- nganjuk_linnet
linmarks <- marks(nganjuk_linnet_marks$lines)
summary(linmarks)
```

b. Eksplorasi data

```
### Plot point pattern data kecelakaan 2018-2020
#####
coordinates(dataall)<-c("Long", "Lat")
proj4string(dataall)<-CRS("+proj=longlat +datum=WGS84")
dataall.UTM<-spTransform(dataall,CRS("+proj=utm +zone=49 ellps=WGS84"))
kall<-cbind(dataall.UTM$Long,dataall.UTM$Lat)
kall1<-0.1*kall #mengubah koordinat lokasi kecelakaan menjadi per dam
nganjuk_linnet_rescaled <- rescale(nganjuk_linnet_marks, 10, c("meter",
"dam")) #mengubah unit observation window menjadi per 1/10 meter
kecall<-lpp(kall1,nganjuk_linnet_rescaled)
plot(kecall, col="black", cols="red", cex=0.8, main="2018-2020")
```

```

linmarks <- linmarks %>%
  mutate(Kerb_R = case_when(
    Kerb_R == 0 ~ "No",
    Kerb_R == 1 ~ "Yes",
    TRUE ~ as.character(Kerb_R)))

linmarks <- linmarks %>%
  mutate(Kerb_L = case_when(
    Kerb_L == 0 ~ "No",
    Kerb_L == 1 ~ "Yes",
    TRUE ~ as.character(Kerb_L)))

linmarks <- linmarks %>%
  mutate(n_Jalur = case_when(
    n_Jalur == 1 ~ "1 path",
    n_Jalur == 2 ~ "2 paths",
    TRUE ~ as.character(n_Jalur)))

funcListScaled <- lapply(linmarks,
  function(z){function(x,y,seg,tp){z[seg]}})
linfunList <- lapply(funcListScaled, function(z, net){linfun(z, net)},
  net=nganjuk_linnet_rescaled)
vnames <- names(linfunList)
Jenis_Jalan <- linfunList$Jenis_Jalan
n_Lajur<-linfunList$n_Lajur
n_Jalur<-linfunList$n_Jalur
Kerb_R<-linfunList$Kerb_R
Kerb_L<-linfunList$Kerb_L
str(Min_Kec)

### Plot Mark Linnet (kovariat)
plot(Jenis_Jalan, main="Jenis Jalan")
plot(n_Lajur, main="Jumlah Lajur")
plot(n_Jalur, main="Jumlah Jalur")
plot(Kerb_R, main="Kerb Kanan Jalan")
plot(Kerb_L, main="Kerb Kiri Jalan")

### Plot covariate (lampu lalu lintas) pada observation window
coordinates(datat1)<-c("Long","Lat")
proj4string(datat1)<-CRS("+proj=longlat +datum=WGS84")
datat1.UTM<-spTransform(datat1,CRS("+proj=utm +zone=49 ellps=WGS84"))
t1<-cbind(datat1.UTM$Long,datat1.UTM$Lat)
t1<-0.1*t
t1<-lpp(t1,nganjuk_linnet_rescaled)
l1 <- connected(t1)
t12<-t11[[1]]
t1fix<-unmark(t12)

### kovariat jarak titik kecelakaan ke lokasi lampu lalu lintas
terdekat dalam bentuk function
f <- distfun.lpp (t1fix)
plot(f, main = " ")

### Uji homogenitas point pattern dengan CDF Test
cdf.test.lpp(kecfix_all, f)
cdf.test.lpp(kecfix_all, Min_Kec)
cdf.test.lpp(kecfix_all, n_Lajur)
cdf.test.lpp(kecfix_all, n_Jalur)
cdf.test.lpp(kecfix_all, Kerb_R)
cdf.test.lpp(kecfix_all, Kerb_L)

```

```

### Korelasi antar kovariat numerik
covariates_im <- list(
  f_im = as.linim(f),
  n_Lajur_im = as.linim(n_Lajur),
  n_Jalur_im = as.linim(n_Jalur)
)

