Estimasi Parameter Pada Model Negatif Binomial Generalized Autoregressive Moving Average (GARMA) Dengan Algoritma IRLS (Studi Kasus Peramalan Jumlah Kecelakaan Di Jalan Tol Gempol-Surabaya)

Mada Aqil Habibi dan Laksmi Prita Wardhani
Departemen Matematika, Fakultas Matematika, Komputasi, dan Sains Data,
Institut Teknologi Sepuluh Nopember (ITS)
e-mail : madahabibi17@gmail.com, Laksmiprita61@gmail.com

***Abstrak*—Model yang sering digunakan untuk data *time series* adalah model ARIMA. Untuk data time series yang merupakan data *count*, pada model klasik Gaussian tidak selalu tepat. Pada penelitian ini, data jumlah kecelakaan yang digunakan yaitu jumlah kecelakaan di jalan Tol Gempol-Surabaya. Data tersebut bersifat *underdispersion* (nilai varians lebih kecil dari pada nilai rata-rata variabel responnya) sehingga pada kasus ini tidak memenuhi asumsi *equidispersion* (nilai variansi dan nilai rata-rata variabel respon adalah sama). Untuk memenuhi asumsi *equidispersion* dibentuk suatu model peramalan data count dengan pendekatan distribusi Negatif Binomial yaitu Model Negatif Binomial GARMA (1,1). Model tersebut didapatkan berdasarkan identifikasi model ARIMA. Penerapan model Negatif Binomial GARMA(1,1) menggunakan algoritma IRLS untuk memperoleh estimasi parameter. Parameter tersebut digunakan untuk mendapatkan hasil peramalan pada model Negatif Binomial GARMA(1,1). Hasil peramalan yang diperoleh dapat dikatakan akurat dengan RMSE sebesar 0,4231 (dibandingkan dengan model ARIMA(1,0,1).**

***Kata Kunci*—Data *count*, Distribusi Negatif Binomial, GARMA, *Equisdispersion*, Model Negatif Binomial GARMA (*p*,*q*).**

I. PENDAHULUAN

METODE *time series* adalah metode peramalan dengan menggunakan analisa plot hubungan antara variabel yang akan diperkirakan dengan variabel waktu. Model *time series* lebih sering digunakan untuk suatu peramalan atau prediksi dalam bidang industri, ekonomi, bisnis, teknik dan ilmu-ilmu sosial. Salah satu model yang sering digunakan untuk memodelkan data deret waktu adalah model *Autoregressive Integrated Moving Average* (ARIMA). Model tersebut merupakan kombinasi dari model *Autoregressive* (AR), *Integrated* (I), dan *Moving Average* (MA). Model ARIMA dapat digunakan untuk mengolah data runtun waktu yang univariat. Metode ARIMA hanya terbatas pada data yang berdistribusi normal dan bersifat stasioner baik rata-rata maupun variansi [1].

Untuk data *time series* yang merupakan data jumlahan (*count*), model klasik Gaussian tidak selalu tepat dan perlu dipertimbangkan pada kasus-kasus yang tidak linier. Sehingga kajian pemodelan peramalan deret waktu dikembangkan dan salah satunya diterapkan pada data jumlahan (*count data*). *Generalized Linear Model* (GLM) digunakan untuk menganalisis data terhitung dan tipe-tipe data diskrit. Data *count* adalah data yang tidak negatif {0,1,2,3,.. } dan nilainya merupakan bilangan bulat [2]. Selain proses perhitungan dalam interval waktu, data *count* dapat pula dihasilkan dari suatu proses perhitungan dalam suatu luasan atau area tertentu [3].

Model yang umum digunakan dalam menganalisis suatu data jumlahan adalah Regresi Poisson. Salah satu asumsi mendasar pada Regresi Poisson yaitu distribusi peluang variabel respon count adalah dari Distribusi Poisson. Pada Distribusi Poisson terdapat asumsi bahwa data harus bersifat *equisdispersion* (nilai varian variabel respon sama dengan rata-rata), namun kondisi tersebut tidak selalu terpenuhi. Untuk data jumlahan yang memiliki nilai nol atau mempunyai rata-rata taksiran yang rendah, sebuah konstanta dapat ditambahkan pada data [2]. Akan tetapi, model Gaussian dengan transformasi dari proses penambahan data tersebut tetap menghasilkan prediksi distribusi yang tidak tepat. Salah satu pendekatan yang tepat adalah model diasumsikan mengikuti distribusi Negatif Binomial (non-Gaussian).

Model stokastik yang telah dikembangkan untuk data-data yang mengikuti distribusi non-Gaussian seperti distribusi poisson dan Negatif Binomial adalah *Generalized Autoregressive Moving Average* (GARMA). Model GARMA merupakan hasil penggabungan komponen ARMA dengan variabel prediktor ke transformasi parameter rata-rata dari distribusi data dengan menggunakan fungsi link (*link function*). Fungsi link ini digunakan untuk memastikan bahwa distribusi data tetap dalam domain bilangan riil positif, sehingga memiliki ketepatan prediksi yang lebih akurat. Model GARMA sangat fleksibel untuk memodelkan data jumlahan dengan struktur *autoregressive* dan atau *moving average* dan dapat digunakan untuk variabel respon dengan waktu dependen [2]. Untuk mengestimasi parameter-parameter model GARMA menggunakan MLE dengan optimalisasi pendekatan *Iteratively Reweighted Least Square* (IRLS) [4]. Pada penelitian ini akan dilakukan analisis mengenai model GARMA menggunakan distribusi Negatif Binomial dengan studi kasus yaitu data jumlahan kecelakaan di jalan tol Gempol-Surabaya.

