



**TUGAS AKHIR - KS184822**

**PEMODELAN BANJIR GENANGAN DI KOTA SURABAYA  
BERDASARKAN DATA BMKG DENGAN MENGGUNAKAN  
REGRESI INVERSE GAUSSIAN DAN REGRESI GAMMA**

**ANGGA MAHENDRA  
NRP 062115 4000 0035**

**Dosen Pembimbing:  
Dr.rer.pol. Dedy Dwi Prastyo, S.Si., M.Si.**

**PROGRAM STUDI SARJANA  
DEPARTEMEN STATISTIKA  
FAKULTAS MATEMATIKA, KOMPUTASI, DAN SAINS DATA  
INSTITUT TEKNOLOGI SEPULUH NOPEMBER  
SURABAYA 2019**







**TUGAS AKHIR - KS184822**

**PEMODELAN BANJIR GENANGAN DI KOTA SURABAYA  
BERDASARKAN DATA BMKG DENGAN MENGGUNAKAN  
REGRESI INVERSE GAUSSIAN DAN REGRESI GAMMA**

**ANGGA MAHENDRA  
NRP 062115 4000 0035**

**Dosen Pembimbing:  
Dr.rer.pol. Dedy dwi Prastyo,S.Si.,M.Si.**

**PROGRAM STUDI SARJANA  
DEPARTEMEN STATISTIKA  
FAKULTAS MATEMATIKA, KOMPUTASI, DAN SAINS DATA  
INSTITUT TEKNOLOGI SEPULUH NOPEMBER  
SURABAYA 2019**





**FINAL PROJECT - KS184822**

**MODELING OF FLOOD IN SURABAYA CITY BASED ON  
BMKG DATA USING INVERSE GAUSSIAN REGRESSION  
AND GAMMA REGRESSION APPROACHES**

**ANGGA MAHENDRA  
NRP 062115 4000 0035**

**Supervisor:  
Dr.rer.pol. Dedy dwi Prastyo,S.Si.,M.Si.**

**UNDERGRADUATE PROGRAMME  
DEPARTMENT OF STATISTICS  
FACULTY OF MATHEMATICS, COMPUTING, AND DATA SCIENCE  
INSTITUT TEKNOLOGI SEPULUH NOPEMBER  
SURABAYA 2019**



## LEMBAR PENGESAHAN

### PEMODELAN BANJIR GENANGAN DI KOTA SURABAYA BERDASARKAN DATA BMKG DENGAN MENGUNAKAN REGRESI INVERSE GAUSSIAN DAN REGRESI GAMMA

#### TUGAS AKHIR

Diajukan untuk Memenuhi Salah Satu Syarat  
Memperoleh Gelar Sarjana Statistika

pada

Program Studi Sarjana Departemen Statistika  
Fakultas Matematika, Komputasi, dan Sains Data  
Institut Teknologi Sepuluh Nopember

Oleh :

**Angga Mahendra**  
NRP. 062115 4000 0035

Disetujui oleh Pembimbing:

**Dr. rer. Pol. Dedy Dwi Prastyo, S.Si., M.Si.**  
NIP. 19831204 200812 2 002



Mengetahui,  
Kepala Departemen Statistika

**Dr. Suhartono**  
NIP. 19710929 199512 1 001

SURABAYA, JULI 2019

*(Halaman ini sengaja dikosongkan)*



**PEMODELAN BANJIR GENANGAN DI KOTA  
SURABAYA BERDASARKAN DATA BMKG DENGAN  
MENGUNAKAN REGRESI INVERSE GAUSSIAN DAN  
REGRESI GAMMA**

**Nama Mahasiswa** : Angga Mahendra  
**NRP** : 062115 4000 0035  
**Departemen** : Statistika  
**Dosen Pembimbing** : Dr.rer.pol. Dedy Dwi Prastyo, S.Si.,  
M.Si

**Abstrak**

*Secara Geografis Indonesia terletak diantara dua benua (Benua Asia dan Benua Australia) dan dua samudra (Pasifik dan Hindia). Indonesia dilalui oleh angin monsun yaitu angin monsun Barat dan angin monsun Timur. Dampak negatif yang ditimbulkan akibat intensitas curah hujan yang tinggi di Surabaya adalah banjir. Penelitian ini akan dilakukan pemodelan curah hujan menggunakan data curah hujan harian mulai tahun 2010 sampai dengan 2018 di 6 stasiun Kota Surabaya yaitu Gunungsari, Wonokromo, Wonorejo, Keputih, Kedung Cowek, dan Gubeng. Prediksi curah hujan dilakukan dengan pendekatan regresi Inverse Gaussian dan regresi Gamma. Variabel prediktor yang digunakan merupakan variabel yang memengaruhi curah hujan seperti kelembaban rata rata, kecepatan angin rata rata, dan suhu udara (temperature) rata rata. Pemodelan dilakukan secara univariat dengan multivariate. Pemodelan curah hujan univariate menghasilkan pemodelan yang terbaik yaitu berupa pasangan reespon-prediktor CH Keputih – Juanda, CH Kd. Cowek – Perak2, CH. Gubeng- Juanda, CH. Wonorejo – Juanda, CH. Wonokromo – Juanda, dan CH – Gn. Sari-Juanda. Kesimpulan yang didapatkan bahwa hasil prediksi curah hujan maksimum sangat rendah (tertinggi pada balai hujan Gubeng sebesar 4.121) sehingga tidak ada saluran yang berpotensi banjir di kota Surabaya.*

**Kata Kunci** : *Inverse Gaussian, Regresi Gamma, Fungsi Link, Curah Hujan*

*(Halaman ini sengaja dikosongkan)*

# MODELING OF FLOOD IN SURABAYA CITY BASED ON BMKG DATA USING INVERSE GAUSSIAN REGRESSION AND GAMMA REGRESSION APPROACHES

**Name** : Angga Mahendra  
**Student Number** : 062115 4000 0035  
**Department** : Statistics  
**Supervisor** : Dr.rer.pol. Dedy Dwi Prastyo.S.Si.,M.Si.

## **Abstract**

*Indonesia is located between two continents (Continental Asia and Continental Australia) and two oceans (Pacific and Indian). Indonesia is traversed by the monsoon namely the West monsoon and the East monsoon. The negative impact caused by the intensity of high rainfall in Surabaya is flooding. Rainfall modeling in this work is conducted using daily rainfall data from 2010 to 2018 in six stations, namely Gunungsari, Wonokromo, Wonorejo, Keputih, Kedung Cowek, and Gubeng. The rainfall prediction is done using Inverse Gaussian regression and Gamma regression approaches. The predictor used is variable that affects rainfall such as average humidity, average wind speed, and average air temperature. The modeling is done by multivariate and univariate in term of number of prediktor. Univariate rainfall modeling produces the best modeling with pair of rainfall-prediktor's station as CH Keputih - Juanda, CH Kd. Cowek - Perak2, CH. Gubeng-Juanda, CH. Wonorejo - Juanda, CH. Wonokromo - Juanda, and CH Gunung Sari - Juanda. The result concludes that the maximum rainfall prediction are very low (the highest in the Gubeng rainforest hall is 4,121 mm) so there are no potentially flooded channels in the city of Surabaya.*

**Keywords** : *Inverse Gaussian, Gamma Regression, Link Function, Rainfall*

*(Halaman ini sengaja dikosongkan)*

## KATA PENGANTAR

Segala puji syukur dipanjatkan kehadirat Tuhan Yang Maha Esa atas segala kasih, hikmat, serta rahmat-Nya sehingga laporan Tugas Akhir yang berjudul **“Pemodelan Banjir Genangan di Kota Surabaya Berdasarkan Data BMKG menggunakan Regresi Inverse Gaussian dan Regresi Gamma”** ini dapat diselesaikan dengan baik dan lancar.

Penulis menyadari bahwa selama proses penyelesaian laporan Tugas Akhir ini tidak lepas dari bantuan dan bimbingan dari berbagai pihak. Pada kesempatan kali ini, dengan kerendahan hati penulis mengucapkan terima kasih kepada:

1. Bapak Dr. Suhartono selaku Kepala Departemen Statistika FMKSD ITS
2. Bapak Dr. Sutikno, M.Si. selaku Ketua Program Studi S1 Departemen Statistika FMKSD ITS
3. Bapak Dr. rer. pol. Dedy Dwi Prastyo S.Si., M.Si. selaku dosen pembimbing Tugas Akhir yang telah meluangkan waktu dan dengan sabar memberikan bimbingan, saran, motivasi dan dukungan selama proses penulisan laporan Tugas Akhir
4. Seluruh dosen dan *staff* pengajar Program Studi Sarjana Departemen Statistika FMKSD ITS yang telah membantu penulis selama menempuh masa kuliah.
5. Ayah, Ibu, Rindy Antika dan Deden Dwi Ralisandu selaku keluarga penulis yang telah memberikan bantuan doa dan motivasi bagi penulis.
6. Teman-teman Statistika ITS angkatan 2015, Vivacious, yang selalu memberikan motivasi kepada penulis
7. Tim bimbingan Tugas Akhir Pak Dedy yang selalu memberikan informasi dan telah bekerja sama dengan baik dengan penulis dalam menyelesaikan Tugas Akhir ini.
8. Risdha Ikfina Putri yang selalu memberikan Motivasi, Dukungan, Bantuan, dan Semangat bagi penulis dalam menyelesaikan Tugas Akhir ini.
9. Semua teman, relasi, dan berbagai pihak yang tidak bisa penulis sebutkan namanya satu persatu yang telah membantu

dalam keberhasilan proses yang panjang ini, baik secara langsung maupun tidak langsung, semoga Tuhan Yang Maha Esa selalu memberkati.

Penulis menyadari bahwa dalam penulisan dan penyusunan laporan Tugas Akhir ini masih banyak kekurangan dan kelemahan. Oleh karena itu, kritik dan saran membangun akan sangat membantu penulis dalam memperbaikinya di masa yang akan xiiemato. Penulis berharap semoga laporan Tugas Akhir ini dapat bermanfaat bagi masyarakat dan bagi ilmu pengetahuan.