# Menetapkan nama variabel untuk setiap elemen di list
names(covariates_im) <- c("Jarak_tf", "n_Lajur", "n_Jalur")
cor_matrix <- cor.im(covariates_im, use = "na.or.complete")

# Banyak pixel image dalam default
n <- 128*128

calculate_p_value <- function(r) {
  Z <- (sqrt(n - 3) / 2) * log((1 + r) / (1 - r))
  p_value <- 2 * pnorm(-abs(Z))
  return(p_value)
}

# Menghitung matriks p-value
p_value_matrix <- matrix(nrow = nrow(cor_matrix), ncol =
ncol(cor_matrix))

for (i in 1:nrow(cor_matrix)) {
  for (j in 1:ncol(cor_matrix)) {
    p_value_matrix[i, j] <- calculate_p_value(cor_matrix[i, j])
  }
}
print(p_value_matrix)

### Uji independensi antar kovariat kategorik
### ubah kovariat menjadi kategorik

linmarks$Jenis_Jalan <- as.factor(linmarks$Jenis_Jalan)
linmarks$Kerb_R <- as.factor(linmarks$Kerb_R)
linmarks$Kerb_L <- as.factor(linmarks$Kerb_L)
funcListScaled <- lapply(linmarks,
function(z){function(x,y,seg,tp){z[seg]}})
linfunList <- lapply(funcListScaled, function(z, net){linfun(z, net)},
net=nganjuk_linnet_rescaled)
vnames <- names(linfunList)
Jenis_Jalan <- linfunList$Jenis_Jalan
Kerb_R<-linfunList$Kerb_R
Kerb_L<-linfunList$Kerb_L

##### membuat dataframe #####
kerb_L_df <- as.data.frame(Kerb_L)
KL_df <- kerb_L_df[, "values"]
kerb_R_df <- as.data.frame(Kerb_R)
KR_df <- kerb_R_df[, "values"]
Jenis_Jalan_df <- as.data.frame(Jenis_Jalan)
Jalan_df <- Jenis_Jalan_df[, "values"]
dat_cat_df <- data.frame(Jalan_df, KL_df, KR_df)

chi_sq_result1 <- chisq.test(dat_cat_df$KR_df, dat_cat_df$Jalan_df)
print(chi_sq_result1)
chi_sq_result2 <- chisq.test(dat_cat_df$KL_df, dat_cat_df$Jalan_df)
print(chi_sq_result2)
chi_sq_result3 <- chisq.test(dat_cat_df$KR_df, dat_cat_df$KL_df)
print(chi_sq_result3)
fisher.test(table(dat_cat_df$KR_df, dat_cat_df$KL_df))

```

c. Model *inhomogeneous Poisson point process* pendekatan regresi Poisson

```
## convert covariate to factor
linmarks$Jenis_Jalan <- as.factor(linmarks$Jenis_Jalan)
linmarks$kerb_R <- as.factor(linmarks$kerb_R)
linmarks$kerb_L <- as.factor(linmarks$kerb_L)

fitall <- lppm(kecall~Jenis_Jalan + n_Lajur + n_Jalur + Kerb_R + Kerb_L
+f)
# Quadrature scheme on linear network
Q = linequad(kecall, eps = NULL, nd = 1000, random = FALSE)
plot(Q$dummy, col="white", cols="black", cex=0.5, main="2018-2020")
points(Q$data, col = "red", cex=0.5)

# GLM with quadrature scheme
fit.all <- ppm(Q, trend = ~Jenis_Jalan + n_Lajur + n_Jalur + kerb_R +
Kerb_L + f, interaction = NULL, data = NULL, method =
"mpl")

#AIC and system time
AICpois=(-2*(fitall$fit$maxlogpl))+(2*8)
time.Poisson <- system.time(fit.all <- ppm(Q, trend = ~Jenis_Jalan +
n_Lajur + n_Jalur + Kerb_R + Kerb_L +f , interaction =
NULL, data = NULL, method = "mpl"))
```

d. Model *inhomogeneous Poisson point process* dengan regularisasi Lasso

