

SKRIPSI

**PERBANDINGAN ANALISIS REGRESI POISSON DAN
POISSON-INVERSE GAUSSIAN (PIG) PADA STUDI KASUS :
BIKE SHARING 2011-2012**



**STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA**

**ELLIF ZAHARA YASMIN ASADILLAH
22106010008**

**PROGRAM STUDI MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA
YOGYAKARTA**

2026

**PERBANDINGAN ANALISIS REGRESI POISSON DAN
POISSON-INVERSE GAUSSIAN (PIG) PADA STUDI KASUS :
BIKE SHARING 2011-2012**

Skripsi

Untuk memenuhi sebagian persyaratan
mencapai derajat Sarjana S-1
Program Studi Matematika



STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA

diajukan oleh

ELLIF ZAHARA YASMIN ASADILLAH

22106010008

Kepada

PROGRAM STUDI MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA
YOGYAKARTA

2026



SURAT PERSETUJUAN SKRIPSI/TUGAS AKHIR

Hal : Persetujuan Skripsi / Tugas Akhir

Lamp :

Kepada

Yth. Dekan Fakultas Sains dan Teknologi

UIN Sunan Kalijaga Yogyakarta

di Yogyakarta

Assalamu 'alaikum wr. wb.

Setelah membaca, meneliti, memberikan petunjuk dan mengoreksi serta mengadakan perbaikan seperlunya, maka kami selaku pembimbing berpendapat bahwa skripsi Saudara:

Nama : Ellif Zahara Yasmin Asadillah
NIM : 22106010008
Judul Skripsi : ANALISIS REGRESI POISSON-INVERSE GAUSSIAN (PIG)(STUDI KASUS : *BIKE SHARING* 2011-2012)

sudah dapat diajukan kembali kepada Program Studi Matematika Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta sebagai salah satu syarat untuk memperoleh gelar Sarjana Strata Satu dalam Program Studi Matematika.

Dengan ini kami berharap agar skripsi/tugas akhir Saudara tersebut di atas dapat segera dimunaqasyahkan. Atas perhatiannya kami ucapkan terima kasih.

Wassalamu 'alaikum wr. wb.

Yogyakarta, 12 Mei 2026

Pembimbing

Sri Utami Zuliana, S.Si., M.Sc., Ph.D.

NIP. 19741003 200003 2 002



PENGESAHAN TUGAS AKHIR

Nomor : B-1222/Un.02/DST/PP.00.9/06/2026

Tugas Akhir dengan judul : PERBANDINGAN ANALISIS REGRESI POISSON DAN POISSON-INVERSE GAUSSIAN (PIG) PADA STUDI KASUS : BIKE SHARING 2011-2012

yang dipersiapkan dan disusun oleh:

Nama : ELLIF ZAHARA YASMIN ASADILLAH
Nomor Induk Mahasiswa : 22106010008
Telah diujikan pada : Senin, 25 Mei 2026
Nilai ujian Tugas Akhir : A/B

dinyatakan telah diterima oleh Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta

TIM UJIAN TUGAS AKHIR



Ketua Sidang

Sri Utami Zuliana, S.Si., M.Sc., Ph.D.

SIGNED

Valid ID: 6a228a15cdf96



Penguji I

Muhamad Rashif Hilmi, S.Si., M.Sc.

SIGNED

Valid ID: 6a2287971f058



Penguji II

Aulia Khifah Futhona, M.Sc.

SIGNED

Valid ID: 6a2126fd38681



Yogyakarta, 25 Mei 2026

UIN Sunan Kalijaga

Dekan Fakultas Sains dan Teknologi

Prof. Dr. Dra. Hj. Khurul Wardati, M.Si.

SIGNED

Valid ID: 6a228caf7af99

SURAT PERNYATAAN KEASLIAN

Yang bertanda tangan dibawah ini:

Nama : Elif Zahara Yasmin Asadillah
NIM : 22106010008
Program Studi : Matematika
Fakultas : Sains dan Teknologi

Dengan ini menyatakan bahwa isi skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar sarjana di suatu Perguruan Tinggi dan sesungguhnya skripsi ini merupakan hasil pekerjaan penulis sendiri sepanjang pengetahuan penulis, bukan duplikasi atau saduran dari karya orang lain kecuali bagian tertentu yang penulis ambil sebagai bahan acuan. Apabila terbukti pernyataan ini tidak benar, sepenuhnya menjadi tanggung jawab penulis.

Yogyakarta, 13 Mei 2026



Elif Zahara Yasmin Asadillah

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

HALAMAN PERSEMBAHAN

DENGAN PENUH RASA SYUKUR KEPADA ALLAH SWT ATAS SEGALA RAHMAT, NIKMAT, DAN KEMUDAHAN YANG TELAH DIBERIKAN, KARYA SEDERHANA INI PENULIS PERSEMBAHKAN KEPADA:

KEDUA ORANG TUA TERCINTA, YANG SELALU MEMBERIKAN DOA, KASIH SAYANG, DUKUNGAN, PENGORBANAN, SERTA SEMANGAT TANPA HENTI DALAM SETIAP LANGKAH KEHIDUPAN PENULIS.

