TUGAS AKHIR - KA 184801

PEMODELAN KASUS COVID-19 DI JAWA TIMUR MENGGUNAKAN METODE *GENERALIZED POISSON REGRESSION* DAN *NEGATIVE BINOMIAL REGRESSION*

SYAILLENDRA ARDIFASALMA

NRP 06311840000027

Dosen Pembimbing

Ulil Azmi, S.Si, M.Si

NIP 1990201912069

Program Studi Sains Aktuaria

Departemen Aktuaria

Fakultas Sains Dan Analitika Data

Institut Teknologi Sepuluh Nopember

Surabaya

2022

TUGAS AKHIR – KA184801

ANALISIS RISIKO SAHAM SEKTOR PERBANKAN MENGGUNAKAN *VALUE AT RISK* DAN *EXPECTED SHORTFALL* DENGAN PENDEKATAN VARMA-GARCH

SYAILLENDRA ARDIFASALMA

NRP 06311840000027

Dosen Pembimbing

Ulil Azmi, S.Si, M.Si

NIP 1990201912069

Program Studi Sains Aktuaria

Departemen Aktuaria

Fakultas Sains Dan Analitika Data

Institut Teknologi Sepuluh Nopember

Surabaya

2022

FINAL PROJECT – KA184801

MODELING THE CASE OF COVID-19 IN EAST JAVA USING GENERALIZED POISSON REGRESSION AND NEGATIVE BINOMIAL REGRESSION METHODS

SYAILLENDRA ARDIFASALMA

NRP 06311840000027

Advisor

Ulil Azmi, S.Si, M.Si

NIP 1990201912069

Study Program Actuarial Science

Department of Actuarial

Faculty of Science and Data Analytics

Institut Teknologi Sepuluh Nopember

Surabaya

2022

LEMBAR PENGESAHAN

PEMODELAN KASUS COVID-19 DI JAWA TIMUR MENGGUNAKAN METODE *GENERALIZED POISSON REGRESSION* DAN *NEGATIVE BINOMIAL REGRESSION*

TUGAS AKHIR

Diajukan untuk memenuhi salah satu syarat
memperoleh gelar Sarjana Ilmu Aktuaria pada
Program Studi Sarjana Sains Aktuaria
Departemen Aktuaria
Fakultas Sains dan Analitika Data
Institut Teknologi Sepuluh Nopember

Oleh: **SYAILLENDRA ARDIFASALMA**
NRP. 063118 4000 0027

Disetujui oleh Tim Penguji Tugas Akhir:

1. Ulil Azmi, S.Si, M.Si — Pembimbing ( )
2. Dr. Drs Soehardjoepri, M.Si — Penguji ( )
3. Imam Safawi Ahmad, S.Si, M.Si — Penguji

SURABAYA

JULI, 2022

(Halaman ini sengaja dikosongkan)

APPROVAL SHEET

MODELING THE CASE OF COVID-19 IN EAST JAVA USING GENERALIZED POISSON REGRESSION AND NEGATIVE BINOMIAL REGRESSION METHODS

FINAL PROJECT

Submitted to fulfill one of the requirements
for obtaining a degree Bachelor of Actuarial Science at
Undergraduate Study Program of Actuarial Science
Department of Actuarial Science
Faculty of Science and Data Analytics
Institut Teknologi Sepuluh Nopember

By: **SYAILLENDRA ARDIFASALMA**
NRP. 063118 4000 0027

Approved by Final Project Examiner Team:

1. Ulil Azmi, S.Si, M.Si	Advisor	( )
2. Dr. Drs Soehardjoepri, M.Si	Examiner	( )
3. Imam Safawi Ahmad, S.Si, M.Si	Examiner	( )

SURABAYA

JULY, 2022

(Halaman ini sengaja dikosongkan)

PERNYATAAN ORISINALITAS

Yang bertanda tangan di bawah ini:

Nama mahasiswa / NRP	: Syaillendra Ardifasalma / 06311840000027
Departemen	: Aktuaria
Dosen Pembimbing / NIP	: Ulil Azmi, S.Si, M.Si,/ 1990201912069

Dengan ini menyatakan bahwa Tugas Akhir dengan judul "PEMODELAN KASUS COVID-19 DI JAWA TIMUR MENGGUNAKAN METODE *GENERALIZED POISSON REGRESSION* DAN *NEGATIVE BINOMIAL REGRESSION"* adalah hasil karya sendiri, bersifat orisinal, dan ditulis dengan mengikuti kaidah penulisan ilmiah.

Bilamana di kemudian hari ditemukan ketidaksesuaian dengan pernyataan ini, maka saya bersedia menerima sanksi sesuai dengan ketentuan yang berlaku di Institut Teknologi Sepuluh Nopember.

Surabaya, Juli 2022

Mengetahui Dosen Pembimbing	Mahasiswa,
(Ulil Azmi, S.Si, M.Si) NIP. 1990201912069	(Syaillendra Ardifasalma) NRP. 06311840000027

(Halaman ini sengaja dikosongkan)

STATEMENT OF ORIGINALITY

The undersigned below:

Name of student / NRP : Syaillendra Ardifasalma / 06311840000027

Department : Aktuaria

Advisor / NIP : Ulil Azmi, S.Si, M.Si / 1990201912069

Hereby declare that the Final Project with the title of “MODELING THE CASE OF COVID-19 IN EAST JAVA USING GENERALIZED POISSON REGRESSION AND NEGATIVE BINOMIAL REGRESSION METHODS” is the result of my own work, is original, and is written by following the rules of scientific writing.

If in the future there is a discrepancy with statement then I am willing to accept sanctions in accordance with the provisions that apply at Institut Teknologi Sepuluh Nopember.

Surabaya, July 2022

Acknowledge

Advisor

(Ulil Azmi, S.Si, M.Si)

NIP. 1990201912069

Student,

(Syaillendra Ardifasalma)

NRP. 06311840000027

(Halaman ini sengaja dikosongkan)

PEMODELAN KASUS COVID-19 DI JAWA TIMUR MENGGUNAKAN METODE GENERALIZED POISSON REGRESSION DAN NEGATIVE BINOMIAL REGRESSION

Nama Mahasiswa / NRP : Syaillendra Ardifasalma / 063118 4000 0027

Departemen : Aktuaria FSAD - ITS

Dosen Pembimbing : Ulil Azmi, S.Si, M.Si

Abstrak

Virus SARS-CoV-2 atau juga dikenal sebagai COVID-19, pertama kali ditemukan di China pada akhir 2019 dan telah menyebar secara global dan menyebabkan lebih dari 178 juta kasus terkonfirmasi dan sebanyak 3,9 juta jiwa meninggal dunia. Untuk kasus di Jawa Timur sendiri kasus COVID-19, hingga bulan Januari 2022 jumlah kasus di Jawa Timur yang terpapar virus COVID-19 sendiri mencapai 402.879 jiwa, sedangkan jumlah yang sembuh mencapai 371.745 jiwa dan meninggal dunia sebanyak 29.774 jiwa. Analisis regresi adalah analisis statistika yang mempelajari hubungan antara variabel dependen dengan satu atau lebih variabel independen. Sedangkan Regresi Poisson sendiri merupakan model dengan variabel dependen berdistribusi Poisson. Namun dalam model regresi Poisson asumsi sering dilanggar antara estimasi varians yang berada di atas *mean* (overdispersi) atau di bawah *mean* (underdispersi). Salah satu model yang digunakan untuk menangani underdispersi atau overdispersi ini yaitu *Generalized Poisson Regression* dan *Negative Binomial Regression.* Data yang akan digunakan untuk memodelkan jumlah korban jiwa COVID-19 yaitu data jumlah kasus meninggal per hari di Jawa Timur dari bulan Oktober 2020 sampai dengan Januari 2022. Proses analisis data dilakukan dengan menggunakan *software RStudio* dengan faktor yang diduga mempengaruhi yaitu kasus aktif, kasus baru, *Stringency Index*, dan *Bed Occupancy Rate* di Provinsi Jawa Timur. Penelitian ini diharapkan dapat membantu masyarakat Provinsi Jawa Timur dalam mengantisipasi penyebaran kasus COVID-19 berdasarkan faktor – faktor yang berpengaruh signifikan pada penelitian dan juga menambah wawasan mengenai faktor –faktor apa saja yang dapat berpengaruh dengan kasus COVID-19 sehingga masyarakat bisa lebih waspada lagi dalam masa pandemi ini. Hasil penelitian menunjukan bahwa model terbaik adalah model *Generalized Poisson Regression*. Hal ini ditunjukan dari nilai AIC dan nilai BIC pada model *Generalized Poisson Regression* yang lebih kecil daripada model regresi Poisson dan model *Negative Binomial Regression*. Faktor-faktor yang berpengaruh terhadap jumlah kasus meninggal COVID-19, yaitu kasus aktif, kasus baru, *Stringency Index*, dan *Bed Occupancy Rate.*

Kata Kunci: COVID-19, Generalized Poisson Regression, Negative Binomial Regression, AIC

(Halaman ini sengaja dikosongkan)

MODELING THE CASE OF COVID-19 IN EAST JAVA USING GENERALIZED POISSON REGRESSION AND NEGATIVE BINOMIAL REGRESSION METHODS

Student Name / NRP : Syaillendra Ardifasalma / 063118 4000 0027
Department : Actuarial Science FSAD - ITS
Advisor : Ulil Azmi, S.Si, M.Si, M.Sc

Abstract

The SARS-CoV-2 virus, also known as COVID-19, was first discovered in China in late 2019 and has spread globally and caused more than 178 million confirmed cases and 3.9 million deaths. For cases in East Java itself, COVID-19 cases, until January 2022 the number of cases in East Java exposed to the COVID-19 virus itself reached 402,879 people, while the number of people who recovered reached 371,745 people and died as many as 29,774 people. Regression analysis is a statistical analysis that studies the relationship between the dependent variable and one or more independent variables. While the Poisson regression itself is a model with the dependent variable having a Poisson distribution. However, in the Poisson regression model the assumption is often violated between the estimated variance that is above the mean (overdispersion) or below the mean (underdispersion). One of the models used to deal with underdispersion or overdispersion is Generalized Poisson Regression and Negative Binomial Regression. The data that will be used to model the number of COVID-19 fatalities is data on the number of cases dying per day in East Java from October 2020 to January 2022. The data analysis process is carried out using RStudio software with factors suspected of influencing, namely active cases, new cases , Stringency Index, and Bed Occupancy Rate in East Java Province. This research is expected to help the people of East Java Province in anticipating the spread of COVID-19 cases based on factors that have a significant influence on research and also add insight into what factors can affect COVID-19 cases so that people can be more vigilant in dealing with COVID-19 cases. this pandemic period. The results showed that the best model was the Generalized Poisson Regression model. This is indicated by the AIC and BIC values in the Generalized Poisson Regression model which are smaller than the Poisson regression model and the Negative Binomial Regression model. Factors that affect the number of COVID-19 deaths, namely active cases, new cases, Stringency Index, and Bed Occupancy Rate.

Keywords: COVID-19, Generalized Poisson Regression, Negative Binomial Regression, AIC

(Halaman ini sengaja dikosongkan)

KATA PENGANTAR

Puji syukur kehadirat Allah SWT atas segala rahmat-Nya sehingga penulis dapat menyelesaikan laporan penilitian tugas akhir yang berjudul "Pemodelan Kasus COVID-19 di Jawa Timur Menggunakan Metode *Generalized Poisson Regression* dan *Negative Binomial Regression*" Maksud dan tujuan penulisan laporan penelitian pada tugas akhir ini adalah sebagai salah satu syarat untuk memperoleh gelar Sarja Ilmu Aktuaria pada program studi Sains Aktuaria Institut Teknologi Sepuluh Nopember Surabaya.

Penulis menyadari bahwa tugas akhir ini tidak akan dapat terselesaikan dengan baik tanpa bantuan dan kerjasama dari banyak pihak. Oleh karena itu, penulis ingin menyampaikan ucapan terimakasih kepada pihak-pihak berikut:

1. Allah SWT yang telah memberikan rahmat, kemudahan dan kesehatan sehingga penulis dapat menyelesaikan skripsi ini.
2. Bapak Dr. Drs Soehardjoepri, M.Si selaku kepala departemen aktuaria yang selama ini selalu mendukung kegiatan positif di jurusan sehingga bisa sampai tahap terkahir ini.
3. Ibu Ulil Azmi, S.Si, M.Si, M.Sc. sebagai pembimbing penulis yang selalu membimbing penulis sejak awal perkuliahan serta memberikan arahan dan nasehat selama proses penelitian skripsi ini dan juga dosen wali saya selama kuliah di Aktuaria ITS yang selalu memantau perkembangan akademik saya.
4. Bapak Imam Safawi Ahmad, M.Si sebagai dosen penguji pertama yang selalu memberikan masukan terhadap penulisan ini sampai tahap akhir dan Bapak Dr. Drs Soehardjoepri, M.Si sebagai dosen penguji kedua yang selalu memberikan masukan terhadap penulisan ini sampai tahap akhir.
5. Kepada dosen dan tendik aktuaria lainnya yang tidak bisa saya tuliskan satu persatu untuk ilmu yang sangat berharga.
6. Orang Tua dan segenap keluarga yang tidak pernah berhenti memberikan doa, dukungan, semangat, dan kasih sayang untuk penulis selama ini.
7. Nurunnisa Fadila Ulya, Alfian Hafiz, dan seluruh teman – teman Program Studi Sarjana (S1) Sains Aktuaria Angkatan 2018 yang telah memberikan dukungan, bantuan, doa, dan semangat, kepada penulis.

Penulis menyadari bahwa dirinya memiliki keterbatasan dan kemampuan yang beragam, sehingga penulisan skripsi ini masih jauh dari sempurna dan membutuhkan kritik maupun saran untuk perbaikan. Penulis berharap para pembaca dapat mengambil manfaat dan terinspirasi dari skripsi ini.

Surabaya, Juli 2022

Penulis

(Halaman ini sengaja dikosongkan)

DAFTAR ISI

DAFTAR GAMBAR

(Halaman ini sengaja dikosongkan)

DAFTAR TABEL

(Halaman ini sengaja dikosongkan)

DAFTAR LAMPIRAN

(Halaman ini sengaja dikosongkan)

BAB I
PENDAHULUAN

1.1 Latar Belakang

Saat ini kita sedang mengalami pandemi, yang disebabkan oleh mewabahnya virus yang bernama COVID-19. Virus SARS-CoV-2 atau juga dikenal sebagai COVID-19, pertama kali ditemukan di China pada akhir 2019 dan telah menyebar secara global dan menyebabkan lebih dari 178 juta kasus terkonfirmasi dan sebanyak 3,9 juta jiwa meninggal dunia. Wuhan merupakan tempat di mana infeksi COVID-19 pertama dicatat. Dalam beberapa bulan terakhir, para ilmuwan telah mencapai konsensus luas bahwa penyebaran virus adalah hasil dari "*zoonosis spillover*" atau "penularan virus dari hewan yang terinfeksi ke manusia" sebelum menjadi penularan yang sangat tinggi dari manusia ke manusia. COVID-19 merupakan ancaman kesehatan masyarakat yang paling serius dari gangguan pernapasan penyakit sejak pandemi flu Spanyol 1918 (Velavan dan Meyer, 2020).

Bukti tentang betapa mematikannya virus COVID-19 terus berkembang. Pada akhir Mei 2020, Pusat Pengendalian dan Pencegahan Penyakit AS (CDC) memperkirakan bahwa sekitar sepertiga dari infeksi virus corona tidak menunjukkan gejala dan sekitar 0,4% orang yang menunjukkan gejala meninggal, menyiratkan tingkat kematian infeksi (IFR) sebesar 0,3% (CDC 2020). Sebaliknya, perkiraan sebelumnya menunjukkan bahwa COVID-19 hingga 30 kali lebih mematikan daripada influenza musiman dan setidaknya 10 kali lebih menular daripada SARS (WHO 2020; Wilder-Smith, Chiew dan Li 2020). Negara-negara menanggapi wabah COVID-19 dengan langkah-langkah seperti *Lockdown*. Ancaman kesehatan dari COVID-19 sangat spesifik terhadap usia, dengan orang tua jauh lebih rentan terhadap mempertaruhkan. Sebuah studi internasional menemukan bahwa jumlah kematian dari semua penyebab antara Februari dan awal Mei di Inggris dan Wales sekitar 11% di atas biasanya untuk orang yang lebih muda dari 45 tetapi 65% di atas biasanya untuk usia 65-74 tahun dan 82% di atas biasanya untuk mereka yang berusia 75 tahun atau lebih.

Kasus di Jawa Timur sendiri COVID-19 pertama kali ditemukan pada tanggal 19 Maret 2020 sebanyak 9 jiwa, hingga bulan Januari 2022 jumlah kasus di Jawa Timur yang terpapar virus COVID-19 sendiri mencapai 402.879 jiwa, sedangkan jumlah yang sembuh mencapai 371.745 jiwa dan meninggal dunia sebanyak 29.774 jiwa. Dengan jumlah tersebut, provinsi Jawa Timur menjadi salah satu provinsi dengan kasus COVID-19 terbanyak di Indonesia. Dalam kurun waktu kurang dari 2 tahun penyebaran virus di wilayah Jawa Timur juga menjadi yang paling besar, ini menunjukkan bahwa persebaran kasus COVID-19 ini begitu cepat persebarannya. Berdasarkan data yang dikeluarkan oleh Pemerintah Provinsi Jawa Timur menunjukkan bahwa tren yang terus meningkat sejak Oktober 2020.

Pasien COVID-19 yang meninggal berada pada rentang usia sekitar 45 hingga 65 tahun (Tosepu dkk, 2020), di Italia sendiri sebanyak 85,6% pasien yang meninggal akibat COVID-19 berusia di atas 70 tahun (Zhu dkk, 2020). Selain itu, penelitian oleh Ji Yunpeng dkk (2020) menyebutkan bahwa kematian akibat COVID-19 juga erat kaitannya dengan ketersediaan sumber daya perawatan kesehatan. Sumber daya kesehatan tenaga medis seperti dokter dan perawat serta fasilitas yang tersedia yang ada di setiap daerah (Ilpaj dan Nurwati, 2020). Lalu, melalui penelitian (Hogan dkk, 2020) dalam artikelnya menyebutkan bahwa penyakit bawaan (Kormobid) dapat meningkatkan risiko kematian pada pasien COVID-19 sehingga hal ini berdampak pada meningkatnya kasus meninggal akibat virus ini. Namun, berdasarkan penelitian sebelumnya belum ada penelitian yang menggunakan pemodelan untuk banyaknya kasus meninggal akibat COVID-19.

Analisis regresi adalah analisis statistika yang mempelajari hubungan antara variabel dependen dengan satu atau lebih variabel independen. Analisis regresi menggunakan variabel dependen sebagai variabel acak kontinu untuk menganalisis data. Namun, terdapat beberapa data yang dianalisis melalui variabel dependennya yaitu variabel random diskrit. Salah satu model regresi yang dapat digunakan untuk menganalisis hubungan antara variabel dependen Y berupa data diskrit adalah model regresi Poisson (Consul dan Famoye, 1992).