Surabaya, Juli 2019

Penulis

# DAFTAR ISI

Halaman

|  |              |
|--|--------------|
| <b>HALAMAN JUDUL</b> .....                                   | <b>i</b>     |
| <b>LEMBAR PENGESAHAN</b> .....                               | <b>v</b>     |
| <b>KATA PENGANTAR</b> .....                                  | <b>xi</b>    |
| <b>DAFTAR ISI</b> .....                                      | <b>xiii</b>  |
| <b>DAFTAR GAMBAR</b> .....                                   | <b>xvi</b>   |
| <b>DAFTAR TABEL</b> .....                                    | <b>xxi</b>   |
| <b>DAFTAR LAMPIRAN</b> .....                                 | <b>xxiii</b> |
| <b>BAB I PENDAHULUAN</b> .....                               | <b>1</b>     |
| 1.1 Latar Belakang.....                                      | 1            |
| 1.2 Rumusan Masalah.....                                     | 3            |
| 1.3 Tujuan.....  | 4            |
| 1.4 Manfaat.....   | 4            |
| <b>BAB II TINJAUAN PUSTAKA</b> .....                         | <b>5</b>     |
| 2.1 Statistika Deskriptif.....                               | 5            |
| 2.2 <i>Principle Component Analysis</i> (PCA).....           | 5            |
| 2.3 Keluarga Distribusi Eksponensial.....                    | 7            |
| 2.4 <i>Generalized Linier Model</i> .....                    | 10           |
| 2.4.1. Bentuk umum GLM.....                                  | 11           |
| 2.4.2. Fungsi Hubung.....                                    | 11           |
| 2.4.3. Fungsi Hubung <i>Square Root</i> .....                | 12           |
| 2.5 <i>Regresi Inverse Gaussian</i> .....                    | 13           |
| 2.6 Estimasi Parameter Regresi <i>Inverse Gaussian</i> ..... | 13           |
| 2.7 Regresi Gamma.....                                       | 18           |
| 2.8 Estimasi parameter.....                                  | 19           |
| 2.9 Metode Newton-Raphson.....                               | 20           |
| 2.10 Uji <i>Goodness of Fit</i> .....                        | 21           |
| 2.11 Uji Signifikansi Parameter dengan Uji Wald.....         | 22           |
| 2.12 Metode Rasional untuk Prediksi Banjir.....              | 23           |
| 2.13 Metode Mononobe.....                                    | 24           |
| 2.14 Uji Korelasi.....                                       | 24           |
| 2.15 <i>Root Mean Square Error</i> .....                     | 25           |
| 2.16 Tinjauan Non Statistik.....                             | 25           |
| 2.10.1 Curah Hujan.....                                      | 25           |
| 2.10.2 Banjir Genangan.....                                  | 26           |

|                |  |           |
|----------------|--|-----------|
| <b>BAB III</b> | <b>METODOLOGI PENELITIAN .....</b>   | <b>29</b> |
| 3.1            | Sumber Data.....   | 29        |
| 3.2            | Variabel Penelitian.....   | 29        |
| 3.3            | Variabel dan Definisi Operasional.....   | 30        |
| 3.4            | Struktur Data.....   | 31        |
| 3.5            | Langkah Penelitian.....  | 32        |
| <b>BAB IV</b>  | <b>HASIL DAN PEMBAHASAN .....</b>  | <b>35</b> |
| 4.1            | <i>Time Series Plot</i> .....  | 35        |
| 4.1.1          | <i>Time Series Plot</i> Temperatur di Setiap Stasiun...35                                      |           |
| 4.1.2          | <i>Time Series Plot</i> Kelembapan di Setiap Stasiun .36                                       |           |
| 4.1.3          | <i>Time Series Plot</i> Kecepatan Angin di Setiap Stasiun<br>.....                             | 36        |
| 4.2            | Boxplot.....   | 37        |
| 4.2.1          | <i>Boxplot</i> Temperatur di Setiap Stasiun.....   | 37        |
| 4.2.2          | <i>Boxplot</i> Kelembapan di Setiap Stasiun .....  | 38        |
| 4.2.3          | <i>Boxplot</i> Kecepatan Angin di Setiap Stasiun.....  | 39        |
| 4.3            | Curah Hujan .....  | 40        |
| 4.4            | <i>Scatter Plot</i> .....  | 41        |
| 4.5            | Pemodelan <i>Univariate</i> Curah Hujan di Surabaya dengan<br>Variabel di Setiap Stasiun ..... | 51        |
| 4.5.1          | Pemodelan Curah Hujan Keputih.....   | 51        |
| 4.5.2          | Pemodelan Curah Hujan Gubeng .....   | 75        |
| 4.5.3          | Pemodelan Curah Hujan Kedung Cowek.....  | 82        |
| 4.5.4          | Pemodelan Curah Hujan Wonorejo.....  | 87        |
| 4.5.5          | Pemodelan Curah Hujan Wonokromo .....  | 94        |
| 4.5.6          | Pemodelan Curah Hujan Gunung Sari .....  | 100       |
| 4.6            | Pemodelan <i>Multivariat</i> Curah Hujan dengan Variabel<br>Stasiun BMKG.....                  | 107       |
| 4.6.1          | Pemodelan <i>Multivariat</i> Curah Hujan Keputih<br>dengan variabel Stasiun BMKG.....          | 107       |
| 4.6.2          | Pemodelan <i>Multivariat</i> Curah Hujan Kedung<br>Cowek dengan variabel Stasiun BMKG .....    | 114       |
| 4.6.3          | Pemodelan <i>Multivariat</i> Curah Hujan Gubeng<br>dengan variabel Stasiun BMKG.....           | 117       |
| 4.6.4          | Pemodelan <i>Multivariat</i> Curah Hujan Wonokromo<br>dengan Variabel Stasiun BMKG.....        | 119       |
| 4.6.5          | Pemodelan <i>Multivariat</i> Curah Hujan Wonorejo<br>dengan variabel Stasiun BMKG.....         | 121       |

|              |   |            |
|--------------|---|------------|
| 4.6.6        | Pemodelan <i>Multivariat</i> Curah Hujan Gunung Sari dengan variabel Stasiun BMKG ..... | 124        |
| 4.7.         | Pemilihan Model Terbaik .....   | 126        |
| 4.8          | Prediksi Curah Hujan dan Banjir di Kota Surabaya ..                                     | 127        |
| 4.8.1        | Prediksi Banjir di Keputih.....   | 127        |
| 4.8.2        | Prediksi Banjir di Gubeng.....  | 128        |
| 4.8.3        | Prediksi Banjir di Kd. Cowek .....  | 129        |
| 4.8.4        | Prediksi Banjir di Wonorejo.....  | 130        |
| 4.8.5        | Prediksi Banjir di Wonokromo .....  | 131        |
| 4.8.6        | Prediksi Banjir di Gunung Sari .....  | 132        |
| <b>BAB V</b> | <b>KESIMPULAN DAN SARAN .....</b>   | <b>135</b> |
| 5.1          | Kesimpulan .....  | 135        |
| 5.2          | Saran   | 136        |
|              | <b>DAFTAR PUSTAKA .....</b>   | <b>137</b> |
|              | <b>LAMPIRAN.....</b>  | <b>139</b> |
|              | <b>BIODATA PENULIS</b>  |            |

## DAFTAR GAMBAR

|   |    |
|---|----|
| <b>Gambar 2.1</b> Distribusi Gamma.....   | 19 |
| <b>Gambar 4.1</b> <i>Time Series</i> plot temperature di setiap stasiun.....                                      | 35 |
| <b>Gambar 4.2</b> <i>Time Series</i> plot kelembapan di setiap stasiun .....                                      | 36 |
| <b>Gambar 4.3</b> <i>Time Series</i> plot Kecepatan angin di setiap stasiun .....                                 | 37 |
| <b>Gambar 4.4</b> Boxplot Temperature di setiap stasiun .....   | 38 |
| <b>Gambar 4.5</b> <i>Boxplot</i> Kelembapan di setiap Stasiun .....   | 38 |
| <b>Gambar 4.6</b> <i>Boxplot</i> Kecepatan Angin di setiap Stasiun.....   | 38 |
| <b>Gambar 4.7</b> <i>Time series Plot</i> Curah Hujan Gabungan Balai Surabaya                                     | 40 |
| <b>Gambar 4.8</b> <i>Boxplot</i> Curah Hujan Gabungan di Balai Surabaya.....                                      | 41 |
| <b>Gambar 4.9</b> <i>Scatter Plot</i> Curah Hujan Keputih (vertikal) dengan variabel BMKG (Horizontal) .....      | 42 |
| <b>Gambar 4.10</b> <i>Scatter Plot</i> Curah Hujan Gubeng (Vertikal) dengan variabel BMKG (Horizontal) .....      | 44 |
| <b>Gambar 4.11</b> <i>Scatter Plot</i> Curah Hujan Kedung Cowek (Vertikal) dengan variabel BMKG (Horizontal)..... | 45 |
| <b>Gambar 4.12</b> <i>Scatter Plot</i> Curah Hujan Wonorejo (Vertikal) dengan variabel BMKG ((Horizontal).....    | 47 |
| <b>Gambar 4.13</b> <i>Scatter Plot</i> Curah Hujan Wonokromo (Vertikal) dengan variabel BMKG (Horizontal) .....   | 49 |
| <b>Gambar 4.14</b> <i>Scatter Plot</i> Curah Hujan Gunung Sari (Vertikal) dengan variabel BMKG (Horizontal) ..... | 50 |
| <b>Gambar 4.15</b> Prediksi Curah Hujan Keputih dengan tiga Variabel.....   | 53 |
| <b>Gambar 4.16</b> Prediksi Curah Hujan dengan Temperature dan <i>dummy</i> .....                                 | 55 |
| <b>Gambar 4.17</b> Prediksi CH dengan Ketiga Variabel dan <i>dummy</i> .....                                      | 56 |
| <b>Gambar 4.18</b> Prediksi CH Keputih dengan Variabel Temperatur, <i>dummy</i> dan lag1 .....                    | 56 |
| <b>Gambar 4.19.</b> Prediksi Curah Hujan Keputih dengan ketiga variabel Juanda, <i>dummy</i> dan lag1 .....       | 57 |
| <b>Gambar 4.20</b> PACF Residual semua Metode .....   | 58 |
| <b>Gambar 4.21</b> Prediksi CH Semua Variabel, <i>Dummy</i> , lag 1 dan lag2 ...                                  | 59 |
| <b>Gambar 4.22</b> Prediksi CH Temperatur, <i>Dummy</i> , lag 1 dan lag2 .....                                    | 59 |
| <b>Gambar 4.23</b> Prediksi Data Testing Juanda .....   | 60 |
| <b>Gambar 4.24</b> Prediksi Curah Hujan Keputih dengan tiga Variabel.....   | 62 |
| <b>Gambar 4.25</b> Prediksi CH Keputih dengan tiga Variabel dan <i>dummy</i> .....                                | 63 |
| <b>Gambar 4.26</b> Prediksi CH Keputih PC1 (Perak1) dan <i>Dummy</i> .....  | 64 |
| <b>Gambar 4.27</b> Prediksi Curah Hujan Keputih dengan ketiga variabel (Perak1), lag1 dan <i>Dummy</i> .....      | 65 |
| <b>Gambar 4.28</b> PACF Residual semua Metode .....   | 65 |