```
## Diskritisasi dataset hasil GLM
fmla_all <- fit.all$fmla
glmData_all <- fit.all$glmdata
wts.all <- glmData_all$mpl.w
Y.all <- glmData_all$mpl.Y
X.all <- data.matrix(glmData_all[, 3:(ncol(glmData_all)-1)])

# One-hot encoding for categorical variables
X.all_matrix <- model.matrix(~ . - 1, data = X.all)
X.all_matrix <- X.all_matrix[, -3]

# Assuming Y.all and wts.all are vectors and don't need encoding
# Check and handle NAs
Y.all_cleaned <- na.omit(Y.all)
wts.all_cleaned <- na.omit(wts.all)

# Ensure that Y.all_cleaned and wts.all_cleaned match the number of
rows in X.all_matrix
X.all_matrix <- X.all_matrix[complete.cases(X.all), ]

glm_lasso_a1 <- glmnet::cv.glmnet(x = X.all_matrix,
y = Y.all_cleaned,
weights = wts.all_cleaned,
family = "poisson",
alpha = 1)

##parameter
beta2 <- as.matrix(glm_lasso_a1$glmnet.fit$beta)
lambda=as.matrix(glm_lasso_a1$glmnet.fit$lambda)
intersep=as.matrix(glm_lasso_a1$glmnet.fit$a0)
beta.gab=rbind(t(intersep),beta2)
```

```

kov.data<-X.all_matrix[1:Q$data$n,]
kov.dummy<-X.all_matrix[Q$data$n+1:Q$dummy$n,]
kov.all<-X.all_matrix[1:Q$dummy$n,]
I.data<-matrix(1,ncol=1,nrow=Q$data$n)
I.dummy<-matrix(1,ncol=1,nrow=Q$dummy$n)
I.all<-rbind(I.data,I.dummy)
kov.data <- cbind(I.data,kov.data)
kov.dummy <- cbind(I.dummy,kov.dummy)
kov.all <- rbind(kov.all,kov.all)

## Tuning parameter selection
bic2=list()
aic2=list()
loglike=list()

for (i in 1:dim(lambda)[1]){
  beta.cur=as.matrix(beta.gab[,i])
  rho.data=exp(kov.data%%beta.cur)
  rho.dummy=exp(kov.dummy%%beta.cur)
  rho.all=exp(rbind(rho.data,rho.dummy))
  loglike[[i]]=sum(log(rho.data))-sum(rho.all*wts.all)
  bic2[[i]] = -2*loglike[[i]]+log(Q$data$n+Q$dummy$n)*sum(beta.cur!=0)
  aic2[[i]]=-2*loglike[[i]]+(2*8)
}

loglike <- as.data.frame(do.call(rbind, loglike))
bic2 <- as.data.frame(do.call(rbind, bic2))
aic2 <- as.data.frame(do.call(rbind, aic2))
min(aic2)
which.min(as.matrix(bic2))
beta.hat2=beta.gab[,which.min(as.matrix(bic2))]
par.beta=as.data.frame(beta.hat2)

```

Lampiran 7. Output hasil analisis data pendekatan regresi Poisson

a. Output Uji homogenitas

```
> cdf.test.lpp(kecfix_all, f)

      Spatial Kolmogorov-Smirnov test of CSR in linear network

data: covariate 'f' evaluated at points of 'kecfix_all'
      and transformed to uniform distribution under CSR
D = 0.22735, p-value < 2.2e-16
alternative hypothesis: two-sided

> cdf.test.lpp(kecfix_all, Min_Kec)

      Spatial Kolmogorov-Smirnov test of CSR in linear network

data: covariate 'Min_Kec' evaluated at points of 'kecfix_all'
      and transformed to uniform distribution under CSR
D = 0.32376, p-value < 2.2e-16
alternative hypothesis: two-sided

> cdf.test.lpp(kecfix_all, n_Lajur)

      Spatial Kolmogorov-Smirnov test of CSR in linear network

data: covariate 'n_Lajur' evaluated at points of 'kecfix_all'
      and transformed to uniform distribution under CSR
D = 0.27583, p-value < 2.2e-16
alternative hypothesis: two-sided