Model regresi Poisson telah banyak digunakan untuk menganalisis menghitung data di mana rata-rata sampel dan varians sampel hampir sama. Juga diketahui bahwa hitungan sering kali menunjukkan hal yang substantial variasi di mana varians sampel lebih besar atau lebih kecil dari rata-rata sampel dan diklasifikasikan sebagai overdispersi atau underdispersi masing-masing. Berbagai model telah disarankan oleh penulis yang berbeda untuk menangani underdispersi atau overdispersi ini (Manton dkk., 1981), (Williams, 1982), (Hinde, 1982), (Cox, 1983), dan (Stein dan Juritz, 1988) telah membuat studi rinci tentang model regresi binomial negatif untuk menghitung data dengan over-dispersi. Namun, tidak satupun dari model yang dibahas oleh berbagai peneliti sedemikian rupa sehingga dapat mengakomodasi dispersi bawah dan dispersi berlebih (Consul dan Famoye, 1992).

Model regresi Poisson sendiri merupakan model dengan variabel dependen berdistriusi Poisson. Namun dalam model regresi Poisson asumsi sering dilanggar antara estimasi varians yang berada di atas *mean (over-dispersion)* atau di atas *mean (under-dispersion)*. Dispersi yang berlebihan menyebabkan heterogenitas sehingga asumsi residual dalam model regresi tidak terpenuhi. Berbagai model telah disarankan oleh penulis yang berbeda untuk menangani underdispersi atau overdispersi ini, salah satu model yang digunakan yaitu *Generalized Poisson Regression* dan *Negative Binomial Regression*.

Berbagai penelitian terdahulu untuk mengatasi kasus overdispersi pada model regresi telah banyak dilakukan. Beberapa penelitian tersebut diantaranya penelitian dari Adams dkk., (2020) penelitian ini bertujuan untuk mencari model Regresi yang sesuai dengan kasus COVID-19 di Nigeria. Variabel yang digunakan pada penelitian tersebut yaitu kasus kematian total, kasus terkonfirmasi, kasus aktif, dan kasus kritis COVID-19. Hasil dari penelitian tersebut menunjukkan jika dilihat melalui nilai kriteria pemilihan model menggunakan nilai -2logL, AIC, dan BIC menetapkan *Generalized Poisson Regression* (GPR) sebagai model terbaik karena memiliki nilai terkecil dari ketiga kriteria seleksi. Penelitian dari Asrul dan Naing (2012) yang bertujuan untuk membandingkan regresi Poisson dan Binomial Negatif pada data tingkat kematian AIDS dengan menggunakan variabel jenis kelamin, kebangsaan, ras, status perkawinan, pekerjaan, dan cara penularan dan hasilnya menunjukkan bahwa nilai *Pearson chi-square* dari *Negative Binomial Regression* lebih kecil dari Regresi Poisson (0,7106 dan 7,9705) masing-masing. Berdasarkan nilai ini, kita dapat menunjukkan bahwa regresi binomial negatif lebih baik dalam mengatasi kasus overdispersi. Penelitian lain oleh Banapon dkk. (2020) pemodelan jumlah kasus Tuberkulosis dilakukan untuk menekan peningkatan jumlah kasus Tuberkulosis di Indonesia dengan pendekatan regresi Binomial Negatif. Hasil dari penelitian menyatakan bahwa model regresi binomial negative dapat mengatasi masalah overdispersi yang terjadi pada Kasus Penderita Tuberkulosis. Pada penelitian yang dilakukan Arisandi dkk. (2018) menggunakan model *Generalized Poisson Regression* untuk mengatasi overdispersi pada Jumlah Penderita Demam Berdarah Dengue di Kota Makassar tahun 2016. Hasilnya menunjukkan bahwa *Generalized Poisson Regression* mampu mengatasi terjadinya overdispersi yang terjadi pada pemodelan regresi Poisson. Sebanyak 67% jumlah penderita Demam Berdarah Dengue ditentukan oleh persentase tempat-tempat umum memenuhi syarat kesehatan, persentase penduduk yang memiliki akses air minum layak, persentase rumah tangga berprilaku hidup bersih dan sehat dan persentase rumah yang memenuhi syarat kesehatan. Selebihnya 33% ditentukan oleh faktor lain.

Berdasarkan pemaparan dan penelitian terdahulu, peneliti tertarik untuk menerapkan metode *Generalized Poisson Regression* dan *Negative Binomial Regression* pada pemodelan kasus COVID-19 di Provinsi Jawa Timur dengan menggunakan data harian di Jawa Timur dari bulan Oktober 2020 sampai dengan Januari 2022. Faktor yang diduga mempengaruhi jumlah korban COVID-19 yaitu kasus aktif COVID-19, kasus baru COVID-19, indeks keketatan respons pemerintah, dan *Bed Occupancy Rate* di Provinsi Jawa Timur.

1.2 Perumusan Masalah

Berdasarkan latar belakang tersebut, maka dapat diketahui bahwa rumusan permasalahan dari penelitian ini adalah:

1. Bagaimana mendeskripsikan variabel jumlah kasus meninggal COVID-19 di Jawa Timur?
2. Apa faktor yang mempengaruhi jumlah kasus meninggal COVID-19 di Jawa Timur dengan metode regresi Poisson, *Generalized Poisson Regression* dan *Negative Binomial Regression* ?
3. Bagaimana pemodelan jumlah kasus meninggal COVID-19 pada Provinsi Jawa Timur dengan metode regresi Poisson, *Generalized Poisson Regression* dan *Negative Binomial Regression*?

1.3 Tujuan Penelitian

Berdasarkan permasalahan yang dirumuskan, tujuan yang ingin dicapai dari penelitian ini adalah sebagai berikut.

1. Mendeskripsikan variabel jumlah kasus meninggal COVID-19 di Jawa Timur dan faktor – faktor yang mempengaruhinya dengan metode regresi Poisson, *Generalized Poisson Regression* dan *Negative Binomial Regression.*
2. Mengetahui faktor yang mempengaruhi jumlah kasus meninggal COVID-19 di Jawa Timur dengan metode regresi Poisson, *Generalized Poisson Regression* dan *Negative Binomial Regression*
3. Mengetahui pemodelan jumlah kasus meninggal COVID-19 pada Provinsi Jawa Timur dengan metode regresi Poisson, *Generalized Poisson Regression* dan *Negative Binomial Regression*.

1.4 Manfaat Penelitian

Manfaat yang ingin dicapai dari hasil penelitian ini adalah sebagai berikut.

1. Sebagai referensi yang dapat digunakan untuk melihat pengaruh dari variabel independen terhadap variabel dependen, dimana jika terjadi overdispersi pada regresi Poisson dapat menggunakan metode *Generalized Poisson Regression* dan *Negative Binomial Regression*.
2. Dapat menambah wawasan mengenai faktor –faktor apa saja yang dapat berpengaruh dengan kasus COVID-19 sehingga masyarakat bisa lebih waspada lagi dalam menghadapi penyebaran COVID-19.

1.5 Batasan Penelitian

Batasan masalah dari penelitian ini adalah sebagai berikut.

1. Data pasien COVID-19 yang digunakan dalam penelitian ini menggunakan variabel jumlah korban COVID-19 yang meninggal, kasus aktif COVID-19, dan kasus baru COVID-19, *Stingency Index*, BOR pada periode Oktober 2020 – Januari 2022
2. Data yang digunakan merupakan data harian dari periode tahun Oktober 2020 – Januari 2022
3. Model yang digunakan pada penelitian ini menggunakan metode *Generalized Poisson Regression* dan *Negative Binomial Regression*

(Halaman ini sengaja dikosongkan)

BAB II
TINJAUAN PUSTAKA

2.1 Penelitian Terdahulu

Penelitian ini menggunakan beberapa penelitian terdahulu sebagai acuan sehingga penelitian ini tidak dapat terlepas dari penelitian-penelitian tersebut. Penelitian terdahulu yang menjadi acuan dan referensi dalam penelitian ini tertulis dalam Tabel 2.1 sebagai berikut

Tabel 2.1 Penelitian Terdahulu

Peneliti	Judul	Metode Penelitian	Hasil Penelitian
Adams dkk., (2020)	*Modeling COVID-19 Cases in Nigeria Using Some Selected Count Data Regression Models*	*Poisson Regression, Negative Binomial Regression* dan *Generalized Poisson Regression*	*Generalized Poisson Regression* (GPR) sebagai model terbaik karena memiliki nilai terkecil pada -2logL, AIC, dan BIC
Moh Asrul dan Naing (2012)	*Comparison Between Negative Binomial and Poisson Death Rate Regression Analysis: AIDS Mortality Co-Infection Patients*	*Poisson Regression* dan *Negative Binomial Regression*	Nilai *Pearson chi-square* dari *Negative Binomial Regression* lebih kecil dari Regresi Poisson
Banapon dkk. (2020)	Penerapan Regresi Binomial Negatif Untuk Mengatasi Pelanggaran Overdispersi Pada Regresi Poisson (Studi Kasus Penderita *Tuberculosis* di Provinsi Jawa Barat Tahun 2017)	*Poisson Regression* dan *Negative Binomial Regression*	Model regresi binomial negatif dapat mengatasi masalah overdispersi yang terjadi pada Kasus Penderita Tuberkulosis.
Arisandi dkk. (2018)	Aplikasi Generalized Poisson Regression dalam Mengatasi Overdispersi pada Data Jumlah Penderita Demam Berdarah Dengue	*Generalized Poisson Regression*	*Generalized Poisson Regression* mampu mengatasi terjadinya overdispersi yang terjadi pada pemodelan regresi Poisson.

2.2 *Coronavirus Disease 2019* (COVID-19)

Coronavirus Disease 2019 (COVID-19) adalah penyakit menular yang disebabkan oleh *Coronavirus* jenis baru. Penyakit ini diawali dengan kemunculan kasus pneumonia yang tidak diketahui etiologinya di Wuhan, China pada akhir Desember 2019. Berdasarkan hasil penyelidikan epidemologi, kasus yang terjadi diduga berhubungan dengan Pasar *Seafood* yang berada di Wuhan, China. Pada 7 Januari 2020, Pemerintah China mengumumkan bahwa penyakit tersebut disebabkan oleh Coronavirus jenis baru yang selanjutnya diberi nama SARS-CoV-2 (*Severe Acute Respiratory Syndrome Coronavirus 2*). Virus ini berasal dari famili yang sama dengan virus penyebab SARS dan MERS, tetapi SARS-CoV-2 lebih menular (Kemkes RI, 2020).

COVID-19 dapat ditularkan melalui kontak langsung dengan orang yang terinfeksi dan kontak tidak langsung dengan permukaan atau benda yang digunakan pada orang yang terinfeksi. Penularan melalui droplet dapat terjadi ketika seseorang berada pada jarak dekat (dalam 1 meter) dengan seseorang yang memiliki gejala pernapasan (batuk atau bersin) sehingga droplet berisiko mengenai mukosa (mulut dan hidung) atau konjungtiva (mata). Penularan juga dapat terjadi melalui benda dan permukaan yang terkontaminasi droplet di sekitar orang yang terinfeksi. Selain itu, juga memungkinkan penularan melalui udara pada keadaan khusus Di mana prosedur atau perawatan suportif yang menghasilkan aerosol (intubasi endotrakeal, bronkoskopi, suction terbuka, pemberian pengobatan nebulisasi, ventilasi manual sebelum intubasi, mengubah pasien ke posisi tengkurap, memutus koneksi ventilator, ventilasi tekanan positif non-invasif, trakeostomi, dan resusitasi kardiopulmoneer. Akan tetapi, masih diperlukan penelitian lebih lanjut mengenai penularan COVID-19 melalui udara (Kemkes RI, 2020).

Masa inkubasi COVID-19 berkisar 5-6 hari dengan range 1-14 hari, tetapi dapat mencapai 14 hari. Beberapa orang yang terinfeksi tidak merasakan gejala apapun dan tetap merasa sehat, tetapi beberapa lainnya merasakan gejala yang biasanya bersifat ringan dan muncul secara bertahap. Gejala yang paling umum adalah demam, rasa lelah, dan batuk kering. Beberapa orang juga mengalami rasa nyeri dan sakit, hidung tersumbat, pilek, nyeri kepala, konjungtivitis, sakit tenggorokan, diare, kehilangan penciuman dan pembauan, serta ruam pada kulit. Pada kasus berat, pasien akan mengalami *Acute Respirator Distress Syndrome* (ARDS), sepsis dan syok septik, gagal multiorgan, termasuk gagal ginjal atau gagal jantung akut hingga berakibat kematian. Orang dengan usia lanjut dan orang dengan kondisi medis tertentu, seperti tekanan darah tinggi, gangguan jantung, gangguan pari, diabetes, dan kanker berisiko lebih besar mengalami keparahan (Kemkes RI, 2020).

Upaya penanggulangan COVID-19 terus dilakukan secara masif dengan beberapa strategi mengingat pandemi COVID-19 yang berkepanangan telah memberikan dampak besar bagi perekonomian dan kehidupan sosial. Oleh karena itu, diperlukan intervensi tidak hanya dari sisi penerapan protokol kesehatan, tetapi juga intervensi yang efektif untuk memutus rantai penularan penyakit, yaitu melalui vaksinasi. Vaksinasi COVID-19 bertujuan untuk mengurangi transmisi/ penularan COVID-19, menurunkan angka kesakitan dan kematian akibat COVID-19, mencapai kekebalan kelompok di masyarakat (*herd immunity*) dan melindungi masyarakat dari COVID-19 agar tetap produktif secara sosial-ekonomi (Kemkes RI, 2021).

2.3 Keluarga Eksponensial

Keluarga eksponensial adalah keluarga distribusi yang sering digunakan pada statistika parametrik. Distribusi yang masuk dalam keluarga eksponensial terdiri dari distribusi diskrit maupun distribusi kontinu, contohnya distribusi normal, Bernoulli, Binomial, Multinomial, Eksponensial, dan Poisson. Pada tabel 2.1 disajikan fungsi kepekatan peluang, nilai harapan ($E(Y)$), dan ragam ($Var(Y)$) pada distribusi yang masuk dalam keluarga eksponensial (Bickel dan Doksum, 2006)

Tabel 2.2 Keluarga Eksponensial

Distribusi	$f(y)$	$E(Y)$	$Var(Y)$
Normal	$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\|x-\mu\|^2/(2\sigma^2)}$	μ	σ^2
Bernoulli	$\alpha^x(1-\alpha)^{1-x}$	p	$p(1-p)$
Binomial	$\binom{n}{x}\alpha^x(1-\alpha)^{n-x}$	np	$np(1-p)$
Multinomial	$\frac{n!}{x_1!\,x_2!\dots x_n!}\prod_{i=1}^{n}\alpha_i^{x_i}$	np_i	$np_i(1-p_i)$
Eksponensial	$\lambda e^{-\lambda x}$	$\frac{1}{\theta}$	$\frac{1}{\theta^2}$
Poisson	$\frac{e^{-\lambda}}{x!}\lambda^x$	λ	λ

Sebuah keluarga parametrik dari distribusi kontinu univariat dikatakan sebagai keluarga eksponensial jika dan hanya jika fungsi kepadatan probabilitas dari setiap anggota keluarga dapat ditulis sebagai

$$f(x) = h(x)exp[\theta^T t(x) - a(\theta)] \tag{2.1}$$

di mana:

θ = parameter

$t(x)$ = statistik cukup

$h(x)$ = ukuran dasar

$a(\theta)$ = fungsi log normalizer

Contoh bentuk umum keluarga eksponensial dapat dilihat dari fungsi kepekatan peluang bagi sebaran Poisson yang disajikan pada rumus (2.2)

$$f(x) = \frac{1}{x!}exp(x\ log\lambda - \lambda) \tag{2.2}$$

dengan ($log\ \lambda$) sebagai parameter. (x) sebagai statistik cukup. ($\frac{1}{x!}$) sebagai ukuran dasar. ($\lambda = e^{\theta}$) sebagai fungsi log normalizer.

2.4 *Maximum Likelihood Estimation*

Tujuan metode *Maximum Likelihood Estimation* (MLE) adalah menduga parameter dengan memaksimumkan fungsi likelihood. Misalkan $X_1, X_2, X_3, \dots, X_n$ merupakan saınpel raındom yang berasal dari populasi X dengan fungsi densitas probabilitas yang tergantung pada θ, yaitu $f(x;\theta)$. Dikarenakan bahwa $X_1, X_2, X_3, \dots, X_n$ merupakan sampel randorn maka fungsi $f(x;\theta)$ juga merupakan variabel acak. Suanı fungsi yang diamati misalnya $t_n(X_1, X_2, X_3, \dots, X_n)$ disebut sebagai statistik, dan disebut juga suatu penaksir dari θ yang dinotasikan dengan $\hat{\theta}$ (Feller, 1967). Akan tetapi ınenghadapi persoalan mencari penaksir terdapat tiga metode yang populer yaitu metode moınen, metode maksimum likelihood dan metode Bayes. Namun suatu metode menghasilkan penaksir yang baik untuk diselidiki yaitu penaksir maksimum *likelihood*, seperti dinyatakan dalam definisi berikut.

a) Misalkan $f(x_1, x_2, x_3, \ldots, x_n; \theta)$ fungsi kepadatan peluang gabungan untuk variabel random $X_1, X_2, X_3, \ldots, X_n$ dengan nilai sampel $x_1, x_2, x_3, \ldots, x_n$(Bain dan Engelhardt, 1992). Fungsi *likelihood* dari sanıpel nya adalah

$$L(\theta; x_1, x_2, x_3, \ldots, x_n) = f(x_1, x_2, x_3, \ldots, x_n; \theta) \tag{2.3}$$

b) Misalkan $X_1, X_2, X_3, \ldots, X_n$ merupakan sampel random yang berasal dari populasi X dengan fungsi kepadatan peluang yang tergantung pada θ. Yaitu $f(x; \theta)$ dengan θ adalah parameter yang tidak diketahui (Bain dan Engelhardt, 1992). Maka Fungsi *likelihood* gabungannya adalah.

$$L(\theta) = \prod_{i=1}^{n} f(x_i; \theta) \tag{2.4}$$

c) Taksiran maksimum *likelihood* untuk θ adalah nilai yang memaksimumkan fungsi likelihood $L(\theta)$. Dengan kata lain $\hat{\theta} = \hat{\theta}_n(X_1, X_2, X_3, \ldots, X_n)$ sedemikian sehingga $L(\hat{\theta}) = maks\,\{L(\theta)\}$ maka $\hat{\theta} = \hat{\theta}_n(X_1, X_2, X_3, \ldots, X_n)$ disebut penaksir maksimum *likelihood*.

Dalam menentukan penaksir maksimum *likelihood*, cukup hanya nıembutuhkan fungsi *likelihood* dan kemudian memaksimumkan fungsi *likelihood* terhadap parameter yang akan ditaksir. Dalam beberapa kasus, untuk memudahkan memaksimumkan fungsi *likelihood*, kadang kala dikerjakan dengan nıelakukan transformasi logaritma natural (ln) terhadap fungsi *likelihood*, selanjutnya disebut sebagai log fungsi *likelihood*. Prosedur menentukan penaksir maksimum *likelihood* adalah sebagai berikut.

a) Menentukan fungsi likelihood, $L(\theta)$.
b) Melakukan transformasi logaritma natural dari $L(\theta)$.
c) Melakukan diferensial $\ln L(\theta)$ terhadap θ, dan kemudian menyarnakan hasil dengan nol (persamaan *likelihood*).
d) Menentukan akar persamaan *likelihood* (θ), yang disebut sebagai penaksir maksimum *likelihood* (θ)

2.5 Regresi Poisson

Regresi Poisson adalah analisis regresi non linier dari distribusi Poisson yang digunakan untuk memprediksi ketergantungan variabel yang terdiri dari data diskrit atau jumlah yang diberikan satu atau lebih variabel independen. Variabel yang diprediksi disebut variabel dependen (atau terkadang respons, hasil, target, atau variabel kriteria). Variabel yang digunakan untuk memprediksi nilai variabel dependen disebut variabel independen (atau prediktor, variabel penjelas, atau regresi) (Hardin dan Hilbe, 2007). Regresi Poisson dikatakan mengandung overdispersi jika variansinya lebih besar dari nilai rata-ratanya. Jika terjadi overdispersi, estimasi parameter regresi tetap konsisten tetapi tidak efisien. Hal ini digunakan pada dasarnya ketika menghadapi masalah di mana hasil dari proses acak hanya dapat mengambil nilai hitungan. Salah satu distribusi yang memenuhi kriteria tersebut berasal dari keluarga distribusi eksponensial yang merupakan distribusi Poisson.

