

BAB III

METODE PENELITIAN

3.1 Jenis dan Sumber Data

Data yang digunakan dalam penelitian ini adalah data sekunder berupa data curah hujan di kabupaten Bogor. Data yang digunakan merupakan data bulanan dari periode Januari 1985 sampai dengan Desember 2021 yang diperoleh dari *website* resmi Badan Meteorologi, Klimatologi, dan Geofisika (BMKG).

Pada penelitian ini data dibagi menjadi dua, yaitu 410 data *training* dan 34 data *testing*. Data *training* digunakan untuk pelatihan model, sedangkan data *testing* digunakan untuk mengetahui keakuratan pada peramalan.

3.2 Variabel Penelitian

Model runtun waktu yang digunakan merupakan model runtun waktu univariat, sehingga hanya terdapat satu variabel yang digunakan dalam penelitian ini, yaitu data curah hujan di kabupaten Bogor. Data tersebut merupakan data dalam waktu bulanan.

3.3 Langkah Analisis

Langkah-langkah analisis data diolah dengan menggunakan bantuan *software RStudio, Python, dan Microsoft Excel*. Adapun langkah-langkah analisis yang dilakukan pada penelitian ini adalah sebagai berikut:

3.3.1 Uji Linearitas Data

Melakukan uji linearitas data curah hujan di Bogor sebelum mengidentifikasi model SARIMA untuk mengetahui data curah hujan di Bogor berpola linear atau tidak.

3.3.2 Pembentukan Model SARIMA

1. Mengidentifikasi stasioneritas data dalam varians dan rata-rata.
 - a. Melakukan transformasi *Box-Cox* pada data yang belum stasioner dalam varians.
 - b. Melakukan *differencing* non musiman pada data yang belum stasioner terhadap rata-rata non musiman.

- c. Melakukan *differencing* musiman pada data yang belum stasioner terhadap rata-rata musiman.
2. Melakukan identifikasi model SARIMA.
3. Melakukan estimasi parameter.
4. Melakukan uji signifikansi parameter model sementara.
5. Melakukan uji asumsi residual *white noise* dengan *Ljung Box-Pierce*.
6. Menetapkan model ARIMA $[(p, d, q)(P, D, Q)^S]$ terbaik berdasarkan nilai *Akaike Information Criterion* (AIC).
7. Melakukan peramalan pada data curah hujan di Bogor.

3.3.3 Pembentukan Model LSTM

1. Normalisasikan data *training* dengan menggunakan metode *minmax scaler*. Metode ini menghasilkan nilai baru antara 0 – 1. Dengan rumus sebagai berikut:

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (3.1)$$

dengan:

x' : data normalisasi

x : data asli

x_{max} : nilai maksimum dari data perkolom

x_{min} : nilai minimum dari data perkolom

2. Membentuk model pelatihan LSTM dengan menentukan nilai parameter yang optimal dari masing-masing *layer* yang ditentukan terlebih dahulu.
3. Menetapkan model pelatihan LSTM terbaik untuk peramalan data *testing*.
4. *Input* data *testing* dengan menormalisasikannya terlebih dahulu ke dalam model pelatihan LSTM yang telah didapatkan sebelumnya.
5. Lakukan peramalan data *testing*.
6. Mengevaluasi hasil peramalan data *testing* berdasarkan *Mean Absolute Percentage Error* (MAPE) dan *Mean Absolute Error* (MAE).

3.3.4 Model Hybrid SARIMA-LSTM

Penggabungan atau *hybrid* pada model SARIMA dan LSTM ini bertujuan untuk dapat mengatasi permasalahan terkait data linear dan non linear yang ada pada

SARIMA dan LSTM. Berikut merupakan kombinasi model runtun waktu yang memiliki struktur linear dan non linear secara umum:

$$Z_t = L_t + N_t \quad (3.2)$$

dengan:

L_t : komponen data linear pada waktu ke-t.

N_t : komponen data non linear pada waktu ke-t.

Model SARIMA akan digunakan untuk menyelesaikan komponen data linear. Kemudian diperoleh residual yang dapat dituliskan dalam persamaan berikut:

$$\varepsilon_t = Z_t - \hat{L}_t \quad (3.3)$$

dengan:

Z_t : nilai aktual pada waktu ke-t.

\hat{L}_t : nilai ramalan pada waktu ke-t.

