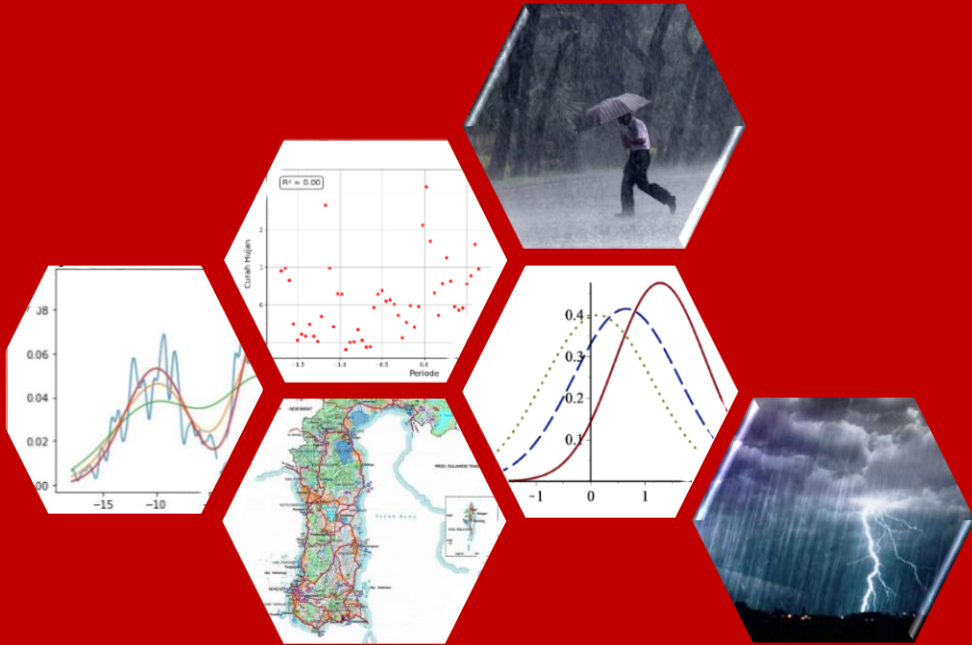


**ESTIMASI MODEL REGRESI NONPARAMETRIK MENGGUNAKAN ESTIMATOR
KERNEL GASSER-MÜLLER GAUSSIAN PADA DATA CURAH HUJAN DI KOTA
MAKASSAR**



**NUR AFIKA
H051201029**



**PROGRAM STUDI STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS HASANUDDIN
2024**

**ESTIMASI MODEL REGRESI NONPARAMETRIK MENGGUNAKAN
ESTIMATOR *KERNEL GASSER-MÜLLER GAUSSIAN* PADA DATA
CURAH HUJAN DI KOTA MAKASSAR**

**NUR AFIKA
H051201029**



**PROGRAM STUDI STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS HASANUDDIN
MAKASSAR
2024**

**ESTIMASI MODEL REGRESI NONPARAMETRIK MENGGUNAKAN
ESTIMATOR *KERNEL GASSER-MÜLLER GAUSSIAN* PADA DATA
CURAH HUJAN DI KOTA MAKASSAR**

NUR AFIKA
H051201029

Skripsi

Diajukan sebagai salah satu syarat untuk mencapai gelar Sarjana Statistika

Program Studi Statistika

pada

**PROGRAM STUDI STATISTIKA DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS HASANUDDIN
MAKASSAR
DESEMBER 2024**

SKRIPSI
ESTIMASI MODEL REGRESI NONPARAMETRIK MENGGUNAKAN
ESTIMATOR *KERNEL GASSER-MÜLLER GAUSSIAN* PADA DATA
CURAH HUJAN DI KOTA MAKASSAR

NUR AFIKA
H051201029

Skripsi,

telah dipertahankan di depan Panitia Ujian Sarjana Statistika pada tanggal 4
Desember 2024 dan dinyatakan telah memenuhi syarat kelulusan pada

Program Studi Statistika
Departemen Statistika
Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Hasanuddin
Makassar

Mengesahkan:
Pembimbing Tugas Akhir,



Dr. Anna Islamiyati, S.Si., M.Si.
NIP. 19770808 200501 2 002

Mengetahui:
Ketua Program Studi,



Dr. Anna Islamiyati, S.Si., M.Si.
NIP. 19770808 200501 2 002

**PERNYATAAN KEASLIAN SKRIPSI
DAN PELIMPAHAN HAK CIPTA**

Dengan ini saya menyatakan bahwa, skripsi berjudul "Estimasi Model Regresi Nonparametrik Menggunakan Estimator *Kernel Gasser-Müller Gaussian* pada Data Curah Hujan di Kota Makassar" adalah benar karya saya dengan arahan dari Dr. Anna Islamiyati, S.Si., M.Si. sebagai Pembimbing Utama. Karya ilmiah ini belum diajukan dan tidak sedang diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka skripsi ini. Apabila di kemudian hari terbukti atau dapat dibuktikan bahwa sebagian atau keseluruhan skripsi ini adalah karya orang lain, maka saya bersedia menerima sanksi atas perbuatan tersebut berdasarkan aturan yang berlaku.

Dengan ini saya melimpahkan hak cipta (hak ekonomis) dari karya tulis saya berupa skripsi ini kepada Universitas Hasanuddin.

Makassar, 4 Desember 2024



Nur Afika
NIM. H051201029

UCAPAN TERIMA KASIH

Puji Syukur penulis panjatkan kepada **Allah SWT.** yang senantiasa memberikan limpahan rahmat dan karunia-Nya, yang senantiasa menjadi penopang dan teman terbaik penulis sepanjang hidupnya. Shalawat serta salam senantiasa tercurahkan kepada baginda **Rasullah Shallallahu 'Alaihi Wa Sallam** beserta keluarga dan sahabatnya.

Penulis menyampaikan penghargaan dan rasa terima kasih yang mendalam kepada **Ibu Dr. Anna Islamiyati, S.Si., M.Si.** selaku Pembimbing Utama yang telah memberikan arahan, ilmu, waktu, dan dukungan penuh selama proses penulisan tugas akhir ini. Terima kasih kepada **Bapak Drs Raupong, M.Si.** dan **Ibu Dr. Erna Tri Herdiani, S.Si., M.Si.** selaku Tim Penguji yang telah memberikan masukan berharga dalam proses penyempurnaan tugas akhir ini. Penghargaan yang sebesar-besarnya juga penulis sampaikan kepada **pimpinan Universitas Hasanuddin, Departemen Statistika**, serta **seluruh dosen dan staf** yang telah memberikan fasilitas, ilmu, dan dukungan yang berharga selama penulis menempuh studi.

Penulis menyampaikan terima kasih yang mendalam kepada kedua orang tua tercinta, **Ratna dan Muhammad Alex**, atas dukungan penuh, pengorbanan, restu, serta doa yang tiada henti mengiringi setiap langkah penulis. Ucapan terima kasih juga penulis sampaikan kepada adik tersayang, **Adiba Azkadinah**, yang selalu menghibur dan memberikan semangat kepada penulis dalam menyelesaikan tugas akhir ini.