Di mana Y menjadi variabel acak (laju di mana kasus COVID-19 terjadi) dan hasil dari kasus menjadi sebuah peristiwa (Walpole, 1995). Variabel dependen dikatakan mengikuti Distribusi Poisson dengan parameter $\lambda \geq 0$ jika memiliki fungsi probabilitas sebagai berikut.

$$P(y, \lambda) = \frac{e^{-\lambda}\lambda^{y}}{y!} \tag{2.5}$$

Keterangan:

λ = rata-rata keberhasilan = np
n = banyaknya kejadian
p = probabilitas sukses
y = variabel dependen diskrit

Dengan λ merupakan rata-rata banyaknya kejadian atas variabel dependen (Y) yang terjadi dalam selang waktu atau daerah tertentu. Distribusi Poisson merupakan salah satu distribusi yang paling sederhana dalam pemodelan data diskrit. Menurut Cameron dan Trivedi (2013), beberapa karakteristik percobaan yang berdistribusi Poisson adalah sebagai berikut.

a) Kejadian yang terjadi pada jumlah anggota populasi yang besar dengan probabilitas yang kecil (*rare event*).
b) Bergantung pada interval waktu tertentu
c) Perulangan dari kejadian yang mengikuti sebaran distribusi binomial.

Model regresi Poisson secara umum disajikan pada persamaan (2.6).

$$y_i = \lambda_i + \varepsilon_i \tag{2.6}$$

Terdapat dua fungsi penghubung yang biasa digunakan dalam regresi Poisson. Pertama adalah penghubung identitas (*identity link*). Kedua adalah penghubung log (*log link*). Fungsi penghubung log adalah fungsi yang lebih cocok digunakan karena fungsi log menjamin bahwa nilai variabel yang diharapkan dari variabel independennya akan bernilai non negatif. Sehingga fungsi penghubung yang digunakan dalam penelitian ini adalah fungsi penghubung log. Model regresi Poisson berganda dapat dituliskan sebagai berikut (Cameron dan Trivedi, 2013).

$$\ln(\lambda_i) = (\beta_0 + \beta_1 X_{i1} + \cdots + \beta_p X_{ip}) \tag{2.7}$$

$$\lambda_i = exp\,(\beta_0 + \beta_1 X_{i1} + \cdots + \beta_p X_{ip}) \tag{2.8}$$

Model regresi Poisson mensyaratkan *equidispersi*, yaitu kondisi di mana nilai *mean* dan variansnya dari variabel dependen bernilai sama. Namun, adakalanya terjadi *overdispersion* berarti varians lebih besar daripada *mean*. Hal ini mengindasikan bahwa model regresi Poisson tidak cocok untuk data tersebut. Fenomena *overdispersion* dinyatakan dengan $Var(Y)$>$E(Y)$ (McCullagh dan Nelder, 1989).

2.5.1 Multikolinearitas

Multikolinearitas merupakan situasi yang menunjukkan adanya korelasi yang tinggi di antara variabel independen yang digunakan dalam model. Variabel independen dapat dikatakan saling bebas jika matriks korelasi antar variabel independen membentuk matriks identitas. Dalam model regresi, adanya korelasi antar variabel independen menyebabkan taksiran parameter regresi yang dihasilkan memiliki *error* yang sangat besar. Kasus multikolinearitas dapat dilihat dengan beberapa cara, yaitu jika nilai koefisien korelasi *Pearson* antar variabel independen > 0.95, maka terdapat korelasi antar variabel tersebut. Cara lain yang dapat digunakan untuk melihat situasi multikolinearitas adalah melalui nilai VIF (*Varians Inflation*

Factor). Nilai VIF (*Varians Inflation Factor*) yang bernilai >10 menunjukkan adanya multikolinearitas pada variabel independen.

a) *Variance Inflation Factor* (VIF)

Suatu model regresi menunjukan adanya kasus multikolinieritas antar variabel independen jika nilai VIF lebih dari 10. Nilai VIF dinyatakan dengan rumus sebagai berikut.

$$VIF = \frac{1}{1-R_j^2} \tag{2.9}$$

Di mana R_j^2 merupakan nilai koefisien determinasi dari salah satu variabel independen dengan $(p-1)$ variabel independen lainnya

b) Nilai Koefisien Korelasi *Pearson.*

Koefisien korelasi Pearson merupakan salah satu uji pada statistic parametric yang menunjukkan hubungan dan derajat hubungan dari dua variabel. Perhitungan dalam teknik korelasi ini mensyaratkan bahwa populasi asal sampel mempunyai dua varian dan berdistribusi normal (Draper dan Smith, 1998). Rumus nilai koefisien korelasi *Pearson* adalah sebagai berikut.

$$r = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\sqrt{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2}\sqrt{n\sum_{i=1}^{n} y_i^2 - (\sum_{i=1}^{n} y_i)^2}} \tag{2.10}$$

Nilai koefisien korelasi *Pearson* (r) dapat bernilai positif atau negatif dan berada pada rentang -1 dan 1. Jika (r) bernilai negatif berarti antara variabel independen dan variabel dependen terdapat hubungan linier (garis lurus) yang negatif (berlawanan arah). Jika nilai variabel independen bertambah besar maka nilai variabel dependen akan bertambah kecil. Kemudian ketika nilai variabel independen bertambah kecil maka nilai variabel dependen bertambah besar. Jika nilainya mendekati 0, maka hubungan keeratan dua variabel semakin lemah. Jika (r) bernilai positif berarti antara variabel independen dan variabel dependen terdapat hubungan linier (garis lurus) yang positif (searah) yaitu nilai variabel independen bertambah besar maka nilai variabel dependen akan bertambah besar. Demikian juga jika nilai variabel independen bertambah kecil maka nilai variabel dependen akan bertambah kecil(Holmes dkk., 2019).

Jika terjadi kasus multikolinearitas, maka hal pertama yang dapat dilakukan adalah mengeluarkan variabel independen yang mengalami multikolinearitas. Namun, jika variabel yang terjadi multikolinearitas tersebut signifikan terhadap model, maka variabel tersebut dapat dipertimbangkan kembali untuk dipertahankan(Hocking, 1996).

2.5.2 Estimasi Parameter Model Regresi Poisson

Estimasi parameter model regresi Poisson menggunakan metode MLE (*Maximum Likelihood Estimation*). Metode ini dilakukan dengan memaksimumkan fungsi *likelihood* (Agresti, 2002). Pada umumnya, metode MLE (*Maximum Likelihood Estimation*) digunakan jika distribusi data yang dimodelkan diketahui. Prosedur estimasi parameter model regresi Poisson menggunakan metode MLE sebagai berikut.

a) Menentukan bentuk umum fungsi likelihood untuk model regresi Poisson dapat dituliskan sebagai berikut.

$$\ln L(\beta) = ln\left(\prod_{i=1}^{n} \frac{exp(-\lambda_i)\lambda_i^{y_i}}{y_i!}\right)$$
$$= \sum_{i=1}^{n} ln\left(\frac{exp(-\lambda_i)\lambda_i^{y_i}}{y_i!}\right)$$
$$= \sum_{i=1}^{n}\left(\ln\left(e^{-\lambda_i}\right) + \ln({\lambda_i}^{y_i}) - \ln(y_i!)\right)$$
$$= \sum_{i=1}^{n}(-\lambda_i + y_i \ln \lambda_i - \ln(y_i!))$$
$$= \sum_{i=1}^{n}\left(-e^{\boldsymbol{x}_i^{\boldsymbol{T}}\boldsymbol{\beta}} + y_i \ln e^{\boldsymbol{x}_i^{\boldsymbol{T}}\boldsymbol{\beta}} - \ln(y_i!)\right) \tag{2.11}$$

b) Kemudian, melakukan turunan parsial fungsi *ln likelihood* pada persamaan di atas terhadap parameter yang akan diestimasi. Fungsi *ln likelihood* dapat dituliskan sebagai berikut.

$$\ln L(\beta) = -\sum_{i=1}^{n}\left(exp\left(\boldsymbol{x}_i^{\boldsymbol{T}}\boldsymbol{\beta}\right)\right) + \sum_{i=1}^{n} y_i\, ln\left(exp\left(\boldsymbol{x}_i^{\boldsymbol{T}}\boldsymbol{\beta}\right)\right) - \sum_{i=1}^{n} \ln\,(y_i!) \tag{2.12}$$

c) Parameter model regresi Poisson yang diestimasi dengan MLE dinyatakan dengan $\hat{\beta}_j$, dapat diperoleh dengan mencari turunan pertama fungsi *ln likelihood* terhadap β_j. Selanjutnya, diturunkan terhadap $\boldsymbol{\beta}^T$ kemudian menyamakan hasil dengan nol. Sebagai berikut.

$$\frac{\partial lnL(\beta)}{\partial \boldsymbol{\beta}^T} = -\sum_{i=1}^{n} \boldsymbol{x}_i\, exp\left(\boldsymbol{x}_i^{\boldsymbol{T}}\boldsymbol{\beta}\right) + \sum_{i=1}^{n} y_i \boldsymbol{x}_i \tag{2.13}$$
$$\frac{\partial^2 lnL(\beta)}{\partial \boldsymbol{\beta}^T \partial \boldsymbol{\beta}} = -\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i^{\boldsymbol{T}}\, exp\left(\boldsymbol{x}_i^{\boldsymbol{T}}\boldsymbol{\beta}\right)$$
$$= -\sum_{i=1}^{n} \boldsymbol{x}_i\, exp\left(\boldsymbol{x}_i^{\boldsymbol{T}}\boldsymbol{\beta}\right) \tag{2.14}$$

Akan tetapi, penyelesaian dengan metode MLE *(Maximum Likelihood Estimation*) tidak dapat dilakukan dengan metode numerik biasa, sehingga alternatif lain yang dapat digunakan untuk mendapatkan penyelesaian dari MLE adalah dengan iterasi numerik, yaitu Metode *Newton-Raphson* dengan langkah-langkah sebagai berikut.

(a) Menentukan nilai estimasi awal untuk parameter $\widehat{\boldsymbol{\beta}}_{(\mathbf{0})}$,

$$\widehat{\boldsymbol{\beta}}_{(\mathbf{0})} = (\boldsymbol{X^T X})^{-\mathbf{1}}\boldsymbol{X^T Y} \tag{2.15}$$

dengan,

$$X = \begin{bmatrix} 1 & x_{11} & \cdots & x_{p1} \\ 1 & x_{12} & \cdots & x_{p2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1i} & \cdots & x_{pi} \end{bmatrix}$$
$$\boldsymbol{Y} = [y_1 \quad y_2 \quad \cdots \quad y_p]^T$$

(b) Membentuk vektor gradien $\boldsymbol{g}$.

$$\boldsymbol{g}(\boldsymbol{\beta}_{(\boldsymbol{m})})_{(\boldsymbol{j}+\mathbf{1})\times\mathbf{1}} = \left(\frac{\partial lnL(\boldsymbol{\beta})}{\partial \beta_0}, \frac{\partial lnL(\boldsymbol{\beta})}{\partial \beta_1}, \dots, \frac{\partial lnL(\boldsymbol{\beta})}{\partial \beta_j}\right)^{\boldsymbol{T}}_{\boldsymbol{\beta}=\boldsymbol{\beta}_{(\boldsymbol{m})}}$$

dengan *j* merupakan banyaknya parameter yang diestimasi (variabel independen)

(c) Membentuk matriks Hessian $\boldsymbol{H}$

$$\boldsymbol{H}(\boldsymbol{\beta}_{(\boldsymbol{m})})_{(\boldsymbol{j}+\mathbf{1})\times(\boldsymbol{j}+\mathbf{1})} = \begin{bmatrix} \frac{\partial^2 lnL(\boldsymbol{\beta})}{\partial \beta_0{}^2} & \frac{\partial^2 lnL(\boldsymbol{\beta})}{\partial \beta_0 \partial \beta_1} & \cdots & \frac{\partial^2 lnL(\boldsymbol{\beta})}{\partial \beta_0 \partial \beta_j} \\ \frac{\partial^2 lnL(\boldsymbol{\beta})}{\partial \beta_1 \partial \beta_0} & \frac{\partial^2 lnL(\boldsymbol{\beta})}{\partial \beta_0{}^2} & \cdots & \frac{\partial^2 lnL(\boldsymbol{\beta})}{\partial \beta_1 \partial \beta_j} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 lnL(\boldsymbol{\beta})}{\partial \beta_j \partial \beta_0} & \frac{\partial^2 lnL(\boldsymbol{\beta})}{\partial \beta_j \partial \beta_1} & \cdots & \frac{\partial^2 lnL(\boldsymbol{\beta})}{\partial \beta_j{}^2} \end{bmatrix}_{\boldsymbol{\beta}=\boldsymbol{\beta}_{(\boldsymbol{m})}}$$

Setelah mendapatkan matriks Hessian $\boldsymbol{H}$, vektor gradien $\boldsymbol{g}$, dan nilai estimasi awal parameter $\widehat{\boldsymbol{\beta}}_{(\mathbf{0})}$, kemudian nilai estimasi awal $\widehat{\boldsymbol{\beta}}_{(\mathbf{0})}$ dimasukkan ke dalam elemen-elemen vektor $\boldsymbol{g}(\widehat{\boldsymbol{\beta}}_{(\mathbf{0})})$dan matriks $\boldsymbol{H}(\widehat{\boldsymbol{\beta}}_{(\mathbf{0})})$.

(d) Melakukan iterasi dengan persamaan sebagai berikut.

$$\boldsymbol{\beta}_{(m+1)} = \boldsymbol{\beta}_{(m)} - \boldsymbol{H}^{-1}(\boldsymbol{\beta}_{(m)})\boldsymbol{g}(\boldsymbol{\beta}_{(m)}) \tag{2.16}$$

Dimulai dari $m = 0$. Nilai $\boldsymbol{\beta}_{(m)}$ merupakan sekumpulan pengestimasi parameter yang konvergen pada iterasi ke-m

Jika parameter yang didapatkan belum konvergen, maka dilakukan kembali langkah ke (d) hingga iterasi ke $m = m + 1$. Iterasi berhenti pada keadaan konvergen, dimana $\left\|\boldsymbol{\beta}_{(m+1)} = \boldsymbol{\beta}_{(m)}\right\| < \varepsilon$, ε merupakan bilangan yang sangat kecil sekali (Agresti, 2002).

2.5.3 Pengujian Parameter Model Regresi Poisson

Pengujian parameter model regresi Poisson dilakukan secara serentak dan parsial. Pengujian parameter model regresi Poisson secara serentak bertujuan untuk mengetahui pengaruh variabel secara keseluruhan terhadap model dengan tingkat signifikansi tertentu. Berbeda dengan pengujian parameter regresi Poisson secara parsial bertujuan untuk mengetahui efek atau pengaruh setiap variabel terhadap model.

Salah satu metode yang digunakan dalam penentuan uji statistik dalam pengujian parameter model regresi Poisson adalah MLRT (*Maximum Likelihood Ratio Test*). Oleh karena itu, sebelum menentukan uji statistik terlebih dahulu menentukan dua buah fungsi likelihood yang berhubungan dengan model regresi yang diperoleh. Fungsi likelihood tersebut adalah $L(\hat{\Omega})$ yaitu nilai maksimum likelihood untuk model lengkap dengan melibatkan variabel independen. Selain itu, $L(\hat{\omega})$ yang merupakan nilai maksimum likelihood untuk model sederhana tanpa melibatkan variabel independen. Hipotesis yang digunakan adalah sebagai berikut.

H_0: tidak ada parameter yang signifikan ($\beta_1 = \beta_2 = \beta_3 = \cdots = \beta_p = 0$)

H_1: terdapat setidaknya satu parameter yang signifikan $\left(\beta_j \neq 0\, ; j = 1, 2, \ldots, p\right)$

Statistik uji yang digunakan adalah sebagai berikut

$$D(\hat{\beta}) = -2ln\left[\frac{L(\hat{\omega})}{L(\hat{\Omega})}\right] = 2(\ln\,(L(\hat{\Omega}) - L(\hat{\omega})) \tag{2.17}$$

Keterangan.

$L(\hat{\Omega})$ = nilai maksimum likelihood dengan variabel independen

$L(\hat{\omega})$ = nilai maksimum likelihood tanpa melibatkan variabel independen

Keputusan yang diambil akan tolak H_0 apabila $D(\hat{\beta}) > X^2_{v;\alpha}$ dengan v adalah banyaknya parameter model di bawah populasi dikurangi dengan banyaknya parameter model di bawah H_0. Tolak H_0 berarti terdapat salah satu variabel yang berpengaruh signifikan terhadap model sehingga dilanjutkan dengan pengujian secara parsial dengan hipotesis sebagai berikut:

H_0: β_j=0

H_1: $\beta_j \neq 0$, $j = 1, 2, \ldots, p$

Statistik uji yang digunakan adalah

$$z = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)} \tag{2.18}$$

Dengan $se(\hat{\beta}_j)$ adalah tingkat kesalahan $\hat{\beta}_j$. Keputusan yang diambil adalah tolak H_0 jika dengan $|z_{hit}| > z_{\frac{\alpha}{2}}$ dengan α adalah tingkat signifikansi.

2.5.4 Pengujian Overdispersi

Metode regresi Poisson mensyaratkan kondisi di mana nilai *mean* dan varians variabel dependen bernilai sama atau kondisi *equidispersion* (Khoshgotaar dkk., 2004). Akan tetapi, dalam penerapannya kondisi tersebut sulit dipenuhi oleh data atau sering terjadi *overdispersion* dalam data yang dimodelkan. Kondisi *overdispersion* dapat ditulis *Var(Y)>E(Y)* (Myers, 1990). Sebaliknya, data yang memiliki varians variabel dependen lebih kecil dari *mean* variabel dependen dapat disebut sebagai kondisi *underdispersion*. Kondisi overdispersi dapat disebabkan oleh adanya sumber keragaman yang tidak teramat. Overdispersi dapat pula disebabkan karena adanya *missing* pada variabel independen, perlunya interaksi dalam model, variabel perlu ditransformasi, atau kesalahan spesifik *link function*. Overdispersi akan menyebabkan model dengan estimasi parameter yang bias dan juga akan berdampak pada nilai *standard error estimator* yang semakin kecil (*underestimate*) yang selanjutnya dapat menyebabkan kesalahan dalam inferensi untuk parameter tersebut (Cameron dan Trivedi, 2013).