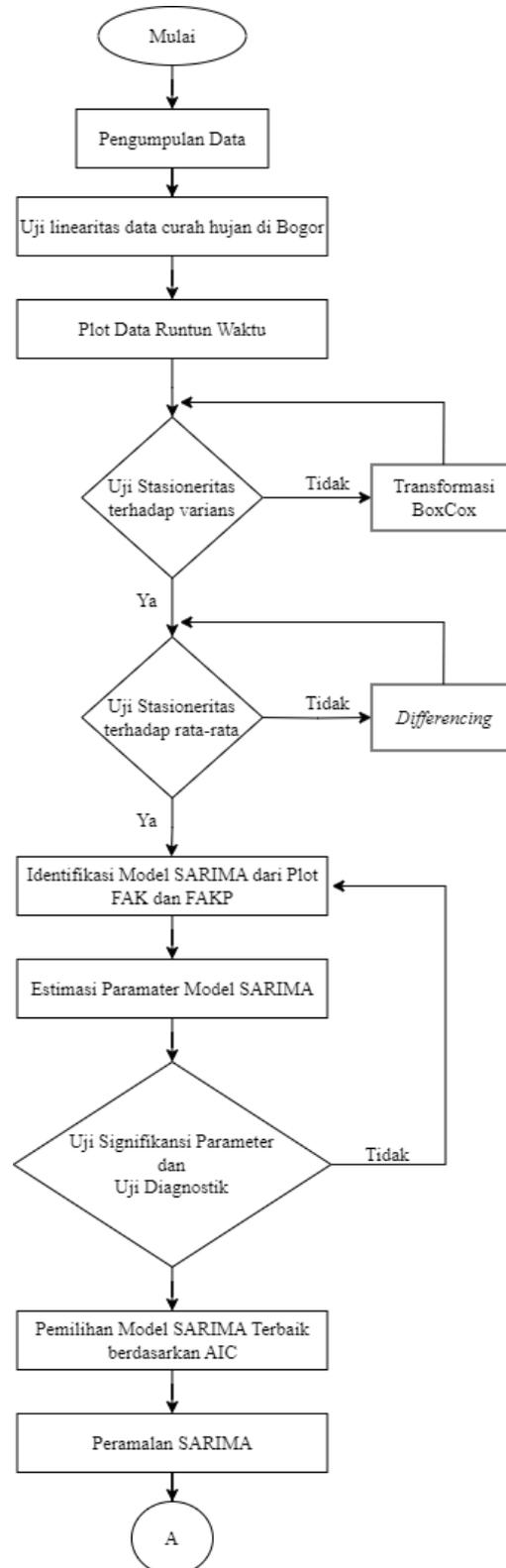
ε_t : nilai residual pada waktu ke-t

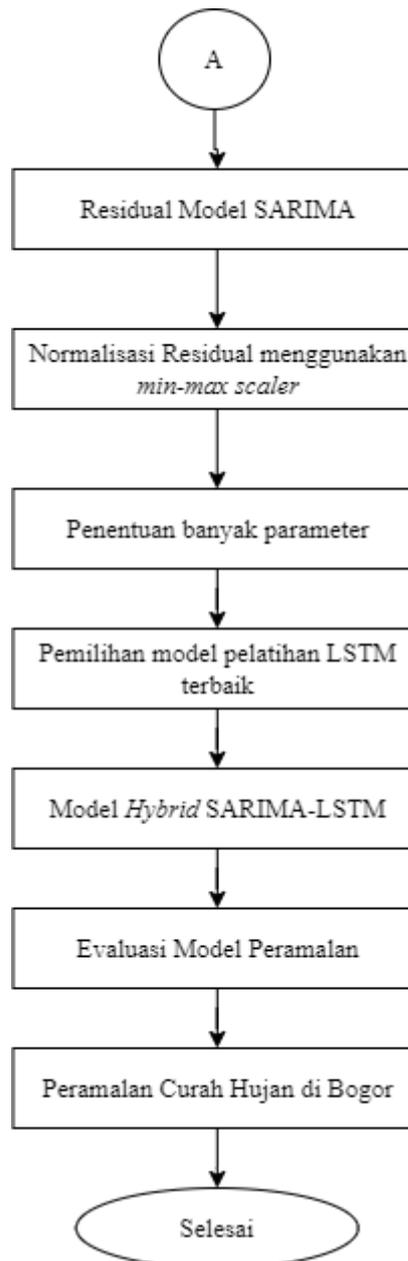
Setelah diperoleh model SARIMA, selanjutnya memodelkan residual model SARIMA dengan menggunakan LSTM. Hasil ramalan dari metode LSTM di gabungkan dengan hasil ramalan SARIMA. Hasil ramalan keseluruhan secara matematis dapat dituliskan dalam persamaan berikut:

$$Z_t = \hat{L}_t + N_t \quad (3.4)$$

3.5 Alur Penelitian

Gambar 3.1 menunjukkan alur penelitian untuk menggambarkan langkah-langkah yang akan dilakukan dalam menganalisis data.