> cdf.test.lpp(kecfix_all, n_Jalur) #homogen

      Spatial Kolmogorov-Smirnov test of CSR in linear network

data: covariate 'n_Jalur' evaluated at points of 'kecfix_all'
      and transformed to uniform distribution under CSR
D = 0.016332, p-value = 0.6823
alternative hypothesis: two-sided

> cdf.test.lpp(kecfix_all, Marka)

      Spatial Kolmogorov-Smirnov test of CSR in linear network

data: covariate 'Marka' evaluated at points of 'kecfix_all'
      and transformed to uniform distribution under CSR
D = 0.33367, p-value < 2.2e-16
alternative hypothesis: two-sided

> cdf.test.lpp(kecfix_all, Kerb_R) #homogen

      Spatial Kolmogorov-Smirnov test of CSR in linear network

data: covariate 'Kerb_R' evaluated at points of 'kecfix_all'
      and transformed to uniform distribution under CSR
D = 0.024699, p-value = 0.1899
alternative hypothesis: two-sided

> cdf.test.lpp(kecfix_all, Kerb_L)

      Spatial Kolmogorov-Smirnov test of CSR in linear network

data: covariate 'Kerb_L' evaluated at points of 'kecfix_all'
      and transformed to uniform distribution under CSR
D = 0.043093, p-value = 0.001547
alternative hypothesis: two-sided
```

b. *Output Pearson correlation dan uji independensi*

```
> cor_matrix
      Jarak_tf      n_Lajur      n_Jalur
Jarak_tf 1.00000000 -0.26615599 0.03595996
n_Lajur  -0.26615599 1.00000000 -0.01973429
n_Jalur   0.03595996 -0.01973429 1.00000000

> print(chi_sq_result1)

      Pearson's Chi-squared test

data:  dat_cat_df$KR_df and dat_cat_df$Jalan_df
X-squared = 45.293, df = 2, p-value = 1.461e-10

> print(chi_sq_result2)

      Pearson's Chi-squared test

data:  dat_cat_df$KL_df and dat_cat_df$Jalan_df
X-squared = 71.863, df = 2, p-value = 2.483e-16

> chi_sq_result1$observed
      dat_cat_df$Jalan_df
dat_cat_df$KR_df Jalan Arteri Jalan Kolektor Jalan Lokal
      0           94           593           1027
      1           12            7            19

> chi_sq_result1$expected
      dat_cat_df$Jalan_df
dat_cat_df$KR_df Jalan Arteri Jalan Kolektor Jalan Lokal
      0 103.700913      586.9863 1023.31279
      1  2.299087      13.0137  22.68721

> chi_sq_result2$observed
      dat_cat_df$Jalan_df
dat_cat_df$KL_df Jalan Arteri Jalan Kolektor Jalan Lokal
      0           93           595           1033
      1           13            5            13

> chi_sq_result2$expected
      dat_cat_df$Jalan_df
dat_cat_df$KL_df Jalan Arteri Jalan Kolektor Jalan Lokal
      0 104.124429      589.38356 1027.49201
      1  1.875571      10.61644  18.50799

> chi_sq_result3$observed
      dat_cat_df$KL_df
dat_cat_df$KR_df  0  1
      0 1712  2
      1  9  29

> chi_sq_result3$expected
      dat_cat_df$KL_df
dat_cat_df$KR_df  0  1
      0 1683.67237 30.3276256
      1  37.32763  0.6723744

> fisher.test(table(dat_cat_df$KR_df, dat_cat_df$KL_df))

      Fisher's Exact Test for Count Data

data:  table(dat_cat_df$KR_df, dat_cat_df$KL_df)
p-value < 2.2e-16
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 5.251282e+02 4.503600e+15
sample estimates:
odds ratio
 2395.338
```

c. *Output model inhomogeneous Poisson point process tanpa Lasso*

```
> fitall
Point process model on linear network
  Fitted to point pattern dataset 'kecall'

