BAB III

Censored Autoregressive Model with Exogenous Variables

Bab ini akan berisi mengenai *Censored Autoregressive Model with Exogenous Variables* (CARX) dan tahapan analisis yang digunakan pada penelitian ini.

3.1 *Autoregressive Exogenous (ARX)*

Model *Autoregressive Exogenous* (ARX) merupakan pengembangan dari model AR, dimana model ARX nilai pengamatan saat ini dipengaruhi atau bergantung pada nilai–nilai pengamatan waktu–waktu sebelumnya serta dipengaruhi oleh faktor luar yaitu variabel eksogen. Model ARX pertama kali dikembangkan oleh Isaksson pada tahun 1993. Dalam penelitiannya, diperoleh bahwa untuk model ARX diperlukan asumsi bahwa data yang akan di uji berdistribusi normal.

Model ARX pada prinsipnya sama dengan model AR hanya ditambah variabel eksogen yang merupakan variabel tambahan yang memiliki pengaruh terhadap variabel yang lain namun tidak dipengaruhi oleh variabel lain dalam model. Model ARX dapat ditulis sebagai berikut:

$$Z_t = \emptyset_1 Z_{t-1} + \emptyset_2 Z_{t-2} + \cdots + \emptyset_p Z_{t-p} + \varphi X_t + a_t \tag{3.1}$$

dengan $\alpha \sim N(0,\sigma^2)$ (Qin, Nishii, Nakagawa, & Nakamoto, 2010).

a_t diasumsikan berdistribusi normal dengan rata rata nol dan varians σ^2.

Keterangan:

Z_t : variabel acak waktu ke t

Z_{t-p} : variabel acak endogen waktu ke t

X_t : variabel acak eksogen waktu ke t

$\emptyset$: parameter variabel endogen

φ : parameter variabel eksogen

a_t : sesatan dalam model

Kelebihan dari model ARX dibandingkan dengan model AR pada metode box jenkin’s adalah dengan adanya variabel eksogen. Variabel eksogen ini dapat membantu keakuratan dalam tahap peramalan oleh model tersebut dan dapat menjelaskan adanya variabel lain selain variabel dependen yang mempengaruhi hasil peramalan suatu variabel. Sedangkan, kekurangan dari model ini adalah jika data yang di uji tidak diketahui distribusinya dan terdapat asumsi tambahan yaitu data tersensor maka model ARX tidak dapat digunakan.

3.2 Penaksiran Model ARX dengan Metode OLS (*Ordinary Least Square*)

Metode OLS merupakan salah satu penaksiran parameter untuk model linear. Dengan diketahuinya model ARX berdistribusi normal maka penaksiran yang dapat digunakan dengan metode OLS. Berikut uraian penaksiran parameter ARX dengan metode OLS:

1. Model ARX

   Model kali ini yang dipilih adalah ARX dengan orde 1 atau ARX(1) yang nantinya hasil penaksiran dapat diperluas secara umum, sebagai berikut:

   $$Z_t = \emptyset_1 Z_{t-1} + \varphi X_t + a_t \tag{3.2}$$

   Dengan asumsi a_t ~ N(0,σ^2).

2. Dari persamaan (3.2) bentuk a_t dapat dibentuk menjadi

   $$a_t = Z_t - \emptyset_1 Z_{t-1} - \varphi X_t \tag{3.3}$$

3. Persamaan (3.3) dikuadratkan kedua ruas dan dimisalkan dengan J

   $$J = (a_t)^2 = (Z_t - \emptyset_1 Z_{t-1} - \varphi X_t)^2 \tag{3.4}$$

4. Turunkan persamaan (3.4) terhadap $\emptyset_1$

   $$\frac{\partial J}{\partial \emptyset_1} = -2Z_{t-1}(Z_t - \emptyset_1 Z_{t-1} - \varphi X_t) \tag{3.5}$$

5. Turunkan persamaan (3.5) terhadap φ

   $$\frac{\partial J}{\partial \emptyset_1} = -2X_t(Z_t - \emptyset_1 Z_{t-1} - \varphi X_t) \tag{3.6}$$