KELUARGA DAN SAUDARA-SAUDARA YANG SENANTIASA MEMBERIKAN MOTIVASI, PERHATIAN, DAN DUKUNGAN SELAMA PROSES PENYUSUNAN SKRIPSI INI.

DOSEN PEMBIMBING DAN SELURUH DOSEN YANG TELAH MEMBERIKAN ILMU, ARAHAN, SERTA BIMBINGAN KEPADA PENULIS SELAMA MASA PERKULIAHAN.

SAHABAT DAN TEMAN-TEMAN SEPERJUANGAN YANG SELALU MENEMANI, MEMBANTU, DAN MEMBERIKAN SEMANGAT DALAM SUKA MAUPUN DUKA SELAMA PROSES PENYELESAIAN SKRIPSI.

SEMUA PIHAK YANG TIDAK DAPAT DISEBUTKAN SATU PER SATU YANG TELAH MEMBERIKAN DUKUNGAN DAN DOA KEPADA PENULIS.

MOTTO

”Sesungguhnya bersama kesulitan ada kemudahan.”

(QS. Al Insyirah:6)

”Tidak semua hal berjalan sesuai keinginan, tetapi semua yang terjadi selalu membawa pelajaran dan menjadi bagian dari takdir terbaik yang telah Allah tetapkan.”



PRAKATA

Puji syukur kehadiran Allah SWT atas segala rahmat, nikmat, dan karunia-Nya yang tidak pernah berhenti mengalir kepada penulis. Atas izin-Nya, penulis dapat menyelesaikan skripsi ini hingga akhir. Shalawat serta salam semoga selalu tercurah kepada Nabi Muhammad SAW yang telah menjadi teladan terbaik dalam menuntut ilmu, bersabar, dan menjalani kehidupan.

Penyusunan skripsi ini menjadi salah satu perjalanan yang penuh proses bagi penulis. Tidak hanya tentang menyelesaikan penelitian, tetapi juga tentang belajar bertahan, menerima banyak hal, dan memahami bahwa setiap proses memiliki waktunya sendiri. Dalam proses tersebut, penulis menyadari bahwa skripsi ini tidak akan selesai tanpa bantuan, dukungan, doa, dan kehadiran banyak pihak.

Oleh karena itu, dengan penuh rasa hormat dan terima kasih, penulis ingin menyampaikan apresiasi kepada:

1. Allah SWT yang selalu memberikan kekuatan, kemudahan, dan jalan terbaik kepada penulis hingga mampu berada di titik ini.
2. Kedua orang tua dan keluarga penulis yang tidak pernah berhenti memberikan doa, kasih sayang, dukungan, dan semangat dalam setiap langkah penulis.
3. Ibu Dr. Khurul Wardati, M.Si., selaku Dekan Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta.
4. Ibu Sri Utami Zuliana, S.Si., M.Sc., Ph.D., selaku dosen pembimbing yang telah membimbing penulis dengan penuh kesabaran, memberikan banyak arahan, masukan, serta motivasi selama proses penyusunan skripsi ini.

5. Titi Prihartati, yang telah banyak menemani, mendengarkan, memberikan dukungan, dan menjadi tempat berbagi cerita selama proses perkuliahan hingga penyusunan skripsi ini.
6. Teman-teman seperbimbingan, terutama Dea Iswari dan Robit Ma'rufi Sabila, yang telah berjuang bersama, saling membantu, saling mengingatkan, dan menemani proses bimbingan skripsi hingga akhir.
7. Teman-teman KKN yang tetap menjaga kedekatan dan kebersamaan meskipun masa KKN telah usai, serta selalu memberikan dukungan dan semangat kepada penulis.
8. Teman-teman Matematika angkatan 2022 yang telah menjadi bagian dari perjalanan perkuliahan penulis dengan banyak cerita, pengalaman, dan kebersamaan yang berharga.
9. Semua pihak yang tidak dapat penulis sebutkan satu per satu yang telah membantu dan mendukung penulis dalam proses penyusunan skripsi ini.

Penulis menyadari bahwa skripsi ini masih jauh dari kata sempurna. Oleh karena itu, penulis dengan terbuka menerima segala kritik dan saran yang membangun demi perbaikan di masa mendatang. Semoga skripsi ini dapat memberikan manfaat bagi pembaca dan menjadi salah satu bentuk kontribusi kecil dalam pengembangan ilmu pengetahuan.