|  |    |
|--|----|
| <b>Gambar 4.29</b> Prediksi Curah Hujan Keputih dengan ketiga variabel (Perak1), lag1, lag2 dan <i>Dummy</i> ..... | 66 |
| <b>Gambar 4.30</b> Prediksi Data Testing (Perak 1) .....   | 67 |
| <b>Gambar 4.30</b> Prediksi Curah Hujan Keputih dengan tiga Variabel .....   | 69 |
| <b>Gambar 4.31</b> Prediksi CH Keputih dengan tiga Variabel dan <i>dummy</i> .....                                 | 70 |
| <b>Gambar 4.32</b> Prediksi Curah Hujan Keputih PC1 dan <i>Dummy</i> .....   | 71 |
| <b>Gambar 4.33</b> Prediksi Curah Hujan Keputih, PC1, lag1 dan <i>Dummy</i> .....                                  | 72 |
| <b>Gambar 4.34</b> Prediksi CH Keputih variabel Perak 2, lag1 dan <i>dummy</i> .....                               | 72 |
| <b>Gambar 4.35</b> PACF Residual Model dengan lag 1 .....  | 73 |
| <b>Gambar 4.36</b> Prediksi CH variabel Perak 2, lag1, lag2 dan <i>Dummy</i> .....                                 | 73 |
| <b>Gambar 4.37</b> Pemodelan Data Testing (perak 2) .....  | 74 |
| <b>Gambar 4.38</b> Estimasi CH Gubeng dengan variabel <i>dummy</i> dan <i>lag1</i> .....                           | 75 |
| <b>Gambar 4.39</b> Estimasi CH Gubeng dengan <i>dummy</i> , <i>lag1</i> dan <i>lag2</i> .....                      | 76 |
| <b>Gambar 4.40</b> Prediksi Data Testing (Juanda) .....  | 77 |
| <b>Gambar 4.41</b> Estimasi CH dengan Ketiga Variabel, <i>dummy</i> dan <i>lag1</i> .....                          | 78 |
| <b>Gambar 4.42</b> Estimasi CH dengan Ketiga Variabel, <i>dummy</i> , <i>lag1</i> dan <i>lag2</i> .....            | 78 |
| <b>Gambar 4.43</b> Prediksi Data Testing (Perak 1) .....   | 79 |
| <b>Gambar 4.44</b> Estimasi CH dengan Ketiga Variabel, <i>dummy</i> dan <i>lag1</i> .....                          | 80 |
| <b>Gambar 4.45</b> Estimasi CH dengan Ketiga Variabel <i>dummy lag1</i> dan <i>lag2</i> .....                      | 80 |
| <b>Gambar 4.46</b> Prediksi Data Testing .....   | 81 |
| <b>Gambar 4.47</b> Estimasi CH dengan Ketiga Variabel, <i>dummy</i> dan <i>lag1</i> .....                          | 82 |
| <b>Gambar 4.48</b> Estimasi CH dengan Ketiga Variabel, <i>dummy</i> , <i>lag1</i> dan <i>lag2</i> .....            | 83 |
| <b>Gambar 4.49</b> Prediksi Data Testing .....   | 84 |
| <b>Gambar 4.50</b> Estimasi CH dengan Ketiga Variabel, <i>dummy</i> dan <i>lag1</i> .....                          | 84 |
| <b>Gambar 4.51</b> Estimasi CH dengan Ketiga Variabel <i>dummy lag1</i> dan <i>lag2</i> .....                      | 85 |
| <b>Gambar 4.52</b> Prediksi Data Testing (Perak 1) .....   | 86 |
| <b>Gambar 4.53</b> Estimasi CH Kd. Cowek dengan Ketiga Variabel, <i>dummy</i> dan <i>lag1</i> .....                | 86 |
| <b>Gambar 4.54</b> Prediksi Data Testing .....   | 87 |
| <b>Gambar 4.55</b> Estimasi Curah Hujan Wonorejo dengan variabel <i>dummy</i> dan <i>lag1</i> .....                | 88 |
| <b>Gambar 4.56</b> Estimasi CH dengan Ketiga Variabel, <i>dummy</i> , <i>lag1</i> dan <i>lag2</i> .....            | 88 |
| <b>Gambar 4.57</b> Pemodelan Data Testing .....  | 89 |
| <b>Gambar 4.58</b> Estimasi CH dengan Ketiga variabel, <i>dummy</i> dan <i>lag1</i> .....                          | 90 |

|   |     |
|---|-----|
| <b>Gambar 4.59</b> Estimasi CH dengan Ketiga variabel, <i>dummy lag1</i> dan <i>lag2</i> .....                    | 90  |
| <b>Gambar 4.60</b> Prediksi Data Testing .....  | 91  |
| <b>Gambar 4.61</b> Estimasi CH dengan Ketiga Variabel, <i>dummy</i> dan <i>lag1</i> .....                         | 92  |
| <b>Gambar 4.62</b> Estimasi CH dengan Ketiga Variabel, <i>dummy, lag1</i> dan <i>lag2</i> .....                   | 92  |
| <b>Gambar 4.63</b> Prediksi Data Testing .....  | 93  |
| <b>Gambar 4.64</b> Estimasi CH dengan Ketiga Variabel, <i>dummy</i> dan <i>lag1</i> .....                         | 94  |
| <b>Gambar 4.65</b> Estimasi CH Ketiga Variabel, <i>dummy, lag1</i> dan <i>lag2</i> ....                           | 95  |
| <b>Gambar 4.66</b> Prediksi Data Testing .....  | 96  |
| <b>Gambar 4.67</b> Estimasi CH dengan Ketiga Variabel, <i>dummy</i> dan <i>lag1</i> .....                         | 97  |
| <b>Gambar 4.68</b> Estimasi CH dengan Ketiga Variabel, <i>dummy, lag1</i> dan <i>lag2</i> .....                   | 97  |
| <b>Gambar 4.69</b> Prediksi Data Testing .....  | 98  |
| <b>Gambar 4.70</b> Estimasi Curah Hujan dengan <i>dummy</i> dan <i>lag1</i> .....                                 | 99  |
| <b>Gambar 4.71</b> Estimasi Curah Hujan dengan <i>dummy, lag1</i> dan <i>lag2</i> ....                            | 99  |
| <b>Gambar 4.72</b> Pemodelan Data Testing.....  | 100 |
| <b>Gambar 4.73</b> Estimasi CH dengan Ketiga Variabel, <i>dummy</i> dan <i>lag1</i> .....                         | 101 |
| <b>Gambar 4.74</b> Estimasi CH dengan Ketiga Variabel, <i>dummy lag1</i> dan <i>lag2</i> .....                    | 101 |
| <b>Gambar 4.75</b> Pemodelan Data Testing.....  | 103 |
| <b>Gambar 4.76</b> Estimasi Curah Hujan dengan <i>dummy</i> dan <i>lag1</i> .....                                 | 103 |
| <b>Gambar 4.77</b> Estimasi Curah Hujan dengan <i>dummy, lag1</i> dan <i>lag2</i> ..                              | 104 |
| <b>Gambar 4.78</b> Pemodelan Data Testing.....  | 104 |
| <b>Gambar 4.79</b> Estimasi Curah Hujan dengan <i>dummy</i> dan <i>lag1</i> .....                                 | 105 |
| <b>Gambar 4.80</b> Estimasi Curah Hujan dengan <i>dummy lag1</i> dan <i>lag2</i> ...                              | 105 |
| <b>Gambar 4.81</b> Prediksi Data Testing .....  | 106 |
| <b>Gambar 4.82</b> Prediksi Curah Hujan Keputih dengan variabel PC.....   | 109 |
| <b>Gambar 4.83</b> Prediksi Curah Hujan dengan variabel PC dan <i>dummy</i> .....                                 | 110 |
| <b>Gambar 4.84</b> Prediksi Curah Hujan dengan ketiga variabel PC.....  | 110 |
| <b>Gambar 4.85.</b> Prediksi Curah Hujan dengan ketiga PC dan <i>dummy</i> ..                                     | 111 |
| <b>Gambar 4.86</b> Prediksi CH Keputih dengan variabel PC Temperatur, <i>dummy</i> dan <i>lag1</i> .....          | 112 |
| <b>Gambar 4.87</b> Prediksi CH dengan ketiga PC, <i>dummy</i> dan <i>lag1</i> .....                               | 112 |
| <b>Gambar 4.88.</b> PACF Residual Regresi Gamma “sqrt”.....   | 113 |
| <b>Gambar 4.89.</b> Prediksi Curah Hujan Keputih dengan PC Temperatur, <i>dummy, lag 1</i> dan <i>lag 2</i> ..... | 113 |
| <b>Gambar 4.90</b> Prediksi Data Testing .....  | 114 |

|   |     |
|---|-----|
| <b>Gambar 4.91</b> Prediksi Curah Hujan Kd. Cowek dengan ketiga variabel PC, <i>dummy</i> dan lag1 .....    | 115 |
| <b>Gambar 4.92</b> Prediksi CH dengan ketiga variabel PC, <i>dummy</i> , lag1 dan lag2.....                 | 115 |
| <b>Gambar 4.93</b> Pemodelan Data Testing .....   | 116 |
| <b>Gambar 4.94</b> Prediksi CH dengan ketiga PC, <i>dummy</i> dan lag1 .....                                | 117 |
| <b>Gambar 4.95</b> Prediksi Curah Hujan Gubeng dengan PC, <i>dummy</i> , lag1 dan lag2 .....                | 118 |
| <b>Gambar 4.96</b> Pemodelan Data Testing (Gubeng).....   | 119 |
| <b>Gambar 4.97</b> Prediksi CH dengan ketiga PC, <i>dummy</i> dan lag1 .....                                | 119 |
| <b>Gambar 4.98.</b> Prediksi CH dengan PC, <i>dummy</i> , lag1 dan lag2 .....                               | 120 |
| <b>Gambar 4.99</b> Pemodelan Data Testing .....   | 121 |
| <b>Gambar 4.100</b> Prediksi CH dengan Ketiga Variabel PC, <i>dummy</i> dan lag1 .....                      | 122 |
| <b>Gambar 4.101</b> Prediksi CH Keputih dengan Ketiga PC, <i>dummy</i> , lag 1 dan lag 2 .....              | 122 |
| <b>Gambar 4.102</b> Prediksi Data Testing.....  | 123 |
| <b>Gambar 4.103</b> Prediksi Curah Hujan Gunung Sari dengan ketiga variabel PC, <i>dummy</i> dan lag1 ..... | 124 |
| <b>Gambar 4.104</b> Prediksi Curah Hujan Gunung Sari dengan Ketiga PC, <i>dummy</i> , lag1 dan lag2 .....   | 125 |
| <b>Gambar 4.105.</b> Pemodelan Data <i>Testing</i> .....  | 126 |