6. Turunan pertama dari persamaan (3.5) dan (3.6) harus dama dengan nol untuk meminimumkan *error*nya

   $$-2Z_{t-1}(Z_t - \emptyset_1 Z_{t-1} - \varphi X_t) = 0$$

   $$Z_t Z_{t-1} - \emptyset Z_{t-1}^2 - \varphi X_t Z_{t-1} = 0$$

   $$\emptyset Z_{t-1}^2 + \varphi X_t Z_{t-1} = Z_t Z_{t-1} \tag{3.7}$$

Dan juga

$$-2X_t(Z_t - \emptyset_1 Z_{t-1} - \varphi X_t) = 0$$

$$Z_t X_t - \emptyset_1 Z_{t-1} X_t - \varphi X_t^2 = 0$$

$$\emptyset_1 Z_{t-1} X_t + \varphi X_t^2 = Z_t X_t \qquad (3.8)$$

7. Dicari nilai $\emptyset$ dan φ dari persamaan (3.7) dan (3.8) menggunakan matriks:

$$\emptyset Z_{t-1}^2 + \varphi X_t Z_{t-1} = Z_t Z_{t-1}$$

$$\emptyset_1 Z_{t-1} X_t + \varphi X_t^2 = Z_t X_t$$

$$\begin{bmatrix} Z_{t-1}^2 & X_t Z_{t-1} \\ Z_{t-1} X_t & X_t^2 \end{bmatrix} \begin{bmatrix} \emptyset \\ \varphi \end{bmatrix} = \begin{bmatrix} Z_t Z_{t-1} \\ Z_t X_t \end{bmatrix}$$

$$\begin{bmatrix} \emptyset \\ \varphi \end{bmatrix} = \begin{bmatrix} Z_{t-1}^2 & X_t Z_{t-1} \\ Z_{t-1} X_t & X_t^2 \end{bmatrix}^{-1} \begin{bmatrix} Z_t Z_{t-1} \\ Z_t X_t \end{bmatrix}$$

$$\begin{bmatrix} \emptyset \\ \varphi \end{bmatrix} = \frac{1}{Z_{t-1}^2 X_t - (X_t Z_{t-1})(X_t Z_{t-1})} \begin{bmatrix} X_t^2 & -X_t Z_{t-1} \\ -Z_{t-1} X_t & Z_{t-1}^2 \end{bmatrix} \begin{bmatrix} Z_t Z_{t-1} \\ Z_t X_t \end{bmatrix}$$

$$\begin{bmatrix} \emptyset \\ \varphi \end{bmatrix} = \begin{bmatrix} \frac{X_t^2 Z_t Z_{t-1} - (X_t Z_{t-1}) Z_t X_t}{Z_{t-1}^2 - (X_t Z_{t-1})(X_t Z_{t-1})} \\ \frac{(-Z_t X_t)(Z_t Z_{t-1}) + Z_{t-1}^2 Z_t X_t}{Z_{t-1}^2 X_t^2 - (X_t Z_{t-1})(X_t Z_{t-1})} \end{bmatrix} \qquad (3.9)$$

Maka untuk t = 2,3,..., n didapatkan

$$\hat{\emptyset} = \sum_{t=2}^{n} \frac{X_t^2 Z_t Z_{t-1} - (X_t Z_{t-1}) Z_t X_t}{Z_{t-1}^2 - (X_t Z_{t-1})(X_t Z_{t-1})}$$

$$\hat{\varphi} = \sum_{t=2}^{n} \frac{(-Z_t X_t)(Z_t Z_{t-1}) + Z_{t-1}^2 Z_t X_t}{Z_{t-1}^2 X_t^2 - (X_t Z_{t-1})(X_t Z_{t-1})}$$

3.3 *Censored Autoregressive Exogenous* (CARX)

Model CARX (*Censored Autoregressive Exogenous*) pertama kali dikembangkan oleh Chao Wang dan Kung-Sik Chan pada tahun 2017. CARX adalah model deret waktu dimana data yang digunakan merupakan data yang tersensor yang menyatakan bahwa nilai pengamatan pada waktu saat ini dipengaruhi pada nilai pengamatan pada waktu – waktu sebelumnya serta dipengaruhi faktor luar yaitu variabel eksogen.