II. TINJAUAN PUSTAKA

A. Model ARMA

Model *Autoregressive Moving Average* (ARMA) merupakan model ARIMA tanpa proses pembedaan atau ARIMA (*p*,*0*,*q*) .secara matematis model ARMA (*p*,*q*) ditulis sebagai berikut [5]:

$$Y_t = \phi_1 Y_{t-1} + \ldots + \phi_p Y_{t-p} + e_t - \theta_1 e_{t-1} - \ldots - \theta_q e_{t-q} \quad (1)$$

dengan:

Y_t : data ke t
ϕ_1 : parameter $autoregressive$
θ_1 : parameter $moving\ average$
e_t : nilai galat pada data ke t

Adapun prosedur Box-Jenkins digunakan untuk memilih model ARMA yang sesuai pada data *time series* sebagai berikut:

1) Identifikasi Model

Identifikasi model ARMA dapat dilakukan dengan melihat plot deret waktu, plot ACF, dan plot PACF. Plot ACF dan PACF digunakan untuk menentukan orde p dan q dari model ARMA(*p*,*q*).

2) Estimasi Parameter

Dari model yang diidentifikasi pada tahap pertama,selanjutnya perlu diestimasi parameter dari model. metode yang digunakan untuk mengestimasi parameter adalah *maximum likelihood estimation*.

3) Uji Signifikasi Parameter

Setelah melakukan perhitungan estimasi parameter dilakukan uji signifikasi parameter. Uji ini digunakan untuk mengetahui apakah parameter AR(*p*) dan MA(*q*) signifikan atau tidak. Jika parameter tersebut signifikan maka model layak digunakan.

Pengujian Signifikasi parameter ϕ meliputi :

Hipotesa :

$H_0: \phi_p = 0$ (paramter ϕ tidak signifikan dalam model)

$H_1: \phi_p \neq 0$ (parameter ϕ signifikan dalam model).

Statistika Uji:

$$t_{hitung} = \frac{\hat{\phi}_p}{SE(\hat{\phi}_p)}. \quad (2)$$

kriteria pengujian :

Jika $|t_{hitung}| > t_{\frac{\alpha}{2}-n-1}$ atau $P_{value} < \alpha$ maka H_0 ditolak.

4) Cek Diagnosa (Diagnostic Check)

Pembentukan model *time series* adalah sebuah prosedur iteratif yang didahului dengan identifikasi model dan penaksiran parameter. Selanjutnya perlu diperiksa kecukupan model dengan memeriksa apakah asumsi model sudah dipenuhi. Beberapa uji yang dilakukan untuk memeriksa kesesuaian model adalah uji asumsi *white noise* dan uji asumsi distribusi normal

5) Overfitting

Salah satu prosedur diagnostik cek yang dikemukakan Box Jenkins adalah *overfitting*, yaitu menambah satu atau lebih parameter dalam model yang dihasilkan pada tahap identifikasi. Model yang dihasilkan dari *overfitting* dijadikan sebagai model alternatif yang kemudian dicari model terbaik diantara model-model yang signifikan.

6) Peramalan (Forecasting)

Tahapan terakhir dari analisis time series adalah peramalan.

B. Model Negatif Binomial untuk Data Count

Distribusi bersyarat dari hasil observasi y_t, untuk $t = 1,2,3,\ldots,n$ diberikan pada himpunan $H_t = \{x_t,\ldots,x_1,y_{t-1},\ldots,y_1,\mu_{t-1},\ldots,\mu_1\}$, yang merupakan keluarga eksponensial. Fungsi kepadatan peluang dari distribusi Negatif Binomial sebagai berikut [4]:

$$f(y_t,\mu,\alpha) = \frac{\Gamma\left(y_t + \frac{1}{\alpha}\right)}{\Gamma\left(\frac{1}{\alpha}\right) y_t!}\left(\frac{1}{1+\alpha\mu_t}\right)^{\frac{1}{\alpha}}\left(\frac{\alpha\mu_t}{1+\alpha\mu_t}\right)^{y_t} \quad (2)$$

Dengan $y_t = 0,1,2,\ldots$

Saat $\alpha \to 0$ maka distribusi Negatif Binomial memiliki varians $Var(y_t) = \mu_t$, Distribusi Negatif Binomial akan mendekati suatu distribusi poisson yang mengansumsikan *mean* dan varians sama yaitu $E(y_t) = Var(y_t) = \mu_t$. Fungsi distribusi keluarga eksponensial dari distribusi Negatif Binomial adalah [2]:

$$f(y_{t,}\mu,\alpha) = \exp\left\{y_t\, ln\left(\frac{\alpha\mu_t}{1+\alpha\mu_t}\right) + \frac{1}{\alpha} ln\left(\frac{1}{1+\alpha\mu_t}\right) + ln\left(\frac{\Gamma\left(y+\frac{1}{\alpha}\right)}{\Gamma\left(\frac{1}{\alpha}\right) y_t!}\right)\right\} \quad (3)$$