*(Halaman ini sengaja dikosongkan)*

## DAFTAR TABEL

|   |     |
|---|-----|
| <b>Tabel 4.1</b> Summary CH Keputih dengan Temperatur .....                     | 52  |
| <b>Tabel 4.2</b> Summary CH Keputih dengan Kelembapan.....                      | 52  |
| <b>Tabel 4.3</b> Summary CH Keputih dengan Kecepatan Angin .....                | 52  |
| <b>Tabel 4.4</b> Summary CH Keputih dengan Tiga Variabel.....                   | 53  |
| <b>Tabel 4.5</b> VIF untuk Semua Model (Juanda).....                            | 54  |
| <b>Tabel 4.6</b> Analisis Komponen Utama (Juanda) .....                         | 54  |
| <b>Tabel 4.7.</b> Pemodelan Curah Hujan dengan Temperatur dan <i>dummy</i> ...  | 55  |
| <b>Tabel 4.8.</b> Pemodelan CH dengan Ketiga Variabel dan <i>dummy</i> .....    | 55  |
| <b>Tabel 4.9</b> Pemodelan CH dengan Ketiga Variabel, <i>dummy</i> dan lag1 ... | 56  |
| <b>Tabel 4.10</b> Pemodelan CH dengan Temperatur, <i>dummy</i> dan lag1.....    | 57  |
| <b>Tabel 4.12</b> Pemodelan CH Terbaik (Ketiga Variabel Juanda) .....           | 59  |
| <b>Tabel 4.13</b> Summary CH Keputih 1 dengan Kelembapan (Perak1).....          | 61  |
| <b>Tabel 4.14</b> Summary CH Keputih dengan Kecepatan Angin (Perak1)            | 61  |
| <b>Tabel 4.15</b> Summary CH Keputih dengan variabel Stasiun Perak 1 ....       | 61  |
| <b>Tabel 4.16</b> VIF untuk Semua Model (Perak 1).....                          | 62  |
| <b>Tabel 4.17</b> Analisis Komponen Utama (Perak 1).....                        | 63  |
| <b>Tabel 4.18</b> Summary CH dengan Ketiga Variabel dan <i>dummy</i> .....      | 64  |
| <b>Tabel 4.19</b> Summary CH dengan Ketiga Variabel <i>dummy</i> dan lag1 ...   | 66  |
| <b>Tabel 4.20</b> Summary CH Keputih dengan Temperatur (Perak2) .....           | 67  |
| <b>Tabel 4.21</b> Summary CH Keputih dengan Kelembapan (Perak2).....            | 68  |
| <b>Tabel 4.22</b> Summary CH Keputih dengan Kecepatan Angin (Perak2)            | 68  |
| <b>Tabel 4.23</b> Summary CH Keputih dengan variabel Stasiun Perak 2 ....       | 68  |
| <b>Tabel 4.24</b> VIF untuk Semua Model (Perak 2).....                          | 69  |
| <b>Tabel 4.25</b> Analisis Komponen Utama (Perak 2).....                        | 70  |
| <b>Tabel 4.26</b> Pemodelan curah Hujan dengan PC (Perak 2).....                | 71  |
| <b>Tabel 4.27</b> Pemodelan curah Hujan dengan Ketiga Variabel (Perak 2)        | 71  |
| <b>Tabel 4.28</b> Summary Model Terbaik (Perak2).....                           | 74  |
| <b>Tabel 4.29</b> Summary Model Terbaik (Juanda) .....                          | 76  |
| <b>Tabel 4.30</b> Summary Model Terbaik Gamma"sqrt" (perak1) .....              | 78  |
| <b>Tabel 4.31</b> Summary Model Terbaik (Juanda) .....                          | 81  |
| <b>Tabel 4.32</b> Summary Model Terbaik (Juanda).....                           | 83  |
| <b>Tabel 4.33</b> Summary Model Terbaik Gamma "sqrt" (Perak1) .....             | 85  |
| <b>Tabel 4.35</b> Summary Model Terbaik Gamma "sqrt" (Juanda).....              | 89  |
| <b>Tabel 4.36.</b> Summary Model Terbaik Gamma "sqrt" (Perak1).....             | 91  |
| <b>Tabel 4.37</b> Summary Model Terbaik Gamma "sqrt" (Perak2).....              | 93  |
| <b>Tabel 4.38</b> Summary Model Terbaik Gamma "sqrt" (Juanda).....              | 95  |
| <b>Tabel 4.39</b> Summary Model Terbaik Gamma "sqrt" (perak 1) .....            | 97  |
| <b>Tabel 4.40</b> Summary Model Terbaik Gamma "sqrt" (Perak 2).....             | 100 |

|                   |   |     |
|-------------------|---|-----|
| <b>Tabel 4.41</b> | Summary Model Terbaik Gamma “sqrt” (Juanda).....      | 102 |
| <b>Tabel 4.42</b> | Summary Model Terbaik Gamma “identity” (Perak 2) .... | 106 |
| <b>Tabel 4.43</b> | Summary PCA Temperature dari Ketiga Stasiun.....      | 108 |
| <b>Tabel 4.44</b> | Summary PCA Kelembapan dari Ketiga Stasiun.....       | 108 |
| <b>Tabel 4.45</b> | Summary PCA Kecepatan Angin dari Ketiga Stasiun ..... | 108 |
| <b>Tabel 4.46</b> | Pemodelan Curah Hujan dengan PC1 (Temperature) .....  | 108 |
| <b>Tabel 4.47</b> | Pemodelan Curah Hujan dengan PC1 (Kelembapan) .....   | 109 |
| <b>Tabel 4.48</b> | Pemodelan Curah Hujan dengan PC1 (Kecepatan angin) .  | 109 |
| <b>Tabel 4.49</b> | Pemodelan Curah Hujan Keputih dengan semua PC .....   | 111 |
| <b>Tabel 4.50</b> | Summary Model Terbaik .....                           | 113 |
| <b>Tabel 4.51</b> | Summary Model Terbaik .....                           | 116 |
| <b>Tabel 4.52</b> | Summary Model Terbaik Gamma “sqrt” .....              | 118 |
| <b>Tabel 4.53</b> | Summary Model Terbaik Gamma “sqrt” .....              | 120 |
| <b>Tabel 4.54</b> | Summary Model Terbaik Gamma “sqrt” .....              | 123 |
| <b>Tabel 4.55</b> | Summary Model Terbaik Gamma “sqrt” .....              | 125 |
| <b>Tabel 4.56</b> | Summary RMSE Testing Setiap stasiun .....             | 126 |
| <b>Tabel 4.57</b> | Curah Hujan Maksimum .....                            | 127 |
| <b>Tabel 4.58</b> | Prediksi Banjir Balai Keputih .....                   | 127 |
| <b>Tabel 4.59</b> | Prediksi Banjir Balai Gubeng.....                     | 128 |
| <b>Tabel 4.60</b> | Prediksi Banjir Balai Kd. Cowek .....                 | 129 |
| <b>Tabel 4.61</b> | Prediksi Banjir Balai Wonorejo .....                  | 130 |
| <b>Tabel 4.62</b> | Prediksi Banjir Balai Wonokromo .....                 | 131 |
| <b>Tabel 4.63</b> | Prediksi Banjir Balai Gunung Sari .....               | 132 |

## DAFTAR LAMPIRAN

|   |     |
|---|-----|
| <b>Lampiran 1.</b> <i>Syntax</i> Penggabungan Data Stasiun Juanda, Perak 1 dan Perak2.....        | 139 |
| <b>Lampiran 2.</b> <i>Syntax Time Series Plot</i> Temperatur, Kelembapan dan Kecepatan Angin..... | 139 |
| <b>Lampiran 3.</b> <i>Syntax BoxPlot</i> Curah Hujan.....   | 140 |
| <b>Lampiran 4.</b> <i>Syntax ScatterPlot</i> CH dengan Prediktor .....                            | 142 |
| <b>Lampiran 5.</b> <i>Syntax</i> <i>Pemodelan Curah Hujan dengan Juanda</i> .....                 | 143 |
| <b>Lampiran 6.</b> <i>Syntax</i> <i>Pemodelan Curah Hujan dengan Perak1</i> .....                 | 145 |
| <b>Lampiran 7.</b> <i>Syntax</i> <i>Pemodelan Curah Hujan dengan Perak2</i> .....                 | 147 |
| <b>Lampiran 8.</b> <i>Syntax</i> <i>Pemodelan Curah Hujan PC</i> .....                            | 149 |
| <b>Lampiran 9.</b> <i>Syntax</i> <i>Prediksi data Testing</i> .....                               | 151 |
| <b>Lampiran 10.</b> <i>Syntax</i> <i>Prediksi data Testing Multivariat PC</i> ....                | 152 |
| <b>Lampiran 11.</b> <i>Syntax output</i> <i>Pemodelan</i> .....                                   | 153 |
| <b>Lampiran 12.</b> <i>Prediksi Curah Hujan Gubeng</i> .....                                      | 153 |
| <b>Lampiran 13.</b> <i>Prediksi Curah Hujan Kedung Cowek</i> .....                                | 154 |
| <b>Lampiran 14.</b> <i>Prediksi Curah Hujan Wonorejo</i> .....                                    | 155 |
| <b>Lampiran 15.</b> <i>Prediksi Curah Hujan Wonokromo</i> .....                                   | 157 |
| <b>Lampiran 16.</b> <i>Prediksi Curah Hujan Gunung Sari</i> .....                                 | 158 |
| <b>Lampiran 17.</b> Surat Keterangan Pengambilan Data (BMKG). .....                               | 160 |
| <b>Lampiran 18.</b> Surat Keterangan Pengambilan Data .....                                       | 161 |