Adanya overdispersi dapat diketahui melalui nilai devians (D) (Hilbe, 2011) pada persamaan (2.18) atau nilai *Pearson Chi-Square* (χ^2) (Ismail dan Jemain, 2007) pada persamaan (2.19). Jika $\frac{D}{db}$ atau $\frac{\chi^2}{db}$ menghasilkan nilai yang lebih besar dari satu, maka dapat disimpulkan bahwa terjadi overdispersi pada data.

$$\theta_1 = \frac{D^2}{db} > 1; D^2 = 2\sum_{i=1}^{n}\left\{y_i \log\frac{y_i}{\mu_i}\right\} \tag{2.19}$$

$$\theta_2 = \frac{\chi^2}{db} > 1; \chi^2 = \sum_{i=1}^{n}\frac{(y_i-\mu_i)^2}{\mu_i} \tag{2.20}$$

Keterangan:

D = nilai Devians
χ^2 = nilai *Pearson Chi-Square*
y_i = nilai variabel dependen dari pengamatan ke-i
db = derajat bebas = $n-(p+1)$
μ_i = penduga bagi respon rata – rata ke-i = $(\exp(\beta_0 + \beta_1 X_{i1} + \cdots + \beta_p X_{ip}))$
n = banyaknya pengamatan

Jika θ_1 atau θ_2 bernilai lebih dari 1 maka terjadi overdispersi pada data.

2.6 *Generalized Poisson Regression*

Generalized Poisson Regression (GPR) dikembangkan untuk menangani pelanggaran asumsi ekuidispersi pada model regresi Poisson. Hal ini terutama digunakan untuk menyesuaikan adanya overdispersi yaitu dalam situasi di mana $Var\ (y_i)$ > $E(y_i)$ serta underdispersi ($Var(y_i)$ < $E(y_i)$). Model GPR dengan parameter (λ, θ) mirip dengan model regresi Poisson, tetapi diasumsikan bahwa komponen didistribusikan secara acak ke Poisson umum. Dalam analisis GPR, jika θ sama dengan 0 maka modelnya adalah model Poisson. Jika θ lebih dari 0 maka model GPR juga mewakili data yang mengandung overdispersi dan jika θ kurang dari 0 mewakili data yang mengandung underdispersi. Itu menyarankan bahwa Ketika y_i merupakan variabel dependen hitungan dan mengikuti distribusi Poisson Umum (Famoye, 2004), probabilitas fungsi kepadatan diberikan bahwa $i = 1, 2, \ldots, n$, maka.

$$f(y_i, \lambda_i; \theta) = \left[\frac{\lambda_i}{1+\theta\lambda_i}\right]^{y_i} \frac{(1+\theta y_i)^{y_i-1}}{y_i!} exp\left[\frac{-\lambda_i(1+\theta y_i)}{1+\theta\lambda_i}\right] \tag{2.21}$$

Di mana, *mean* sebagai $\lambda_i = E(y_i)$, varians sebagai $Var(y_i) = \lambda_i(1+\theta\lambda_i)^2$ dan θ disebut sebagai parameter dispersi.

2.6.1 Estimasi Parameter Model *Generalized Poisson Regression*

Estimasi parameter model *Generalized Poisson Regression* dilakukan dengan metode *Maximum Likelihood Estimastion*. Prosedur estimasi parameter untuk model *Generalized Poisson Regression* menggunakan MLE adalah sebagai berikut.

a) Menentukan bentuk fungsi *likelihood* untuk model GPR adalah sebagai berikut.

$$L(\boldsymbol{\beta}, \theta) = \prod_{i=1}^{n} \left[\frac{\lambda_i}{1+\theta\lambda_i}\right]^{y_i} \prod_{i=1}^{n} \frac{(1+\theta y_i)^{y_i-1}}{y_i!} exp\left[\sum_{i=1}^{n} \frac{-\lambda_i(1+\theta y_i)}{1+\theta\lambda_i}\right] \tag{2.22}$$

b) Selanjutnya membentuk fungsi logaritma natural dari fungsi *likelihood* sebagai berikut.

$$lnL(\boldsymbol{\beta}, \theta) = ln\left\{\prod_{i=1}^{n} \left[\frac{\lambda_i}{1+\theta\lambda_i}\right]^{y_i} \prod_{i=1}^{n} \frac{(1+\theta y_i)^{y_i-1}}{y_i!} exp\left[\sum_{i=1}^{n} \frac{-\lambda_i(1+\theta y_i)}{1+\theta\lambda_i}\right]\right\}$$

$$\begin{aligned} lnL(\boldsymbol{\beta}, \theta) = \sum_{i=1}^{n} \Big\{ & y_i \ln(\lambda_i) \\ & - y_i \ln(1+\theta\lambda_i) + (y_i - 1)\ln(1+\theta y_i) - ln(y_i!) \\ & - \frac{\lambda_i(1+\theta y_i)}{1+\theta\lambda_i}\Big\} \end{aligned} \tag{2.23}$$

Mensubtitusikan nilai $\lambda_i = \exp\left(\boldsymbol{X_i^T\beta}\right)$ maka diperoleh persamaan sebagai berikut.

$$lnL(\boldsymbol{\beta}, \theta) = \sum_{i=1}^{n}\Big\{y_i \ln\big(\exp(\boldsymbol{X_i^T\beta})\big) - y_i \ln\big(1 + \theta\exp(\boldsymbol{X_i^T\beta})\big) + (y_i - 1)\ln(1 + \theta y_i) - ln(y_i!) - \frac{\exp(\boldsymbol{X_i^T\beta})(1+\theta y_i)}{1+\theta\exp(\boldsymbol{X_i^T\beta})}\Big\}$$

$$lnL(\boldsymbol{\beta}, \theta) = \sum_{i=1}^{n}\Big\{y_i(\boldsymbol{X_i^T\beta}) - y_i \ln\big(1 + \theta\exp(\boldsymbol{X_i^T\beta})\big) + (y_i - 1)\ln(1 + \theta y_i) - ln(y_i!) - \exp(\boldsymbol{X_i^T\beta})(1+\theta y_i)(1+\theta\exp(\boldsymbol{X_i^T\beta}))^{-1}\Big\} \tag{2.24}$$

c) Kemudian dari fungsi ln *likelihood* diturunkan terhadap $\boldsymbol{\beta}^T$ dan disamadengankan nol. Bentuk turunan yang dihasilkan adalah sebagai berikut.

$$\frac{\partial lnL(\boldsymbol{\beta},\theta)}{\partial \boldsymbol{\beta}^{\boldsymbol{T}}} = \sum_{i=1}^{n} \Big\{ y_i \boldsymbol{X}_i - \theta y_i exp\big(\boldsymbol{X}_{\boldsymbol{i}}^{\boldsymbol{T}} \boldsymbol{\beta}\big)\big(1 + \theta \exp\big(\boldsymbol{X}_{\boldsymbol{i}}^{\boldsymbol{T}} \boldsymbol{\beta}\big)\big)^{-1} - (1 + \theta y_i) \Big\{ \boldsymbol{X}_i \exp\big(\boldsymbol{X}_{\boldsymbol{i}}^{\boldsymbol{T}} \boldsymbol{\beta}\big) \big(1 + \theta \exp\big(\boldsymbol{X}_{\boldsymbol{i}}^{\boldsymbol{T}} \boldsymbol{\beta}\big)\big)^{-1} - \theta \boldsymbol{X}_i \big(\exp\big(\boldsymbol{X}_{\boldsymbol{i}}^{\boldsymbol{T}} \boldsymbol{\beta}\big)\big)^2 \big(1 + \theta \exp\big(\boldsymbol{X}_{\boldsymbol{i}}^{\boldsymbol{T}} \boldsymbol{\beta}\big)\big)^{-2} \Big\} \Big\} \quad (2.25)$$

d) Persamaan 2.23 diturunkan terhadap θ dan disamadengankan nol. Sehingga bentuk turunan yang dihasilkan sebagai berikut.

$$\frac{\partial lnL(\boldsymbol{\beta},\theta)}{\partial \boldsymbol{\theta}} = \sum_{i=1}^{n} \Big\{ y_i exp\big(\boldsymbol{X}_{\boldsymbol{i}}^{\boldsymbol{T}} \boldsymbol{\beta}\big)\big(1 + \theta \exp\big(\boldsymbol{X}_{\boldsymbol{i}}^{\boldsymbol{T}} \boldsymbol{\beta}\big)\big)^{-1} + y_i(y_i - 1)(1 + \theta y_i)^{-1} - exp\big(\boldsymbol{X}_{\boldsymbol{i}}^{\boldsymbol{T}} \boldsymbol{\beta}\big) \Big\{ y_i \big(1 + \theta \exp\big(\boldsymbol{X}_{\boldsymbol{i}}^{\boldsymbol{T}} \boldsymbol{\beta}\big)\big)^{-1} - (1 + \theta y_i) \exp\big(\boldsymbol{X}_{\boldsymbol{i}}^{\boldsymbol{T}} \boldsymbol{\beta}\big) \big(1 + \theta \exp\big(\boldsymbol{X}_{\boldsymbol{i}}^{\boldsymbol{T}} \boldsymbol{\beta}\big)\big)^{-2} \Big\} \Big\} \quad (2.26)$$

Penyelesaian dengan metode MLE (*Maximum Likelihood Estimation*) tidak dapat dilakukan dengan metode numerik biasa, sehingga digunakan iterasi numerik untuk mendapatkan penyelesaian dari MLE dengan metode *Newton-Raphson*

2.6.2 Pengujian Parameter *Generalized Poisson Regression*

Pengujian parameter model *Generalized Poisson Regression* sama dengan pengujian parameter yang digunakan dalam model regresi Poisson yaitu dengan metode *Maximum Likelihood Ratio Test* (MLRT). Hipotesis yang digunakan adalah
H_0: tidak ada parameter yang signifikan ($\beta_1 = \beta_2 = \beta_3 = \cdots = \beta_p = 0$)
H_1: terdapat setidaknya satu parameter yang signifikan ($\beta_j \neq 0\ ; j = 1, 2, \dots, p$)

Statistik uji yang digunakan adalah $D(\hat{\beta})$ pada persamaan (2.17). Keputusan yang diambil akan tolak H_0 jika $D(\hat{\beta}) > X_{v;\alpha}^2$ dengan v adalah banyaknya parameter model dibawah populasi dikurangi dengan banyaknya parameter model dibawah H_0. Tolak H_0 berarti ada salah satu variabel yang berpengaruh signifikan terhadap model sehingga dilanjutkan dengan pengujian secara parsial dengan hipotesis sebagai berikut.
H_0: β_j=0
H_1: $\beta_j \neq 0$, $j = 1, 2, \dots, p$

Statistik uji yang digunakan seperti pada persamaan (2.18) dengan daerah penolakan menolak H_0 jika $|z_{hit}| > z_{\frac{\alpha}{2}}$ dengan α adalah tingkat signifikansi.

2.7 *Negative Binomial Regression*

Negative Binomial Regression adalah model non-linear yang berasal dari distribusi Binomial Negatif (*Poisson–gamma mixture*) yang merupakan penerapan dari GLM yang menggambarkan hubungan antar variabel dependen dengan variabel independen. Regresi Binomial Negatif biasanya digunakan untuk memodelkan data dengan variabel dependen berupa data diskrit. Regresi Binomial Negatif digunakan sebagai alternatif dari model regresi Poisson yang digunakan untuk mengatasi masalah overdispersi dalam data.

Overdispersi terjadi ketika ada varians yang besar dalam kumpulan data. Lebih lanjut dapat dijelaskan bahwa ini terjadi ketika varians yang diamati lebih tinggi dari varians model teoritis, maka overdispersi dikatakan terjadi. Di sisi lain, underdispersi hanya berarti bahwa ada lebih sedikit variasi dalam data daripada yang diperkirakan. Overdispersi adalah karakteristik yang sangat umum dalam analisis data terapan karena, dalam praktiknya, populasi adalah sering heterogen (tidak seragam). Salah satu metode untuk menangani overdispersi adalah melakukan pemodelan regresi Binomial Negatif. Fungsi probabilitas pada sebaran Binomial Negatif ditetapkan sebagai berikut.

$$P(y,\mu;\theta) = \frac{\Gamma\left(y_i+\frac{1}{\theta}\right)}{\Gamma\left(\frac{1}{\theta}\right)y_i!}\left(\frac{\theta\mu}{1+\theta\mu}\right)^{y_i}\left(\frac{1}{1+\theta\mu}\right)^{\frac{1}{\theta}} \tag{2.27}$$

Keterangan:
y = variabel dependen bernilai diskrit
θ = derajat overdispersi
μ = parameter

Di mana, *mean* diberikan sebagai, $\lambda_i = E(y_i)$ sedangkan varians y_i dari diberikan sebagai; $var(y_i) = (\lambda_i, \theta\lambda_i)^2$ dan $\theta > 0$ model akan disebut sebagai parameter dispersi. Regresi Binomial Negatif tidak mengasumsikan persamaan *mean* dan varians tetapi mengoreksi overdispersi yang muncul ketika varians lebih besar dari rata-rata.

2.7.1 Estimasi Parameter Model *Negative Binomial Regression*

Metode yang digunakan untuk estimasi parameter dari *Negative Binomial Regression* yaitu metode *Maximum Likelihood Estimate*. Prosedur estimasi parameter untuk model *Negative Binomial Regression* menggunakan MLE adalah sebagai berikut.

a) Fungsi *likelihood* dari model *Negative Binomial Regression* dinyatakan sebagai.

$$L(\mu,\theta) = \prod_{i=1}^{n}\frac{\Gamma\left(y_i+\frac{1}{\theta}\right)}{\Gamma\left(\frac{1}{\theta}\right)y_i!}\left(\frac{\theta\mu}{1+\theta\mu}\right)^{y_i}\left(\frac{1}{1+\theta\mu}\right)^{\frac{1}{\theta}} \tag{2.28}$$

b) Selanjutnya menyusun persamaan logaritma natural dari fungsi *likelihood* sebagai berikut.

$$lnL(\mu,\theta) = \sum_{i=1}^{n}\left[ln\left(\frac{\Gamma\left(y_i+\frac{1}{\theta}\right)}{\Gamma\left(\frac{1}{\theta}\right)y_i!} + y_i\ln\left(\frac{\theta\mu}{1+\theta\mu}\right) + \frac{1}{\theta}\left(\frac{1}{1+\theta\mu}\right)\right)\right]$$

$$lnL(\mu,\theta) = \sum_{i=1}^{n}\left[ln\left(\frac{\Gamma\left(y_i+\frac{1}{\theta}\right)}{\Gamma\left(\frac{1}{\theta}\right)y_i!} + y_i ln(\theta\mu) - \left(y_i+\frac{1}{\theta}\right)ln(1+\theta\mu)\right)\right]$$

$$lnL(\mu,\theta) = \sum_{i=1}^{n}\left[ln\left(\frac{\Gamma\left(y_i+\frac{1}{\theta}\right)}{\Gamma\left(\frac{1}{\theta}\right)y_i!} + \ln(y_i!) + y_i ln\theta + y_i ln\mu - \left(y_i+\frac{1}{\theta}\right)ln(1+\theta\mu)\right)\right] \tag{2.29}$$

Bentuk $\frac{\Gamma\left(y_i+\frac{1}{\theta}\right)}{\Gamma\left(\frac{1}{\theta}\right)y_i!}$ Dapat dijabarkan sebagai berikut.

$$\frac{\Gamma\left(y_i+\frac{1}{\theta}\right)}{\Gamma\left(\frac{1}{\theta}\right)y_i!} = \left(\frac{1}{\theta}+y_i-1\right)\left(\frac{1}{\theta}+y_i-2\right)\ldots.\left(\frac{1}{\theta}+1\right)\left(\frac{1}{\theta}\right)$$

Sehingga.

$$ln\left\{\frac{\Gamma\left(y_i+\frac{1}{\theta}\right)}{\Gamma\left(\frac{1}{\theta}\right)y_i!}\right\} = \sum_{v=0}^{y_i-1}\left(\frac{1}{\theta}+v\right) \tag{2.30}$$

Subtitusi persamaan (2.32) ke persamaan (2.31) maka didapatkan

$$lnL(\mu,\theta) = \sum_{i=1}^{n}\left[\left(\sum_{v=0}^{y_i-1}\ln\left(\frac{1}{\theta}+v\right)\right) - \ln(y_i!) + y_i ln\theta + y_i ln\mu - (y_i + \frac{1}{\theta})\ln(1\theta\mu)\ \right] \quad (2.31)$$

c) Estimasi untuk μ dan θ diturunkan terhadap $\boldsymbol{\beta}^T$ dan disamadengankan nol. Sehingga bentuk turunan yang dihasilkan sebagai berikut.

$$\frac{\partial lnL(\mu,\theta)}{\partial\mu} = \sum_{i=1}^{n}\left(\frac{y_i}{\mu} - \frac{\theta\, y_i+1}{1+\theta\mu}\right) \quad (2.32)$$

$$\frac{\partial lnL(\mu,\theta)}{\partial\mu} = \sum_{i=1}^{n}\left[\left(\sum_{v=0}^{y_i-1} - \frac{1}{\theta(1+\theta v)}\right) + \frac{y_i}{\theta} + \left(\frac{\ln(1+\theta\mu)}{\theta^2} - \frac{\mu(y_i\theta+1)}{\theta(1+\theta\mu)}\right)\right] \quad (2.33)$$

Estimasi parameter model regresi binomial negatif dengan MLE tidak bisa menghasilkan solusi secara langsung sehingga perlu dilakukan iterasi dengan menggunakan metode *Newton-Raphson*.

2.7.2 Pengujian Parameter Model *Negative Binomial Regression*

Pengujian parameter model *Negative Binomial Regression* sama dengan pengujian parameter yang digunakan dalam model regresi Poisson yaitu dengan metode *Maximum Likelihood Ratio Test* (MLRT). Hipotesis yang digunakan adalah
H_0: tidak ada parameter yang signifikan ($\beta_1 = \beta_2 = \beta_3 = \cdots = \beta_p = 0$)
H_1: terdapat setidaknya satu parameter yang signifikan ($\beta_j \neq 0\ ; j = 1, 2, \dots, p$)

Statistik uji yang digunakan adalah $D(\hat{\beta})$ pada persamaan (2.17). Keputusan yang diambil akan tolak H_0 jika $D(\hat{\beta}) > X^2_{v;\alpha}$ dengan *v* adalah banyaknya parameter model dibawah populasi dikurangi dengan banyaknya parameter model dibawah H_0. Tolak H_0 berarti ada salah satu variabel yang berpengaruh signifikan terhadap model sehingga dilanjutkan dengan pengujian secara parsial dengan hipotesis sebagai berikut.
H_0: $\beta_j=0$
H_1: $\beta_j \neq 0$, $j = 1, 2, \dots, p$

Statistik uji yang digunakan seperti pada persamaan (2.18) dengan daerah penolakan menolak H_0 jika $|z_{hit}| > z_{\frac{\alpha}{2}}$ dengan α adalah tingkat signifikansi.

2.8 Pemilihan Model

Akaike Information Criterion (AIC) dan *Bayesian Information Criterion* (BIC), kriteria pemilihan model harus diterapkan pada tiga (3) model data hitungan untuk menentukan model terbaik untuk Kasus COVID-19 di Jawa Timur.