Nonstationary Poisson process

Log intensity: ~Jenis_Jalan + n_Lajur + n_Jalur + Kerb_R + Kerb_L + f

Fitted trend coefficients:
(Intercept) Jenis_Jalan1 Jenis_Jalan2      n_Lajur      n_Jalur      Kerb_RYes      Kerb_LYes
-4.008970505  0.980289493  1.565581336  0.275297609 -0.105068787 -0.618636268  0.522134993
      f
-0.001974592

      Estimate      S.E.      CI95.lo      CI95.hi Ztest      Zval
(Intercept) -4.008970505  0.6092587605 -5.203095732 -2.814845277 *** -6.580079
Jenis_Jalan1  0.980289493  0.0559718217  0.870586738  1.089992247 ***  17.513982
Jenis_Jalan2  1.565581336  0.0871149421  1.394839187  1.736323485 ***  17.971444
n_Lajur       0.275297609  0.0333463808  0.209939904  0.340655314 ***   8.255697
n_Jalur      -0.105068787  0.3031999613 -0.699329792  0.489192217   -0.346533
Kerb_RYes    -0.618636268  0.2030242604 -1.016556506 -0.220716030 **  -3.047105
Kerb_LYes    0.522134993  0.1951419299  0.139663839  0.904606147 **   2.675668
f            -0.001974592  0.0001365927 -0.002242309 -0.001706876 *** -14.456065

Domain: Linear network with 11547 vertices and 11736 lines
Line segments are marked
Enclosing window: rectangle = [58095.81, 62898.79] x [-86008.44, -81929.18] dam
```

d. *Output glmnet dengan regularisasi Lasso*

```
> glm_lasso_a1

Call: glmnet::cv.glmnet(x = X.all_matrix, y = Y.all_cleaned, weights = wts.all_cleaned,
family = "poisson", alpha = 1)

Measure: Poisson Deviance

      Lambda Index Measure      SE Nonzero
min 0.000076   66  0.1753 0.002785      7
1se 0.003780   24  0.1779 0.002867      4
```

e. *Output $\hat{\theta}$ dengan nilai γ optimal*

```
> par.beta
      beta.hat2
      -4.205170643
Jenis_Jalan1  0.844578506
Jenis_Jalan2  1.481853725
n_Lajur       0.274809741
n_Jalur       0.000000000
Kerb_RYes     0.000000000
Kerb_LYes     0.000000000
f            -0.001492924
```

BIODATA PENULIS



Penulis dilahirkan di Gresik, 7 Januari 2000, merupakan anak ketiga dari tiga bersaudara dari pasangan Ayah Drs. Moh. Syafi'i M.Pd.I dan Ibu Alifah, S.Pd.I. Penulis menyelesaikan pendidikan formal jenjang sekolah dasar di MI. Nurul Ulum Gumeno tahun 2012. Kemudian menyelesaikan pendidikan berikutnya di SMP Negeri 1 Gresik pada tahun 2015 dan di SMA Negeri 1 Gresik pada tahun 2018. Penulis kemudian melanjutkan pendidikan jenjang S1 melalui jalur SNMPTN di Departemen Statistika FSAD - ITS dan menyelesaikan pada tahun 2022. Setelah kelulusannya, penulis melanjutkan studi magister di Departemen Statistika FSAD – ITS dengan beasiswa *fresh graduate*. Dalam kesehariannya selama menjalani studi, penulis juga bekerja sebagai asisten tenaga ahli statistika di Pusat Kajian Potensi Daerah dan Pemberdayaan Masyarakat (PDPM) di bawah DRPM ITS. Di sisi lain, selama melanjutkan studi di jenjang magister, penulis terlibat sebagai asisten dosen di mata kuliah Analisis Data Departemen Statistika dan Multivariat lanjutan Departemen Statistika Bisnis Institut Teknologi Sepuluh Nopember. Selain itu, penulis juga aktif terlibat sebagai pengurus di organisasi ISNU dan PMII Kota Surabaya. Bagi para pembaca yang ingin memberikan kritik, saran, atau pertanyaan mengenai penelitian ini, dapat menghubungi penulis melalui email: vandafitriyanah@gmail.com.