Penghargaan yang tulus dan ucapan terima kasih juga penulis sampaikan kepada teman-teman seperjuangan penulis, **Ciwi-Ciwi Among (Aish, Aliah, Aul, Ayu, Cynthia, Ica, Jihan, Krisna, Laurine, Nadia, Nahla, Radia, Rahmi, Rfidah, Parida, Peby, dan Putri)**, **Among Us (Alip, Lili, Yoel, Theo, Faldi, Ryval, Fadlan, Fahmi, Rais, Bahar, dan Hakam)**, dan **Data Divas (Aliah dan Nadia)** yang telah memberikan dukungan, motivasi, dan semangat yang tiada henti sepanjang masa perkuliahan penulis. Terima kasih atas kebersamaan yang telah terjalin dengan baik, atas segala kontribusi yang telah diberikan, baik dalam bentuk ide, dukungan moral, maupun kerja sama yang produktif.

Penulis menyadari bahwa masih terdapat banyak kekurangan dalam penyusunan tugas akhir ini. Oleh karena itu, dengan penuh kerendahan hati, penulis memohon maaf yang sebesar-besarnya. Penulis berharap karya ini dapat bermanfaat sebesar-besarnya bagi para penuntut ilmu.

Makassar, 4 Desember 2024



Nur Afika

ABSTRAK

Nur Afika. **Estimasi Model Regresi Nonparametrik Menggunakan Estimator *Kernel Gasser-Müller Gaussian* pada Data Curah Hujan di Kota Makassar** (dibimbing oleh Anna Islamiyati).

Latar Belakang. Curah hujan di Indonesia mengalami fluktuasi yang signifikan akibat perubahan iklim global dan pola cuaca yang tidak stabil sehingga penting untuk melakukan identifikasi pola curah hujan guna meminimalisir dampak yang ditimbulkan. Regresi nonparametrik *kernel* merupakan metode yang dapat mengestimasi pola data dengan efektif dan fleksibel. Estimator yang dapat mengestimasi fungsi regresi *kernel* dengan stabil dan memiliki kemampuan adaptabilitas yang baik terhadap data adalah *Gasser-Müller Gaussian*. **Tujuan.** Penelitian ini bertujuan untuk memperoleh nilai *bandwidth* optimal dan bentuk estimasi model regresi nonparametrik *kernel* pada data curah hujan di Kota Makassar. **Metode.** Penelitian ini menggunakan metode *kernel Gasser-Müller Gaussian*. **Hasil.** Model estimator *kernel Gasser-Müller Gaussian* menghasilkan estimasi curah hujan menggunakan periode, kecepatan angin maksimum, dan arah angin saat kecepatan maksimum dengan nilai *Mean Squared Error* (MSE) secara berturut-turut sebesar 0,24, 0,73, dan 0,87. **Kesimpulan.** Bentuk estimasi model regresi nonparametrik *kernel* diperoleh dengan *bandwidth* optimal 0,05 (*Generalized Cross Validation*), 0,34 (*Silverman's Rule of Thumb*), dan 0,33 (*Silverman's Rule of Thumb*).

Kata Kunci: *Kernel, Gasser-Müller, Gaussian, GCV, Silverman's Rule of Thumb, Scott's Rule.*

ABSTRACT

Nur Afika. **Estimation of Nonparametric Regression Model Using Gasser-Müller Gaussian Kernel Estimator on Rainfall Data in Makassar City** (supervised by Anna Islamiyati).

Background. Rainfall in Indonesia experiences significant fluctuations due to global climate change and unstable weather patterns, so it is important to identify rainfall patterns to minimize the impact. Kernel nonparametric regression is a method that can estimate data patterns effectively and flexibly. The estimator that can estimate the kernel regression function stably and has good adaptability to the data is Gasser-Müller Gaussian. **Objective.** This study aims to obtain the optimal bandwidth value and estimation form of the kernel nonparametric regression model on rainfall data in Makassar City. **Methods.** This research uses the Gasser-Müller Gaussian kernel method. **Results.** The Gasser-Müller Gaussian kernel estimator model produces rainfall estimates using period, maximum wind speed, and wind direction at maximum speed with Mean Squared Error (MSE) values of 0.24, 0.73, and 0.87, respectively. **Conclusion.** The estimation form of the kernel nonparametric regression model is obtained with an optimal bandwidth of 0.05 (Generalized Cross Validation), 0.34 (Silverman's Rule of Thumb), and 0.33 (Silverman's Rule of Thumb).

Kata Kunci: Kernel, Gasser-Müller, Gaussian, GCV, Silverman's Rule of Thumb, Scott's Rule.

DAFTAR ISI

	Halaman
HALAMAN JUDUL	i
HALAMAN PENGAJUAN	ii
HALAMAN PENGESAHAN	iii
PERNYATAAN KEASLIAN SKRIPSI DAN PELIMPAHAN HAK CIPTA	iv
UCAPAN TERIMA KASIH	v
ABSTRAK	vi
ABSTRACT	vii
DAFTAR ISI	viii
DAFTAR TABEL	ix
DAFTAR GAMBAR	x
DAFTAR LAMPIRAN	xi
DAFTAR LAMBANG/SINGKATAN	xii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Tujuan dan Manfaat	2
1.3 Batasan Masalah	3
1.4 Teori	3
1.4.1 Analisis Regresi	3
1.4.2 Regresi Nonparametrik	4
1.4.3 Standarisasi Data	5
1.4.4 Metode <i>Kernel</i>	5
1.4.5 Fungsi <i>Kernel</i>	7
1.4.6 Estimator Gasser-Müller	8
1.4.7 <i>Bandwidth</i>	8
1.4.8 <i>Mean Squared Error</i>	10
1.4.9 Curah Hujan	10
BAB II METODE PENELITIAN	12
2.1 Sumber Data	12
2.2 Tahapan Penelitian	12
BAB III HASIL DAN PEMBAHASAN	14
3.1 Hasil	14
3.1.1 Deskripsi Data	14
3.1.2 Standarisasi Data	16
3.1.3 Pemilihan <i>Bandwidth</i> Optimum	17
3.1.4 Estimator Gasser-Müller	18
3.1.5 Pemilihan Model Terbaik	24
BAB IV KESIMPULAN	28
4.1 Kesimpulan	28
4.2 Saran	28
DAFTAR PUSTAKA	29
LAMPIRAN	32

DAFTAR TABEL

Nomor Urut	Halaman
Tabel 1. Variabel respon dan prediktor	12
Tabel 2. Statistik deskriptif variabel penelitian	15
Tabel 3. Data penelitian setelah standarisasi.....	16
Tabel 4 . Nilai <i>bandwith</i> variabel penelitian	17
Tabel 5. Nilai MSE dari kombinasi variabel dan <i>bandwidth</i>	25