2.8.1 *Akaike Information Criterion* (AIC)

AIC bertujuan untuk mendapatkan model terbaik untuk proses pembangkitan data sebenarnya yang tidak diketahui. Salah satu kriteria informasi yang paling umum digunakan yaitu AIC diperkenalkan oleh Akaike. Idenya adalah untuk memilih model yang meminimalkan kemungkinan jumlah kuadrat galat (Gujarati, 2004). Model regresi terbaik memiliki nilai AIC yang terkecil. Perhitungan nilai AIC dilakukan berdasarkan persamaan berikut (Gujarati, 2004).

$$AIC = exp\left(\frac{2k}{n}\frac{\sum_{i=1}^{n}\varepsilon_i^2}{n}\right) \quad (2.34)$$

Keterangan:
k = banyaknya parameter
$\frac{2k}{n}$ = faktor penalti
ε_i = galat ke-i
n = banyak observasi

2.8.2 *Bayesian Information Criterion* (BIC)

BIC berbeda dari AIC hanya dalam istilah pertama yang tergantung pada ukuran sampel n. Model dengan nilai BIC terkecil merupakan model terbaik. Dari perspektif Bayesian, BIC dirancang untuk menemukan model yang paling mungkin dari data yang diberikan, itu adalah salah satu kriteria informasi yang paling banyak digunakan (Cavanaugh, 2005). Tidak seperti AIC, BIC adalah perkiraan faktor Bayes untuk dua model yang digunakan. BIC ditentukan oleh persamaan berikut.

$$BIC = -2 ln f\left(y\middle|\hat{\theta}_k\right) + k \ln n \tag{2.35}$$

keterangan
$f\left(y\middle|\hat{\theta}_k\right)$ = fungsi maksimum *likelihood*
k = banyaknya parameter model
n = banyaknya observasi

BAB III
METODOLOGI PENELITIAN

3.1 Sumber Data

Data yang digunakan dalam penelitian ini adalah data sekunder. Data sekunder merupakan data yang dikumpulkan oleh pihak lain sehingga sudah mendapatkan data jadi ataupun data setengah jadi dari pihak lain. Data yang digunakan untuk permodelan jumlah korban jiwa COVID-19 yaitu data harian di Jawa Timur dari bulan Oktober 2020 sampai dengan Januari 2022. Data diperoleh dari website COVID-19 indonesia (https://covid19.go.id) dan Dinas Kesehatan Provinsi Jawa Timur.

3.2 Variabel Penelitian

Pada penelitian ini, variabel yang digunakan adalah jumlah korban COVID-19 yang meninggal, kasus aktif COVID-19, kasus baru COVID-19, *Stringency Index,* dan *Bed Occupancy Rate* pada periode Oktober 2020 – Januari 2022, serta memenuhi persyaratan untuk uji *Poisson Regression,* dengan metode *Generalized Poisson Regression* dan *Negative Binomial Regression*. Data yang digunakan merupakan data harian dengan periode Oktober 2020 – Januari 2022. Variabel penelitian yang akan digunakan pada penelitian ini ditunjukkan pada Tabel 3.1.

Tabel 3. 1 Variabel Penelitian

Variabel	Nama Variabel	Skala
Y	Jumlah kasus kematian COVID-19	Rasio
X_1	Jumlah kasus aktif COVID-19	Rasio
X_2	Jumlah kasus baru COVID-19	Rasio
X_3	*Stringency Index*	Rasio
X_4	*Bed Occupancy Rate*	Rasio

Definisi dari variabel penelitian dijelaskan sebagai berikut.

1. Jumlah kasus meninggal COVID-19 di Jawa Timur (Y)
   Jumlah kasus meninggal setiap harinya akibat COVID-19 di Jawa Timur pada periode Oktober 2020 hingga Januari 2022
2. Jumlah kasus aktif COVID-19 di Jawa Timur (X_1)
   Jumlah kasus yang aktif terjadi pada setiap harinya akibat penyebaran COVID-19 di Jawa Timur pada periode Oktober 2020 hingga Januari 2022
3. Jumlah kasus baru COVID-19 di Jawa Timur (X_2)
   Pertambahan kasus baru yang terjadi pada setiap harinya akibat penyebaran COVID-19 di Jawa Timur pada periode Oktober 2020 hingga Januari 2022
4. *Stringency Index* (X_3)
   Persentase perhitungan indeks keketatan pemerintah dengan menggunakan 9 respon, yaitu penutupan sekolah, penutupan tempat kerja, pembatasan acara publik, pembatasan pertemuan publik, pembatasan kendaraan umum, persyaratan tinggal di rumah, kampanye informasi publik, pembatasan pergerakan, dan pembatasan perjalanan internasional

5. *Bed Occupancy Rate* (X_4)
   Persentase pemakaian tempat tidur pada satuan waktu tertentu dalam penelitian ini dari Oktober 2020 hingga Januari 2022. Indikator ini memberikan gambaran tinggi rendahnya tingkat pemanfaatan tempat tidur rumah sakit

3.3 Stuktur Data

Struktur data yang digunakan dalam penelitian mengenai pemodelan jumlah kasus meninggal akibat COVID-19 di Provinsi Jawa Timur adalah sebagai berikut.

Tabel 3.2 Struktur Data Penelitian

Observasi	$\boldsymbol{X_1}$	$\boldsymbol{X_2}$	$\boldsymbol{X_3}$	$\boldsymbol{X_4}$
$\boldsymbol{Y_1}$	X_{11}	X_{21}	X_{31}	X_{41}
$\boldsymbol{Y_2}$	X_{12}	X_{22}	X_{32}	X_{42}
$\boldsymbol{Y_n}$	X_{1n}	X_{2n}	X_{3n}	X_{4n}

3.4 Tahapan Analisis

Metode analisis yang digunakan dalam penelitian ini adalah metode Regresi Poisson menggunakan *Generalized Poisson Regression* dan *Negative Binomial Regression* untuk mengetahui model terbaik untuk kasus COVID-19 di Jawa Timur pada periode Oktober 2020 – Januari 2022. Proses analisis data dilakukan dengan menggunakan *software RStudio*. Berikut merupakan tahapan analisis dalam penelitian ini.

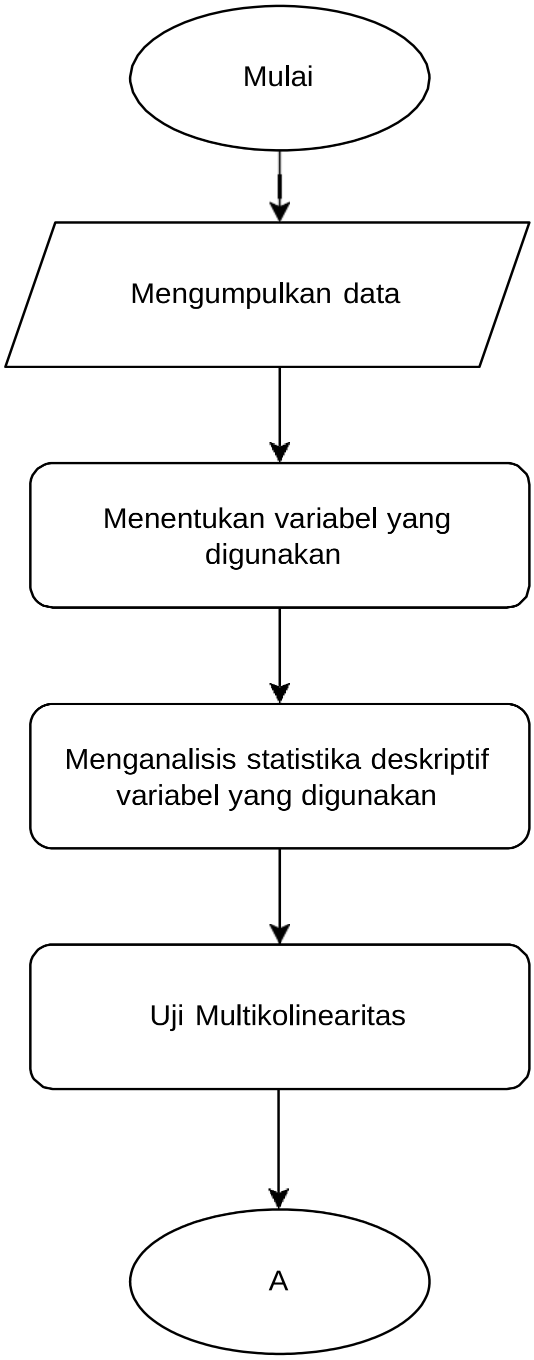
1. Mengumpulkan data variabel yang akan digunakan dalam pemodelan jumlah kasus meninggal akibat COVID-19 di Provinsi Jawa Timur melalui website https://covid19.go.id dan Dinas Kesehatan Provinsi Jawa Timur periode Oktober 2020 – Januari 2022
2. Menentukan variabel dependen (variabel *Y*) dan variabel independen (variabel *X)* dari data COVID-19 tersebut
3. Mendeskripsikan karakteristik kasus COVID-19 di Jawa Timur dan faktor-faktor yang diduga mempengaruhi dengan menggunakan analisa statistika deskriptif untuk variabel dependen dan variabel independen.
4. Menganalisis korelasi antar variabel-variabel independen untuk mendeteksi adanya kasus multikolinieritas dengan menggunakan nilai VIF (*Variance Inflation Factors*).
5. Mendapatkan estimasi model terbaik untuk regresi Poisson dengan langkah sebagai berikut.
   i. Mengestimasi parameter model regresi Poisson nerdasarkan persamaan (2.9)
   ii. Melakukan uji signifikansi secara serentak berdasarkan persamaan (2.12) dan uji parsial berdasarkan persamaan (2.13) terhadap parameter model regresi Poisson.
   iii. Memperoleh estimasi model regresi Poisson serta mendapatkan nilai AIC berdasarkan persamaan (2.18) dan nilai BIC berdasarkan persamaan (2.19).
6. Mendeteksi adanya kasus overdispersi maupun underdispersi pada data dengan melihat nilai *Pearson Chi-squares* berdasarkan persamaan (2.14) dan *Deviance* berdasarkan persamaan (2.15) yang dibagi derajat bebasnya.
7. Mendapatkan estimasi model terbaik menggunakan *Generalized Poisson Regression*. Berikut adalah langkah-langkah pemodelan dengan *Generalized Poisson Regression*.

i. Mengestimasi parameter model *Generalized Poisson Regression* berdasarkan persamaan (2.21).
ii. Melakukan uji signifikansi secara serentak berdasarkan persamaan (2.19) dan uji parsial berdasarkan persamaan (2.20) terhadap parameter model *Generalized Poisson Regression*.
iii. Memperoleh estimasi model *Generalized Poisson Regression* serta mendapatkan nilai AIC berdasarkan persamaan (2.18) dan nilai BIC berdasarkan persamaan (2.19).

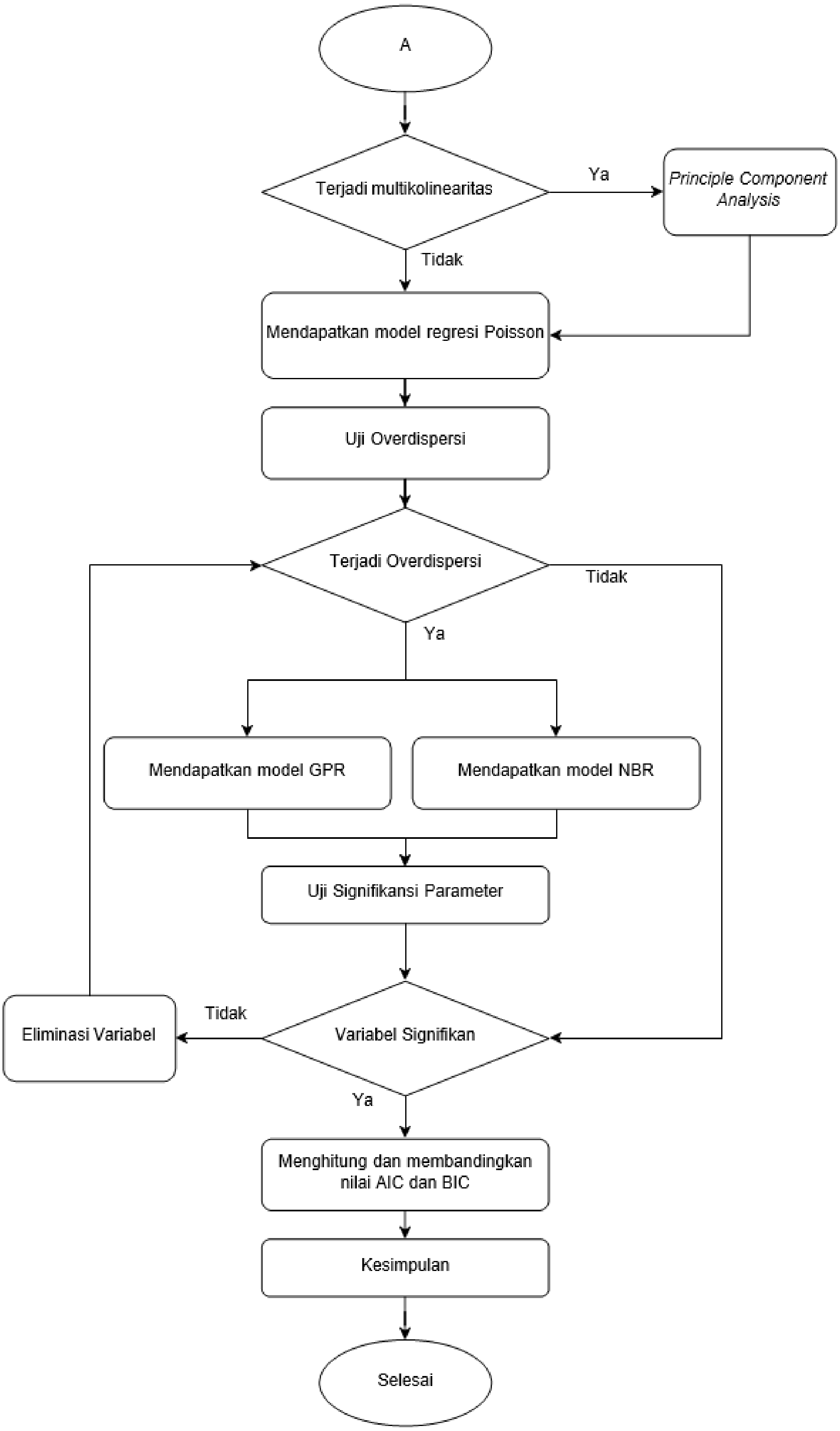
8. Mendapatkan estimasi model terbaik menggunakan *Negative Binomial Regression*. Berikut adalah langkah-langkah pemodelan dengan *Negative Binomial Regression.*
   i. Mengestimasi parameter model *Negative Binomial Regression* berdasarkan persamaan (2.26).
   ii. Melakukan uji signifikansi secara serentak berdasarkan persamaan (2.12) dan uji parsial berdasarkan persamaan (2.13) terhadap parameter model *Negative Binomial Regression*.
   iii. Memperoleh estimasi model *Negative Binomial Regression* serta mendapatkan nilai AIC berdasarkan persamaan (2.18) dan nilai BIC berdasarkan persamaan (2.19).
9. Membandingkan model terbaik hasil *Generalized Poisson Regression* dan *Negative Binomial Regression* menggunakan nilai AIC dan nilai BIC.

3.5 Diagram Alir

Diagram alir pada penelitian "Pemodelan Kasus COVID-19 di Jawa Timur Menggunakan Metode *Generalized Poisson Regression* Dan *Negative Binomial Regression"* sebagai alur perjalanan dalam analisis data pada Gambar 3.1 sebagai berikut.



Gambar 3.1 Diagram Alir Analisis Data



Gambar 3.1 (Lanjutan)

(Halaman ini sengaja dikosongkan)

BAB IV
HASIL DAN PEMBAHASAN

4.1 Ringkasan Kasus COVID-19

Provinsi Jawa Timur merupakan salah satu wilayah yang terdampak cukup parah oleh COVID-19, menurut data yang diambil dari Dinas Kesehatan Jawa Timur sendiri hingga 31 Januari 2022 setidaknya sudah ada 402.879 total kasus dengan 29.774 korban jiwa. Hal ini membuktikan bahwa Provinsi Jawa Timur merupakan salah satu Provinsi dengan jumlah kasus maupun jumlah korban jiwa COVID-19 terbanyak di Indonesia. Banyak faktor yang dapat menyebabkan kasus kematian akibat COVID-19 di Jawa Timur ini meningkat. Salah satu faktor yang dirasa mempengaruhi jumlah harian kumulatif kasus kematian COVID-19, yaitu jumlah harian kumulatif dari kasus aktif, jumlah harian kumulatif dari kasus baru, Stringency Index dan Bed Occupancy Rate. Dari keempat variabel yang telah disebutkan, diambil data yang sesuai periode penelitian, yaitu dalam rentang 1 Oktober 2020 hingga 31 Januari 2022. Data yang digunakan merupakan data sekunder. Kemudian data tersebut dianalisis dengan menggunakan model hitungan seperti Regresi Poisson, Generalized Poisson Regression, dan Negative Binomial Regression.

Tabel 4.1 Statistik Deskripsi Kasus COVID-19

Variabel	**Rata-rata**	**Std. Dev.**	**Min.**	**Max.**
Meninggal	54.47	84.801	0	431
Kasus Aktif	6618.08	11801.784	91	57126
Kasus Baru	735.93	1264.638	5	8230
Stringency Index	66.5437	6.368	50.46	72.69
BOR	41.9905	29.10754	1.40	91.60

Ringkasan statistika deskriptif seperti yang ditunjukkan pada Tabel 4.1 menunjukkan bahwa kasus yang diamati berjumlah 488. Rata-rata jumlah kasus kematian harian akibat COVID-19 di Jawa Timur berjumlah 54 kasus dengan standar deviasi sebesar 85 kasus. Dengan kasus meninggal paling sedikit berjumlah 0 kasus terjadi pada 2 Desember 2021 dan kasus meninggal terbanyak berjumlah 431 kasus yang terjadi pada tanggal 11 Agustus 2021. Hal ini menunjukkan jika kasus yang terjadi pada periode penelitian tergolong fluktuasi dikarenakan jumlah kasus kematian per harinya cenderung terjadi sebuah periode tertentu.

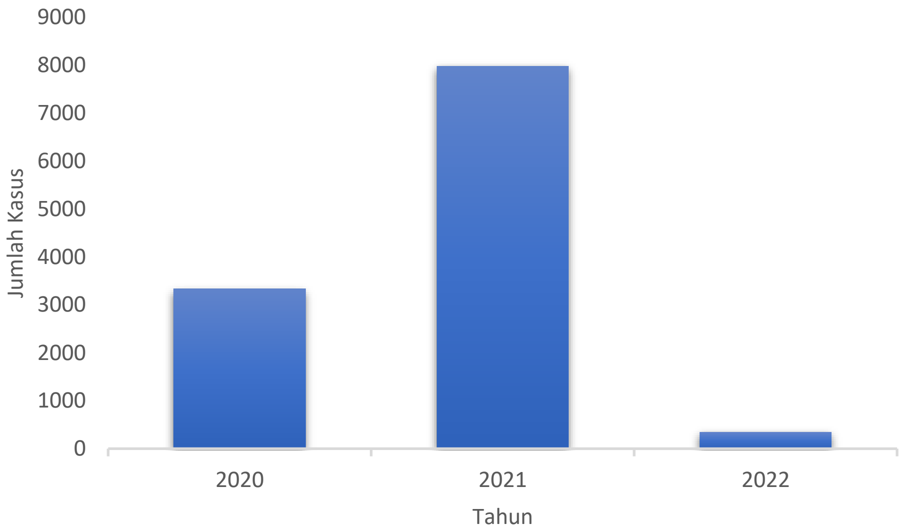
Sedangkan yang terjadi pada kasus aktif, memiliki rata-rata dan standar deviasi kasus didapat dengan jumlah 6618 kasus dan 11802 kasus. Dengan kasus aktif paling sedikit sebesar 91 kasus yang terjadi pada 29 Desember 2021 dan kasus aktif terbanyak sebesar 57126 kasus terjadi pada tanggal 29 Juli 2021. Hal ini menunjukkan jika kasus yang terjadi pada periode penelitian juga mengalami suatu periode dengan periode tertinggi berada pada pertengahan tahun dan melandai pada akhir tahun.