Misal $\{Y_t^*\}$ nilai runtun waktu. Misal C ⊂ R daerah sensor ∋ Y_t^* tidak diobservasi jika $Y_t^* \in$ C. Daerah sensor berada di interval (-∞, c) disebut sensor kiri, (c, ∞) disebut sensor kanan, dan interval berhingga (c_t, c_u) disebut sensor interval. Pada prakteknya,

a. Data dengan subjek sensor kiri dengan sensor limit c bernilai Y_t = max {c, Y_t^*}
b. Data dengan subjek sensor kanan dengan sensor limit c bernilai Y_t = min {c, Y_t^*}
c. Data dengan subjek sensor interval dengan sensor limit c bernilai $Y_t = Y_t^* \times I(Y_t^* \notin \mathrm{C}) + \mathrm{c} \times \mathrm{I}(Y_t^* \in C)$ dimana $c = (c_t + c_u)/2$

Misal X_t adalah vektor kovariat yang berhubungan regresi linier dengan Y, dengan kesalahan diasumsikan mengikuti model Autoregresif (p). $\{X_t\}$ diasumsikan selalu terobservasi. Misal v^T adalah transpose dari matriks v. Berikut persamaan AR dan regresi untuk memodelkan CARX:

$$Y_t^* = X_t^T \beta + \eta_t \tag{3.10}$$

$$\text{Dimana } \eta t = \sum_{i=1}^{p} \varphi_i \, \eta_{t-1} + \varepsilon_t \tag{3.11}$$

Selanjutnya substitusikan persamaan (3.11) pada persamaan (3.10) sehingga diperoleh persamaan model CARX sebagai berikut:

$$Y_t^* = X_t^T \beta + \varphi_1 \eta_{t-1} + \varphi_2 \eta_{t-2} + \cdots + \varphi_p \eta_{t-p} + \varepsilon_t \tag{3.12}$$

Keterangan:

Y_t^*	: variabel acak waktu ke t
η_{t-1}	: variabel acak endogen tersensor waktu ke t
X_t^T	: variabel acak eksogen waktu ke t
φ	: parameter variabel endogen tersensor
β	: parameter variabel eksogen
ε_t	: sesatan dalam model CARX

Dari persamaan (3.12), didapat bentuk model lain menggunakan operator *backshift* namun masih memiliki pengertian yang sama seperti persamaan (3.12) adalah sebagai berikut:

$$Y_t^* = X_t^T\beta + \eta_t$$

$$Y_t^* = X_t^T\beta + \sum_{i=1}^{p} \varphi_i\ \eta_{t-1} + \varepsilon_t$$

$$Y_t^* - X_t^T\beta - \sum_{i=1}^{p} \varphi_i\ \eta_{t-1} = \varepsilon_t$$

$$Y_t^* - X_t^T\beta - (\ \varphi_1\eta_{t-1} + \varphi_2\eta_{t-2} + \ldots + \varphi_p\eta_{t-p}) = \varepsilon_t$$

$$Y_t^* - X_t^T\beta - (\ \varphi_1 B + \ \varphi_2 B^2 + \ldots + \varphi_p B^p)\ \eta_t = \varepsilon_t$$

$$Y_t^* - X_t^T\beta - \varphi(B)\eta_t = \varepsilon_t$$

$$Y_t^* - X_t^T\beta - \varphi(B)(Y_t^* - X_t^T\beta\ ) = \varepsilon_t$$

$$(1 - \varphi(B))\ (Y_t^* - X_t^T\beta\ ) = \varepsilon_t \qquad (3.13)$$

Persamaan (3.13) ini juga dapat disebut Model *Autoregressive Exogenous* Data Tersensor. Dimana ε_t diasumsikan berdistribusi normal dengan mean = 0 dan varian = σ^2. Parameter yang perlu ditaksir pada model ini adalah φ, β dan σ. Ketiga parameter tersebut di misalkan θ untuk mempermudah dalam proses penaksirannya.