DAFTAR GAMBAR

Nomor Urut	Halaman
Gambar 1. <i>Scatter plot</i> antara rata-rata curah hujan dengan periode	14
Gambar 2. <i>Scatter plot</i> antara rata-rata curah hujan dengan rata-rata kecepatan angin maksimum	14
Gambar 3. <i>Scatter plot</i> antara rata-rata curah hujan dengan rata-rata arah angin saat kecepatan maksimum	15
Gambar 4. Kurva data aktual dan data estimasi rata-rata curah hujan dan periode dengan metode GCV	20
Gambar 5. Kurva data aktual dan data estimasi rata-rata curah hujan dan periode dengan metode SROT	20
Gambar 6. Kurva data aktual dan data estimasi rata-rata curah hujan dan periode dengan metode <i>Scott's Rule</i>	20
Gambar 7. Kurva data aktual dan data estimasi rata-rata curah hujan dan rata-rata kecepatan angin maksimum dengan metode GCV	21
Gambar 8. Kurva data aktual dan data estimasi rata-rata curah hujan dan rata-rata kecepatan angin maksimum dengan metode SROT	22
Gambar 9. Kurva data aktual dan data estimasi rata-rata curah hujan dan rata-rata kecepatan angin maksimum dengan metode <i>Scott's Rule</i>	22
Gambar 10. Kurva data aktual dan data estimasi rata-rata curah hujan dan rata-rata arah angin saat kecepatan maksimum dengan metode GCV	23
Gambar 11. Kurva data aktual dan data estimasi rata-rata curah hujan dan rata-rata arah angin saat kecepatan maksimum dengan metode SROT	23
Gambar 12. Kurva data aktual dan data estimasi rata-rata curah hujan dan rata-rata arah angin saat kecepatan maksimum dengan metode <i>Scott's Rule</i>	24
Gambar 13. Kurva data aktual dan data estimasi rata-rata curah hujan dan periode dengan nilai <i>bandwidth</i> optimal 0,05.....	25
Gambar 14. Kurva data aktual dan data estimasi rata-rata curah hujan dan rata-rata kecepatan angin maksimum dengan nilai <i>bandwidth</i> optimal 0,34	26
Gambar 15. Kurva data aktual dan data estimasi rata-rata curah hujan dan rata-rata arah angin saat kecepatan maksimum dengan nilai <i>bandwidth</i> optimal 0,33	27

DAFTAR LAMPIRAN

Lampiran 1. Data rata-rata bulanan curah hujan, periode, kelembapan rata-rata, dan kecepatan angin maksimum sebelum standarisasi	32
Lampiran 2. Data rata-rata bulanan curah hujan, periode, kelembapan rata-rata, dan kecepatan angin maksimum setelah standarisasi	35
Lampiran 3. <i>Source code</i> perhitungan koefisien determinasi (R^2) pada data curah	38
Lampiran 4. <i>Source code</i> perhitungan koefisien determinasi (R^2) pada data curah hujan dan kecepatan angin maksimum	39
Lampiran 5. <i>Source code</i> perhitungan koefisien determinasi (R^2) pada data curah hujan dan arah angin saat kecepatan maksimum	40
Lampiran 6. <i>Source code</i> model estimator Gasser-Müller Gaussian	41
Lampiran 7. <i>Source code</i> metode <i>Generalized Cross Validation</i> (GCV).....	42
Lampiran 8. <i>Source code</i> metode <i>Silverman's Rule of Thumb</i> (SROT).....	43
Lampiran 9. <i>Source code</i> metode <i>Scott's Rule</i>	44

DAFTAR LAMBANG/SINGKATAN

Lambang/Singkatan	Arti dan Penjelasan
erf	<i>Error Function</i>
GCV	<i>Generalized Cross Validation</i>
mm	Milimeter
MSE	<i>Mean Squared Error</i>
R^2	Koefisien Determinasi
SROT	<i>Silverman's Rule of Thumb</i>
m/s	<i>Meter/Second</i>
$\hat{f}_h(x)$	Fungsi kerapatan <i>kernel</i>
$\hat{m}^{GM}(x)$	Fungsi estimator Gasser-Müller
\hat{y}_i	Data hasil estimasi
$W_i(x_{ij}, h)$	Fungsi berbobot
y_i	Nilai variabel respon pada pengamatan ke- i
$\hat{m}(x)$	Fungsi estimator <i>kernel</i>
\bar{x}	Nilai rata-rata
x_i	Nilai variabel prediktor pada pengamatan data ke- i
x_i^*	Nilai standarisasi data ke- i
x_j	Variabel prediktor pada pengamatan ke- j
ε_i	Sisaan atau kesalahan acak dengan $E(\varepsilon_i) = 0$, variansi (σ^2) dan kovariansi nol ($var(\varepsilon_i) = 0$ dan $cov(\varepsilon_i, \varepsilon_j) = 0$ untuk semua i, j , dan $i \neq j$), ε_i dan ε_j tidak berkorelasi
o	Derajat
h	<i>Bandwidth</i>
$E(Y X)$	Fungsi ekspektasi bersyarat dari variabel respon relatif terhadap variabel prediktor
I	Fungsi indikator
IQR	<i>Inter Quartil Range</i> ($Q_3 - Q_1$)
K	Fungsi <i>kernel</i>
$f(x, y)$	Fungsi probabilitas gabungan
$g(x)$	Fungsi probabilitas marginal dari x
$m(x_i)$	Fungsi regresi yang tidak diketahui
n	Ukuran sampel
r	Derajat momen
s	Standar deviasi
W	Matriks <i>smoothing</i> atau matriks hat ($n \times n$) dari $W_i(x_{ij}, h)$
α, β	Parameter

BAB I PENDAHULUAN

1.1 Latar Belakang

Indonesia memiliki letak geografis wilayah perairan seluas 70% dengan region lintang rendah (*low latitudes*). Secara umum, pola musim di Indonesia dikenal sebagai pola Monsun yang dipengaruhi oleh angin monsun. Adanya angin ini menyebabkan Indonesia mengalami dua musim, yakni musim hujan yang puncaknya terjadi pada bulan Desember, Januari, dan Februari, serta musim kemarau yang puncaknya terjadi pada bulan Juni, Juli, dan Agustus (Ihwan, 2013). Beberapa tahun terakhir, curah hujan di Indonesia mengalami fluktuasi yang signifikan akibat perubahan iklim global dan pola cuaca yang tidak stabil, termasuk banjir dan longsor yang sering terjadi di beberapa daerah. Oleh karena itu, untuk mengatasi fluktuasi curah hujan yang signifikan dan dampaknya terhadap sektor-sektor vital dalam kehidupan masyarakat di Indonesia, penting untuk mengidentifikasi pola-pola yang kompleks dalam data curah hujan.

Terdapat beberapa metode yang dapat digunakan untuk melakukan identifikasi pola data. Mulai model regresi linier sederhana yang cocok untuk data yang memiliki perilaku linier, hingga teknik yang lebih canggih yang mempertimbangkan persamaan yang lebih kompleks (Cusi & Bernal, 2022). Tantangan dalam melakukan proyeksi data adalah memodelkan perilaku yang kompleks dan realistis, yang menyebabkan pembagian metode estimasi menjadi tradisional yang berfokus pada linearitas dan asumsi serta non-tradisional yang membutuhkan pengetahuan teknik statistik yang dapat diimplementasikan secara empiris menggunakan kemajuan teknologi dan komputasi.