Kontribusi variabel prediktor dalam model Regresi Negatif Binomial dinyatakan dalam bentuk kombinasi linier antar parameter (η) dengan parameter regresi yang akan diestimasi yaitu:

$$\eta_t = \beta_0 + \beta_1 x_{t-1} + \cdots + \beta_k x_{t-k} \quad (4)$$

Atau dalam matriks dituliskan dalam bentuk

$$\eta = X^T\beta \quad (5)$$

Nilai ekspektasi dari vaiabel respon *Y* adalah diskrit dan bernilai positif. Maka, untuk mentrasformasikan nilai η_t (bilangan riil) ke rentang yang sesuai dngan rentang pada respon *Y* diperlukan suatu fungsi link *g(.)* yaitu [6]:

$$g(\mu_t) = \ln \mu_t = X^T\beta \quad (6)$$

C. Generalized Linear Models (GLM)

Generalized linear models (GLM) merupakan salah satu kelompok model statistika yang menghubungkan kombinasi linier antara variabel respon dengan variabel prediktor.

D. Model GARMA

Model GARMA dikenalkan pertama kali oleh Benjamin dkk. (2003). Salah satu bentuk $Z_{t-1}^T\beta$ yang digunakan pada model GARMA (p,q) yang mempunyai bentuk sebagai berikut [4]:

$$g(\mu) = Z_{t-1}^T\beta = X_{t-1}^T\beta + \tau_t \quad (7)$$

dengan

$$\tau_t = \sum_{j=1}^{p} \phi_j \mathcal{A}(y_{t-j}, x_{t-j,}\beta) + \sum_{j=1}^{q} \theta_j \mathcal{M}(y_{t-j}, \mu_{t-j}) \quad (8)$$

τ_t : komponen AR dan MA
A : fungsi yang mempresentasikan bentuk autoregressive
$\mathcal{M}$: fungsi yang mempresentasikan bentuk moving average
ϕ^T : parameter autoregressive ; $\phi^T = (\phi_1, \phi_2, \ldots, \phi_p)$
θ^T : parameter moving average ; $\boldsymbol{\theta}^T = (\theta_1, \theta_2, \ldots, \theta_p)$

Bentuk sub model parsimoni dari model GARMA (p,q) yang didefinikan oleh Benjamin dkk.(2003) adalah sebagai berikut :

$$g(\mu) = \boldsymbol{X}_{t-1}^T\boldsymbol{\beta} + \sum_{j=1}^{p} \phi_j\{g(y_{t-j}) - X_{t-j}^T\beta\} + \sum_{j=1}^{q} \theta_j \{g(y_{t-j}) - \eta_{t-j}\beta\} \quad (9)$$

E. *Negatif Binomial GARMA(p,q)*

Model Negatif Binomial GARMA (p,q) dikembangkan berdasarkan model GARMA (p,q) dengan asumsi data berdistribusi Negatif Binomial. Diberikan $y^T = (y_t, y_{t+1}, \ldots, y_{t+n})$ adalah model data deret waktu. Jika $y_t \sim Negatif\ Binomial(\mu_t, \alpha)$ maka $E(y_t) = Var(y_t) = \mu_t$ dan $\alpha \to 0$. Jika g adalah fungsi logaritma natural persamaan 9 maka :

$$\ln(\mu_t) = \boldsymbol{X}_{t-1}^T\boldsymbol{\beta} + \sum_{j=1}^{p} \phi_j \{\ln y_{t-j}^* - \boldsymbol{X}_{t-1}^T\boldsymbol{\beta}\} + \sum_{j=1}^{q} \theta_j\{\ln(y_{t-j}^*) - \ln(\mu_{t-j})\} \quad (10)$$

dimana $y_{t-j}^* = \max(y_{t-j}, c)$ dan $0 < c \leq 1$. Jika terdapat nilai 0 pada nilai y_{t-j} maka akan diganti dengan c.

F. *Maximum Likielihood Estimation(MLE)*

Metode *Maximum Likelihood Estimation* adalah metode pendugaan yang memaksimumkan fungsi likelihood. Dalam penelitian ini metode MLE digunakan untuk menduga parameter Poisson GARMA. Adapun fungsi Likelihood $L(\theta)$ sebagai berikut[8]:

$$L(\theta) = \prod_{t=1}^{n} f(x_i|\theta)$$

Untuk mempermudah perhitungan secara matematis, umumnya digunakan fungsi log-*likelihoood.*

$$l(\theta) = \ln L(\theta) = \sum_{i=1}^{n} \ln f(x_i|\theta) \quad (11)$$

Syarat cukup agar fungsi dari *partial log-likelihood* maksimum adalah :

$$l\frac{\partial l(\mu_t)}{\partial \beta_0} = 0\,; \frac{\partial l(\mu_t)}{\partial \phi_j} = 0\,; \frac{\partial l(\mu_t)}{\partial \theta_j} = 0 \quad (12)$$

G. *Algoritma Iteratively Reweighted Least Square (IRLS)*

Algoritma IRLS digunakan ketika metode MLE yang telah digunakan sebelumnya tidak bisa diselesaikan atau disebut tidak *close form* menurut persamaan 12. Adapun langkah-langkah algoritma IRLS untuk mendapatkan taksiran parameter menurut Kedem dan Fokianos (2002) adalah sebagai berikut:

1. Inisialisasi, memilih nilai awal untuk $\hat{\beta}_{(0)}$ dengan menggunakan metode *Ordinary Least Square* (OLS) untuk $t = 2,3,4 \ldots, T$.
$$\hat{\beta}_{(0)} = (x^T x)^{-1}(x^T y_t^*)$$

x dan y didefinisikan sebagai berikut :

$$x = \begin{bmatrix} 1 & x_{11} & \ldots & x_{k1} \\ 1 & x_{12} & \ldots & x_{k2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & \ldots & x_{kn} \end{bmatrix} \text{dan } y = \begin{bmatrix} \ln y_1^* \\ \ln y_2^* \\ \vdots \\ \ln y_{T-1}^* \end{bmatrix}$$

2. Menghitung nilai $\hat{\boldsymbol{\mu}}^{(\mathbf{0})}$
$$\hat{\boldsymbol{\mu}}^{(\mathbf{0})} = \exp(\boldsymbol{X_t^T\beta})$$
$$\hat{\boldsymbol{\mu}}^{(\mathbf{0})} = \hat{\beta}_{(0)}^{(0)} + \hat{\beta}_{(1)}^{(0)} \ln y_{t-1}^*$$
3. Menghitung nilai W
$$W^{(1)} = \begin{bmatrix} \frac{1}{\hat{\boldsymbol{\mu}}_{\mathbf{0}}{}^{(\mathbf{0})}} & & & \\ & \frac{1}{\hat{\boldsymbol{\mu}}_{\mathbf{1}}{}^{(\mathbf{0})}} & & \\ & & \ddots & \\ & & & \frac{1}{\hat{\boldsymbol{\mu}}_{\boldsymbol{T-1}}{}^{(\mathbf{0})}} \end{bmatrix}$$
4. Menghitung nilai z yakni sebagai berikut :
$$z^{(1)} = \boldsymbol{\eta}_t^{(1)} + \frac{y_t + \hat{\boldsymbol{\mu}}_{\mathbf{0}}}{\hat{\boldsymbol{\mu}}_{\mathbf{0}}}$$
$$\boldsymbol{\eta}_t^{(1)} = \boldsymbol{X}\hat{\beta}^{(1)}$$
5. Menghitung estimasi
$$\widehat{\boldsymbol{\beta}}^{(k+1)} = \left(\boldsymbol{X^T W^{(k+1)} X}\right)^{-1}\boldsymbol{X^T W^{(k+1)}} z^{(k+1)}$$
6. Ulangi langkah sampai dengan 5 menggunakan nilai $\widehat{\boldsymbol{\beta}}^{(\mathbf{1})}$ sehingga akan diperoleh nilai baru $\widehat{\boldsymbol{\beta}}^{(\mathbf{2})}$
7. Update k ke $k+1$ dan ulangi langkah 6 sampai memenuhi syarat berikut :
$$\left|\hat{\beta}^{(t)} - \hat{\beta}^{(t-1)}\right| < \varepsilon_\beta$$

III. METODOLOGI PENELITIAN

Pada penelitian ini, data yang digunakan adalah data jumlah kejadian kecelakaan di jalan Tol Surabaya-Gempol ruas Waru-Sidoarjo dari bulan Januari tahun 2012 hingga Agustus tahun 2017 dengan jumlah data sebesar 68. Data tersebut diperoleh dari pihak pengelola jalan tol P.T. Jasa Marga Variabel Y yang digunakan dalam penelitian ini adalah jumlah kecelakaan di jalan Tol Gempol-Surabaya.

Table 1
Variabel Penelitian

Tahun	Bulan	Jumlah Kecelakaan (Y)
2012	Januari	Y_1
	Februari	Y_2
	⋮	⋮
	Desember	Y_{12}
2013	Januari	Y_{13}
	Februari	Y_{14}
	⋮	⋮
	Desember	Y_{24}
⋮	⋮	⋮

2017	Januari	Y_{61}
	Februari	Y_{62}
	⋮	
	Agustus	Y_{68}

Langkah-langkah analisis yang digunakan untuk mencapai tujuan penelitian yakni sebagai berikut:

1. Pengumpulan data
   Pada tahap ini dilakukan pengumpulan data jumlah kecelakaan di jalan Tol Surabaya-Gempol ruas Waru-Sidoarjo dari P.T. Jasa Marga.
2. Studi Literatur
   Pada tahap ini dilakukan studi literatur untuk mendukung penelitian. Bahan-bahan referensi yang digunakan berupa buku, jurnal, tugas akhir, thesis dan juga media elektronik (internet) yang sesuai dan berhubungan dengan permasalahan yang dibahas.
3. Aplikasi model ARMA(p,q) pada data kecelakaan
4. Mengkaji cara mendapatkan penaksir parameter model Negatif Binomial GARMA (p,q) menggunakan metode MLE dengan optimasi Algoritma IRLS.
5. Melakukan peramalan model Negatif Binomial GARMA (p,q) pada data jumlah kecelakaan di jalan tol Gempol-Surabaya