Gambar 3.1 Prosedur Peramalan *Hybrid* SARIMA-LSTM

BAB IV

HASIL DAN PEMBAHASAN

Bab ini membahas penerapan model *Hybrid Seasonal Autoregressive Integrated Moving Average – Long Short Term Memory* (SARIMA-LSTM) untuk data curah hujan di kabupaten Bogor periode Januari 1985 sampai Desember 2021. Data yang digunakan pada penelitian ini diperoleh dari *website* resmi Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) <https://dataonline.bmkg.go.id/>.

4.1 Uji Linearitas

Uji linearitas digunakan untuk mengetahui apakah data yang digunakan bersifat linear atau tidak. Model SARIMA memiliki kekurangan, yaitu mengabaikan model non linear, sedangkan data curah hujan yang akan digunakan belum tentu bersifat linear. Maka dari itu perlu untuk dilakukan uji linearitas terhadap data curah hujan. Uji linearitas yang akan digunakan pada penelitian ini, yaitu menggunakan uji *Terasvirta*. Berikut hasil uji linearitas *Terasvirta* pada data curah hujan di Bogor periode Januari 1985 - Desember 2021:

Tabel 4.1 Uji *Terasvirta* Data Curah Hujan Bogor

Data	<i>P-value</i>
Curah hujan di Bogor	0.0001

Berdasarkan tabel 4.1 dapat dilihat bahwa nilai *p-value* yang diperoleh dari uji *Terasvirta* yaitu sebesar 0.0001, karena nilai *p-value* kurang dari tingkat signifikansi $\alpha = 0,05$ maka dapat disimpulkan bahwa H_0 ditolak. Hal ini menunjukkan bahwa data curah hujan di Bogor memiliki pola yang bersifat non linear. Selanjutnya, berdasarkan hasil uji *terasvirta* yang menunjukkan bersifat non linear, maka peramalan curah hujan dapat dilakukan menggunakan model peramalan *hybrid* yang dapat mengatasi permasalahan tersebut. Untuk informasi lebih rinci mengenai hasil uji linearitas menggunakan uji *Terasvirta*, dapat dilihat pada lampiran 2.

4.2 Pembentukan model SARIMA

4.2.1 Prosedur *Input Data* dan *Split Data*

Data curah hujan di Bogor yang berupa file csv dimasukkan ke dalam prosedur sebagai berikut:

```
rm(list = ls())
library(readxl)
library(starma)
library(tseries)
library(MASS)
library(forecast)
library(lmtest)
library(TSA)
library(Metrics)
library(tidyverse)
library(FinTS)

#INPUT DATA CURAH HUJAN
setwd("C:\\Users\\Asus\\Downloads\\CITEKO NO 8888")
data <- read.csv("sumciteko.csv", sep=",")
hujan <- ts(data)
```

Gambar 4.1 Prosedur *Input Data* SARIMA

Setelah data curah hujan di masukkan, kemudian data curah hujan dibagi menjadi 2 yakni data *training* dan data *testing*. Dalam penelitian ini, data dibagi menjadi 410 data *training* dan 34 data *testing* dengan prosedur sebagai berikut:

```
#PLOT DATA CURAH HUJAN
ts.plot(hujan, col="blue",
        xlab="Jan 1985 - Des 2021", ylab="Curah Hujan",
        main="Plot Data Curah Hujan Per Bulan")
```

Gambar 4.2 Prosedur *Split Data* SARIMA

4.2.2 Identifikasi Data Curah Hujan

Dalam melakukan analisis data runtun waktu, langkah pertama yang dilakukan adalah mengidentifikasi data. *Identifikasi* data dilihat berdasarkan plot data curah hujan. Berikut adalah *syntax* pada Rstudio untuk memperoleh plot data curah hujan di Bogor pada periode Januari 1985 sampai Desember 2021.