Salah satu pendekatan non-tradisional yang dapat digunakan adalah regresi nonparametrik. Regresi nonparametrik merupakan metode yang mengestimasi bentuk kurva regresi tanpa membuat asumsi tentang pola data. Beberapa teknik untuk mengestimasi bentuk kurva regresi nonparametrik antara lain estimator *kernel*, estimator *spline*, estimator deret *Fourier*, estimator *wavelet*, k-NN, estimator histogram, estimator deret orthogonal, dan lain-lain (Abdy, 2019).

Estimator *kernel* adalah metode regresi nonparametrik yang menggunakan fungsi pembobot *kernel* (fungsi *kernel*) untuk mengestimasi nilai ekspektasi bersyarat dari variabel respon berdasarkan variabel prediktor. Nilai estimasi dinyatakan sebagai jumlah terbobot dari variabel respon pada setiap titik data yang digunakan dalam proses estimasi (Adityaningrum dkk., 2023). Hal ini menyebabkan Estimator *kernel* sangat fleksibel dan dapat mencapai tingkat kekonvergenan yang relatif cepat. Jenis-jenis metode estimator regresi *kernel* yang dapat digunakan untuk mengestimasi fungsi *kernel*, yaitu Nadaraya-Watson, Priestley-Chao, dan Gasser-Müller. Perbedaan mendasar antara estimator Nadaraya-Watson, estimator Priestley-Chao, dan estimator Gasser-Müller terletak pada strategi penentuan bobot atau kontribusi relatif dari setiap titik data dalam proses estimasi fungsi regresi. Pendekatan Nadaraya-Watson menggunakan metode rata-rata tertimbang dengan bobot yang secara proporsional bergantung pada jarak dari titik estimasi, sementara pendekatan Priestley-Chao menekankan bobot yang lebih sensitif terhadap jarak dan

interval antara titik data serta titik estimasi. Sementara itu, Pendekatan Gasser-Müller mengadopsi metode yang lebih kompleks dengan mengintegrasikan fungsi *kernel* di sekitar setiap titik data untuk menentukan bobotnya, sehingga menghasilkan estimasi yang lebih halus dan fleksibel (Herawati dkk., 2022). Dalam estimator kernel, parameter pemulus yang digunakan untuk mengontrol kemulusan kurva regresi adalah nilai *bandwidth*. Metode yang dapat digunakan untuk memilih *bandwidth* optimal adalah *Generalized Cross Validation (GCV)*, *Silverman's Rule of Thumb (SROT)*, dan *Scott's Rule*.

Penelitian sebelumnya telah dilakukan oleh Anisa dkk., (2019) dengan mengestimasi model regresi nonparametrik *kernel* menggunakan estimator Nadaraya-Watson pada data Indeks Pembangunan Manusia di Indonesia. Penelitian ini memperoleh hasil bahwa regresi *kernel* nonparametrik dengan estimator Nadaraya-Watson memiliki kemampuan estimasi yang sangat baik serta efektif dalam mengatasi kasus distribusi data yang tidak diketahui bentuk fungsinya. Penelitian lain dilakukan oleh Nurjaina dkk., (2022) dengan mengestimasi kurva regresi nonparametrik menggunakan estimator Priestley-Chao dengan fungsi *kernel Triangle* pada data *sunspot* matahari. Penelitian ini memperoleh hasil estimasi kurva yang hampir mengikuti pola kurva data asli yang sebenarnya.

Metode yang sering digunakan pada penelitian-penelitian sebelumnya untuk estimasi dalam analisis *kernel* nonparametrik adalah estimator Nadaraya-Watson. Namun, pada penelitian ini, penulis akan melakukan pendekatan yang lebih mendalam dengan melakukan estimasi kurva regresi nonparametrik menggunakan metode estimator Gasser-Müller. Berdasarkan penelitian yang dilakukan oleh Peristera & Kostaki, (2005) yang mengevaluasi kinerja dari ketiga estimator *kernel* menggunakan data mortalitas kelulusan, memperoleh hasil bahwa dibandingkan dengan estimator Nadaraya-Watson dan Priestley-Chao, estimator Gasser-Müller unggul dalam hal stabilitas dan adaptabilitas terhadap pola data dibandingkan dengan dua estimator sebelumnya.

Penelitian lain telah dilakukan oleh Herawati dkk., (2022) dengan mengestimasi data curah hujan di Kota Lampung menggunakan estimator Nadaraya-Watson, Priestley-Chao, dan Gasser-Müller. Penelitian tersebut menggunakan fungsi *kernel Epanechnikov* dengan pemilihan *bandwidth* berdasarkan nilai *Generalized Cross Validation (GCV)* untuk ketiga estimator. Oleh karena itu, pada penelitian ini dilakukan estimasi model regresi nonparametrik menggunakan estimator Gasser-Müller dengan fungsi *kernel Gaussian* dan menentukan nilai *bandwidth* berdasarkan nilai *GCV*, *Silverman's Rule of Thumb (SROT)*, dan *Scott's Rule* pada data curah hujan di Kota Makassar.

1.2 Tujuan dan Manfaat

Tujuan dilakukannya penelitian ini adalah:

1. Memperoleh nilai *bandwidth* optimal pada estimator Gasser-Müller *Gaussian* dengan menggunakan metode *Generalized Cross Validation (GCV)*, *Silverman's Rule of Thumb (SROT)*, dan *Scott's Rule* pada data curah hujan di Kota Makassar.

2. Memperoleh bentuk estimasi model regresi nonparametrik melalui estimator *Kernel Gasser-Müller Gaussian* pada data curah hujan di Kota Makassar.

Manfaat yang ingin dicapai penelitian ini antara lain:

1. Memperdalam wawasan mengenai konsep, teori, dan metode yang terkait dengan regresi nonparametrik menggunakan estimator *Kernel Gasser-Müller Gaussian*.
2. Memberikan informasi mengenai pola curah hujan di Kota Makassar yang diharapkan dapat mendukung perencanaan yang lebih baik dalam pengelolaan sumber daya air, pertanian, atau mitigasi risiko bencana di Kota Makassar.

1.3 Batasan Masalah

Batasan permasalahan dalam penelitian ini yaitu:

1. Data yang digunakan merupakan data curah hujan di Kota Makassar pada tahun 2018-2023 yang diperoleh dari Stasiun Meteorologi Maritim Paotere yang terletak di wilayah Makassar.
2. Fungsi *kernel* yang digunakan adalah fungsi *kernel Gaussian*.
3. Teknik akurasi estimasi yang digunakan berdasarkan nilai *Mean Squared Error* (MSE).

1.4 Teori

1.4.1 Analisis Regresi

Regresi pertama kali diperkenalkan pada tahun 1886 oleh Francis Galton, seorang antropolog dan ahli meteorologi, dalam artikelnya yang berjudul "*Family Likeness in Stature*". Penelitiannya memperkenalkan konsep hukum regresi universal yang menyatakan bahwa distribusi tinggi badan dalam suatu populasi tidak mengalami perubahan signifikan antar generasi. Hal ini didasarkan pada pengamatan bahwa tinggi rata-rata anak dari orang tua yang memiliki tinggi tertentu cenderung mendekati rata-rata tinggi seluruh populasi (*regress*).