IV. ANALISIS DAN PEMBAHASAN

A. *Pemodelan ARMA*

Langkah pertama dalam merumuskan model ARMA adalah proses identifikasi dengan melakukan uji stasioneritas data terhadap varian dan *mean*. Data yang akan diuji adalah data kecelakaan *in sample* sebanyak 60 data. Untuk melihat kestasioneran data terhadap varian dapat menggunakan plot Box-Cox dan plot tren anilisis. Apabila data tersebut sudah stasioner terhadap varian (*rounded value* $= 1$) maka tidak perlu dilakukan transformasi, namun apabila belum stasioner (*rounded value* $\neq 1$) terhadap varian harus dilakukan proses transformasi sampai data stasioner (*rounded value* $= 1$). Sedangkan untuk melihat kestasioneran terhadap *mean* dapat dilihat dari plot tren anilisis data, apabila data menunjukkan pola yang teratur serta tidak terjadi fluktuasi yang signifikan maka data sudah stasioner dalam mean, namun jika belum maka perlu dilakukan proses *differencing*. Setelah data stasioner terhadap varian dan *mean*, dilakukan plot ACF dan PACF untuk mendapatkan model sementara. Bentuk umum model ARIMA (p,d,q) adalah sebagai berikut :

$$Y_t = \phi_1 Y_{t-1} + \ldots + \phi_p Y_{t-p} + e_t - \theta_1 e_{t-1} - \ldots - \theta_q e_{t-q}$$

dengan:

Y_t :data ke t
ϕ_1 : parameter $autoregressive$
θ_1 : parameter $moving\ average$
e_t : nilai kesalahan pada waktu ke t

Langkah berikutnya penaksiran dan pengujian parameter. Parameter model ARMA akan di estimasi menggunakan metode *Maximum Likelihood* dengan bantuan *software* Minitab dan selanjutnya akan dilakukan pengujian menggunakan uji *t*.

Langkah selanjutnya adalah diagnostik cek pada residual yang meliputi asumsi residual bersifat *white noise* dan berdistribusi normal. Pengujian residual bersifat *white noise* dilakukan dengan uji Ljung-Box dan pengujian residual berdistribusi normal dilakukan dengan uji Komogorov-Smirnov. Selanjutnya akan dilakukan proses *overfitting* sebagai model alternatif.

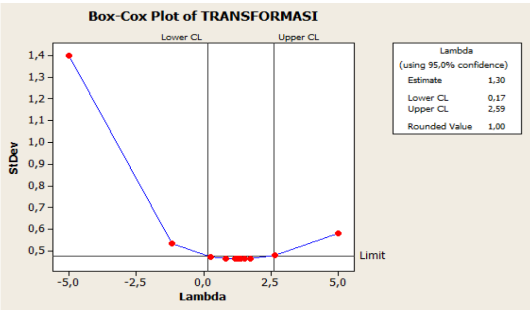
Model terbaik ARIMA akan dipilih berdasarkan nilai RMSE terkecil dari semua model yang memenuhi semua asumsi yaitu parameter signifikan, residual bersifat *white noise* dan berdistribusi normal. Persamaan RMSE didefinisikan sebagai berikut.

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(Y_t - \hat{Y}_t\right)^2}$$

dengan :

n : banyaknya data *out sample*
Y_t : data aktual pada waktu ke-t
$\hat{Y}_t$: data ramalan pada waktu ke-t

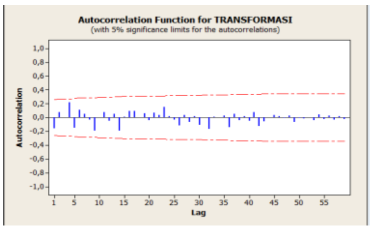
Langkah pertama dalam merumuskan model ARMA adalah proses identifikasi dengan melakukan uji stasioneritas data terhadap varian dan *mean*. Transformasi Box-Cox pada Gambar 1 menunjukkan data sudah stasioner terhadap varian.



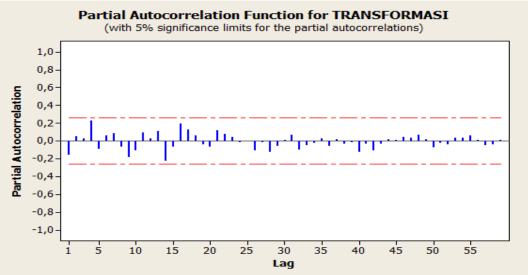
Gambar 1. Blox-Plot Data Transformasi Kecelakaan

Pada Gambar 2 Terlihat bahwa plot ACF memiliki pola *tentative* karena tidak ada lag yang keluar dari *significant limit*. Maka dapat dikatakan bahwa plot ACF telah memenuhi asumsi stasioneritas terhadap varians dan *mean* maka langkah selanjutnya adalah mengidentifikasi plot PACF yang dapat dilihat seperti pada Gambar 3.

Dengan melihat plot ACF dan PACF yang diperoleh dari proses transformasi *Box-Cox* maka kita dapat meyimpulkan bahwa nilai d bernilai 0 karena proses *differencing* tidak perlu dilakukan. data transformasi *Box-Cox* telah memenuhi asumsi stasioneritas terhadap *mean*. Terlihat bahwa pada plot PACF tidak ada lag-lag yang keluar dari *significant limit* maka diperoleh model smentara ARMA yaitu ([1,0,1]).