Dengan ε_t diasumsikan berdistribusi normal dengan mean = 0 dan varian = σ^2 diperoleh bentuk distribusi normalnya sebagai berikut:

$$\ell(\mathrm{Y_t^*}|\mathrm{F_t^*};\theta) = -\frac{1}{2}\log(\sigma^2) - \{(Y_t^* - X_t^T\beta) - \sum_{j=1}^{p} \varphi_j(Y_{t-j}^* - X_{t-j}^T\beta)\}^2\ /2\sigma^2$$

dimana $Z_{t1,t2}(\theta) = E_\theta[\mathrm{Y_t^*}|\mathrm{F_t^*}]$ dan $\sum(\theta) = cov(\mathrm{Y_t^*}|\mathrm{F_t^*};\theta)$

Kelebihan dari model ini adalah dapat menyelesaikan permasalahan jika terdapat data tersensor *time series* untuk peramalan model AR dengan penambahan variabel eksogen sehingga peramalan yang dapat dilakukan lebih akurat.

Sedangkan kelemahan dari model ini adalah terbatasnya metode penaksiran yang dapat dilakukan yaitu hanya dengan metode *quasi likelihood* karena data tersensor membuat data tidak dapat diketahui ditribusinya.

3.4 Penaksiran Model CARX dengan Metode *Quasi Likelihood*

Metode *Quasi Likelihood* merupakan penaksiran parameter untuk model CARX karena model CARX belum diketahui distribusinya maka penaksiran yang dapat digunakan dengan metode *Quasi Likelihood*. Berikut uraian penaksiran parameter CARX dengan metode *Quasi Likelihood*:

1. Model CARX

   Model kali ini yang dipilih adalah CARX adalah sebagai berikut:

$$(1 - \varphi(B))\ (Y_t^* - X_t^T \beta\ ) = \varepsilon_t \tag{3.14}$$

   Dimana ε_t diasumsikan berdistribusi normal dengan mean = 0 dan varian = σ^2.

2. Menentukan fungsi *likelihood* dari model CARX

   Misal *{$\mathbf{Y_t^*}$} tersensor dan {ε_t} merupakan sesatan dari proses AR tersensor maka fungsi likelihood untuk* $\mathrm{Y}_{n=1}^*$ *dengan syarat* $X_{n=1}$ adalah sebagai berikut:

$$\ell(\mathrm{Y}_{n=1}^*|\mathrm{X}_{n=1};\theta) = \sum_{t=p+1}^{n} \ell(\mathrm{Y}_t^*|\mathrm{F}_t^*;\theta) + \ell\left(\mathrm{Y}_t^*\middle|\mathrm{X}_{p=1};\theta\right) \tag{3.15}$$

   Dimana

$$\ell(\mathrm{Y}_t^*|\mathrm{F}_t^*;\theta) = \log \mathrm{f}_\theta\big((1 - \varphi(B)\big)\ (Y_t^* - X_t^T \beta)) \tag{3.16}$$

3. *Dengan menginisialisasi parameter pada persamaan* (3.15), *persamaan ini dapat disederhanakan menjadi:*

$$\ell_*(\theta) = \sum_{t=p+1}^{n} \ell(\mathrm{Y}_t^*|\mathrm{F}_t^*;\theta) \tag{3.17}$$

4. Menurunkan persamaan (3.17) terhadap θ dan turunannya harus sama dengan nol untuk memaksimumkan:

$$0 = \partial \ell_*(\theta) = \sum_{t=p+1}^{n} \nabla \ell(\mathrm{Y}_t^*|\mathrm{F}_t^*; \theta) \qquad (3.18)$$

dengan ∂ notasi turunan parsial untuk θ.

Misal $S(\mathrm{Y}_t^*|\mathrm{F}_t^*; \theta) = \nabla \ell(\mathrm{Y}_t^*|\mathrm{F}_t^*; \theta)$, karena persamaan (3.18) merupakan persamaan tak bias maka *ganti* $S(\mathrm{Y}_t^*|\mathrm{F}_t^*; \theta)$ *dengan ekspektasinya sehingga diperoleh persamaan:*

$$\sum_{t=p+1}^{n} \mathrm{E}_\theta(S_t(\theta)|\mathrm{Y}_{t-k}, \mathrm{X}_{t-k}, k = 0, \dots, p) = 0 \qquad (3.19)$$