Saat ini istilah regresi telah mengalami pergeseran makna dari definisi awal yang diutarakan oleh Galton. Analisis regresi dipahami sebagai suatu metode untuk mengukur ketergantungan antara suatu variabel terhadap variabel prediktor, dengan tujuan membuat estimasi atau prediksi nilai rata-rata variabel respon berdasarkan nilai variabel prediktor yang diketahui (Basuki & Prawoto, 2016). Selain itu, analisis regresi juga diartikan sebagai pendekatan dalam statistika yang bertujuan untuk menguji sejauh mana hubungan antar variabel atau menilai bagaimana suatu variabel memengaruhi variabel lainnya (Sarbaini dkk., 2022).

Bentuk umum fungsi regresi linear dapat dituliskan sebagai berikut (Basuki & Prawoto, 2016):

$$y_i = \alpha + \beta x_i + \varepsilon_i, \text{ dengan } i = 1, 2, \dots, n \quad (1)$$

dengan

y_i : nilai variabel respon pada pengamatan ke- i ,

α, β : parameter,

x_i : nilai variabel prediktor pada pengamatan data ke- i ,
 ε_i : sisaan atau kesalahan acak dengan $E(\varepsilon_i) = 0$, variansi (σ^2) dan kovariansi nol ($var(\varepsilon_i) = 0$ dan $cov(\varepsilon_i, \varepsilon_j) = 0$ untuk semua j , dan $i \neq j$), ε_i dan ε_j tidak berkorelasi.

Analisis regresi berperan penting dalam mengidentifikasi pola atau hubungan antara variabel respon dan variabel prediktor melalui kurva regresi. Dalam analisis regresi, bentuk pola yang terbentuk antara variabel respon dan variabel prediktor tidak selalu dapat ditangani dengan pendekatan parametrik yang mengasumsikan bentuk kurva regresi tertentu, seperti linier, kuadratik, atau kubik. Dalam beberapa kasus, pola yang muncul memerlukan pendekatan semiparametrik atau nonparametrik. Pendekatan semiparametrik digunakan ketika sebagian pola pada kurva regresi diketahui, sementara sebagian lainnya tidak. Sebaliknya, jika pola kurva regresi diasumsikan tidak diketahui, misalnya data tersebar secara acak, terkumpul di titik-titik tertentu, atau menunjukkan fluktuasi yang bervariasi maka pendekatan nonparametrik diperlukan untuk analisis regresi (Dani & Ni'matuzzahroh, 2022).

1.4.2. Regresi Nonparametrik

Regresi nonparametrik merupakan salah satu metode regresi yang digunakan ketika pola hubungan antara variabel respon dan prediktor tidak diketahui. Metode ini hanya mengasumsikan bahwa kurva regresi yang menunjukkan pola hubungan antara variabel respon dan prediktor bersifat mulus (*smooth*). Hal ini menyebabkan kurva regresi tersebut tidak terikat pada bentuk fungsional tertentu dan memiliki tingkat fleksibilitas yang tinggi (Yulianti & Sihombing, 2020). Dalam regresi nonparametrik, fleksibilitas yang tinggi memungkinkan data untuk menentukan bentuk pola estimasinya sendiri tanpa dipengaruhi oleh preferensi subjektif dari peneliti (Dani dkk., 2020).

Perbedaan utama antara analisis regresi parametrik dan analisis regresi nonparametrik terletak pada kebutuhan asumsi model. Analisis regresi parametrik memiliki beberapa asumsi yang harus dipenuhi, sedangkan analisis regresi nonparametrik tidak mengharuskan adanya asumsi spesifik. Hal ini menyebabkan metode ini lebih tepat untuk menganalisis data yang sangat fluktuatif (Suparti & Santoso, 2023). Macam-macam model regresi nonparametrik yakni histogram, deret orthogonal, *kernel*, *spline*, deret Fourier, *polynomial local*, dan *wavelet* (Rahayu dkk., 2021).

Model umum regresi nonparametrik dapat ditulis sebagai berikut (Yulianti & Sihombing, 2020):

$$Y_i = m(x_i) + \varepsilon_i, \text{ dengan } i = 1, 2, \dots, n \quad (2)$$

dengan

y_i : nilai variabel respon pada pengamatan ke- i ,
 x_i : nilai variabel prediktor pada pengamatan data ke- i ,
 $m(x_i)$: fungsi regresi yang tidak diketahui,
 ε_i : sisaan atau residual ke- i , dengan $\varepsilon_1, \dots, \varepsilon_n$ i.i.d dan $E(\varepsilon_i) = 0$.

1.4.3 Standarisasi Data

Standarisasi data merupakan metode yang digunakan untuk melakukan perubahan skala data. Standarisasi sangat disarankan ketika variabel yang dianalisis memiliki skala atau satuan yang berbeda secara signifikan, dengan tujuan untuk mengurangi perbedaan nilai antar variabel. Persamaan standarisasi data dapat ditulis sebagai berikut (Rizkia dkk., 2024):

$$x_i^* = \frac{x_i - \bar{x}}{s} \quad (3)$$

dengan

- x_i^* : nilai standarisasi data ke- i ,
- x_i : data ke- i ,
- \bar{x} : nilai rata-rata,
- s : standar deviasi.

1.4.4 Metode Kernel

Metode *kernel* adalah teknik estimasi regresi nonparametrik yang menggunakan fungsi pemulus *kernel* untuk menghasilkan estimasi regresi. Pendekatan ini berfokus pada penerapan rata-rata terbobot dari data untuk menghasilkan nilai hasil regresi, yang memungkinkan model untuk menangkap struktur yang kompleks dalam data. Metode *kernel* memanfaatkan fungsi *kernel* untuk mengidentifikasi korelasi *non-linear* antara variabel respon dan variabel prediktor, serta mengestimasi nilai ekspektasi bersyarat dari variabel acak (Yuliati & Sihombing, 2020). Persamaan ekspektasi bersyarat dari variabel respon relatif terhadap variabel prediktor dapat ditulis sebagai berikut (Natalia, 2014):

$$E(Y|X) = m(X) = \int_{-\infty}^{\infty} y \cdot \frac{f(x, y)}{g(x)} dy \quad (4)$$

dengan $f(x, y)$ adalah fungsi probabilitas gabungan, $g(x)$ adalah fungsi probabilitas marginal dari x , dan m mewakili suatu fungsi yang tidak diketahui, yang memerlukan penentuan dan penggunaan bobot yang tepat. Dalam analisis regresi *kernel*, terdapat berbagai jenis estimator yang dapat diterapkan seperti, estimator Priestley-Chao, estimator Nadaraya-Watson, dan estimator Gasser-Muller. Persamaan ketiga estimator yaitu (Putri dkk., 2022):