*(Halaman ini sengaja dikosongkan)*

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Secara Geografis Indonesia terletak diantara dua benua (Benua Asia dan Benua Pasifik) dan dua samudra (Pasifik dan Hindia). Letak geografis tersebut menyebabkan Indonesia memiliki tiga iklim utama yaitu iklim monsun (musim), iklim panas (tropis), dan iklim laut. Iklim monsun atau musim terjadi karena adanya pengaruh angin monsun yang bertiup berganti arah setiap setengah tahun sekali. Angin monsun merupakan gerakan massa udara yang terjadi karena adanya perbedaan tekanan udara yang mencolok antara daratan dan juga samudra. Angin monsun terbagi menjadi dua yaitu, angin monsun barat dan angin monsun timur. Angin monsun barat bertiup dari kawasan benua Asia ke kawasan benua Australia yang megandung banyak uap air sehingga pada bulan Oktober hingga April Indonesia mengalami musim penghujan. Angin monsun timur bertiup dari kawasan benua Australia ke kawasan benua Asia. Angin bertiup melewati daerah gurun pasir yang terletak di bagian utara benua Australia yang bersifat kering dan melalui lautan yang sempit. Akibatnya angin tidak mengandung uap air sehingga pada bulan April hingga Oktober kawasan di negara Asia termasuk Indonesia akan mengalami musim kemarau.

Dampak dari angin monsun barat adalah terjadi curah hujan yang cukup tinggi. Curah hujan merupakan ketinggian air hujan yang terkumpul dalam tempat yang datar, tidak menguap, tidak meresap dan tidak mengalir. Curah hujan satu millimeter berarti dalam luasan satu meter persegi pada tempat yang datar tertampung air setinggi satu millimeter. Indonesia merupakan negara yang memiliki angka curah hujan yang bervariasi dikarenakan daerahnya yang berada pada ketinggian yang berbeda-beda salah satunya adalah Kota Surabaya. Surabaya merupakan kota terbesar yang berada di Jawa Timur. Menurut Badan Pusat Statistik (BPS), Surabaya memiliki rata-rata curah hujan berkisar 1.321 mm pertahun. Kota Surabaya mengalami musim kemarau pada umumnya pada bulan Mei sampai Oktober, dan musim hujan dari bulan November hingga April.

Dampak negatif yang ditimbulkan akibat intensitas curah hujan yang tinggi di Surabaya adalah banjir. Banjir dalam pengertian umum adalah debit aliran air sungai dalam jumlah yang tinggi, atau debit aliran air di sungai secara relatif lebih besar dari kondisi normal akibat hujan yang turun di hulu atau di suatu tempat tertentu terjadi secara terus menerus, sehingga air tersebut tidak dapat ditampung oleh alur sungai yang ada, maka air melimpah keluar dan menggenangi daerah sekitarnya (Peraturan Dirjen RLPS No.04 thn 2009). Terjadinya bencana banjir juga disebabkan oleh rendahnya kemampuan infiltrasi tanah, sehingga menyebabkan tanah tidak mampu lagi menyerap air. Banjir dapat terjadi akibat naiknya permukaan air lantaran curah hujan yang diatas normal, perubahan suhu, tanggul/bendungan yang bobol, pencairan salju yang cepat, terhambatnya aliran air di tempat lain (Ligal, 2008). Seperti dilansir detiknews.com (2019) memberitakan bahwa akibat hujan deras mengakibatkan perumahan warga di Surabaya menjadi tergenang. Tidak hanya itu pada januari 2019 di wilayah Surabaya Barat yang mengakibatkan banyak mobil tenggelam.

Kerugian yang ditimbulkan akibat banjir terhitung tidak sedikit, baik kerugian yang memakan korban jiwa maupun secara materiil. Berdasarkan hasil kajian dari BNPB, jumlah jiwa terpapar risiko bencana banjir tersebar di beberapa pulau di Indonesia dengan jumlah melebihi 170 juta jiwa dan nilai aset terpapar melebihi Rp 750 triliun. Untuk provinsi yang memiliki risiko bencana banjir tertinggi di Indonesia, Jawa Timur merupakan daerah yang memiliki dampak risiko terbesar. Surabaya sebagai kota metropolitan menjadi kota yang paling dirugikan dari bencana banjir. Bagaimana tidak pusat perekonomian di provinsi Jawa Timur terletak di kota Surabaya. Apabila perekonomian di Kota Surabaya mengalami kelumpuhan atau gangguan akan menyebabkan ketidakstabilan ekonomi di provinsi Jawa Timur maupun tingkat pertumbuhan ekonomi di tingkat nasional.

Pengukuran curah hujan harian dan variabel klimatologi lainnya dilakukan di tujuh rayon yang tersebar di tingkat kecamatan kota Surabaya. Salah satu peran dari BMKG adalah mengumpulkan informasi atau data data yang berasal dari masing masing stasiun. Berdasarkan data data hasil pengukuran tersebut diharapkan mampu menjadi *based* atau acuan dalam memprediksi

banjir. Salah satu cara untuk dapat mengelola resiko terjadinya banjir adalah dengan memperkirakan kapan suatu daerah tersebut akan berpotensi terkena banjir. Analisa ini dapat dilakukan dengan memperkirakan potensi terjadinya hujan lebat atau curah hujan tinggi di Surabaya atau daerah lokal berdasarkan variabel yang diturunkan dari data klimatologi.

Maka dari itu pada penelitian ini akan dilakukan pemodelan curah hujan harian menggunakan data Balai Buntung Peketingan untuk memprediksi resiko banjir di kota Surabaya dengan pendekatan regresi *Inverse Gaussian*. Metode regresi *Inverse Gaussian* sesuai karena data curah hujan termasuk kedalam distribusi yang berdomain positif. Model regresi inverse Gaussian merupakan model regresi dengan peubah respons kontinu non negatif yang berdistribusi inverse Gaussian. Regresi inverse Gaussian dipilih dengan alasan bahwa memuat nilai peubah respons kontinu non negatif dan termasuk dalam distribusi keluarga eksponensial sehingga model GLM dapat digunakan untuk menyelesaikan permasalahan banjir dengan peubah respons kontinu non negatif dan tidak memerlukan pemenuhan asumsi seperti pada model regresi linear klasik. Kriteria kebaikan model yang digunakan adalah berdasarkan nilai RMSEP dan  $R^2_{predict}$ . Nilai RMSEP merupakan nilai dari error hasil taksiran sehingga model dengan RMSEP minimum yang menandakan nilai taksiran mendekati nilai sebenarnya sedangkan semakin besar nilai  $R^2_{predict}$ , maka semakin baik pula model yang didapatkan karena mampu menjelaskan lebih banyak data (Drapper dan Smith, 1996).

## 1.2 Rumusan Masalah

Tingginya curah hujan menyebabkan peristiwa yang dinamakan dengan banjir. Selain dari curah hujan yang sangat tinggi banjir juga dipengaruhi oleh beberapa faktor seperti kurangnya daerah resapan air dan tersumbatnya saluran drainase. Berbagai upaya untuk mengatasi dampak yang terjadi akibat banjir terus dilakukan bahkan juga sudah mulai adanya pengembangan dalam segi pencegahan banjir. Salah satu pencegahan yang dilakukan untuk mengatasi atau mengurangi terjadinya banjir adalah dengan melakukan prediksi terhadap datangnya curah hujan. Dengan adanya prediksi curah hujan mampu mengetahui

potensi curah hujan yang tinggi di waktu yang akan datang. Prediksi curah hujan dilakukan dengan variabel klimatologi. Hasil dari prediksi curah hujan akan digunakan dalam menentukan prediksi banjir genangan. Salah satu ukuran dikatakan mengalami banjir genangan adalah tinggi air atau genangan mencapai 30 cm. Oleh karena itu digunakan metode regresi Inverse Gaussian untuk melakukan prediksi curah hujan karena curah hujan merupakan variabel yang nonnegatif yang kemudian dapat pula dilihat hasil prediksi banjir dari model yang dihasilkan.

### **1.3 Tujuan**

Berdasarkan rumusan masalah yang telah diuraikan di atas, tujuan yang ingin dicapai dalam penelitian ini adalah sebagai berikut:

1. Mendeskripsikan karakteristik data curah hujan dan variabel klimatologi.
2. Memprediksi curah hujan berdasarkan pemodelan dengan variabel klimatologi.
3. Memprediksi banjir di Surabaya berdasarkan analisis pemodelan curah hujan.

### **1.4 Manfaat**

Pada penelitian ini dengan penerapan ilmu statistika, dapat diketahui karakteristik pasien setelah melakukan kemoterapi, faktor-faktor yang mempengaruhi keberhasilan suatu kemoterapi, serta hasil klasifikasi keberhasilan kemoterapi pada pasien. Diharapkan hasil dari penelitian ini bisa dimanfaatkan dalam dunia kesehatan menjadi salah satu pertimbangan untuk mengoptimalkan metode pengobatan kanker menggunakan kemoterapi. Khususnya pada kemoterapi menggunakan senyawa 5-FU atau *5-Fluorouracil*. Dalam bidang keilmuan penelitian ini juga dapat dimanfaatkan untuk menambah wawasan keilmuan statistika sebagai pengembangan dari penerapan metode regresi probit bivariat sebagai bahan referensi untuk penelitian-penelitian selanjutnya.