Untuk menyelesaikan persamaan (3.19) dilakukan metode *Quasi Likelihood.* Jika data bukan merupakan data tersensor, maka cukup sampai menyelesaikan persamaan (3.18). Berikut adalah skema untuk memecahkan persamaan (3.19) (Wang & Chan, 2017):

a. Inisialisasi estimasi parameter yang dilambangkan dengan $\theta^{(0)}$.
   Estimasi awal $\theta^{(0)}$ diasumsikan sebagai estimator yang konsisten sehingga ruang parameter Θ dibatasi pada beberapa lingkungan kompak Θ dari θ yang merupakan parameter sebenarnya. Estimasi $\theta^{(0)}$ dibatasi agar iterasi yang dilakukan tidak terlalu banyak sehingga hasilnya dalam keadaan konvergen.
b. Untuk setiap k=1,...., maka $\theta^{(k)}$ menjadi

$$\theta^{(k)} = argmax_\theta Q(\theta|\theta^{(k-1)})$$

   Dimana $Q\left(\theta\middle|\theta^{(k-1)}\right) = \sum_{t=p+1}^{n} Q_t(\theta|\theta^{(k-1)})$

$$Q_t\left(\theta\middle|\theta^{(k-1)}\right) = E_\theta(\mathrm{S}(\mathrm{Y}_t^*|\mathrm{F}_t^*; \theta)|\mathrm{G}_t)$$

c. Iterasi langkah (b) sehingga diperoleh penaksiran untuk $\hat{\theta}$ adalah $\left\|\theta^k - \theta^{(k-1)}\right\| / \|\theta^{k-1}\| < \varepsilon$. Biarkan $\hat{\theta}$ menjadi estimasi yang diperoleh.

3.5 Tahapan Analisis Data

Tahap 1: pengumpulan data

Pada tahap ini dilakukan pengumpulan data curah hujan dan kecepatan angin di kota Bandung. Data yang digunakan merupakan data sekunder yang diperoleh dari website data Bandung yang bekerjasama dengan BMKG. Data tersebut harus memenuhi syarat sebagai data tersensor interval, sehingga dalam penelitian ini dilakukan pembatasan dengan memilih data curah hujan dan kecepatan angin yang mengikuti hujan munson barat pada bulan Januari 2014 hingga Desember 2019 dengan batas sensor atas dan batas sensor bawah sehingga data dapat di bentuk menggunakan model CARX.

Tahap 2: Uji Stasioneritas dan Uji Normalitas

Pada tahap ini data yang diperoleh dilakukan uji stasioneritas menggunakan uji uji *Augmented dickey fuller* (ADF) dan untuk uji normalitas menggunakan uji *Shapiro wilk.* Kedua uji tersebut dilakukan pada *software* R. Kriteria pengujian untuk uji ADF adalah tolak H_0 jika p-value > α, sedangkan uji *Shapiro wilk* adalah terima H_0 jika p-value > α.

Tahap 3: Transformasi Data

Tahap ini dilakukan jika pada tahap 2 diperoleh hasil uji stasioneritas bahwa data yang akan dilakukan belum stasioner. Maka data perlu dilakukan transformasi dalam bentuk logaritma.

Tahap 4: Identifikasi model

Pada tahap ini dilakukan identifikasi model dengan menggunakan model ACF dan PACF sehingga diperoleh signifikansi pada lag ke p. Jika plot PACF signifikan pada lag ke p dan plot ACF menurun secara eksponensial, maka model yang digunakan adalah AR(p), namun data yang digunakan adalah data tersensor, maka model yang digunakan adalah model AR(p) tersensor.

Tahap 5: verifikasi model AR

Model-model AR yang terpilih selanjutnya dilakukan verifikasi dengan cara di cek uji keberartian koefisisen, uji kecocokan model dan uji variansi sesatannya.

Tahap 6: penambahan variabel eksogen pada model AR(p) tersensor

Setelah di peroleh model AR(p) tersensor dilihat dari plot ACF dan PACF, maka tahap selanjutnya adalah penambahan variabel eksogen yaitu variabel kecepatan angin ke dalam model AR(p) tersensor yang merupakan hasil dari data curah hujan sehingga model berubah menjadi model CARX(p).