Yogyakarta, 18 Mei 2026

Ellif Zahara Yasmin Asadillah

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN	iv
HALAMAN PERSEMBAHAN	v
HALAMAN MOTTO	vi
PRAKATA	vii
DAFTAR ISI	ix
DAFTAR TABEL	xii
DAFTAR GAMBAR	xiii
DAFTAR LAMBANG	xiv
INTISARI	xv
ABSTRACT	xvi
I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	5
1.3. Batasan Masalah	5
1.4. Tujuan Penelitian	6
1.5. Manfaat Penelitian	6
1.6. Tinjauan Pustaka	6
1.7. Sistematika Penulisan	8
II DASAR TEORI	10
2.1. Distribusi Normal	11
2.2. Analisis Regresi	12

2.3. Regresi Poisson	13
2.4. Estimasi Parameter Regresi Poisson	17
2.5. Overdispersi	20
2.6. Distribusi Inverse Gaussian	21
2.7. Distribusi Poisson–Inverse Gaussian	23
2.8. Regresi Poisson-Inverse Gaussian	28
2.9. Estimasi Parameter Regresi Poisson Inverse Gaussian	30
2.10. Uji Hipotesis Parameter Regresi Poisson dan Poisson Inverse Gaussian	32
2.11. Pemilihan Model Terbaik	34
2.12. Faktor-Faktor yang Mempengaruhi Jumlah Penyewaan Sepeda	35
III METODE PENELITIAN	38
3.1. Pendekatan Penelitian	38
3.2. Variabel Data Penelitian	38
3.3. Langkah-langkah Analisis	39
3.4. Diagram Analisis Data (<i>Flowchart</i>)	39
IV HASIL DAN PEMBAHASAN	41
4.1. Analisis Deskriptif	41
4.2. Pemodelan Regresi Poisson	42
4.3. Uji Overdispersi	45
4.4. Pemodelan Regresi Poisson-inverse Gaussian	46
4.5. Pemilihan Model Terbaik	50
4.6. Interpretasi Model	50
4.7. Simulasi Prediksi <i>Bike Sharing</i>	51
4.7.1. Prediksi pada Observasi ke-278	52
V PENUTUP	54

	xi
5.1. Kesimpulan	54
5.2. Saran	55
DAFTAR PUSTAKA	56
LAMPIRAN	59
A Data Penelitian	59
B Tabel Tinjauan Pustaka	60
C Source Code Program R	63
D Biodata Penulis	73



DAFTAR TABEL

3.1	Variabel Penelitian	38
4.1	Descriptive Statistics	41
4.2	Hasil Estimasi Parameter Regresi Poisson	42
4.3	Hasil Estimasi Parameter Regresi Poisson	44
4.4	Hasil Estimasi Parameter μ Regresi Poisson Inverse Gaussian	46
4.5	Hasil Estimasi Parameter Model Regresi Poisson–Inverse Gaussian	48
4.6	Perbandingan Nilai Akaike Information Criterion (AIC)	50
1.1	Ringkasan Variabel Utama	59
2.1	Tinjauan Pustaka	60



DAFTAR GAMBAR

2.1	Kurva Distribusi Normal	12
3.1	Flowchart Penelitian	40



DAFTAR LAMBANG

y_i	= variabel respon ke- i
μ_i	= nilai harapan (mean) dari y_i
σ^2	= variansi
λ	= parameter dispersi
β	= vektor parameter regresi
β_0	= intercept model regresi
β_j	= koefisien regresi ke- j
x_i	= vektor variabel independen ke- i
x_{ij}	= variabel independen ke- j pada observasi ke- i
n	= jumlah observasi
$L(\cdot)$	= fungsi likelihood
$\ell(\cdot)$	= fungsi log-likelihood
$\exp(\cdot)$	= fungsi eksponensial
$\ln(\cdot)$	= logaritma natural
AIC	= Akaike Information Criterion
X_1	= suhu yang dirasakan
X_2	= suhu aktual
X_3	= kelembaban udara
X_4	= kecepatan angin
Y	= jumlah penyewaan sepeda

INTISARI

PERBANDINGAN ANALISIS REGRESI POISSON DAN POISSON-INVERSE GAUSSIAN (PIG) PADA STUDI KASUS : BIKE SHARING 2011-2012

Ellif Zahara Yasmin Asadillah

22106010008

Jumlah penyewaan sepeda merupakan *count data* yang dapat dimodelkan menggunakan regresi Poisson. Namun, regresi Poisson mensyaratkan asumsi equidispersi, yaitu nilai rata-rata sama dengan variansi. Pada praktiknya, *count data* sering mengalami overdispersi sehingga asumsi tersebut tidak terpenuhi dan dapat menyebabkan estimasi parameter menjadi kurang efisien. Salah satu metode yang dapat digunakan untuk mengatasi permasalahan tersebut adalah regresi Poisson–Inverse Gaussian yang merupakan pengembangan dari regresi Poisson dengan memasukkan parameter dispersi.

Penelitian ini bertujuan untuk memperoleh model regresi Poisson dan regresi Poisson–Inverse Gaussian (PIG) dalam memodelkan jumlah penyewaan sepeda pada data *Bike Sharing* tahun 2011–2012 serta menentukan model terbaik berdasarkan nilai *Akaike Information Criterion* (AIC). Data yang digunakan merupakan *Bike Sharing Dataset* yang diperoleh dari *UCI Machine Learning Repository*. Estimasi parameter dilakukan menggunakan metode *Maximum Likelihood Estimation* (MLE) dan pengolahan data dilakukan dengan perangkat lunak R versi 4.5.2.