## **BAB II**

### **TINJAUAN PUSTAKA**

Dalam bab ini akan dibahas beberapa teori statistika maupun non statistika yang mendasari penelitian tentang hasil prediksi banjir yang terjadi. Penelitian ini diawali dengan penjelasan mengenai teori statistik yang kemudian dilanjutkan dengan penjelasan mengenai curah hujan dan banjir. Teori statistik yang mendasari penelitian ini dimulai dari fungsi distribusi yang berkaitan dengan analisis regresi Inverse Gaussian, distribusi keluarga eksponensial, estimasi parameter engan *MLE* dan metode Newton Raphson.

#### **2.1 Statistika Deskriptif**

Distribusi Statistika deskriptif merupakan metode-metode yang berkaitan dengan pengumpulan dan penyajian data sehingga memberikan informasi yang berguna. Metode ini bertujuan untuk menguraikan tentang sifat-sifat atau karakteristik dari suatu keadaan dan membuat gambaran yang sistematis dan akurat mengenai fakta-fakta dan sifatsifat dari fenomena yang diselidiki. Penyajian data dapat dilakukan dalam bentuk tabel, grafik dan gambar. Ukuran pemusatan memberikan ukuran mengenai nilai maksimum, nilai minimum, rata-rata, median, modus. Ukuran penyebaran meliputi simpangan rata-rata, standar deviasi, jangkauan kuartil dan jangkauan persentil (Walpole, 1995).

#### **2.2 *Principle Component Analysis (PCA)***

Prosedur PCA pada dasarnya adalah bertujuan untuk menyederhanakan variabel yang diamati dengan cara menyusutkan (mereduksi) dimensinya. Hal ini dilakukan dengan cara menghilangkan korelasi diantara variabel bebas melalui transformasi variabel bebas asal ke variabel baru yang tidak berkorelasi sama sekali atau yang biasa disebut dengan principal component. Setelah beberapa komponen hasil PCA yang bebas multikolinearitas diperoleh, maka komponen-komponen tersebut menjadi variabel bebas baru yang akan diregresikan atau dianalisa pengaruhnya terhadap variabel tak bebas (Y) dengan menggunakan analisis regresi dengan sedikit faktor , sebesar mungkin varians  $X_1$ .

Principal Component yang pertama dinyatakan dengan  $PC_1$  mengandung jumlah terbesar dari total variasi data.  $PC_1$  sebagai kombinasi linier dari variabel  $X_j$  dengan  $j = 1, 2, \dots, k$  dapat dinyatakan sebagai berikut.

$$PC_1 = a_{11}X_1 + a_{12}X_{12} + \dots + a_{1k}X_{1k} \quad (2.1)$$

Dimana jika  $a_{1i}$  dipilih, akan memaksimalkan rasio dari varians  $PC_1$  terhadap total varians, dengan  $\sum_{j=1}^k a_{1j}^2 = 1$ .

Pembentukan regresi komponen utama melalui analisis komponen utama ada dua cara. Pertama, pembentukan komponen utama berdasarkan *matriks kovariansi*. Kedua, pembentukan komponen utama berdasarkan *matriks korelasi*.

#### 4.2.1. Komponen Utama Berdasarkan Matriks Kovarians

Proses mereduksi data dengan analisis komponen utama diuraikan melalui data asal  $X_{n \times p}$  akan dicari matriks varians kovarians  $\sum$  dimana unsur unsurnya adalah

$$S_{jk} = \frac{1}{n-1} \sum_{i=1}^n (X_{ij} - \bar{X}_j)(X_{ik} - \bar{X}_k) \quad (2.2)$$

Kemudian dari matriks varians kovarians tersebut dicari nilai eigen  $\lambda_i$  dengan  $i = 1, 2, \dots, n$ , yang diperoleh dari bentuk persamaan determinan:  $|S - \lambda_i I| = 0$  dari nilai eigen tersebut.

Setelah mendapatkan faktor yang terbentuk melalui proses reduksi, maka perlu dicari persamaannya, dalam bentuk  $Y = F(X_1^*, X_2^*)$  yang merupakan model baru dengan

$X_1^*$  = variabel komponen 1

$X_2^*$  = variabel komponen 2

$X_k^*$  = variabel komponen k

Model di atas lebih sederhana dibandingkan model regresi multiple awal yang berbentuk :

$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik} + \epsilon_i$  atau  $Y = F(X_1, X_2, \dots, X_k)$

Proporsi total varians populasi yang dijelaskan oleh komponen utamake-k adalah

$$\frac{\lambda_k}{tr(\Sigma)} = \frac{\lambda_k}{\lambda_1 + \lambda_2 + \dots + \lambda_k} \text{ dengan } j = 1, 2, \dots, k$$

Dengan  $\lambda$  merupakan nilai dari *eigen value*.

#### 4.2.2. Komponen Utama Berdasarkan Matriks Korelasi

Komponen utama ke- $i$  atau  $W_i$  yang dibentuk berdasarkan variabel-variabel yang telah dibakukan  $Z' = (Z_1, Z_2, \dots, Z_k)$ , dengan  $cov(Z) = \rho$  didefinisikan sebagai berikut :

$$w_i = e_{i1}z_1 + e_{i2}z_2 + \dots + e_{ik}z_k \text{ dimana } j = 1, 2, \dots, k$$

Sementara itu, proporsi total variansi yang dapat dijelaskan oleh komponen ke- $j$  berdasarkan variabel bebas yang telah dibakukan didefinisikan sebagai berikut:

$$\frac{\lambda_j}{k}$$

Dengan  $\lambda_j$  adalah eigen dari  $\rho$ , dan  $j = 1, 2, \dots, k$

### 2.3 Keluarga Distribusi Eksponensial

Suatu variabel acak  $Y$  dengan fungsi kepadatan peluang  $f(y)$  dan parameter  $\theta$  dikatakan menjadi anggota distribusi keluarga eksponensial jika dapat dinyatakan dalam bentuk umum distribusi keluarga eksponensial. Bentuk umum dari distribusi keluarga eksponensial menurut Jong & Heller (2008:35) adalah:

$$f(y | \theta, \phi) = c(y, \phi) \exp\left(\frac{y\theta - b(\theta)}{a(\phi)}\right)$$

Suatu variabel acak  $Y$  dengan fungsi kepadatan peluang  $f(y)$  dan Parameter  $\theta$  dikatakan menjadi anggota distribusi keluarga eksponensial jika dapat dinyatakan dalam bentuk umum distribusi keluarga eksponensial. Bentuk umum distribusi keluarga eksponensial pada GLM menurut Mc Cullagh & Nelder (1989: 28) adalah

$$f(y | \theta, \phi) = \exp\left\{\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)\right\} \quad (2.3)$$

Dengan  $\theta$  dan  $\phi$  adalah parameter kanonik dan parameter dispersi. Pemilihan fungsi  $b(\theta)$  dan  $c(y, \phi)$  yaitu dengan

menetapkan fungsi tersebut sebagai fungsi peluang seperti normal, Binomial, atau gamma. Nilai harapan atau ekspektasi  $E(Y)$  dan ragam  $Var(Y)$  fungsi  $b(\theta)$  dari distribusi keluarga eksponensial dapat dicari dengan rumus:

$$E(Y) = b'(\theta) \quad (2.4)$$

$$Var(Y) = \phi b''(\theta) \quad (2.5)$$

Dengan  $b'(\theta)$  dan  $b''(\theta)$  adalah turunan pertama dan kedua dari  $b(\theta)$  terhadap  $\theta$ .

### 2.3.1. Nilai Ekspektasi

Untuk menunjukkan hubungan-hubungan pada persamaan (2.4) ditetapkan  $f'(y)$  sebagai turunan pertama dari  $f(y)$  pada persamaan (2.5) yang berkenaan dengan  $\theta$ .

$$f'(y) = f(y) \left( \frac{y\theta - b(\theta)}{\phi} \right), \quad f''(y) = f(y) \left( \frac{y\theta - b(\theta)}{\phi} \right)^2 - f(y) \frac{b'(\theta)}{\phi}$$

Masing-masing kedua ruas diintegrasikan berkenaan dengan  $y$  dan didapatkan seperti dibawah

$$\begin{aligned} \int_{-\infty}^{\infty} f'(y) dy &= \int_{-\infty}^{\infty} f(y) \left( \frac{y\theta - b(\theta)}{\phi} \right) dy \\ \frac{\partial}{\partial(\theta)} \int_{-\infty}^{\infty} f(y) dy &= \int_{-\infty}^{\infty} \frac{yf(y)}{\phi} dy - \int_{-\infty}^{\infty} \frac{b'(\theta)f(y)}{\phi} dy \quad (2.6) \end{aligned}$$

Karena  $\int_{-\infty}^{\infty} f(y) dy = 1$  dan  $E(Y) = \int_{-\infty}^{\infty} yf(y) dy$  maka persamaan

(2.6) menjadi,

$$\begin{aligned} \frac{\partial}{\partial(\theta)} &= \frac{E(Y)}{\phi} - \frac{b'(\theta)}{\phi} \\ 0 &= \frac{E(Y)}{\phi} - \frac{b'(\theta)}{\phi} \end{aligned}$$

$$\frac{E(Y)}{\phi} = \frac{b'(\theta)}{\phi}$$

$$E(Y) = b'(\theta)$$

### 2.3.2. Nilai Varians

Untuk menunjukkan hubungan-hubungan pada persamaan (2.3) ditetapkan  $f''(y)$  sebagai turunan kedua dari  $f(y)$  pada persamaan (2.3) yang berkenaan dengan  $\theta$ .