1. Estimator Nadaraya-Watson

$$\hat{m}(x) = \frac{\sum_{i=1}^n K\left(\frac{x_j - x_i}{h}\right) y_i}{\sum_{i=1}^n K\left(\frac{x_j - x_i}{h}\right)} \quad (5)$$

2. Estimator Priestley-Chao

$$\hat{m}(x) = \frac{1}{h} \sum_{i=1}^n (x_j - x_{i-1}) y_i K\left(\frac{x_j - x_i}{h}\right) \quad (6)$$

3. Estimator Gasser-Muller

$$\hat{m}(x) = \frac{1}{h} \sum_{i=1}^n y_i \int_{S_{i-1}}^{S_i} K\left(\frac{x_j - x_i}{h}\right) dx \quad (7)$$

dengan $S_0 = 0, S_n = n$ dan $S_i = \left(\frac{x_i + x_{i+1}}{2}\right)$ untuk $i = 1, 2, \dots, n - 1$. Keterangan untuk ketiga estimator yaitu:

- $\hat{m}(x)$: fungsi estimator *kernel*,
- y_i : variabel respon pada pengamatan ke- i ,
- x_j : variabel prediktor pada pengamatan ke- j ,
- x_i : variabel prediktor pada pengamatan ke- i ,
- K : fungsi *kernel*,
- n : ukuran sampel,
- h : *bandwidth*.

Metode *kernel* umumnya digunakan dalam mengestimasi kurva regresi nonparametrik karena memiliki kemampuan fleksibilitas yang tinggi dengan perhitungan yang relatif mudah. Metode estimasi *kernel* merupakan pengembangan dari metode histogram yang pertama kali diperkenalkan oleh Rosenblatt (1956) dan Parzen (1962). Karena kontribusi keduanya, penaksir ini juga dikenal sebagai penaksir densitas *kernel* Rosenblatt-Parzen (Yulianti & Sihombing, 2020). Dalam pendekatan nonparametrik, estimator *kernel* sangat umum digunakan karena memiliki beberapa keunggulan, yaitu (Herawati dkk., 2022):

1. Memiliki bentuk yang fleksibel dan mudah diolah secara matematis
2. Memiliki rata-rata konvergensi yang cukup cepat.

secara umum, fungsi kerapatan *kernel* yaitu:

$$\hat{f}_h(x) = \frac{1}{n} \sum_{i=1}^n K_h(x_j - x_i) \quad (8)$$

kemudian persamaan *kernel* K dengan *bandwidth* (h) didefinisikan sebagai berikut:

$$K_h(x_j - x_i) = \frac{1}{h} K\left(\frac{x_j - x_i}{h}\right) \quad (9)$$

untuk $-\infty < x < \infty$, $i, j = 1, 2, \dots, n$, dan $h > 0$. Sehingga fungsi kerapatan *kernel* $\hat{f}_h(x)$ dapat dituliskan sebagai berikut:

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x_j - x_i}{h}\right) \quad (10)$$

dengan K merupakan fungsi *kernel* kontinu. Estimasi sangat bergantung pada pemilihan fungsi *kernel* dan parameter yang digunakan. Fungsi *kernel* memengaruhi bentuk bobot *kernel*, sementara parameter digunakan untuk menentukan ukuran bobotnya. Parameter pemulus yang digunakan disebut dengan *bandwidth*. *Bandwidth* sebagai parameter dalam penajaman atau penyesuaian fungsi *kernel* berperan penting dalam mendefinisikan dan menentukan variasi serta bias dalam estimasi. Secara grafis, suatu fungsi dianggap kontinu pada interval $x \in [a, b]$ jika grafik fungsi f tidak terputus pada titik (a, b) . Fungsi kernel atau K dalam Persamaan 7 diasumsikan memenuhi kondisi regularitas tertentu sebagai berikut:

1. $K(x) \geq 0$, untuk semua x

2. $\int_{-\infty}^{\infty} K(x)dx = 1$
3. $\int_{-\infty}^{\infty} x^2 K(x)dx = \sigma^2 > 0$
4. $\int_{-\infty}^{\infty} x_i K(x)dx = \begin{cases} 1, i = 0 \\ 0, 1 \leq i < r \\ \neq 0, i = r \end{cases}$, dengan r adalah derajat momen
5. $K(x)$ simetris terhadap nol.

1.4.5 Fungsi Kernel

Fungsi *kernel* dalam regresi *kernel* harus memenuhi beberapa syarat. Fungsi tersebut harus kontinu, simetris, terbatas, dan memiliki nilai nyata atau nilai riil. Fungsi umum dari *kernel* K dengan parameter penghalus *bandwith* didefinisikan sebagai berikut (Herawati dkk., 2022):

$$K_h(x_j - x_i) = \frac{1}{h} K\left(\frac{x_j - x_i}{h}\right), \text{ untuk } -\infty < x < \infty \text{ dan } h > 0 \quad (11)$$

dengan

K : fungsi *kernel*,

h : *bandwidth*,

x_j : nilai variabel prediktor pada pengamatan data ke- j ,

x_i : nilai variabel prediktor pada pengamatan data ke- i .

terdapat beberapa tipe dari fungsi *kernel*, yaitu sebagai berikut:

1. *Uniform* : $K(u) = \frac{1}{2}I$, : $|u| \leq 1$
2. *Epanechnikov* : $K(u) = \frac{3}{4}(1 - u^2)I$, : $|u| \leq 1$
3. *Biweight* (kuadratik) : $K(u) = \frac{15}{16}(1 - u^2)^3I$, : $|u| \leq 1$
4. *Triweight* : $K(u) = \frac{35}{32}(1 - u^2)^3I$, : $|u| \leq 1$
5. *Cosinus* : $K(u) = \frac{\pi}{4} \cos\left(\frac{\pi}{2}u\right)$, : $|u| \leq 1$
6. *Gaussian* : $K(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(u^2)\right)$, : $-\infty < u < \infty$
7. *Triangular* : $K(u) = (1 - |u|)I$, : $|u| \leq 1$

dengan I merupakan sebuah fungsi indikator:

$$I(u) = \begin{cases} 1, \text{ jika } |u| \leq 1 \\ 0, \text{ jika } |u| > 1 \end{cases}$$

dengan

$$u = \frac{x_j - x_i}{h}$$

penelitian ini menggunakan fungsi *kernel Gaussian*. Bentuk umum dari fungsi *kernel Gaussian* didefinisikan sebagai berikut:

$$K(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(u^2)\right) \quad (12)$$

fungsi *kernel Gaussian* adalah fungsi *kernel* yang sering digunakan dalam penelitian, karena memiliki sifat yang fleksibel dan *smooth* (Dani, 2021)