Dari hasil analisis diperoleh model regresi Poissonnya adalah $\ln(\mu_i) = 8,192982 + 1,677465X_1 - 0,672520X_3 - 1,034128X_4$ dan model regresi Poisson Inverse Gaussiannya diperoleh $\mu_i = \exp(8,3398 + 2,2177X_1 - 1,0864X_3 - 1,8261X_4)$. Berdasarkan perbandingan nilai AIC, model regresi Poisson–Inverse Gaussian menghasilkan nilai AIC yang lebih kecil dibandingkan model regresi Poisson. Oleh karena itu, regresi Poisson–Inverse Gaussian dipilih sebagai model terbaik untuk memodelkan jumlah penyewaan sepeda pada data *Bike Sharing* tahun 2011–2012.

Kata kunci: *Bike Sharing*, *count data*, overdispersi, regresi Poisson, regresi Poisson–Inverse Gaussian.

ABSTRACT

COMPARISON OF POISSON REGRESSION AND POISSON-INVERSE GAUSSIAN (PIG) ANALYSIS IN A CASE STUDY: BIKE SHARING

2011-2012

Ellif Zahara Yasmin Asadillah

22106010008

Bike rental counts are a form of count data that can be modeled using Poisson regression. However, Poisson regression assumes equidispersion, meaning that the mean and variance of the response variable are equal. In practice, count data often exhibit overdispersion, causing this assumption to be violated and resulting in inefficient parameter estimates. One method that can be used to address this issue is Poisson–Inverse Gaussian (PIG) regression, which extends Poisson regression by incorporating a dispersion parameter.

This study aims to obtain Poisson regression and Poisson–Inverse Gaussian (PIG) regression models for modeling bike rental counts in the 2011–2012 Bike Sharing dataset and to determine the best model based on the Akaike Information Criterion (AIC). The data used in this study were obtained from the Bike Sharing Dataset available in the UCI Machine Learning Repository. Parameter estimation was carried out using the Maximum Likelihood Estimation (MLE) method, and the analysis was conducted using R software version 4.5.2.

Based on the analysis results, the Poisson regression model obtained is $\ln(\mu_i) = 8.192982 + 1.677465X_1 - 0.672520X_3 - 1.034128X_4$, while the Poisson–Inverse Gaussian regression model is given by $\mu_i = \exp(8.3398 + 2.2177X_1 - 1.0864X_3 - 1.8261X_4)$. Based on the comparison of the Akaike Information Criterion (AIC) values, the Poisson–Inverse Gaussian regression model yields a lower AIC value than the Poisson regression model. Therefore, the Poisson–Inverse Gaussian regression model is selected as the best model for describing the number of bicycle rentals in the *Bike Sharing* dataset for the years 2011–2012.

Keywords: Bike Sharing, count data, overdispersion, Poisson regression, Poisson–Inverse Gaussian regression.

BAB I

PENDAHULUAN

1.1. Latar Belakang

Regresi linier adalah salah satu metode analisis yang paling sering digunakan dalam pengolahan data. Ini adalah teknik statistik yang digunakan untuk membangun model yang menggambarkan hubungan antara variabel dependen dan satu atau lebih variabel independen (Montgomery et al., 2021). Secara umum, analisis regresi digunakan untuk menganalisis hubungan dalam data di mana variabel dependen bersifat acak dan kontinu. Namun, terdapat kasus di mana variabel dependen yang dianalisis merupakan variabel acak yang berupa *count data*. Untuk situasi semacam itu, digunakan regresi Poisson, karena metode ini dirancang khusus untuk data dengan variabel dependen yang berupa *count data* (Wibawa et al., 2023).

Dalam statistika terapan, variabel respon yang dianalisis tidak selalu berbentuk data kontinu, namun dapat berupa *count data*, yaitu data diskrit yang menyatakan jumlah kejadian dalam suatu unit pengamatan dan bernilai bilangan bulat tak negatif. *Count data* memiliki distribusi yang tidak simetris dan cenderung menceng ke kanan, sehingga secara teoretis tidak sesuai dimodelkan menggunakan regresi linear klasik yang mengasumsikan galat berdistribusi normal dan variansi konstan. Penerapan model regresi yang tidak selaras dengan sifat dasar data dapat menghasilkan estimasi parameter dan inferensi statistik yang bias, sehingga diperlukan pendekatan regresi yang secara khusus dirancang untuk menangani *count data*.