Diketahui bahwa turunan pertama dan kedua dari  $f(y)$  adalah

$$f'(y) = f(y) \left( \frac{y\theta - b(\theta)}{\phi} \right), \quad f''(y) = f(y) \left( \frac{y\theta - b(\theta)}{\phi} \right)^2 - f(y) \frac{b'(\theta)}{\phi}$$

Masing-masing kedua pada ruas turunan kedua diintegrasikan berkenaan dengan  $y$

$$\int_{-\infty}^{\infty} f''(y) dy = \int_{-\infty}^{\infty} \left[ f(y) \left( \frac{y\theta - b(\theta)}{\phi} \right)^2 - f(y) \frac{b'(\theta)}{\phi} \right] dy$$

$$\frac{\partial^2}{\partial(\theta^2)} \int_{-\infty}^{\infty} f(y) dy = \int_{-\infty}^{\infty} f(y) \left( \frac{y\theta - b(\theta)}{\phi} \right)^2 dy - \int_{-\infty}^{\infty} f(y) \frac{b'(\theta)}{\phi} dy \quad (2.7)$$

Karena,

$$\int_{-\infty}^{\infty} f(y) dy = 1, \quad \mu = E(Y) \quad \text{dan} \quad \text{Var}(Y) = \int_{-\infty}^{\infty} (y - \mu)^2 f(y) dy \text{ maka}$$

persamaan (2.7) menjadi

$$\frac{\partial^2}{\partial(\theta^2)} = \frac{\text{Var}(Y)}{\phi^2} - \frac{b''(\theta)}{\phi}$$

$$0 = \frac{\text{Var}(Y)}{\phi^2} - \frac{b''(\theta)}{\phi}$$

$$\frac{\text{Var}(Y)}{\phi^2} = \frac{b''(\theta)}{\phi}$$

$$\text{Var}(Y) = \phi b''(\theta)$$

sehingga disimpulkan  $E(Y) = b'(\theta)$  dan  $\text{Var}(Y) = \phi b''(\theta)$

Pada **Tabel 2.1** akan ditampilkan distribusi keluarga eksponensial dengan parameternya.

**Tabel 2.1** Distribusi keluarga eksponensial dan parameternya

| Distribusi                                  | $\theta$                  | $b(\theta)$                       | $\phi$        | $E(Y)$ | $Var(\mu)$    |
|---|---------------------------|-----------------------------------|---------------|--------|---------------|
| Binomial,<br>$B(n, \pi)$                    | $\ln \frac{\pi}{1-\pi}$   | $n \ln(1 + e^\theta)$             | 1             | $n\pi$ | $n\pi(1-\pi)$ |
| Poisson,<br>$P(\mu)$                        | $\ln(\mu)$                | $e^\theta$                        | 1             | $\mu$  | $\mu$         |
| Normal,<br>$G(\mu, v)$                      | $\mu$                     |                                   |               | $\mu$  | 1             |
| Gamma,<br>$N(\mu, \sigma^2)$                | $-\frac{1}{\mu}$          | $-\ln(-\theta)$                   | $\frac{1}{v}$ | $\mu$  | $\mu^2$       |
| Inverse<br>Gaussian,<br>$IG(\mu, \sigma^2)$ | $-\frac{1}{2\mu^2}$       | $-\sqrt{-2\theta}$                | $\sigma^2$    | $\mu$  | $\mu^3$       |
| Negative<br>Binomial,<br>$NB(\mu, k)$       | $\ln \frac{k\mu}{1+k\mu}$ | $-\frac{1}{k} \ln(1 - ke^\theta)$ | 1             | $\mu$  | $\mu(1+k\mu)$ |

Keterangan:

- $\theta$  : parameter kanonik pada variabel respon  
 $\phi$  : parameter dispersi pada variabel respon  
 $E(Y)$  : nilai harapan pada variabel respon  
 $Var(\mu)$  : varians pada variabel respon

#### 2.4 Generalized Linier Model

GLM digunakan untuk menilai dan mengukur hubungan antara variabel respons dengan variabel penjelas. Menurut Jong & Heller (2008: 64), GLM merupakan perluasan dari proses pemodelan linear yang memungkinkan penentuan model dari suatu data dengan variabel acak tidak harus menyebar normal, asalkan sebaran tersebut termasuk dalam distribusi keluarga eksponensial, antara lain yaitu Inverse Gaussian, Gamma dan Poisson. Uji hipotesis yang diterapkan pada GLM tidak memerlukan asumsi kenormalan dari variabel responsnya ataupun kehomogenan ragam. Sehingga dengan demikian GLM tak hanya dapat digunakan pada variabel respons yang berdistribusi Normal tetapi

juga untuk variabel respons yang berdistribusi selain Normal dan ragam yang tidak konstan atau homogen.

### 2.4.1. Bentuk umum GLM

Generalized Linear Model (GLM) merupakan bentuk umum atau general dari Model Linear. Diketahui vektor  $y$  memiliki  $n$  komponen, yang merupakan realisasi dari sebuah matrik respon  $Y$ . Setiap komponennya independen dan berdistribusi dengan mean atau  $E(Y) = \mu$ . Jika model yang terbentuk memiliki prediktor  $X$ , dengan beberapa parameter yang tidak diketahui  $\beta_1, \dots, \beta_k$ , maka

modelnya adalah berupa kombinasi linear  $\mu = \sum_{j=1}^k x_j \beta_j$  atau jika

dituliskan dalam bentuk matriks menjadi  $\mu = \mathbf{X}'\beta$ . Sebagai transisi dari model Linear ke Generalized Linear model, maka dijabarkan bentuk melalui tiga buah komponen, yaitu:

i. Variabel dependen (variabel respon)  $y_1, y_2, \dots, y_n$  dengan mean

$E(y_i) = \mu_i$ . Variabel dependen diasumsikan sebagai keluarga eksponensial, yaitu distribusi normal, Poisson, Binomial, gamma atau Invers Gaussian.

ii. Systematic Component, yaitu kombinasi linear dari variabel  $\mathbf{X}$  dengan parameter  $\beta$  yang dilambangkan dengan  $\mu = \mathbf{X}\beta$

iii. Ada fungsi link  $g$  yang monoton dan *differentiable* yang menentukan hubungan antara mean dengan prediktor linier seperti berikut.

$$g(\mu_i) = \eta_i \text{ atau secara lengkapnya}$$

$$g(\mu_i) = g[E(\mathbf{Y}_i)] = \beta_0 + \beta_1 \mathbf{X}_{i1} + \beta_2 \mathbf{X}_{i2} + \dots + \beta_p \mathbf{X}_{ip}$$

Secara umum Generalized Linier Models (GLM) dapat dimodelkan sebagai berikut.

$$g(\mu_i) = \mathbf{X}'_i \beta \quad (2.8)$$

### 2.4.2. Fungsi Hubung

Menurut Mc Cullagh & Nelder (1989: 32), fungsi hubung adalah suatu fungsi yang menghubungkan fungsi penduga linear dengan nilai harapan. Dalam model linear klasik, fungsi hubung bisa berupa fungsi yang identik atau kanonik. Suatu fungsi hubung

dikatakan fungsi penghubung kanonik bila parameter kanoniknya sama dengan fungsi hubungannya, yaitu :

$$\eta = \theta \quad (2.9)$$

Dengan  $\theta$  adalah parameter kanonik. Berikut fungsi penghubung kanonik untuk beberapa distribusi.

**Tabel 2.2** Fungsi hubung (Jong & Heller, 2008: 67)

| Fungsi Hubung      | $g(\mu)$                | Penghubung kanonik untuk distribusi                       |
|--------------------|-------------------------|---|
| <i>Identity</i>    | $\mu$                   | <i>Normal</i>   |
| <i>Log</i>         | $\ln \mu$               | <i>Poisson</i>  |
| <i>Power</i>       | $\mu^p$                 | Gamma ( $p = 1$ )<br><i>Inverse Gaussian</i> ( $p = -2$ ) |
| <i>Square root</i> | $\sqrt{\mu}$            |   |
| <i>Logit</i>       | $\ln \frac{\mu}{1-\mu}$ | <i>Binomial</i>   |

### 2.4.3. Fungsi Hubung *Square Root*

Fungsi hubung adalah suatu fungsi yang menghubungkan fungsi penduga linear dengan nilai harapan. Fungsi hubung “sqrt” tidak memiliki penghubung kanonik untuk distribusi tertentu atau dengan kata lain bisa digunakan dalam pemodelan secara umum. Berikut bentuk turunannya.

Diketahui bahwa :

$$\eta = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k \quad (2.10)$$

Dengan fungsi link sqrt adalah  $\sqrt{\mu}$  maka,

$$\begin{aligned} \eta &= \sqrt{\mu} \\ \eta^2 &= \mu \end{aligned} \quad (2.11)$$

Sehingga estimasi model yang terbentuk untuk fungsi link “sqrt” adalah sebagai berikut.

$$\hat{\mu} = (\hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \dots + \hat{\beta}_k X_k)^2 \quad (2.12)$$

Dengan :

- $\beta_j$  : parameter ke- $j$
- $X_j$  : Variabel prediktor ke- $j$ ,  $j = 1, 2, 3$

## 2.5 Regresi Inverse Gaussian

Model regresi inverse Gaussian termasuk dalam GLM dengan variabel respons berdistribusi inverse Gaussian. Salah satu tujuan dari analisis regresi adalah untuk menentukan pola hubungan antara variabel respons dengan variabel penjelas, maka dalam regresi inverse Gaussian hubungan tersebut dapat dituliskan dalam bentuk.

$$E(Y_i) = \mu_i = \beta_0 + \beta_1 \mathbf{X}_{i1} + \beta_2 \mathbf{X}_{i2} + \dots + \beta_k \mathbf{X}_{ik} \quad (2.13)$$

Regresi inverse Gaussian merupakan model regresi yang berdistribusi inverse Gaussian dengan variabel respons berbentuk eksponensial. Fungsi hubung dikatakan fungsi penghubung kanonik bila parameter kanoniknya sama dengan fungsi hubungannya, yaitu  $\eta = \theta$  dengan  $\theta$  adalah parameter kanonik.

Pada regresi inverse Gaussian diketahui bahwa  $\theta = -\frac{1}{2\mu_i^2}$ ,

sehingga karena nilai  $\mu_i > 0$  (harus positif) maka digunakan fungsi hubung  $g(\mu_i) = \eta_i = \mu_i^{-2}$  atau  $\eta_i = \mu_i^{-2} = \mathbf{X}'\beta$ . Dengan demikian model regresi inverse Gaussian dapat ditulis dalam bentuk

$$E(Y_i) = \mu_i = \mathbf{X}'\beta, \text{ dan } g(\mu_i) = \mu_i^{-2} \quad (2.14)$$

dengan  $\beta$  merupakan parameter yang tidak diketahui dalam model dan harus diestimasi.

## 2.6 Estimasi Parameter Regresi Inverse Gaussian

Estimasi parameter pada analisis variabel respons kontinu non negatif dengan regresi inverse Gaussian dapat dicari dengan menggunakan metode MLE. Untuk mencari nilai estimasi parameter dengan MLE, perlu diketahui terlebih dahulu fungsi kepadatan peluang regresi inverse Gaussian.