Untuk variabel kasus baru dengan rata-rata berjumlah 736 kasus dan standar deviasi dengan jumlah 1265 kasus yang disajikan pada hasil menunjukkan bahwa bertambahnya kasus baru pada periode penelitian juga cenderung fluktuasi. Ini juga ditunjukkan melalui jumlah kasus baru yang mencapai 8230 kasus yang terjadi pada tanggal 15 Juli 2021, namun pada periode penelitian juga sempat turun hingga hanya terjadi pertambahan 5 kasus pada tanggal

26 Desember 2021. Sama seperti yang terjadi pada kasus meninggal maupun kasus aktif, untuk kasus baru juga memiliki periode yang sama.

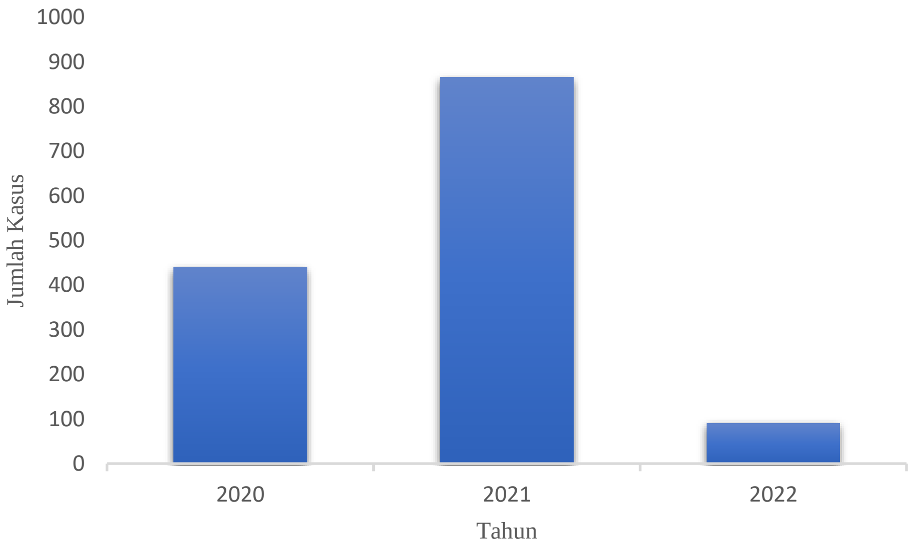
Pemerintah juga masih memberlakukan peraturan yang ketat seperti Pemberlakuan Pembatasan Kegiatan Masyarakat (PPKM) untuk mengurangi jumlah kasus COVID-19. Ini terlihat dari *Stringency Index* dengan jumlah rata-rata 66,543 % dengan standar deviasi sebesar 6,368 % dengan peraturan paling longgar terjadi pada tanggal 11 November 2020 dengan sebesar 50,46 % dan peraturan paling ketat terjadi pada tanggal 5 Oktober 2020 dengan sebesar 72,69 %. Hal ini menggambarkan bahwa pemerintah cenderung masih waspada dalam menangani kasus COVID-19 ini dikarenakan pada periode penelitian kasus COVID-19 sendiri tergolong masih tinggi.

Pada *Bed Occupancy Rate* (BOR) kasus COVID-19 yang disajikan menunjukkan bahwa ketersediaan tempat tidur rumah sakit di Jawa Timur sendiri cenderung di bawah ambang batas aman *Bed Occupancy Rate* yang ditetapkan oleh Organisasi Kesehatan Dunia (WHO) sebesar 60 %, ini dapat ditunjukkan dengan rata-rata sebesar 41.9905 %. Dengan standar deviasi sebesar 29.10754 %. BOR paling sedikit terjadi pada tanggal 3 Januari 2022 dengan jumlah keterisian sebesar 1.4 %. Sedangkan untuk BOR paling tinggi berada pada tanggal 28 Juni 2021 dengan jumlah keterisian sebesar 91.6 %. Ini menggambarkan bahwa *Bed Occupancy Rate* khususnya di Jawa Timur cukup berhasil dalam mengatur keterisian tempat tidur, dapat ditunjukkan pada rata-rata *Bed Occupancy Rate.* Akan tetapi keterisian *Bed Occupancy Rate* sempat mengkhawatirkan dikarenakan pernah mencapai 91.6 %.



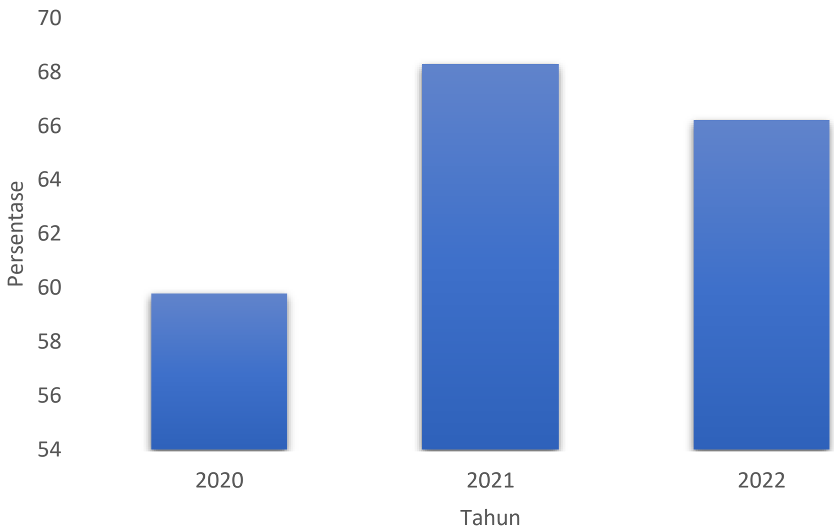
Gambar 4.1. Histogram Rata-Rata Jumlah Kasus Aktif COVID-19 per Tahun

Gambar 4.1 menunjukkan bahwa pada tahun 2021 memiliki rata-rata kasus aktif tertinggi, dengan kasus per harinya berada pada angka 8000 kasus. Tahun 2022 memiliki rata-rata kasus aktif yang terjadi perharinya paling sedikit dibandingkan dengan tahun 2020 dan 2021 dengan jumlah tidak mencapai 1000 kasus per harinya. Tahun 2020 memiliki rata-rata kasus perharinya berada pada kisaran 3000 hingga 4000 kasus. Hal ini juga dapat terjadi dikarenakan jumlah amatan pada tahun 2021 yang lebih banyak disbanding amatan pada tahun 2020 dan tahun 2022. Amatan pada tahun 2020 dimulai pada bulan Oktober 2020 hingga Desember 2020. Amatan pada tahun 2021 dimulai pada Januari 2021 hingga Desember 2021. Tahun 2022 merupakan tahun dengan jumlah amatan paling sedikit dikarenakan hanya memakai data pada bulan Januari 2022 saja.



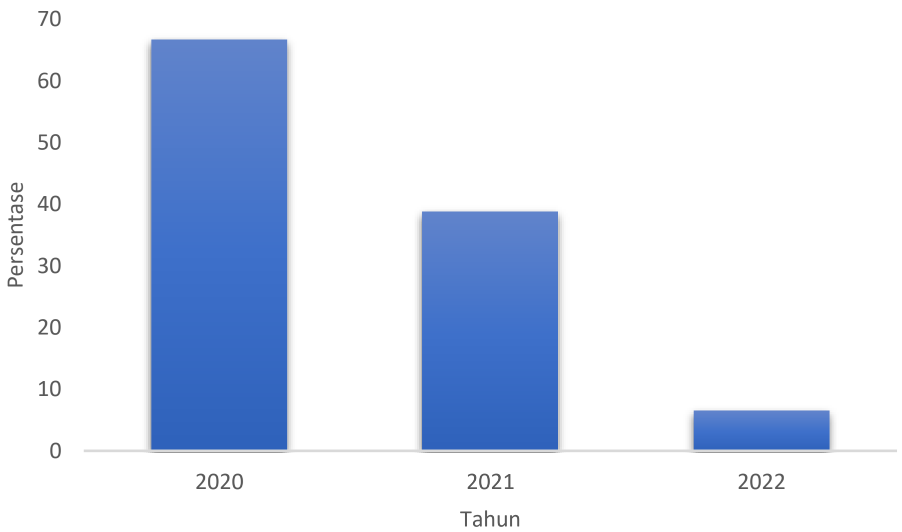
Gambar 4.2 Histogram Rata-Rata Jumlah Kasus Baru COVID-19 per Tahun

Gambar 4.2 menunjukkan bahwa pada tahun 2021 memiliki rata-rata kasus baru tertinggi, dengan kasus baru per harinya berada pada angka 400 kasus hingga 500 kasus. Tahun 2022 memiliki rata-rata kasus baru yang terjadi perharinya paling sedikit dibandingkan dengan tahun 2020 dan 2021 dengan jumlah tidak mencapai 100 kasus per harinya. Tahun 2020 memiliki rata-rata kasus baru perharinya berada pada kisaran 400 hingga 500 kasus. Hal ini juga dapat terjadi dikarenakan jumlah amatan pada tahun 2021 yang lebih banyak disbanding amatan pada tahun 2020 dan tahun 2022. Amatan pada tahun 2020 dimulai pada bulan Oktober 2020 hingga Desember 2020. Amatan pada tahun 2021 dimulai pada Januari 2021 hingga Desember 2021. Tahun 2022 merupakan tahun dengan jumlah amatan paling sedikit dikarenakan hanya memakai data pada bulan Januari 2022 saja**.**



Gambar 4.3 Histogram Rata-Rata *Stringency Index* COVID-19 per Tahun

Gambar 4.3 menunjukkan bahwa pada tahun 2021 memiliki rata-rata *Stringency Index* tertinggi, dengan rata-rata persentase per harinya berada pada 68%. Tahun 2020 memiliki rata-rata *Stringency Index* yang terjadi perharinya paling sedikit dibandingkan dengan tahun 2021 dan 2022 dengan persentase tidak mencapai 60% per harinya. Tahun 2021 memiliki rata-rata *Stringency Index* perharinya berada pada kisaran 66%. Hal ini menunjukkan pada tahun 2020 dimana merupakan fase awal masuknya COVID-19 di Jawa Timur, pemerintah belum begitu serius dalam menangani penyebaran kasus COVID-19 ini, dapat dilihat dengan rata-rata *Stringency Index* yang tidak tinggi**.** Berbanding terbalik pada tahun 2021 yang meningkat jika dibandingkan dengan tahun 2020. Ini dapat dijelaskan jika pemerintah Jawa Timur semakin serius dalam mencegah penyebaran kasus COVID-19. Pada tahun 2022 hal ini berlanjut di mana rata-rata persentase berada pada jumlah 66%.



Gambar 4.4 Histogram Rata-Rata BOR COVID-19 per Tahun

Gambar 4.4 menunjukkan bahwa rata-rata dari BOR pada kasus COVID-19 tertinggi berada pada tahun 2020 dengan persentase rata-rata BOR berada pada nilai 60% hingga 70%, nilai ini mengalami penurunan tahun 2021 dengan persentase rata- rata BOR sebesar 38%. Nilai rata-rata pada tahun 2022 mengalami penurunan yang cukup signifikan dengan persentase rata-rata BOR tidak sampai 10%, ini juga dapat disebabkan pada jumlah amatan yang hanya pada bulan Januari 2022. Tingkat keterisian tempat tidur dengan ini menunjukkan bahwa setiap tahunnya mengalami penurunan, hal ini baik dikarenakan pada persentase rata-rata BOR ditahun 2020 pun tidak jauh melebihi nilai BOR yang disarankan oleh WHO sebesar 60%.

4.2 Pemodelan Kasus COVID-19 Menggunakan Regresi Poisson

Kasus kematian yang dikonfirmasi akibat COVID-19 di Jawa Timur merupakan data acuan, dan jenis data ini mengikuti distribusi Poisson. Pemodelan dengan analisis Regresi Poisson dilakukan untuk mengetahui faktor-faktor yang mempengaruhi jumlah kematian terkait COVID-19. Metode *Maximum Likelihood Estimation* (MLE) digunakan untuk mendapatkan estimasi parameter dari model Regresi Poisson.

4.2.1 Uji Multikolinearitas Kasus COVID-19

Salah satu syarat untuk pemodelan regresi adalah dengan tidak terjadinya multikolinearitas pada data. Multikolinearitas dapat terjadi jika nilai VIF lebih besar dari 10. Nilai VIF untuk masing masing variabel independent dapat dilihat pada Tabel 4.2.

Tabel 4.2 Nilai VIF dari Variabel Independen

Variabel Prediktor	**VIF**
Aktif	5,695
Kasus Baru	6,039
Stringency Index	1,125
BOR	1,198

Berdasarkan nilai VIF masing-masing variabel independen yang disajikan pada Tabel 4.2 dapat disimpulkan bahwa semua variabel independen memiliki nilai $VIF < 10$. Dengan kata lain pada seluruh variabel independen tidak ditemukan adanya kasus multikolinearitas, sehingga asumsi multikolinearitas terpenuhi dan semua variabel independen dapat dimasukkan dalam analisis selanjutnya.

4.2.2 Model Regresi Poisson

Regresi Poisson merupakan metode statistika yang digunakan untuk mengetahui hubungan antara variabel dependen dengan variabel independen. Pemodelan pada regresi Poisson dilakukan dengan cara meregresikan semua kombinasi dari seluruh variabel yang dianalisis. Pemilihan model regresi Poisson terbaik dapat dilihat berdasarkan nilai AIC yang terkecil. Berikut adalah pemodelan regresi Poisson pada kasus COVID-19 di Jawa Timur. Selanjunya dilakukan pengujian secara serentak dan pengujian secara parsial pada data dengan hasil sebagai berikut.

Pengujian serentak dilakukan untuk mengetahui apakah terdapat pengaruh antara variabel independen terhadap variabel dependen dengan menggunakan hipotesis adalah sebagai berikut:
H_0: tidak ada parameter yang signifikan $(\beta_1 = \beta_2 = \beta_3 = \cdots = \beta_p = 0)$
H_1: terdapat setidaknya satu parameter yang signifikan $\left(\beta_j \neq 0\right)$

Perbandingan antara nilai *Likelihood Ratio Chi-square* sebesar 34088 dengan nilai $X^2_{v;\alpha}$ sebesar 9,4877 pada model regresi Poisson, maka keputusan yang didapatkan dari pengujian parameter secara serentak adalah tolak H_0. Dikarenakan nilai $D(\hat{\beta}) > X^2_{v;\alpha}$. Sehingga dapat disimpulkan bahwa terdapat minimal satu variabel independen yang berpengaruh signifikan terhadap variabel dependen. Setelah dilakukan pengujian secara serentak selanjutnya dilakukan pengujian secara parsial dengan hasil pengujian sebagai berikut.

Tabel 4.3 Estimasi Parameter Pada Regresi Poisson

Parameter	**Koefisien**	**Std.eror**	**Z**	**Pvalue**
Konstanta	-1.657	0.1227	-13.506	0.000
Kasus Aktif	4.774×10^{-5}	6.249×10^{-7}	76.404	0.000
Kasus Baru	-2.907×10^{-5}	6.137×10^{-6}	-4.737	0.000
Stringency Index	0.067	1.798×10^{-5}	37.211	0.000
BOR	0.009	2.93×10^{-4}	31.486	0.000

Dari pemodelan data menggunakan regresi Poisson, didapatkan nilai estimasi parameter yang terlampir berdasarkan Tabel 4.3 tersebut dilakukan juga uji parsial. Pengujian parameter secara parsial dilakukan dengan hipotesis sebagai berikut:
H_0: β_j=0
H_1: $\beta_j \neq 0$, $j = 1, 2, \ldots, p$

Mengacu pada Tabel 4.4 yang menunjukkan bahwa dapat dilihat semua nilai parameter Z_{hit} lebih besar dari $-Z\alpha/_2$ atau lebih kecil dari $Z\alpha/_2$, sehingga keputusan yang diambil adalah tolak H_0. Sehingga dapat disimpulkan bahwa parameter berpengaruh signifikan terhadap model. Dari hasil pengujian secara serentak dan juga pengujian parsial tersebut maka estimasi model regresi Poisson yang dihasilkan sebagai berikut.

$$\hat{\lambda}_i = \exp\left(-1.657 + 0.000048\, X_{i1} - 0.000029\, X_{i2} + 0.067\, X_{i3} + 0.009\, X_{i4}\right)$$

Selanjutnya dapat dijelaskan melalui estimasi model tersebut maka variabel kasus aktif, kasus baru, *Stringency Index* dan *Bed Occupancy Rate* berpengaruh terhadap kasus kematian akibat COVID-19 ini. Lebih lanjut dapat dijelaskan bahwa dengan peningkatan setiap 1% pada variabel kasus aktif, *Stringency Index* dan *Bed Occupancy Rate* maka kasus meninggal COVID-19 bertambah masing-masing sebanyak exp(0.000048) atau sebanyak 1 kasus, exp(0.067) atau sebanyak 1.07 kasus dan exp(0.009) atau sebanyak 1 kasus dengan syarat bahwa nilai variabel lain konstan. Sedangkan pada peningkatan setiap 1% pada variabel kasus baru maka jumlah kasus meninggal COVID-19 bertambah sebanyak exp(-0.000029) atau sebanyak 0.99 kasus dengan syarat bahwa nilai variabel lain konstan.

Pengujian serentak dan pengujian parsial menunjukkan bahwa semua variabel berpengaruh signifikan terhadap jumlah kasus meninggal akibat COVID-19 di Jawa Timur dengan peningkatan maupun penurunan jumlah kasus kematian COVID-19 bergantung pada koefisien masing-masing variabel tersebut. Selain, itu dari model regresi Poisson tersebut menghasilkan analisis lain menggunakan kriteria AIC dan BIC. Hasil yang didapatkan kriteria AIC dari estimasi model regresi Poisson sebesar 10217,412 dan didapatkan juga nilai BIC sebesar 10238,363.

Dalam pemodelan regresi Poisson, juga terdapat asumsi ekuidispersi bahwa nilai *mean* dan varians harus sama atau *equidispersion*. Namun, asumsi ini jarang terpenuhi karena overdispersi hadir dalam banyak kasus. Untuk mendeteksi overdispersi, dapat dilihat melalui nilai devians apabila dibagi dengan derajat kebebasan harus lebih besar dari 1 tetapi ketika pembagian nilai deviance dengan derajat kebebasan (d.f.) kurang dari 1 maka akan terjsdi underdispersi.

Tabel 4.4 Nilai Deviasi dan Pearson Pada Model Regresi Poisson

Kriteria	**Nilai**	**d.f.**	**Nilai/df**
Deviasi	7871.759	483	16.298
Pearson Chi-Square	8393.767	483	17.378

Tabel 4.4 menunjukkan bahwa nilai yang didapat dari pembagian antara nilai deviasi dengan derajat kebebasan (d.f) sebesar 16.298 dan nilai *Pearson chi-square* dibagi dengan derajat kebebasan (d.f) sebesar 17.378. Jika dilakukan uji overdispersi, maka didapatkan hasil bahwa kedua nilai tersebut lebih besar dari 1. Sehingga dapat disimpulkan pada model Regresi Poisson, bahwa pemodelan pada kasus meninggal COVID-19 terjadi overdispersi. Overdispersi akan menyebabkan model dengan estimasi parameter yang bias. Untuk mengatasi masalah tersebut, dilakukan pemodelan dengan menggunakan *Negative Binomial Regression* dan *Generalized Poisson Regression* karena kedua metode tersebut dapat mengakomodasi parameter dispersi.