Sebagai konsekuensi dari sifat diskrit dan ketidaksimetrian *count data*, pendekatan regresi klasik berbasis asumsi normalitas tidak lagi memadai untuk melakukan pemodelan dan inferensi statistik. Oleh karena itu, dalam statistika matematis dikembangkan model regresi yang secara eksplisit mendasarkan diri pada distribusi probabilitas diskrit, sehingga struktur peluang dari variabel respons dapat dimodelkan secara tepat. Pendekatan ini kemudian melahirkan berbagai model regresi berbasis distribusi diskrit, dengan distribusi Poisson sebagai fondasi utama dalam pemodelan *count data* dan titik awal pengembangan regresi Poisson sebagai model regresi paling dasar untuk *count data* (Hilbe, 2014).

Regresi Poisson adalah salah satu model statistik yang digunakan untuk menganalisis *count data*. Model ini mengasumsikan bahwa variabel respons mengikuti distribusi Poisson, yang ditandai dengan asumsi dispersi yang sama. Seperti namanya, distribusi Poisson diperkenalkan oleh Siméon Denis Poisson, seorang matematikawan Prancis yang pada abad ke-19 menggunakan regresi Poisson untuk menggambarkan jumlah kasus kecelakaan di Prancis. Dalam studi yang dilakukan oleh (Wibawa et al., 2023), penelitian ini menganalisis faktor-faktor yang mempengaruhi kasus hipertensi di Kota Kendari pada tahun 2019 menggunakan model regresi Poisson. Hasil penelitian menyimpulkan bahwa regresi Poisson merupakan metode yang tepat untuk menganalisis dan memodelkan *count data*.

Namun, pada kenyataannya *count data* seringkali menunjukkan variansi yang lebih tinggi daripada rata-rata (overdispersi). Menggunakan regresi Poisson ketika data mengalami overdispersi dapat menyebabkan kesalahan standar dari estimasi parameter menjadi terlalu kecil (*underestimated*) (Salamah et al., 2022). Untuk mengatasi masalah ini, beberapa model alternatif telah dikembangkan, salah

satunya adalah Regresi Poisson-Inverse Gaussian. Dalam studi yang dilakukan oleh (Salamah et al., 2022) didapatkan kesimpulan bahwa model regresi Poisson–Inverse Gaussian memberikan hasil analitis yang lebih baik untuk *count data* yang mengalami overdispersi dibandingkan dengan model regresi Poisson. Selain itu, model Poisson–Inverse Gaussian mampu mengidentifikasi variabel prediktor yang benar-benar memiliki pengaruh signifikan terhadap model.

Distribusi Poisson-Inverse Gaussian merupakan anggota dari keluarga *mixed Poisson distributions*, yaitu distribusi yang diperoleh dengan mencampurkan distribusi Poisson dengan distribusi kontinu tertentu pada parameter intensitasnya. Secara khusus, pada distribusi Poisson Inverse Gaussian, parameter intensitas Poisson diasumsikan mengikuti distribusi Inverse Gaussian (Hilbe, 2014). Distribusi Poisson Inverse Gaussian memiliki sifat yang menarik karena mempertahankan struktur diskrit dari *count data*, namun dengan fleksibilitas variansi yang tidak lagi terikat secara ketat pada nilai harapan. Oleh karena itu, distribusi ini sering digunakan sebagai alternatif distribusi Poisson ketika data menunjukkan overdispersi.

Berdasarkan distribusi Poisson Inverse Gaussian, dikembangkan model regresi Poisson Inverse Gaussian sebagai perluasan dari regresi Poisson. Dalam model ini, variabel respons diasumsikan mengikuti distribusi Poisson Inverse Gaussian. Pendekatan ini memungkinkan pemodelan hubungan kovariat terhadap rata-rata respons sekaligus mengakomodasi variansi yang lebih besar daripada nilai harapan. Hilbe menjelaskan bahwa regresi Poisson Inverse Gaussian sangat sesuai digunakan ketika *count data* menunjukkan tingkat overdispersi yang tinggi dan pola variasi yang tidak dapat dijelaskan hanya melalui regresi Poisson atau regresi binomial negatif. Model ini memberikan estimasi parameter yang lebih stabil

dibandingkan regresi Poisson ketika asumsi equidispersion dilanggar (Hilbe, 2014).

Model regresi untuk *count data*, seperti regresi Poisson dan Poisson Inverse Gaussian, telah banyak diaplikasikan dalam berbagai bidang. Pada penelitian ini, data yang digunakan adalah data *Bike Sharing* tahun 2011–2012 yang merepresentasikan jumlah penyewaan sepeda dalam suatu periode waktu tertentu. Data layanan penyewaan sepeda, seperti dataset *Bike Sharing* (2011–2012), merupakan *count data*. Polusi udara merupakan salah satu permasalahan lingkungan yang berdampak serius terhadap kesehatan masyarakat. Menurut Portal Sehat Indonesia yang dikelola oleh Kementerian Kesehatan Republik Indonesia, polusi udara berkontribusi terhadap berbagai penyakit seperti gangguan pernapasan, penyakit kardiovaskular, hingga kanker paru-paru, serta meningkatkan risiko kematian dini (Kementerian Kesehatan Republik Indonesia, 2024). Sumber polusi udara berasal dari berbagai aktivitas manusia, seperti emisi kendaraan bermotor, aktivitas industri, pembangkit listrik berbahan bakar fosil, hingga kebakaran hutan dan aktivitas domestik. Kondisi ini menunjukkan bahwa faktor lingkungan dan aktivitas manusia memiliki peran penting dalam memengaruhi kualitas udara dan tingkat paparan polutan. Oleh karena itu, diperlukan suatu pendekatan analisis data yang mampu memodelkan hubungan antara berbagai faktor tersebut dengan tingkat kejadian atau intensitas tertentu yang diamati.