Tahap 7: menaksir parameter model CARX

Setelah diketahui model yang digunakan adalah CARX(p) Selanjutnya adalah melakukan estimasi parameter model CARX dengan menggunakan metode *Quasi Likelihood*. Paramater yang harus diestimasi model CARX(p) adalah β, φ dan σ.

Tahap 8: *Checking diagnosis*

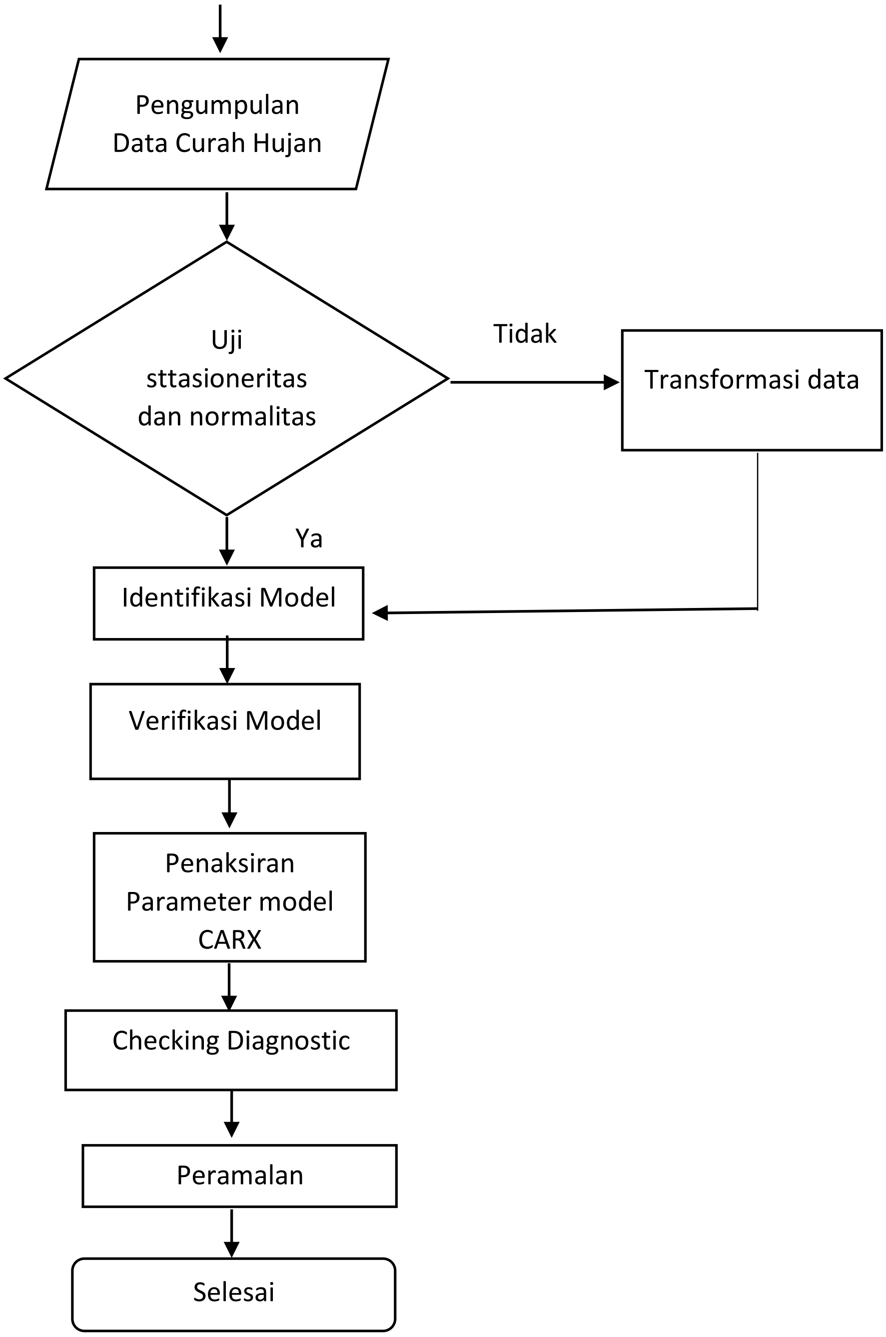
Setelah diperoleh nilai parameter dari model CARX(p), selanjutnya dilakukan *checking diagnosis* untuk untuk mengetahui apakah residual telah memenuhi asumsi residual *white noise* dan berdistribusi normal. Diperoleh dari hasil plot residual serta Ljung-Box dengan menggunakan software R.