4.3 Pemodelan Kasus COVID-19 Menggunakan *Generalized Poisson Regression*

Metode yang dapat digunakan untuk mengatasi overdispersi pada data hitung adalah *Generalized Poisson Regression*. Karena data jumlah kasus meninggal akibat COVID-19 mengalami overdispersi maka model *Generalized Poisson Regression* dapat digunakan. Berikut merupakan hasil untuk pengujian serentak dan pengujian parsial pada model *Generalized Poisson Regression.*

Pengujian serentak dilakukan dengan menggunakan hipotesis sebagai berikut.

H_0: tidak ada parameter yang signifikan ($\beta_1 = \ \beta_2 = \beta_3 = \cdots = \beta_p = 0$)

H_1: terdapat setidaknya satu parameter yang signifikan($\beta_j \neq 0, j = 1, 2, \dots, p$)

Pengujian parameter secara serentak berdasarkan nilai *Likelihood Ratio Chi-square* sebesar 750.6. Dengan melihat nilai *Likelihood Ratio Chi-square* dibandingkan dengan nilai $X^2_{v;\alpha}$ sebesar 9,4877 pada model *Generalized Poisson Regression*, maka keputusan yang diambil adalah tolak H_0. Dikarenakan semua nilai $D(\hat{\beta}) > X^2_{v;\alpha}$. Sehingga dapat disimpulkan bahwa terdapat seluruh variabel independen yang berpengaruh signifikan terhadap variabel dependen.

Tabel 4.5 Estimasi Parameter Pada Generalized Poisson Regression

Parameter	Koefisien	Std.error	Z	Pvalue
Konstanta	-1.559	0.3498	-4.457	0.000
Parameter Dispersi	1.282	0.05838	21.952	0.000
Kasus Aktif	4.957×10^{-5}	2.76×10^{-6}	17.962	0.000
Kasus Baru	-7.214×10^{-5}	2.686×10^{-5}	-2.686	0.007
Stringency Index	0.0369	5.223×10^{-3}	7.079	0.000
BOR	0.01848	1.083×10^{-3}	17.058	0.000

Melalui hasil estimasi dengan menggunakan model *Generalized Poisson Regression* seperti yang terlihat pada Tabel 4.5 yang merupakan hasil estimasi parameter. Selanjutnya dilakukan pengujian parsial untuk model *Generalized Poisson Regression* antara variabel dependen dengan variabel independen. Untuk pengujian parisal menggunakan hipotesis sebagai berikut,

H_0: β_j=0

H_1: $\beta_j \neq 0$; $j = j = 1, 2, \dots, p$

Setelah dilakukannya uji parsial didapatkan hasil bahwa tolak H_0, dapat dilihat melalui nilai Z_{hit} variabel pada Tabel 4.5 model *Generalized Poisson Regression.* Karena tidak terdapat parameter dengan nilai Z_{hit} yang lebih besar dari $-Z\alpha/_2$ atau lebih kecil dari $Z\alpha/_2$. Sehingga dapat ditarik kesimpulkan bahwa parameter berpengaruh signifikan terhadap model. Sehingga dapat disimpulkan bahwa semua variabel berpengaruh signifikan terhadap model. Estimasi model yang didapat dari *Generalized Poisson Regression* sebagai berikut

$$\hat{\lambda}_i = \exp\,(-1.559 + 1.282 + 0.00005\, X_{i1} - 0.00007\, X_{i2} + 0.037\, X_{i3} + 0.018\, X_{i4})$$

Dari pengujian secara parsial dan pengujian secara serentak tersebut mengungkapkan bahwa keempat variabel independent pada kasus kematian COVID-19 berpengaruh signifikan. Ini menegaskan bahwa dalam estimasi model menggunakan *Generalized Poisson Regression*, dalam kasus aktif memiliki pengaruh signifikan pada jumlah kematian yang terkait dengan COVID-19 di Jawa Timur peningkatan 1% dalam kasus aktif dengan syarat variabel lain konstan maka dapat menyebabkan peningkatan sebesar exp (0.00005) atau sebanyak 1 kasus dalam jumlah kematian terkait COVID-19.

Pada kasus baru memiliki pengaruh signifikan pada jumlah kematian yang terkait dengan COVID-19 di Jawa Timur dengan setiap penambahan 1% dalam kasus baru dengan syarat variabel lain konstan maka dapat menyebabkan peningkatan sebesar exp(-0.00007) atau sebanyak 0.99 kasus dalam jumlah kematian terkait COVID-19.

Pada *Stringency Index* memiliki pengaruh signifikan pada jumlah kematian yang terkait dengan COVID-19 di Jawa Timur dengan setiap penambahan 1% dalam *Stringency Index* dengan syarat variabel lain konstan maka dapat menyebabkan peningkatan sebesar exp(0.037) atau sebanyak 1.04 kasus dalam jumlah kematian terkait COVID-19.

Peningkatan 1% dalam *Bed Occupancy Rate* di Jawa Timur selama periode yang diselidiki dapat menyebabkan peningkatan sebesar exp(0.018) atau sebanyak 1.02 kasus jumlah kematian melalui COVID-19 di Jawa Timur, hasilnya menunjukkan bahwa *Bed Occupancy Rate* memiliki pengaruh signifikan terhadap jumlah kematian COVID-19. Dapat disimpulkan bahwa seluruh variabel merupakan variabel yang berpengaruh signifikan terhadap model. Selain itu, dari model regresi tersebut dihasilkan analisis lain menggunakan kriteria AIC dan BIC. Hasil yang didapatkan kriteria AIC dari estimasi model *Generalized Poisson Regression* sebesar 4061,512 dan didapatkan juga nilai BIC sebesar 4086,654.

4.4 Pemodelan Kasus COVID-19 Menggunakan *Negative Binomial Regression*

Untuk mengatasi masalah overdispersi dengan pemodelan Regresi Poisson dari data hitungan, juga dapat dilakukan dengan menggunakan model *Negative Binomial Regression* karena model *Negative Binomial Regression* juga dapat mengakomodasi dan menangkap parameter dispersi dalam kasus ini terjadi overdispersi pada pemodelan data hitungan. Sebelum dilakukan pemodelan data menggunakan *Negative Binomial Regression* dilakukan pengujian secara serentak dan pengujian secara parsial terhadap data.

Pengujian serentak untuk model *Negative Binomial Regression* antara variabel dependen dengan variabel independen, dilakukan dengan menggunakan hipotesis sebagai berikut.
H_0: tidak ada parameter yang signifikan ($\beta_1 = \beta_2 = \beta_3 = \cdots = \beta_p = 0$)
H_1: terdapat setidaknya satu parameter yang signifikan $\left(\beta_j \neq 0, j = 1, 2, \ldots, p\right)$

Pengujian parameter secara serentak dilakukan dengan melihat nilai *Likelihood Ratio Chi-square* pada model *Negative Binomial Regression* sebesar 697.2 dibandingkan dengan nilai $X^2_{v;\alpha}$ dari alpha (0.05) sebesar 9,4877, maka keputusan yang diambil tolak H_0. Dikarenakan nilai $D(\hat{\beta}) > X^2_{v;\alpha}$ sehingga dapat disimpulkan bahwa terdapat minimal satu variabel independen yang berpengaruh signifikan terhadap variabel dependen pada model *Negative Binomial Regression*. Selanjutnya dilakukan pengujian parsial terhadap model *Negative Binomial Regression.*

Tabel 4.6 Estimasi Parameter Pada Negative Binomial Regression

Parameter	**Koefiisien**	**Std.error**	**Z**	**Pvalue**
Konstanta	-1.557	0.349	-4.457	0.000
Kasus Aktif	8.661×10^{-5}	5.968×10^{-6}	14.512	0.000
Kasus Baru	-1.899×10^{-4}	5.737×10^{-5}	-3.31	0.000
Stringency Index	5.382×10^{-2}	5.294×10^{-3}	10.166	0.000
BOR	2.003×10^{-2}	1.191×10^{-3}	16.821	0.000

Hasil estimasi parameter dengan menggunakan model *Negative Binomial Regression* seperti yang terlihat pada Tabel 4.6 dapat dijadikan acuan dalam pengujian parsial. Dengan hipotesis yang digunakan pada pengujian parsial adalah sebagai berikut.
H_0: $\beta_j = 0$
H_1: $\beta_j \neq 0$; $j = 1, 2, \ldots, p$

Setelah dilakukannya uji parsial jika dilihat melalui nilai Z_{hit} pada Tabel 4.6 didapatkan hasil bahwa pada model *Negative Binomial Regression* tolak H_0 karena tidak terdapat parameter dengan nilai Z_{hit} yang lebih besar dari $-Z\alpha/_2$ atau lebih kecil dari $Z\alpha/_2$. Sehingga dapat disimpulkan bahwa semua variabel berpengaruh signifikan terhadap model. Estimasi model yang didapat dari Regresi Binomial Negatif sebagai berikut.

$$\hat{\lambda}_i = \exp\left(-1.557 + 0.000087\, X_{i1} - 0.00019\, X_{i2} + 0.054\, X_{i3} + 0.02\, X_{i4}\right)$$

Pengujian serentak maupun pengujian parsial menegaskan bahwa dalam memperkirakan koefisien menggunakan model *Negative Binomial Regression* seluruh variabel berpengaruh signifikan terhadap model, pada kasus aktif memiliki pengaruh signifikan pada jumlah kematian yang terkait dengan COVID-19 di Jawa Timur dengan setiap penambahan 1% dalam kasus aktif dengan syarat variabel lain konstan maka dapat menyebabkan peningkatan sebesar exp(0.000087) atau sebanyak 1 kasus dalam jumlah kematian terkait COVID-19.

Dalam kasus baru memiliki pengaruh signifikan pada jumlah kematian yang terkait dengan COVID-19 di Jawa Timur peningkatan 1% dalam kasus baru dengan syarat variabel lain konstan maka dapat menyebabkan penurunan sebesar exp(-0.00019) atau sebesar 1 kasus dalam jumlah kematian terkait COVID-19.

Pada *Stringency Index* memiliki pengaruh signifikan pada jumlah kematian yang terkait dengan COVID-19 di Jawa Timur dengan setiap penambahan 1% dalam *Stringency Index* dengan syarat variabel lain konstan maka dapat menyebabkan peningkatan sebesar exp(0.054) atau sebanyak 1.06 kasus dalam jumlah kematian terkait COVID-19.

Pada *Bed Occupancy Rate* memiliki pengaruh yang signifikan, berarti bahwa peningkatan 1% dalam *Bed Occupancy Rate* di Jawa Timur selama periode yang diselidiki dapat menyebabkan peningkatan sebesar exp(0.02) atau sebanyak 1.02 kasus jumlah kematian melalui COVID-19 di Jawa Timur. Selain itu, dari model regresi tersebut menghasilkan analisis lain menggunakan kriteria AIC dan BIC. Hasil yang didapatkan kriteria AIC dari estimasi model *Negative Binomial Regression* sebanyak 4103,979 dan didapatkan juga nilai BIC sebesar 4129,121.

4.5 Pemilihan Model Terbaik

Setelah dilakukan pengujian dan pemodelan dari *Poisson Regression, Generalized Poisson Regression* dan *Negative Binomial Regression* dilakukan pemilihan untuk mengetahui model yang lebih baik dalam pemodelan jumlah kasus meninggal untuk COVID-19 di Jawa Timur. Kriteria yang digunakan untuk menentukan model terbaik dipilih berdasarkan perhitungan nilai AIC dan nilai BIC.

Tabel 4.7 Pemilihan Model Terbaik Berdasarkan Model Kasus COVID-19

Model	**AIC**	**BIC**
PR	10217.412	10238.363
NBR	4103.979	4129.121
GPR	4061.512	4086.654

Pada Tabel 4.7 menunjukkan hasil kriteria untuk ketiga data hitung, yaitu; *Poisson Regression, Negative Binomial Regression*, dan *Generalized Poisson Regression*. Hasil penelitian menunjukkan bahwa model yang paling baik untuk pemodelan kasus kematian COVID-19 di Jawa Timur adalah *Generalized Poisson Regression* dikarenakan memiliki nilai AIC (4061.512) dan BIC (4086.654), paling kecil diantara ketiga model yang digunakan. Hal ini menunjukkan bahwa model terbaik untuk memodelkan kasus overdispersi pada regresi

Poisson untuk jumlah kasus meninggal COVID-19 di Jawa Timur adalah *Generalized Poisson Regression.* Estimasi model *Generalized Poisson Regression* didapatkan sebagai berikut.

$$\hat{\lambda}_i = \exp\left(-1.559 + 1.282 + 0.00005\ X_{i1} - 0.00007\ X_{i2} + 0.037\ X_{i3} + 0.018\ X_{i4}\right)$$

Faktor-faktor yang mempengaruhi jumlah kasus meninggal akibat COVID-19 adalah kasus aktif, kasus baru, *Stringency Index,*dan *Bed Occupancy Rate.* Setiap bertambahnya persentase pada variabel kasus aktif, maka akan meningkatkan jumlah kasus kematian pada COVID-19. Dengan kata lain bahwa variabel kasus aktif berhubungan secara langsung dalam bertambahnya kasus meninggal pada COVID-19 di Jawa Timur. Juga dengan variabel *Stringency Index* setiap pertambahan pada variabel *Stringency Index* maka akan maningkatkan juga kasus kematian pada COVID-19 di Jawa Timur. Hal yang sama berlaku bagi variabel *Bed Occupancy Rate.* setiap pertambahan pada variabel *Bed Occupancy Rate* maka akan maningkatkan juga kasus kematian pada COVID-19 di Jawa Timur. Variabel kasus baru, pada variabel ini pertambahan persentase pada kasus baru juga akan menambahkan jumlah kematian pada kasus COVID-19 di Jawa Timur.

Berdasarkan model jumlah kasus meninggal COVID-19 yang diperoleh dari model *Generalized Poisson Regression* apabila variabel kasus aktif, kasus baru, *Stringency Index* dan *Bed Occupancy Rate* bertambah satu satuan maka jumlah kasus meninggal di Provinsi Jawa Timur akan bertambah dari semula. Jika kasus aktif sebesar 1360 kasus. Kasus baru sebesar 276 kasus *Stringency Index* sebesar 66.2% dan *Bed Occupancy Rate* sebesar 21.95% maka diduga jumlah kasus kematian COVID-19 di Provinsi Jawa Timur bertambah setidaknya sebanyak 14 kasus meninggal.

BAB V
KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil dan pembahasan pada bab sebelumnya, maka didapatkan kesimpulan bahwa,

1. Sebanyak 488 amatan dilakukan dan dihasilkan rata-rata dan standar deviasi pada kasus kematian harian akibat COVD-19 sebanyak 54 kasus dan 85 kasus dengan nilai terkecil sebanyak 0 kasus dan nilai terbesar sebanyak 431 kasus. Untuk rata-rata dan standar deviasi pada kasus aktif akibat COVD-19 didapatkan 6618 kasus dan 11802 kasus dengan nilai terkecil sebanyak 91 kasus dan nilai terbesar sebanyak 57126 kasus. Untuk rata-rata dan standar deviasi kasus baru akibat COVD-19 sebanyak 736 kasus dan 1265 kasus dengan nilai terkecil sebanyak 5 kasus dan nilai terbesar sebanyak 8230 kasus. Untuk rata-rata dan standar deviasi *Stringency Index* COVID-19 sebanyak 66.5437 % dan 6.368 % dengan nilai terkecil sebanyak 50.46 % dan nilai terbesar sebanyak 72.69 %. Untuk rata-rata dan standar deviasi *Bed Occupancy Rate* COVID-19 sebanyak 41.99 % dan 29.1 % dengan nilai terkecil sebanyak 1.4 % dan nilai terbesar sebanyak 91.6 %.
2. Dari penelitian ini ditemukan bahwa pengaruh signifikan positif pada variabel kasus aktif, kasus baru, *Stringency Index* dan *Bed Occupancy Rate* terhadap jumlah kematian COVID-19 di Jawa Timur. Setiap penambahan 1% dalam kasus aktif dengan syarat variabel lain konstan maka dapat menyebabkan peningkatan sebesar 1 kasus dalam jumlah kematian terkait COVID-19. Peningkatan 1% dalam kasus baru dengan syarat variabel lain konstan maka dapat menyebabkan peningkatan sebesar 0.99 kasus dalam jumlah kematian terkait COVID-19. Penambahan 1% dalam *Stringency Index* dapat menyebabkan peningkatan sebesar 1.04 kasus dalam jumlah kematian terkait COVID-19. Bahwa dalam peningkatan 1% dalam *Bed Occupancy Rate* di Jawa Timur selama periode yang diselidiki dapat menyebabkan peningkatan sebesar 1.02 kasus jumlah kematian melalui COVID-19 di Jawa Timur.
3. Hasil analisis menunjukkan pada model Regresi Poisson terdapat adanya overdispersi, sehingga metode *Negative Binomial Regression* dan *Generalized Poisson Regression* dapat digunakan untuk mengatasi overdispersi. Diperlihatkan pula bahwa *Generalized Poisson Regression* adalah model terbaik yang dipilih. untuk menentukan faktor-faktor yang mempengaruhi kematian akibat COVID-19 di Jawa Timur ketika terdapat indikasi adanya overdispersi pada data hitungan. Hal ini disimpulkan dari hasil kriteria pemilihan model yang diterapkan yang meliputi AIC, dan BIC, dengan nilai AIC sebesar 4061.512 dan nilai BIC sebesar 4086.654. dari kedua kriteria tersebut menetapkan *Generalized Poisson Regression* sebagai model terbaik karena memiliki nilai terkecil dari kedua kriteria seleksi.

5.2 Saran

Berdasarkan hasil dan temuan dari penelitian ini, rekomendasi berikut dibuat;

1. Saran yang disampaikan kepada Pemerintah Jawa Timur melalui Satuan Tugas COVID-19 pada COVID-19 sebaiknya memperhatikan peningkatan pada kasus aktif, kasus baru, *Stringency Index* dan *Bed Occupancy Rate* karena keempat faktor tersebut merupakan faktor yang dapat meningkatkan kasus kematian terkait COVID-19 di Jawa Timur.
2. Saran untuk penelitian selanjutnya, dapat dilakukan uji autokorelasi pada data jika data merupakan data timeseries.

(Halaman ini sengaja dikosongkan)

DAFTAR PUSTAKA

Adams, S. O., Bamanga, M. A., Olanrewaju, S. O., Yahaya, H. U., dan Akano, R. O. 2020. Modeling COVID-19 cases in Nigeria using some selected count data regression models. International Journal of Healthcare and Medical Sciences 6, 64-73.

Agresti A., *An Introduction to Categorical Data Analysis Second Edition*, 2nd Edition. Hoboken: John Wiley dan Sons Inc., 2007.

Akaike, H., 1973. Information theory and extension of the maximum likelihood principle, In: B.N. Petrov and F. Csaki (eds), international Symposium on Information Theory. 2nd ed. Budapest: Akademiai Kiado. pp. 267-87.

Arisandi, Arwini dan Herdiani, Erna dan Sahriman, Sitti. (2019). Aplikasi Generalized Poisson Regression dalam Mengatasi Overdispersi pada Data Jumlah Penderita Demam Berdarah Dengue. STATISTIKA: Journal of Theoretical Statistics and Its Applications. 18. 123-130. 10.29313/jstatv18i2.4542.