Dalam penelitian ini digunakan data *Bike Sharing* 2011–2012 yang memuat informasi jumlah penyewaan sepeda serta faktor-faktor yang memengaruhinya. Data ini dipilih karena memiliki karakteristik yang serupa dengan fenomena lingkungan, yaitu adanya variabilitas tinggi yang dipengaruhi oleh banyak faktor eksternal. Selain itu, data jumlah penyewaan sepeda merupakan data diskrit non-negatif yang secara alami sesuai dimodelkan menggunakan pendekatan regresi

untuk *count data*.

Namun demikian, dalam praktiknya *count data* sering kali tidak memenuhi asumsi dasar regresi Poisson, yaitu kesamaan antara nilai rata-rata dan variansi. Namun sering kali *count data* mengalami overdispersi (variansi lebih besar dibanding rata-rata). Untuk mengatasi permasalahan tersebut, digunakan metode Regresi Poisson Inverse Gaussian yang merupakan pengembangan dari regresi Poisson dan mampu menangani overdispersi dengan lebih fleksibel.

1.2. Rumusan Masalah

1. Bagaimana pemodelan jumlah penyewaan sepeda pada data Bike Sharing tahun 2011–2012 menggunakan regresi Poisson?
2. Bagaimana pemodelan jumlah penyewaan sepeda pada data Bike Sharing tahun 2011–2012 menggunakan regresi Poisson–Inverse Gaussian (PIG)?
3. Model manakah yang lebih baik dalam memodelkan jumlah penyewaan sepeda pada data Bike Sharing tahun 2011–2012 berdasarkan kriteria pemilihan model menggunakan nilai AIC?

1.3. Batasan Masalah

Pada penelitian ini terdapat beberapa batasan masalah, dengan tujuan untuk memfokuskan pada tujuan penelitian sesuai dengan yang diharapkan. Adapun batasan masalahnya sebagai berikut:

1. Variabel penelitian menggunakan dataset *Bike Sharing* periode 2011-2012 dari UCI Machine Learning Repository.
2. Menggunakan *software* R versi 4.5.2.

3. Pemilihan model terbaik dilakukan menggunakan nilai AIC (*Akaike Information Criterion*).

1.4. Tujuan Penelitian

1. Memperoleh model regresi Poisson untuk memodelkan jumlah penyewaan sepeda pada data Bike Sharing tahun 2011–2012.
2. Memperoleh model regresi Poisson–Inverse Gaussian (PIG) untuk memodelkan jumlah penyewaan sepeda pada data Bike Sharing tahun 2011–2012.
3. Membandingkan kinerja model regresi Poisson dan Poisson–Inverse Gaussian (PIG) serta menentukan model yang paling sesuai untuk memodelkan jumlah penyewaan sepeda pada data Bike Sharing tahun 2011–2012.

1.5. Manfaat Penelitian

Penelitian ini dapat digunakan sebagai referensi tambahan bagi pembaca yang ingin mendalami topik serupa. Hasil yang diperoleh juga bisa menjadi bahan pertimbangan bagi penelitian berikutnya. Dengan begitu, penelitian ini diharapkan dapat memberi manfaat sederhana berupa tambahan rujukan dalam kajian yang sejenis.

1.6. Tinjauan Pustaka

Tinjauan pustaka berisi tentang hasil penelitian terdahulu yang menjadi dasar dan referensi dalam pelaksanaan penelitian ini.

Wibawa et al. (2023) membahas tentang pemanfaatan regresi Poisson untuk menganalisis data jumlah penderita hipertensi di Kota Kendari tahun 2019. Penelitian ini menunjukkan bahwa regresi Poisson efektif digunakan untuk memodelkan hubungan antara variabel respon yang berupa *count data* dengan beberapa variabel prediktor, Penelitian ini juga menekankan pentingnya uji signifikansi parameter, baik secara serentak maupun parsial, untuk mengidentifikasi variabel yang berpengaruh signifikan terhadap jumlah kasus hipertensi. Hasil uji menunjukkan bahwa variabel pelayanan kesehatan memiliki pengaruh signifikan terhadap kasus hipertensi 2019 di Kota Kendari.