Tahap 9: peramalan

Langkah terakhir dalam penelitian ini adalah peramalan. Sebelum dilakukan peramalan, terlebih dahulu dihitung nilai MAPE pada data yang digunakan untuk melihat ketepatan peramalan dalam suatu pemodelan. Setelah didapat nilai MAPE yang termasuk kriteria baik maka model tersebut dapat dilakukan peramalan untuk data curah hujan pada waktu yang akan datang.

Selanjutnya tahapan analisis dibuat dalam bentuk diagram pada Gambar 3.1 sebagai berikut:

\



Gambar 3.1 Tahapan Analisis Data

1. Prosedur Penggunaan *Software* minitab 7 untuk menentukan plot ACF dan PACF

Perangkat lunak minitab 7 dapat membantu dalam menentukan plot ACF dan PACF. Langkah – langkah penggunaan perangkat lunak minitab 7 adalah sebagai berikut.

1. Aktifkan aplikasi Minitab 17, lalu masukkan data yang akan diolah ke dalam worksheet minitab.
2. Lalu pilih menu *Stat*→*Time Series*→*Time series plot*
3. Pilih *Simple Plot* lalu klik OK.
4. Setelah itu masukan kolom berisi data yang akan kita olah ke dalam kolom Series klik Ok.
5. Untuk memastikan apakah data tersebut stasioner atau tidak dapat kita lakukan dengan *Trend Analysis Plot*. Caranya, pilih menu *Stat*→*Time Series*→*Trend Analysis*
6. Setelah itu akan muncul kotak dialog seperti dibawah ini, lalu masukkan data ke kolom yang disediakan dengan klik data yang akan di cek, lalu klik *select*→ klik *Linier* pada Model *Type*→ Ok.
7. Selain itu perlu perhatikan juga plot dari fungsi autokorelasi (ACF) dan fungsi autokorelasi parsial (PACF) data tersebut.
8. Perhatikan plot ACF dan PACF. Jika plot PACF signifikan pada lag ke p dan plot ACF menurun secara eksponensial, maka model yang digunakan adalah AR(p).

2. Prosedur Penggunaan *Software* R untuk penaksiran parameter

Perangkat lunak R dapat membantu dalam mengestimasi parameter model CARX dan nilai peramalannya. Langkah – langkah penggunaan perangkat lunak R dalam mengestimasi model CARX sebagai berikut:

1. Membuka perangkat lunak R 3.6.0
2. Menentukan package yang akan digunakan untuk model CARX
   ```
   >library(TSA)
   >library(tseries)
   >library(CARX)
   ```
3. Memasukkan data

Data yang digunakan dalam format excel harus diubah terlebih dahulu ke bentuk ".csv" lalu masukan ke dalam perangkat lunak R

>ch<-read.csv(file.choose(), header=TRUE)

>ka<-read.csv(file.choose(), header=TRUE)

>chka<-read.csv(file.choose(), header=TRUE)

4. Stasioneritas dan Normalitas

   Uji stasioneritas dengan uji ADF dan uji normalitas dengan shapiro wilk

   >adf.test(chka$ch)

   >shapiro.test(chka$ch)

5. Mengestimasi model

   Mdl<-carx(y=chka$CH, x=chka$KA, p=2, prmtrX=NULL, prmtrAR=NULL, sigma=NULL).

   Keterangan:

   p : Orde dari *Autoregressive*. Default=1.

   prmtrX : Nilai awal parameter regresi untuk x. default=NULL.

   prmtrAR : Nilai awal parameter untuk koefisien *Autoregressive.* Default=NULL.

   sigma : Nilai awal parameter untuk standar deviasi. Default=NULL.

6. *Checking Diagnostic*

   >tsdiag(mdl,p=2)

7. Peramalan

   > forecast<-predict(mdl,newxreg,n.ahead=4)

   > U=forecast$pred + forecast$se

   > L=forecast$pred – forecast$se