1.4.6 Estimator Gasser-Müller

Persamaan umum untuk estimator *kernel* dapat ditulis sebagai berikut

$$\hat{m}(x) = \sum_{i=1}^n W_i(x_{ij}, h) y_i \quad (13)$$

dengan $W_i(x_{ij}, h)$ merupakan fungsi berbobot atau *weighted function* yang nilainya bergantung pada h, i, x dan K (Pomenkova & Klejmova, 2016). Fungsi regresi $\hat{m}(x)$ yang tidak diketahui dapat diestimasi menggunakan estimator Gasser-Müller dengan *kernel* K berperan sebagai fungsi densitas probabilitas atau alat untuk mengukur seberapa sering nilai-nilai data muncul, sekaligus menentukan *bandwidth* yang mengontrol seberapa lebar atau sempitnya rentang data yang sedang dianalisis. Estimator Gasser-Müller didefinisikan sebagai berikut :

$$\hat{m}(x_j) = \sum_{i=1}^n \left[\int_{S_{i-1}}^{S_i} K_h(x_j - x_i) dx_i \right] y_i \quad (14)$$

untuk $x \in [0,1]$, dengan $s_0 = 0, s_n = 1, x_i \leq S_i \leq x_{i+1}$, dan $S_i = \frac{x_i + x_{i+1}}{2}$ untuk $i = 1, 2, \dots, n-1$. Pada Persamaan 11, fungsi *kernel* dapat dituliskan $K_h(x_j - x_i) = \frac{1}{h} K\left(\frac{x_j - x_i}{h}\right)$, maka estimator Gasser-Müller dapat didefinisikan sebagai berikut (Herawati dkk., 2022):

$$\hat{m}^{GM}(x) = \frac{1}{h} \sum_{i=1}^n \left[\int_{S_{i-1}}^{S_i} K_h\left(\frac{x_j - x_i}{h}\right) dx_i \right] y_i \quad (15)$$

dengan

$\hat{m}^{GM}(x)$: fungsi estimator Gasser-Müller,

y_i : variabel respon pada pengamatan ke- i ,

x_j : variabel prediktor pada pengamatan ke- j ,

x_i : variabel prediktor pada pengamatan ke- i ,

K : fungsi *kernel*,

n : ukuran sampel,

h : *bandwidth*.

1.4.7 Bandwidth

Bandwidth dalam regresi nonparametrik *kernel* merupakan suatu parameter pemulus yang berfungsi untuk mengontrol kemulusan dari kurva yang diestimasi. *Bandwidth* yang terlalu kecil akan menghasilkan kurva yang *undersmoothing* yaitu sangat kasar, dan sebaliknya *bandwidth* yang terlalu lebar akan menghasilkan kurva yang *oversmoothing* yaitu sangat mulus tetapi tidak sesuai dengan pola data. Oleh karena itu, perlu dipilih *bandwidth* optimal untuk mendapatkan kurva yang optimal (Pratiwi dkk., 2020).

Pentingnya pemilihan *bandwidth* dalam regresi *Kernel* melebihi pemilihan fungsi *Kernel* itu sendiri karena *bandwidth* yang optimal harus seimbang antara tingkat bias dan variansi, yang dapat diukur melalui *Mean Square Error* (MSE). Bias mengukur seberapa jauh rata-rata nilai estimasi model dari nilai aktualnya, sehingga

model dengan bias rendah cenderung *overfit*. Sementara itu, variansi mengukur seberapa sensitif model terhadap variasi data, sehingga model cenderung *overfit* jika variansi tinggi.

a. *Generalized Cross Validation*

Dalam penelitian ini, digunakan teknik *Generalized Cross Validation* (GCV) untuk memilih *bandwidth* optimal. Metode GCV memiliki beberapa keunggulan dibandingkan dengan metode lain, seperti *Cross Validation* (CV) dan *Metode Unbiased Risk* (UBR). Secara teoritis, metode GCV memiliki sifat optimal secara asimtotik, persamaannya tidak mengandung varians σ , serta bersifat invarian terhadap transformasi (Dani dkk., 2023). Fungsi GCV ditunjukkan pada Persamaan (13) (Karimuse dkk., 2023):

$$GCV = \frac{MSE}{\left(1 - \frac{\text{tr}(\mathbf{W})}{n}\right)^2} \quad (16)$$

Nilai MSE didefinisikan sebagai berikut (Putri dkk., 2022):

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (17)$$

dengan

n : jumlah data,

y_i : data aktual atau data variabel respon pada pengamatan ke- i ,

\hat{y}_i : data hasil estimasi,

\mathbf{W} : matriks *smoothing* atau matriks hat ($n \times n$) dari $W_i(x_{ij}, h)$.

b. *Silverman's Rule of Thumb*

Silverman (1986) memberikan aturan untuk menentukan *bandwidth* optimal yang dikenal sebagai *Silverman's Rule of Thumb* (SROT). Metode ini merupakan pengembangan dari metode *Normal Rule of Thumb* (NROT). Silverman telah melakukan studi simulasi kecil dengan ukuran sampel 100 dan menemukan bahwa SROT memiliki kinerja yang baik pada densitas dengan *skewness* dan lebih tahan terhadap *outlier* (Harpole dkk., 2014). Metode SROT ditunjukkan pada Persamaan (15).

$$h_{SROT} = 0,9 \min \left\{ s, \frac{IQR}{1,34} \right\} n^{-\frac{1}{5}} \quad (18)$$

dengan

n : jumlah data,

s : standar deviasi,

IQR : *Inter Quartil Range* ($Q_3 - Q_1$).

c. *Scott's Rule*

Scott's Rule merupakan metode penentuan *bandwidth* optimal yang diperkenalkan oleh David W. Scott. Metode *Scott's Rule* memanfaatkan perhitungan yang relatif lebih sederhana dan efisien karena tidak memerlukan proses komputasi yang kompleks (Fuentes-Santos dkk., 2016). Metode ini sangat sesuai untuk diterapkan

dalam situasi di mana kecepatan atau efisiensi menjadi faktor utama, atau ketika menangani data dalam jumlah besar. Metode *Scott's Rule* dihitung dengan persamaan sebagai berikut (Yongjin dkk., 2020):

$$h = 1,059sn^{-\frac{1}{5}} \quad (19)$$

dengan

n : jumlah data,
 s : standar deviasi.

1.4.8 Mean Squared Error

Mean Squared Error (MSE) merupakan metode perhitungan *error* dengan mengkuadratkan nilai *error* pada tiap titik data. Nilai *error* diperoleh dari nilai selisih antara data aktual dan data hasil estimasi. MSE dapat dihitung dengan rumus sebagai berikut (Ngestisari dkk., 2020):

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (20)$$

dengan

n : jumlah data,
 y_i : data aktual atau data variabel respon pada pengamatan ke- i ,
 \hat{y}_i : data hasil estimasi.

1.4.9 Curah Hujan

Kota Makassar yang merupakan salah satu kota di Indonesia berada di wilayah yang dilalui oleh garis khatulistiwa, sehingga memiliki pola iklim yang sama sepanjang tahun. Iklim di wilayah ini didominasi oleh dua musim utama yaitu musim hujan dan musim kemarau. pada musim hujan, suhu udara cenderung hangat dengan kelembapan yang tinggi, sementara musim kemarau ditandai oleh kondisi yang lebih kering. Presipitasi adalah turunnya air dari atmosfer ke permukaan bumi yang berupa hujan, salju, embun, dan sebagainya. Indonesia termasuk daerah tropis sehingga jenis presipitasi yang dominan terjadi adalah hujan (Hidayat & Empung, 2016).