Gambar 2. Plot ACF pada hasil transformasi Box-Cox terhadap data kecelakaan



Gambar 3. Plot PACF Tanpa *Differencing*

Selanjutnya dilakukan penaksiran dan pengujian parameter, uji residual *white noise*, serta uji residual berdistribusi normal. Proses *overfitting* dilakukan dengan menguji semua hasil model ARMA berdasarkan plot ACF dan PACF. Model ARMA terbaik diperoleh berdasarkan nilai RMSE terkecil dari semua model yang memenuhi semua asumsi. Nilai RMSE model ARMA yang memenuhi uji signifikansi, asumsi residual bersifat *white noise* dan residual berdistribusi normal adalah 1,686682.

Dari model ARMA(1,0,1) diperoleh data ramalan untuk 8 data *out-sample* yang dapat dilihat pada Tabel 2.

Tabel 2.
Data Hasil Peramalan Model ARMA(1,0,1)

No	Periode	Ramalan
1	Januari 2017	3,5554
2	Februari 2017	4,5088
3	Maret 2017	3,6869
4	April 2017	4,3842
5	Mei 2017	3,7845
6	Juni 2017	4,2943
7	Juli 2017	3,8566
8	Agustus 2017	4,2293

H. *Estimasi Parameter Model Negatif Binomial GARMA(1,1)*

Menurut Marinho dan Ricardo (2015) $X_{t-1}^{T}\beta$

$$\ln(\mu_t) = \beta_0 + \sum_{j=1}^{p} \phi_j \{\ln y_{t-j}^*\} + \sum_{j=1}^{q} \theta_j \{\ln(y_{t-j}^*) - \ln(\mu_{t-j})\} \quad (13)$$

dimana $y_{t-j}^* = \max(y_{t-j}, c)$ dan $0 < c \leq 1$.

Jika persamaan dibentuk kedalam matriks, maka bentuk matrik sebagai berikut:

$$\begin{bmatrix} \ln(\mu_2) \\ \ln(\mu_3) \\ \ln(\mu_4) \\ \vdots \\ \ln(\mu_T) \end{bmatrix} = \begin{bmatrix} 1 & \ln y_1^* & \dots & \ln y_1^* & \ln \frac{y_1^*}{\mu_1} & \dots & \ln \frac{y_1^*}{\mu_1} \\ 1 & \ln y_2^* & \dots & \ln y_2^* & \ln \frac{y_2^*}{\mu_2} & \dots & \ln \frac{y_2^*}{\mu_2} \\ 1 & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \ln y_{T-1}^* & \dots & \ln y_{T-1}^* & \ln \frac{y_{T-1}^*}{\mu_{T-1}} & \dots & \ln \frac{y_{T-1}^*}{\mu_{T-1}} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \phi_1 \\ \vdots \\ \phi_p \\ \theta_1 \\ \vdots \\ \theta_q \end{bmatrix} \quad (14)$$

Sehingga diperoleh persamaan sebagai berikut:

$$\mu_t = \exp\left[\beta_0 + \sum_{j=1}^{p} \phi_j \{\ln y_{t-j}^*\} + \sum_{j=1}^{q} \theta_j \{\ln(y_{t-j}^*) - \ln(\mu_{t-j})\}\right] \quad (15)$$

Dengan β_0 sebagai parameter yang tidak diketahui.

Dalam mencari estimasi parameter model Negatif Binomial GARMA (p,q) langkah awal yang dilakukan yaitu dengan mengasumsikan bahwa variabel respon y pada data jumlahan mengikuti distribusi Negatif Binomial . bentuk *partial likelihood* dari distribusi Negatif Binomial sebagai berikut:

$$PL(\mu_t, \alpha) = \prod_{t=1}^{N} \left(\frac{1}{1+\alpha\mu_t}\right)^{1/\alpha} \left(\frac{\alpha\mu_t}{\mu_t\alpha + 1}\right)^{y_t} \left(\frac{\Gamma(1/\alpha + y_t)}{y_t!\ \Gamma(1/\alpha)}\right) \quad (16)$$

Bentuk diatas diubah kedalam bentuk partial *log-likehood* seperti berikut:

$$l(\mu_t, \alpha) = \sum_{t=1}^{N} \ln\Gamma\left(\frac{1}{\alpha} + y_t\right) - \sum_{t=1}^{N} \ln\ \Gamma\left(\frac{1}{\alpha}\right) - \sum_{t=1}^{N} \ln y_t! - \frac{1}{\alpha}\sum_{t=1}^{N} \ln(1+\mu_t) + \sum_{t=1}^{N} y_t \ln\alpha\mu_t - \sum_{t=1}^{N} y_t \ln(1+\alpha\mu_t) \quad (17)$$