Salamah et al. (2022) membahas tentang pemanfaatan regresi Poisson-Inverse Gaussian untuk menganalisis data jumlah penderita demam berdarah di Kota Kendari tahun 2020. Penelitian ini menunjukkan bahwa regresi Poisson-Inverse Gaussian lebih akurat digunakan untuk memodelkan hubungan antara variabel respon dengan beberapa variabel prediktor pada data yang mengalami overdispersi dibandingkan dengan regresi Poisson. Hasil uji menunjukkan bahwa variabel rasio fasilitas kesehatan memiliki pengaruh signifikan terhadap kasus demam berdarah tahun 2020 di Kota Kendari.

Pada penelitian Herindrawati et al. (2017) pada data yang mengalami overdispersi dilakukan pemodelan Poisson-Inverse Gaussian. Yang kemudian berhasil mengidentifikasi faktor-faktor penting yang memengaruhi peningkatan kasus HIV di Jawa Tengah tahun 2015. Faktor demografi (usia 25–34 tahun, daerah perkotaan), faktor kesehatan (rasio fasilitas), serta perilaku (penggunaan kondom) terbukti signifikan.

Pada penelitian yang dilakukan Utami et al. (2024) membahas tentang pemanfaatan regresi Poisson Inverse Gaussian untuk kasus pneumonia, kemudian

didapati bahwa data mengalami overdispersi sehingga pemodelan dapat dilakukan dengan metode regresi Poisson Inverse Gaussian. Dari penelitian ini dapat disimpulkan bahwa persentase balita yang memperoleh ASI eksklusif berpengaruh signifikan terhadap penurunan jumlah kasus pneumonia.

Pada penelitian Wulandari et al. (2024) data yang digunakan pada penelitian ini nilai memiliki variansi yang jauh lebih tinggi dibandingkan dengan nilai rata-ratanya. Sehingga untuk mengatasi masalah overdispersi, pemodelan regresi Poisson Inverse Gaussian sangat sesuai. Hasil uji menunjukkan persentase kepadatan penduduk berpengaruh positif terhadap kasus demam berdarah di kabupaten Deli Serdang, sedangkan persentase jumlah penduduk miskin dan persentase jumlah tenaga kesehatan berpengaruh negatif terhadap jumlah kasus demam berdarah di kabupaten Deli Serdang.

1.7. Sistematika Penulisan

Penelitian ini disusun dengan menggunakan sistematika sebagai berikut:

- BAB I** : Membahas latar belakang masalah, batasan masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, tinjauan pustaka, serta sistematika penulisan.
- BAB II** : Membahas landasan teori yang meliputi distribusi Poisson, distribusi Inverse Gaussian, distribusi Poisson–Inverse Gaussian, overdispersi, serta regresi Poisson–Inverse Gaussian, pemilihan model terbaik (AIC).
- BAB III** : Membahas metode penelitian yang mencakup data penelitian, pendekatan penelitian, serta langkah-langkah pengolahan data.
- BAB IV** : Membahas hasil analisis data dan pembahasan berdasarkan metode yang telah diterapkan dalam penelitian.

BAB V : Membahas kesimpulan yang diperoleh dari hasil penelitian serta saran untuk penelitian selanjutnya.



BAB V

PENUTUP

5.1. Kesimpulan

Berdasarkan hasil analisis dan pembahasan yang telah dilakukan pada penelitian ini, diperoleh kesimpulan sebagai berikut:

1. Pemodelan jumlah penyewaan sepeda pada data Bike Sharing tahun 2011–2012 menggunakan regresi Poisson menghasilkan model dengan variabel prediktor X_1 (*apparent temperature*/suhu yang dirasakan), X_3 (*humidity*/kelembapan udara), dan X_4 (kecepatan angin) yang signifikan. Diperoleh persamaan modelnya adalah:

$$\mu_i = 8,192982 + 1,677465X_1 - 0,672520X_3 - 1,034128X_4$$

2. Pemodelan jumlah penyewaan sepeda menggunakan regresi Poisson–Inverse Gaussian (PIG) menghasilkan model yang mampu mengakomodasi overdispersi pada data. Hasil estimasi menunjukkan bahwa variabel X_1 (*apparent temperature*/suhu yang dirasakan), X_3 (*humidity*/kelembapan udara), dan X_4 (kecepatan angin) berpengaruh signifikan terhadap jumlah penyewaan sepeda, sedangkan variabel lainnya tidak signifikan. Diperoleh persamaan modelnya adalah:

$$\mu_i = 8,3398 + 2,2177X_1 - 1,0864X_3 - 1,8261X_4$$

3. Berdasarkan kriteria pemilihan model menggunakan nilai AIC, model regresi Poisson–Inverse Gaussian (PIG) memiliki nilai AIC sebesar 13069,96 lebih

kecil dibandingkan model regresi Poisson yang memiliki nilai AIC sebesar 393421,37. Hal ini menunjukkan bahwa model PIG lebih baik dalam memodelkan jumlah penyewaan sepeda pada data Bike Sharing tahun 2011–2012 karena lebih sesuai dalam menangani overdispersi pada data.