Curah hujan merupakan ketinggian air hujan yang terkumpul dalam tempat yang datar, tidak menguap, tidak meresap, dan tidak mengalir. Satuan curah hujan selalu dinyatakan dalam satuan milimeter atau inci, namun di Indonesia satuan curah hujan yang digunakan adalah satuan milimeter (mm). Curah hujan dalam 1 (satu) milimeter memiliki arti dalam luasan satu meter persegi pada tempat yang datar, tertampung air setinggi satu milimeter atau sebanyak satu liter.

Intensitas curah hujan adalah jumlah curah hujan dalam suatu satuan waktu tertentu, yang biasanya dinyatakan dalam mm/jam, mm/hari, mm/tahun, dan sebagainya. Data yang sering digunakan untuk proses analisis adalah nilai maksimum, minimum dan nilai rata-rata (Ajr & Dwirani, 2019). Berdasarkan BMKG Kota Makassar, kriteria curah hujan Kota Makassar dikategorikan sangat lebat. Secara geomorfologi, Makassar merupakan daerah resapan dengan kerucut gunung api yang mengelilingi dan memanjang di sepanjang jalur utara-selatan melewati

puncak Gunung Lompobattang, sehingga daerah Makassar mempunyai potensi air tanah yang besar (Bongi dkk., 2020).

Secara umum, curah hujan dipengaruhi oleh arah dan kecepatan angin. Angin merupakan pergerakan massa udara secara horizontal yang terjadi akibat perbedaan tekanan antara dua wilayah. Angin bergerak dari daerah dengan tekanan tinggi menuju daerah dengan tekanan rendah. Terdapat 16 arah atau penjuror angin yaitu, Utara (360°), Utara Timur Laut ($22,5^\circ$), Timur Laut (45°), Timur Timur Laut ($67,5^\circ$), Timur (90°), Timur Tenggara ($112,5^\circ$), Tenggara (135°), Selatan Tenggara ($157,5^\circ$), Selatan (180°), Selatan Barat Daya ($202,5^\circ$), Barat Daya (225°), Barat Barat Daya ($247,5^\circ$), Barat (270°), Barat Barat Laut ($292,5^\circ$), Barat Laut (315°), dan Utara Barat Laut ($337,5^\circ$). Ketika angin berhembus dari arah Samudra Indonesia atau Samudra Hindia, maka angin tersebut akan membawa udara lembab yang berpotensi menyebabkan curah hujan yang tinggi (Simbolon dkk., 2022).

BAB II METODE PENELITIAN

2.1 Sumber Data

Data yang digunakan dalam penelitian ini adalah data sekunder yakni data curah hujan di Kota Makassar dari bulan Januari tahun 2018 hingga bulan Desember tahun 2023. Data tersebut diperoleh melalui situs *website* <https://dataonline.bmkg.go.id/> dan akan diolah dengan menggunakan metode *kernel* nonparametrik dengan estimator Gasser-Müller. Adapun jumlah observasi yang digunakan pada penelitian ini adalah sebanyak 72 data atau 72 bulan. Variabel yang digunakan pada penelitian ini adalah variabel respon yang terdiri dari data curah hujan (mm) di Kota Makassar dan variabel prediktor yang terdiri dari data periode (bulan), kecepatan angin maksimum (m/s), dan arah angin saat kecepatan maksimum ($^{\circ}$). Adapun definisi operasional variabel yang digunakan dalam penelitian ini dapat dilihat pada Tabel 1.

Tabel 1. Variabel respon dan prediktor

Variabel	Keterangan	Satuan	Definisi Operasional
Y	Curah hujan	mm	Curah hujan bulanan di Kota Makassar pada bulan Januari 2018 hingga bulan Desember 2023
X_1	Periode	Bulan	Periode curah hujan di Kota Makassar dari bulan Januari 2018 hingga bulan Desember 2023
X_2	Kecepatan angin maksimum	m/s	Kecepatan angin maksimum bulanan di Kota Makassar pada bulan Januari 2018 hingga bulan Desember 2023
X_3	Arah angin saat kecepatan maksimum	$^{\circ}$	Arah angin saat kecepatan maksimum bulanan di Kota Makassar pada bulan Januari 2018 hingga bulan Desember 2023

2.2 Tahapan Penelitian

Penelitian ini menerapkan pendekatan estimator Gasser-Müller dengan menggunakan fungsi *kernel Gaussian*. *Bandwidth* ditentukan dengan mengacu pada nilai optimum *Generalized Cross Validation (GCV)*, *Silverman's Rule of Thumb*

(SROT), dan *Scott's Rule*. Pemrosesan data dilakukan menggunakan *software Microsoft Excel* dan *Python*. Langkah-langkah yang dilakukan adalah sebagai berikut:

1. Mengumpulkan data curah hujan di Kota Makassar pada tahun 2018-2023.
2. Melakukan analisis statistik deskriptif untuk memahami karakteristik data yang akan digunakan.
3. Melakukan standarisasi data agar data yang digunakan memiliki skala data yang sama dengan menggunakan Persamaan 3.

$$x_i^* = \frac{x_i - \bar{x}}{s} \quad (3)$$

4. Membuat *scatter plot* antara variabel respon dan variabel prediktor untuk mengidentifikasi dan memahami pola data serta hubungan antara variabel respon dan prediktor.
5. Membentuk estimasi model regresi nonparametrik *kernel* menggunakan estimator Gasser-Müller dan fungsi *kernel Gaussian* dengan tahapan sebagai berikut:
 - a. Mendefinisikan persamaan regresi nonparametrik *kernel* dengan estimator Gasser-Müller sesuai Persamaan 15.

$$\hat{m}^{GM}(x) = \frac{1}{h} \sum_{i=1}^n \left[\int_{S_{i-1}}^{S_i} K_h \left(\frac{x_j - x_i}{h} \right) dx_i \right] y_i \quad (15)$$

- b. Mendefinisikan fungsi *kernel Gaussian* sesuai persamaan 12.

$$K(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(u^2)\right) \quad (12)$$

- c. Membuat estimasi model regresi nonparametrik dengan menyubstitusi fungsi *kernel Gaussian* ke dalam persamaan regresi nonparametrik *kernel* dengan estimator Gasser-Müller.
 - d. Menghitung nilai bandwidth optimum dengan menggunakan kriteria GCV, SROT, dan *Scott's Rule* yang kemudian digunakan pada persamaan model regresi nonparametrik menggunakan persamaan 16, 18, dan 19.

$$GCV = \frac{MSE}{\left(1 - \frac{tr(W)}{n}\right)^2} \quad (16)$$

$$h_{SROT} = 0,9 \min \left\{ s, \frac{IQR}{1,34} \right\} n^{-\frac{1}{5}} \quad (18)$$

$$h = 1,059sn^{-\frac{1}{5}} \quad (19)$$

6. Menguji kebaikan model dengan menghitung nilai MSE.
7. Membuat plot hasil estimasi model regresi non nonparametrik dengan data asli curah hujan bulanan di Kota Makassar pada bulan Januari 2018 hingga bulan Desember 2023.