Persamaan dibawah ini merupakan hasil subsitusi model khusus Negatif Binomial GARMA(1,1) ke persamaan partial *log-likehood* sehingga didapat persamaan baru sebagai berikut:

$$(\mu_t, \alpha) = \sum_{t=1}^{N} \ln\Gamma\left(\frac{1}{\alpha} + y_t\right) - \sum_{t=1}^{N} \ln\ \Gamma\left(\frac{1}{\alpha}\right) - \sum_{t=1}^{N} \ln y_t! - \frac{1}{\alpha}\sum_{t=1}^{N} \ln(1 + \exp(\beta_0 + \phi_1\{\ln y_{T-1}^*\} + \theta_1\{\ln y_{T-1}^* - \ln\mu_{T-1}\})) + \sum_{t=1}^{N} y_t \ln\alpha \exp(\beta_0 + \phi_1\{\ln y_{T-1}^*\} + \theta_1\{\ln y_{T-1}^* - \ln\mu_{T-1}\}) - \sum_{t=1}^{N} y_t \ln(1 + \alpha(\exp(\beta_0 + \phi_1\{\ln y_{T-1}^*\} + \theta_1\{\ln y_{T-1}^* - \ln\mu_{T-1}\})) \quad (18)$$

Untuk memaksimumkan fungsi pada persamaan (18) maka dilakukan turunan pertama terhadap parameter. Misalkan parameter yang diestimasi adalah $\beta_0, \phi_j\ dan\ \theta_j$. Persamaan hasil turunan pertama tersebut tidak bisa diselesaikan secara analitik karena tidak memenuhi syarat cukup dari metode MLE sehingga untuk mengestimasi $\beta_0, \phi_j\ dan\ \theta_j$ digunakan optimasi algoritma *Iteratively Reweighted Least Square* (IRLS). Iterasi yang dilakukan pada algoritma IRLS berhenti ketika didapatkan parameter yang konvergen. Penaksiran parameter menggunakan metode IRLS dengan menggunakan software matlab2013a.

Sesuai dengan yang telah dijelaskan sebelumnya bahwa model Negatif Binomial GARMA(1,1) diperoleh dari model sementara ARMA(1,1). Pada perumusan model Negatif Binomial GARMA(*p*,*q*) tidak perlu lagi dilakukan uji asumsi berdistribusi normal karena pada data observsi Y sudah diasumsikan mengikuti distribusi negatif binomial sehingga tidak berkaitan dengan asumsi error normal. Model terbaik setelah dilakukan *overfitting* adalah Negatif Binomial GARMA(1,1). Sehingga dapat dilihat bahwa parameter yang dicari untuk model Negatif Binomial GARMA (1,1) adalah $\beta_0, \phi_1, \theta_1$. Dengan inisialisasi orde $p = 1$ dan $q = 1$ didapatkan hasil estimasi model Negatif Binomial GARMA(1,1) dengan menggunakan iterasi algoritma IRLS yang dapat dilihat pada tabel 3.

Tabel 3.
Estimasi Parameter Model Negatif Binomial GARMA (1,1)

Parameter	Koefisien	SE
β_0	2,831	0,000
ϕ_1	0,943	0,056
θ_1	-1,359	0,168

I. Peramalan Model Negatif Binomial GARMA(1,1)

Peramalan Model Negatif Binomial GARMA(1,1) dilakukan dengan menggunakan *software* Matlab. Pada peramalan model negatif Binomial GARMA(1,1) ini dilakukan 60 data sebagai *in-sample* dan 8 data sebagai *out-sample*. Hasil peramalan dari data jumlah kecelakaan dijalan tol Gempol-Surabaya dapat dilihat pada Tabel 4.

Pemilihan model terbaik melalui RMSE (*Root Mean Square Percentage Error*). Model dikatakan baik, jika model tersebut memiliki nilai RMSE yang kecil. Hasil perhitungan RMSE dengan menggunakan *software* matlab menghasilkan nilai RMSE 0,4271.

Pada tahap ini, hasil peramalan yang telah didapatkan pada tahap sebelumnya akan dianalisis tingkat akurasinya. Hasil data peramalan dari metode ARIMA dan GARMA Negatif Binomial (1,1) akan dibandingakn data aktual yang ada. Untuk hasil perbandingan data peramalan dapat dilihat pada Tabel 5.

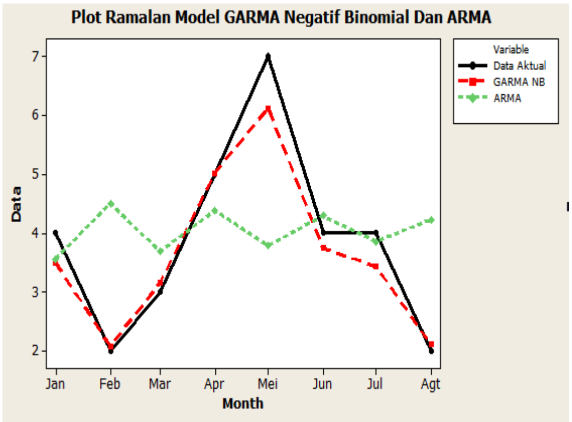
Tabel 4.
Peramalan Model Negatif Binomial GARMA(1,1) Pada Data Kecelakaan Jalan Tol Gempol-Surabaya

No	Periode	Ramalan
1	Januari 2017	3,4932
2	Februari 2017	2,0791
3	Maret 2017	3,1610
4	April 2017	5,0225
5	Mei 2017	6,1219
6	Juni 2017	3,7430
7	Juli 2017	3,4326
8	Agustus 2017	2,1049

Tabel 5.
Perbandingan Nilai Data Peramalan Model Negatif Binomial GARMA (1,1) dan Model ARIMA Pada Data Kecelakaan Jalan Tol Gempol-Surabaya