5.2. Saran

Berdasarkan hasil penelitian yang telah dilakukan, terdapat beberapa saran yang dapat diberikan untuk pengembangan penelitian selanjutnya. Penelitian ini hanya menggunakan metode Regresi Poisson Inverse Gaussian dalam memodelkan data yang mengalami overdispersi. Oleh karena itu, pada penelitian selanjutnya disarankan untuk mempertimbangkan penggunaan metode lain yang juga sesuai untuk *count data* yang mengalami overdispersi. Hal ini bertujuan untuk memperoleh model yang lebih optimal dan sebagai pembanding terhadap hasil yang diperoleh dalam penelitian ini.

Selain itu, pemilihan model terbaik dalam penelitian ini didasarkan pada kriteria *Akaike Information Criterion* (AIC). Pada penelitian selanjutnya, disarankan untuk menggunakan kriteria pemilihan model lainnya dalam menentukan model terbaik.

DAFTAR PUSTAKA

- Agresti, A. (1996). *An introduction to categorical data analysis*, John Wiley & Sons. Inc., Publication.
- Akaike, H. (1998). A bayesian analysis of the minimum aic procedure. In *Selected Papers of Hirotugu Akaike*, pages 275–280. Springer.
- De Jong, P. & Heller, G. Z. (2008). *Generalized Linear Models for Insurance Data*. Cambridge University Press.
- Dean, C., Lawless, J. F., & Willmot, G. E. (1989). A mixed poisson inverse gaussian regression model. *Canadian Journal of Statistics*, 17(2):171–181.
- Diez, D. M., Barr, C. D., & Cetinkaya-Rundel, M. (2012). *OpenIntro Statistics*, volume 4. OpenIntro Boston, MA, USA:.
- Gradshteyn, I. S. & Ryzhik, I. M. (2014). *Table of integrals, series, and products*. Academic press.
- Herindrawati, A. Y., Latra, I. N., & Purhadi, P. (2017). Pemodelan regresi poisson inverse gaussian studi kasus: Jumlah kasus baru hiv di provinsi jawa tengah tahun 2015. *Jurnal Sains Dan Seni ITS*, 6(1):137–143.
- Hilbe, J. M. (2014). *Modeling Count Data*. Cambridge University Press, Cambridge, ISBN: 9781139236065, DOI: 10.1017/CBO9781139236065.
- Karomah, B. (2024). Penerapan metode newton raphson untuk pendugaan parameter regresi non linear cobb-douglas. *Jurnal Jendela Matematika*, 2(02):93–103.

- Kementerian Kesehatan Republik Indonesia (2024). Bahaya polusi udara bagi kesehatan: Dampak, penyebab dan pencegahannya. <https://ayosehat.kemkes.go.id/bahaya-polusi-udara-bagi-kesehatan>. Diakses pada 12 Mei 2026.
- Lane, D. (2025). Online statistics education. In *International encyclopedia of statistical science*, pages 1806–1809. Springer.
- Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). *Introduction to Linear Regression Analysis*. John Wiley & Sons.
- Nugraha, J. (2014). *Pengantar Analisis Data Kategorik: Metode dan Aplikasi Menggunakan Program R*. Deepublish.
- Quach, J. & Malekian, R. (2020). Exploring the weather impact on bike sharing usage through a clustering analysis. *arXiv preprint arXiv:2008.07249*.
- Rennie, N., Cleophas, C., Sykulski, A. M., & Dost, F. (2023). Analysing and visualising bike-sharing demand with outliers. *Discover Data*, 1(1):1.
- Salamah, N. A. H., Ruslan, R., Baharuddin, B., Yahya, I., Makkulau, M., Agusrawati, A., & Mukhsar, M. (2022). Pemodelan faktor-faktor yang mempengaruhi jumlah penderita demam berdarah dengue (dbd) di kota kendari menggunakan regresi poisson inverse gaussian. *Jurnal Matematika Komputasi dan Statistika*, 2(2):115–122.
- Utami, K. N., Sugito, S., & Santoso, R. (2024). Pemodelan jumlah kasus pneumonia pada balita di jawa timur menggunakan metode regresi poisson inverse gaussian dilengkapi gui-r. *Jurnal Gaussian*, 12(4):539–548.

Waruwu, M., Puat, S. N., Utami, P. R., Yanti, E., & Rusydiana, M. (2025). Metode penelitian kuantitatif: Konsep, jenis, tahapan dan kelebihan. *Jurnal Ilmiah Profesi Pendidikan*, 10(1):917–932.

Wibawa, G. N. A., Abapihi, B., et al. (2023). Pemodelan regresi poisson terhadap faktor-faktor yang mempengaruhi terjadinya hipertensi di kota kendari. *Jurnal Matematika Komputasi dan Statistika*, 3(1):255–262.

Wulandari, M., Lubis, R. S., & Aprilia, R. (2024). Penerapan regresi poisson inverse gaussian pada faktor yang mempengaruhi jumlah kasus demam berdarah dengue di kabupaten deli serdang. *Jurnal Pembelajaran dan Matematika Sigma (JPMS)*, 10(2):226–232.