Bain, I Lee & Engelhardt, Max. 1992. *Introduction to Probability of Mathematical Statistics second edition.* Duxbury Press, California

Banapon, A., Putra, M. L. P., dan Widodo, E., 2020, Penerapan Regresi Binomial Negatif Untuk Mengatasi Pelanggaran Overdispersi Pada Regresi Poisson (Studi Kasus Penderita Tuberculosis di Provinsi Jawa Barat Tahun 2017), *Journal Biastatistics Departemen Statistika FMIPA Universitas Padjadjaran*, 14(1), 53-63.

Bickel, P. J. and Doksum, K. (2006). Mathematical Statistics, Basic Ideas and Selected Topics, Vol I, Prentice Hall, Saddle River, NJ.

Cameron, A. C., dan Trivedi, P. K., 2013, *Regression analysis of count data*, Vol. 53, Cambridge university press, New York.

Cavanaugh, J. (2005) *Model Selection, The Schwarz Information Criterion (SIC).* Department Biostatistics. The University of Iowa

Consul, P.C., dan Famoye, F. (1992). Generalized poisson regression model. *Communications in Statistics-theory and Methods, 21*, 89-109.

Draper, N.R. dan Smith, H. (1998). *Applied Regression Analysis.* New York: John Wiley dan Sons.

Famoye, F., 1993. "Restricted generalized poisson regression model. Communications in statistics." Theory and Methods, vol. 22, pp. 1335-1354.

Famoye, F. W., 2004. On The Generalized Poisson Regression Model with an Application to Accident Data. *Journal of Data Science 2*, 287-295.

Feller W. 1967. *An lntroduction Probability Theory and Its Application (vol. 1): 3rd edition*. Willey, New York.

Gujarati, D.N. (2004) Basic Econometrics. 4th Edition, McGraw-Hill Companies.

Gupta dan Kundu, *Generalisasi Exponential Distributions, Austral & Zealand J. Statist.* Vol. 2, No. 41, hal. 173-188, 1999.

Hardin, J. W. dan Hilbe, J. M., 2007. Generalized linier models and extensions. Texas: A Strata Perss Publication.

Hilbe, J. M. (2011). *(2nd ed.). Cambridge University Press.*

Holmes, A., Illowsky, B., dan Dean, S. (2019). Linear Regression and Correlation. In *Introductory Business Statistics*. OSCRice University.

Williamson, B. dan Hogan, A. (2020). Commercialisation and Privatisation in/of Education in the Context of Covid-19. Education International.

Ilpaj, S. M., dan Nurwati, N., 2020, Analisis Pengaruh Tingkat Kematian Akibat COVID-19 Terhadap Kesehatan Mental Masyarakat di Indonesia, Focus: Jurnal Pekerjaan Sosial, 3(1), 16-28.

Ismail, N. dan Jemain, A.A., 2007. Handling Overdispersion with Binomial Negative and Generalized Poisson Regression Models. Malaysia

Kementerian Kesehatan RI. (2020). Pedoman Pencegahan dan Pengendalian Coronavirus Disease (COVID-19). Kementerian Kesehatan RI.

Keputusan Menteri Kesehatan RI tentang Petunjuk Teknis Pelaksanaan Vaksinasi dalam Rangka Penanggulangan Pandemi Corona Virus Disease 2019 (COVID-19), Pub. L. No. HK.01.07/MENKES/4638/2021, Kementerian Kesehatan RI (2021).

Khoshgoftaar T.M., Gao K., and Szabo R. M., "Comparing Software Fault Predictions of Pure and Zero-inflated Poisson Regression Models," *International Journal of Systems Science*, vol. 36, no. 11, pp. 705–715, Sep. 2005, doi: 10.1080/00207720500159995.

McCullagh, P. dan Nelder, J. A., 1983. Generalized linier models. London: Chapmann and Hall.

Moh, M. A., dan Naing, N. N., 2012, Comparison between negative binomial and poisson death rate regression analysis: AIDS mortality co-infection patients, *IOSR Journal of Mathematics*, 3, 34-38.

Myers, R. H. (1990). *Classical and Modern Regression with Application.* Boston: PWS-KENT Publishing Company.

Nishiura, H., Linton, N. M., dan Akhmetzhanov, A. R., 2020. Serial interval of novel coronavirus (COVID-19) infections. International journal of infectious diseases, 93, 284-286.

Philip, N. and Sebastian, N., 2017. "Application of poisson regression on traffic safety. Degree project, in second level mathematical statistics stockholm." Available: https://www.divaportal.org/smash/get/diva2:816402/FULLTEXT01.pdf

Schwarz, G., 1978. "Estimating the dimension of a model." Annals of Statistic, vol. 6, pp. 461-464.

Tohari, A., dan Chamidah, N., 2019, Modeling of HIV and AIDS in Indonesia Using Bivariate Binomial negative Regression. In IOP Conference Series: Materials Science and Engineering,Vol. 546, No. 5, p. 052079.

Tosepu, R., Gunawan, J., Effendy, D. S., Lestari, H., Bahar, H., dan Asfian, P., 2020, Correlation between weather and COVID-19 pandemic in Jakarta, Indonesia, Science of The Total Environment, 725, 138436.

Velavan T.P. and Meyer C.G., "The COVID-19 Epidemic," *Tropical Medicine and International Health*, vol. 25, no. 3, pp. 278–280, Mar. 2020, doi: 10.1111/tmi.13383.

Walpole, R.E., 1995, Pengantar Statistika Edisi ke-3, Jakarta: Gramedia

WHO (World Health Organization). 2020. 'QdanA on Coronavirus (COVID-19)'. 17 April. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-answershub/q-a-detail/q-a-coronaviruses.

Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., ... dan Tan, W., 2020, A novel coronavirus from patients with pneumonia in China, 2019, New England journal of medicine, 382(8), 727–733.

LAMPIRAN

Lampiran 1. Data kasus COVID-19 di Jawa Timur dan Faktor yang Mempengaruhi

Date	**Y**	**X1**	**X2**	**X3**	**X4**
Oct 1, 2020	3222	3455	314	72.69	62.5
Oct 2, 2020	3240	3360	283	72.69	62.15
Oct 3, 2020	3260	3302	308	72.69	60.77
Oct 4, 2020	3280	3215	249	72.69	59.69
Oct 5, 2020	3302	3145	237	72.69	58.87
Oct 6, 2020	3325	3112	282	72.69	59.13
Oct 7, 2020	3350	3125	331	72.69	58.08
Oct 8, 2020	3374	3114	347	72.69	57.64
Oct 9, 2020	3394	3098	310	72.69	56.98
Oct 10, 2020	3404	3093	310	72.69	56.04
Oct 11, 2020	3425	3069	269	72.69	55.34
Oct 12, 2020	3447	3040	296	50.46	54.59
Oct 13, 2020	3468	2983	315	50.46	54.74
Oct 14, 2020	3485	2935	299	50.46	54.75
Oct 15, 2020	3497	2878	267	50.46	54.58
Oct 16, 2020	3516	2804	291	50.46	54.17
Oct 17, 2020	3529	2706	238	50.46	54.46
Oct 18, 2020	3544	2680	242	50.46	55.13
Oct 19, 2020	3562	2563	242	50.46	56.72
Oct 20, 2020	3582	2536	300	50.46	56.6
Oct 21, 2020	3606	2524	327	50.46	54.81
Oct 22, 2020	3619	2432	268	50.46	55.7
Oct 23, 2020	3631	2374	295	50.46	56.16
Oct 24, 2020	3647	2341	289	50.46	54.8
Oct 25, 2020	3663	2351	268	50.46	54.98
Oct 26, 2020	3683	2331	296	50.46	56.32
Oct 27, 2020	3704	2352	289	50.46	55.89
Oct 28, 2020	3724	2345	246	50.46	55.32
Oct 29, 2020	3740	2343	268	50.46	55.17
Oct 30, 2020	3758	2332	222	50.46	55.48
Oct 31, 2020	3768	2319	223	50.46	55.42
……………	……….	……	…...	…	……...
Jan 1, 2022	29746	98	15	66.2	2.12
Jan 2, 2022	29746	98	8	66.2	1.85
Jan 3, 2022	29746	99	10	66.2	1.4
Jan 4, 2022	29746	92	17	66.2	1.91
Jan 5, 2022	29748	91	12	66.2	2.03
Jan 6, 2022	29748	96	25	66.2	2.34
Jan 7, 2022	29749	94	15	66.2	2.49
Jan 8, 2022	29752	104	29	66.2	2.43
Jan 9, 2022	29753	98	16	66.2	2.19

Lampiran 2. (Lanjutan)

Date	**Y**	**X1**	**X2**	**X3**	**X4**
Jan 10, 2022	29754	98	8	66.2	2.44
Jan 11, 2022	29755	115	29	66.2	2.45
Jan 12, 2022	29757	122	18	66.2	2.74
Jan 13, 2022	29757	133	23	66.2	2.54
Jan 14, 2022	29758	134	21	66.2	2.76
Jan 15, 2022	29758	137	23	66.2	3.1
Jan 16, 2022	29759	147	26	66.2	3.26
Jan 17, 2022	29759	141	14	66.2	4.25
Jan 18, 2022	29760	164	44	66.2	4.6
Jan 19, 2022	29762	190	49	66.2	4.83
Jan 20, 2022	29765	250	89	66.2	5.49
Jan 21, 2022	29766	314	92	66.2	6.33
Jan 22, 2022	29768	346	79	66.2	6.65
Jan 23, 2022	29769	370	69	66.2	7.75
Jan 24, 2022	29769	377	60	66.2	8.9
Jan 25, 2022	29770	531	213	66.2	10.66
Jan 26, 2022	29770	664	238	66.2	12.76
Jan 27, 2022	29771	807	255	66.2	14.6
Jan 28, 2022	29774	955	318	66.2	16.42
Jan 29, 2022	29774	1166	363	66.2	17.32
Jan 30, 2022	29774	1330	359	66.2	20.23
Jan 31, 2022	29774	1360	276	66.2	21.95

Lampiran 3. Model Regresi Poisson

```
> anova(pois1)
Analysis of Deviance Table
Model: poisson, link: log
Response: Y
     Df Deviance Resid. Df Resid. Dev
NULL                   487      41959
X1    1  30944.8       486      11014
X2    1    100.7       485      10914
X3    1   2035.4       484       8878
X4    1   1006.6       483       7872

> summary(pois1)
Call:
glm(formula = Y ~ X1 + X2 + X3 + X4, family = "poisson", data = data1)
Deviance Residuals:
   Min      1Q  Median      3Q     Max
-9.559  -3.330  -1.517   1.307  14.454
Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.657e+00  1.227e-01 -13.506  < 2e-16 ***
X1           4.774e-05  6.249e-07  76.404  < 2e-16 ***
X2          -2.907e-05  6.137e-06  -4.737 2.17e-06 ***
X3           6.692e-02  1.798e-03  37.211  < 2e-16 ***
X4           9.225e-03  2.930e-04  31.486  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
(Dispersion parameter for poisson family taken to be 1)
    Null deviance: 41959.3  on 487  degrees of freedom
Residual deviance:  7871.8  on 483  degrees of freedom
AIC: 10217
Number of Fisher Scoring iterations: 5

> lrtest(pois1)
Likelihood ratio test
Model 1: Y ~ X1 + X2 + X3 + X4
Model 2: Y ~ 1
  #Df   LogLik Df Chisq Pr(>Chisq)
1   5  -5103.7
2   1 -22147.5 -4 34088  < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Lampiran 4. Model *Generalized Poisson Regression*

```
> anova(gpr1)
Analysis of Deviance Table (Type II tests)
Model: 'genpoisson0'
Links: 'loglink', 'logitlink'
Response: Y
   Df 2 * LogLik Diff. Resid. Df  LogLik  Pr(>Chi)
X1  1         197.460      971 -2123.5 < 2.2e-16 ***
X2  1           6.369      971 -2027.9   0.01161 *
X3  1          51.878      971 -2050.7 5.906e-13 ***
X4  1         254.454      971 -2152.0 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> summary(gpr1)
Call:
vglm(formula = Y ~ X1 + X2 + X3 + X4, family = genpoisson0, data = data1,
    trace = TRUE)
Coefficients:
               Estimate Std. Error z value Pr(>|z|)
(Intercept):1 -1.559e+00  3.498e-01  -4.457 8.32e-06 ***
(Intercept):2  1.282e+00  5.838e-02  21.952  < 2e-16 ***
X1             4.957e-05  2.760e-06  17.962  < 2e-16 ***
X2            -7.214e-05  2.686e-05  -2.686  0.00724 **
X3             3.697e-02  5.223e-03   7.079 1.45e-12 ***
X4             1.848e-02  1.083e-03  17.058  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Names of linear predictors: loglink(theta), logitlink(lambda)
Log-likelihood: -2024.756 on 970 degrees of freedom
Number of Fisher scoring iterations: 10
No Hauck-Donner effect found in any of the estimates
> lrtest(gpr1)
Likelihood ratio test
Model 1: Y ~ X1 + X2 + X3 + X4
Model 2: Y ~ 1
  #Df  LogLik Df Chisq Pr(>Chisq)
1 970 -2024.8
2 974 -2400.1  4 750.6  < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Lampiran 5. Model *Negative Binomial Regression*

```
> anova(nbr1)
Analysis of Deviance Table
Model: Negative Binomial(2.401), link: log
Response: Y
Terms added sequentially (first to last)
     Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
NULL                   487    2111.65
X1    1  1180.79       486     930.86 < 2.2e-16 ***
X2    1    15.61       485     915.25  7.78e-05 ***
X3    1    86.90       484     828.35 < 2.2e-16 ***
X4    1   259.54       483     568.81 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> summary(nbr1)
Call:
glm.nb(formula = Y ~ X1 + X2 + X3 + X4, data = data1, init.theta = 2.401021949,
    link = log)
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.9504  -0.7792  -0.1810   0.3244   3.3386
Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.557e+00  3.494e-01  -4.457 8.33e-06 ***
X1           8.661e-05  5.968e-06  14.512  < 2e-16 ***
X2          -1.899e-04  5.737e-05  -3.310 0.000934 ***
X3           5.382e-02  5.294e-03  10.166  < 2e-16 ***
X4           2.003e-02  1.191e-03  16.821  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
    Null deviance: 2111.65  on 487  degrees of freedom
Residual deviance:  568.81  on 483  degrees of freedom
AIC: 4104
Number of Fisher Scoring iterations: 1
> odTest(nbr1)
Critical value of test statistic at the alpha= 0.05 level: 2.7055
Chi-Square Test Statistic =  6115.4324 p-value = < 2.2e-16
```

Lampiran 6. Surat Rekomendasi Penelitian

PEMERINTAH PROVINSI JAWA TIMUR
DINAS KESEHATAN
Jl. Jend. A. Yani No.118 Telp. (031) 8280910 – 8280713Fax (031) 8290423
SURABAYA 60231

Surabaya, 23 Mei 2022

Nomor : 440/ 8441 /102.1/2022
Lampiran : -
Perihal : Rekomendasi Penelitian
An. SYAILENDRA ARDIFASALMA

Kepada Yth.
Kepala Departemen Aktuaria
Fakultas Sains dan Analitika Data
Institute Teknologi Sepuluh
Nopember Surabaya
di-
SURABAYA

Sehubungan dengan permohonan izin penelitian atas nama SYAILENDRA ARDIFASALMA dan Rekomendasi Penelitian dari Badan Kesatuan Bangsa dan Politik Pemerintah Provinsi Jawa Timur tanggal 27 April 2022 Nomor 070 / 4124 / 209.4 / 2022, yang bersangkutan dalam melakukan penelitian, wawancara, penyebaran quesioner maupun penggunaan data di Dinas Kesehatan Provinsi Jawa Timur, **diharuskan** melibatkan Bidang/SubBag/Seksi terkait sebagai Pembimbing/Penguji .

Judul penelitian : "Pemodelan Kasus Covid-19 Di Jawa Timur Menggunakan Metode Generalized Poisson Regression Dan Negative Binomial Regression". Rekomendasi akan diberikan apabila peneliti tersebut **bersedia memberikan salinan / softcopy file pdf hasil penelitian** atau sejenisnya dengan terlebih dahulu mengisi form persetujuan dan tunduk pada segala ketentuan serta tata cara yang berlaku di Dinas Kesehatan Provinsi Jawa Timur.

Pelaksanaan penelitian, wawancara ataupun penyebaran quesioner agar tidak disalah gunakan untuk maksud dan tujuan tertentu yang berdampak buruk pada masyarakat Jawa Timur.

Demikian atas perhatian dan kerjasamanya, disampaikan terima kasih.

KEPALA DINAS KESEHATAN
PROVINSI JAWA TIMUR

PEMERINTAH PROVINSI
DINAS KESEHATAN
JAWA TIMUR

Dr. ERWIN ASTHA TRIYONO, dr.,Sp PD.,KPTI
Pembina Tk. I
NIP. 19690420 200501 1 009

Lampiran 7. Surat Rekomendasi Pengambilan Data

DINAS KESEHATAN

Jl. Jend. A. Yani No.118, Telp/Fax (031) 8299056 Surabaya 60231
Web : www.dinke.jatimprov.go.id. Email : info@dinkesjatim.go.id

NOTA DINAS
No : 217 /102.1/2022

Kepada Yth.	:	Kepala Bidang Pelayanan Kesehatan Seksi Pelayanan Kesehatan Rujukan di Dinas Kesehatan Provinsi Jawa Timur
Dari	:	Sekretaris Dinas Kesehatan Provinsi Jawa Timur
Tanggal	:	30 Mei 2022
Sifat	:	Penting
Lampiran	:	- Berkas
Perihal	:	Rekomendasi Untuk Pengambilan Data Penelitian

Sehubungan dengan permohonan izin penelitian atas nama SYAILENDRA ARDIFASALMA dan Rekomendasi Penelitian dari Badan Kesatuan Bangsa dan Politik Pemerintah Provinsi Jawa Timur tanggal 27 April 2022 Nomor 070 / 4124 / 209.4 / 2022, yang bersangkutan akan melakukan penelitian, wawancara, penyebaran quesioner maupun penggunaan data di Dinas Kesehatan Provinsi Jawa Timur. Adapun judul penelitian : "Pemodelan Kasus Covid-19 Di Jawa Timur Menggunakan Metode Generalized Poisson Regression Dan Negative Binomial Regression", mohon dukungan untuk variabel penelitian terkait.

Demikian, atas perhatian dan kerjasama yang baik disampaikan terima kasih.

An. KEPALA DINAS KESEHATAN
PROVINSI JAWA TIMUR
SEKRETARIS

HERTANTO, SKM.MSi
Pembina Tk. I
NIP. 19650516 198903 1010

BIODATA PENULIS

Penulis dilahirkan di Malang, 29 Mei 2000, merupakan anak pertama dari 2 bersaudara. Penulis telah menempuh pendidikan formal, yaitu di TK Permata Bunda Malang, SDN Sawojajar 1 Malang, SMPN 3 Malang, dan SMAN 3 Malang. Setelah lulus dari SMAN 3 Malang pada tahun 2018, penulis mengikuti SBMPTN dan diterima di Departemen Aktuaria FSAD-ITS pada tahun 2018 dan terdaftar dengan NRP 063118 4000 0027.

Di Departemen Aktuaria Penulis sempat aktif di beberapa kegiatan pelatihan yang diselenggarakan oleh Badan Eksekutif Mahasiswa FMKSD dan sebagai Staf Ahli Departemen Internal Himpunan Mahasiswa Aktuaria (Himasakta) pada tahun 2020-2021.