No	Y Aktual	$\hat{Y}$ GARMA	Square of the error GARMA	$\hat{Y}$ ARIMA	Square of the error ARIMA
1	4	3,4932	0,2568	3,5554	0,1977
2	2	2,0791	0,0063	4,5088	6,2941
3	3	3,1610	0,0259	3,6869	0,4718
4	5	5,0225	0,0005	4,3842	0,3792
5	7	6,1219	0,7711	3,7845	10,3394
6	4	3,7430	0,0660	4,2943	0,0866
7	4	3,4326	0,3219	3,8566	0,0206
8	2	2,1049	0,0110	4,2293	4,9689
	RMSE		0,4231		1,6867

Berdasarkan Tabel 5 dapat dilihat bahwa nilai data peramalan Jalan Tol Gempol-Surabaya dengan menggunakan metode Negatif Binomial GARMA menghasilkan nilai yang hampir sama dengan nilai aktual data Kecelakaan. Misalkan data nomor 2, pada data aktual terdapat 2 kecelakaan dan pada peramalan dengan menggunakan model GARMA dihasilkan 2,0791, sedangkan peramalan dengan menggunakan model ARIMA didapatkan nilai 4.5088. Hal tersebut menujukkan bahwa hasil peramalan dengan menggunakan model GARMA relatif lebih mendekati data aktual dari pada hasil peramalan dengan menggunakan model ARIMA. Disisi lain, pada model Negatif Binomial GARMA menghasilkan nilai RMSE sebesar 0,4231 dan pada model ARIMA menghasilkan nilai RMSE sebesar 1,6867. Sehingga berdasarkan Tabel 4.6 menunjukkan bahwa Model Negatif Binomial GARMA lebih baik dibandingkan Model ARMA karena nilai RMSE Model Negatif Binomial GARMA lebih kecil dari Model ARIMA [7]. Hal ini dapat dilihat pada *plot* ramalan dari kedua model seperti yang terlihat pada Gambar 4.



Gambar 4. *Plot* Ramalan Model Negatif Binomial GARMA dan ARMA

V. KESIMPULAN DAN SARAN

A. Kesimpulan

Berdasarkan analisis dan pembahasan, dapat disimpulkan bahwa :

A. Estimasi parameter model Negatif Binomial GARMA (p, q) dapat dilakukan dengan menggunakan iterasi numerik yaitu algoritma IRLS untuk menyelesaikan metode MLE yang *close form*. Iterasi dilakukan sampai parameter tersebut dapat dikatakan konvergen.

B. Model Negatif Binomial GARMA (p, q) yang dibangun dari model terbaik ARMA(1,1) yaitu Negatif Binomial GARMA (1,1) dengan hasil estimasi parameter koefisien regresi bernilai 2,831, *Autoregressive* bernilai 0,943 dan *Moving Average* bernilai 1,359.

C. Perbandingan hasil antara model Negatif Binomial GARMA(1,1) dengan ARIMA(1,0,1) menunjukkan bahwa nilai RMSE pada model Negatif Binomial GARMA lebih kecil daripada model ARIMA(1,0,1) sehingga model Negatif Binomial GARMA(1,1) lebih baik digunakan untuk meramalkan data kecelakaan di Jalan Tol Gempol-Surabaya

B. Saran

1. Data yang digunakan dalam penelitian ini merupakan data suatu kejadian yang jarang terjadi yaitu bersifat jumlahan (*count*) dimana nilainya adalah diskrit non negatif. Namun, dari hasil peramalan yang diperoleh data yang diperoleh *desimal* sehingga perlu dilakukan suatu pembalikan data agar menghasilkan kembali data berupa *count* (diskrit non negatif)
2. Pada penelitian ini tidak melibat variabel prediktor sehingga diharapkan pada penelitian selanjutnya dapat melibatkan variabel prediktor agar dapat mengahasilkan akurasi peramalan yang lebih baik.

DAFTAR PUSTAKA

[1] F. A. Siga, "Studi Pengaruh Outlier Terhadap Outlier Terhadap Orde K-Faktor GARMA dan Aplikasinya untuk Peramalan Beban Listrik," Institut Teknologi Sepuluh Nopember, 2016.

[2] Asriawan, "Model Generalized Seasonal Autoregressive Integrated Moving Average (GSARIMA) untuk Peramalan Jumlah Penderita DBD di kota Surabaya," Institut Teknologi Sepuluh Nopember, 2014.

[3] E. T. Astuti, "Estimator Polinomial Lokal Dalam Model Regresi Non-Parametrik Untuk Data Count," Institut Teknologi Sepuluh Nopember, 2013.

[4] D. . Benjamin, M.A., Rigby, R.A, dan Stasinopoulus, "Generalized Autoregressive Moving Average Models," *J. Am. Stat. Assos.*, 2003.

[5] S. W. Makridakis, *Metode dan Aplikasi Peramalan Edisi Kedua Jilid I*. Erlangga, 1999.

[6] W. W. S. Wei, *Time Series Analysis Univariate and Multivariate Methods*. Pearson Education, Inc, 2006.

[7] T. C. dan T. Montip, "Application of Box-Jenkins (Seasonal ARIMA) and GARCH models to dengue incidence in Thailand," 2018.