### 2.6.1 Fungsi Kepadatan Peluang Regresi Inverse Gaussian

Variabel  $Y_i$  pada model regresi inverse Gaussian merupakan variabel non negatif dengan fungsi kepadatan peluang sebagai berikut  $Y_i \sim IG(\mu_i, \sigma^2)$  (Jong & Heller, 2008: 125).

$$f(y_i) = \frac{1}{\sigma\sqrt{2\pi y_i^3}} \exp\left\{-\frac{1}{2y_i} \left(\frac{y_i - \mu_i}{\mu_i\sigma}\right)^2\right\}, y_i > 0, \sigma > 0, \mu_i > 0$$

(2.15)

dengan  $i = 1, 2, \dots, n$

Keterangan

$Y_i$  : variabel acak pada pengamatan ke- $i$

$f(y_i)$  : fungsi kepadatan peluang distribusi inverse Gaussian

$\mu_i, \sigma$  : parameter pada peubah respons

Nilai estimasi parameter regresi inverse Gaussian dapat dicari dengan jika fungsi kepadatan peluang inverse Gaussian terlebih dahulu diubah ke dalam bentuk umum distribusi keluarga eksponensial. Bentuk khusus fungsi kepadatan peluang dari variabel acak  $Y_i$  yang berdistribusi inverse Gaussian yaitu sebagai berikut:

$$f(y_i) = \frac{1}{\sigma\sqrt{2\pi y_i^3}} \exp\left\{-\frac{1}{2y_i} \left(\frac{y_i - \mu_i}{\mu_i\sigma}\right)^2\right\}$$

$$f(y_i) = \frac{(2\pi y^3)^{-\frac{1}{2}}}{\sigma} \exp\left\{-\frac{1}{2y_i} \left(\frac{y_i^2 - 2\mu_i y_i + \mu_i^2}{\mu_i^2 \sigma^2}\right)\right\}$$

$$f(y_i) = \left\{-\frac{1}{2} \ln(2\pi y^3) - \ln(\sigma)\right\} \exp\left\{\frac{1}{2y_i} \left(\frac{y_i^2 - 2\mu_i y_i + \mu_i^2}{\mu_i^2 \sigma^2}\right)\right\}$$

$$f(y_i) = \left\{-\frac{1}{2} \ln(2\pi y^3) - \ln(\sigma)\right\} \exp\left\{\left(\frac{-y_i^2 + 2\mu_i y_i - \mu_i^2}{2y_i \mu_i^2 \sigma^2}\right)\right\}$$

$$f(y_i) = \left\{-\frac{1}{2} \ln(2\pi y^3) - \ln(\sigma)\right\} \exp\left\{-\frac{y_i}{2\mu_i^2 \sigma^2} + \frac{1}{\mu_i \sigma^2} - \frac{1}{2y_i \sigma^2}\right\}$$

$$f(y_i) = \exp\left\{-\frac{y_i}{2\mu_i^2 \sigma^2} + \frac{1}{\mu_i \sigma^2} - \frac{1}{2y_i \sigma^2} - \frac{1}{2} \ln(2\pi y^3) - \ln(\sigma)\right\}$$

$$f(y_i) = \exp \left\{ -\frac{y_i}{2\mu_i^2} + \frac{1}{\mu_i} - \frac{1}{2y_i\sigma^2} - \frac{1}{2} \ln(2\pi y_i^3) - \ln(\sigma) \right\}$$

$$f(y_i) = \exp \left\{ -\frac{y_i \left( -\frac{1}{2\mu_i^2} \right) - \left( -\sqrt{2 \left( \frac{1}{2\mu_i^2} \right)} \right)}{\sigma^2} - \frac{1}{2y_i\sigma^2} - \frac{1}{2} \ln(2\pi y_i^3) - \ln(\sigma) \right\}$$

$$f(y_i) = \exp \left\{ -\frac{y_i \left( -\frac{1}{2\mu_i^2} \right) - \left( -\sqrt{2\theta} \right)}{\sigma^2} - \frac{1}{2y_i\sigma^2} - \frac{1}{2} \ln(2\pi y_i^3) - \ln(\sigma) \right\}$$

$$f(y_i) = \exp \left\{ \frac{y_i\theta - b(\theta)}{a(\phi)} + c(y_i, \phi) \right\}$$

dimana,  $\theta = -\frac{1}{2\mu_i^2}$ ;  $b(\theta) = -\left(-\sqrt{2\theta}\right)$ ;  $a(\phi) = \phi$ ;

$$c(y_i, \phi) = -\frac{1}{2y_i\sigma^2} - \frac{1}{2} \ln(2\pi y_i^3) - \ln(\sigma)$$

### 2.6.2. Estimasi Parameter dengan MLE

Metode yang digunakan untuk mengestimasi parameter variabel respons distribusi inverse Gaussian pada penulisan ini adalah metode MLE. Diketahui fungsi kepadatan peluang distribusi inverse Gaussian adalah:

$$f(y_i) = \frac{1}{\sigma\sqrt{2\pi y_i^3}} \exp \left\{ -\frac{1}{2y_i} \left( \frac{y_i - \mu_i}{\mu_i\sigma} \right)^2 \right\}, y_i > 0, \sigma > 0, \mu_i > 0$$

Metode MLE dapat dilakukan jika distribusi dari data diketahui. Langkah pertama yang dilakukan adalah menentukan fungsi likelihood dari model regresi inverse Gaussian. Dengan mengasumsikan  $y_1, y_2, y_3, \dots, y_n$  adalah sekumpulan variabel acak

inverse Gaussian yang saling bebas atau  $y_i$  berdistribusi inverse Gaussian, maka diperoleh fungsi likelihood.

a. Fungsi likelihood variabel respons distribusi inverse Gaussian

$$\begin{aligned}
 L(\mu_i, \sigma^2) &= \prod_{i=1}^n \frac{1}{\sigma \sqrt{2\pi y_i^3}} \exp \left\{ -\frac{1}{2y_i} \left( \frac{y_i - \mu_i}{\mu_i \sigma} \right)^2 \right\} \\
 &= \prod_{i=1}^n \frac{1}{\sigma \sqrt{2\pi y_i^3}} e^{\left\{ -\frac{1}{2y_i} \left( \frac{y_i - \mu_i}{\mu_i \sigma} \right)^2 \right\}} \\
 &= \frac{1}{(2\pi y_i^3)^{\frac{n}{2}}} e^{\left( -\frac{1}{2\sigma^2 \mu_i^2 y_i} \right) \sum_{i=1}^n (y_i - \mu_i)^2}
 \end{aligned} \tag{2.16}$$

Dengan  $\mu_i = \mathbf{X}'\beta$ ,  $i = 1, 2, 3, \dots, n$

b. Logaritma natural dari kedua ruas atau disebut fungsi log-likelihood menjadi:

$$\begin{aligned}
 \ln L(\mu_i, \sigma^2) &= -\frac{n}{2} \ln(2\pi y_i^3) - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2 \mu_i^2 y_i} \sum_{i=1}^n (y_i - \mu_i)^2 \\
 &= -\frac{n}{2} \ln(2\pi y_i^3) - \frac{n}{2} \ln \sigma^2 - \frac{1}{2y_i} \sum_{i=1}^n \left( \frac{y_i - \mu_i}{\mu_i \sigma} \right)^2 \\
 &= -\frac{n}{2} \ln(2\pi y_i^3) - \frac{n}{2} \ln \sigma^2 - \sum_{i=1}^n \frac{y_i^2 - 2\mu_i y_i + \mu_i^2}{2\sigma^2 \mu_i^2 y_i} \\
 &= -\frac{1}{2} \sum_{i=1}^n \frac{y_i}{\sigma^2 \mu_i^2} + \sum_{i=1}^n \frac{1}{\sigma^2 \mu_i} - \frac{1}{2} \sum_{i=1}^n \frac{1}{\sigma^2 y_i} - \frac{n}{2} \ln(2\pi y_i^3) - \frac{n}{2} \ln \sigma^2
 \end{aligned}$$

Karena  $\mu_i = \mathbf{X}'\beta$ , maka

$$\begin{aligned}
 \ln L(\sigma, \beta) &= -\frac{1}{2} \sum_{i=1}^n \frac{y_i}{\sigma^2 (\mathbf{X}'\beta)^2} + \sum_{i=1}^n \frac{1}{\sigma^2 (\mathbf{X}'\beta)} - \frac{1}{2} \sum_{i=1}^n \frac{1}{\sigma^2 y_i} - \frac{n}{2} \ln(2\pi y_i^3) \\
 &\quad - \frac{n}{2} \ln \sigma^2
 \end{aligned} \tag{2.17}$$

c. Turunan  $L(\sigma, \beta)$  yang disamadengankan nol

i. Turunan  $L(\sigma, \beta)$  terhadap  $\sigma$

Sebelum menurunkan  $L(\sigma, \beta)$  terhadap  $\sigma$ , persamaan (2.6) diubah terlebih dahulu ke dalam bentuk sebagai berikut.

$$\ln L(\sigma, \beta) = -\frac{1}{2} \sum_{i=1}^n \frac{y_i \sigma^{-2}}{(\mathbf{X}' \beta)^2} + \sum_{i=1}^n \frac{\sigma^{-2}}{(\mathbf{X}' \beta)} - \frac{1}{2} \sum_{i=1}^n \frac{\sigma^{-2}}{y_i} - \frac{n}{2} \ln(2\pi y_i^3) - \frac{n}{2} \ln \sigma^2 \quad (2.18)$$

Turunan persamaan (2.16) terhadap  $\beta$  adalah

$$\begin{aligned} \frac{\partial \ln L(\sigma, \beta)}{\partial \sigma} &= \sum_{i=1}^n \frac{y_i \sigma^{-3}}{(\mathbf{X}' \beta)^2} - 2 \sum_{i=1}^n \frac{\sigma^{-3}}{(\mathbf{X}' \beta)} + \sum_{i=1}^n \frac{\sigma^{-3}}{y_i} - \frac{1}{\sigma^2} = 0 \\ 0 &= \sum_{i=1}^n \frac{y_i}{(\mathbf{X}' \beta)^2 \sigma^3} - 2 \sum_{i=1}^n \frac{1}{(\mathbf{X}' \beta) \sigma^3} + \sum_{i=1}^n \frac{1}{y_i \sigma^3} - \frac{1}{\sigma^2} \end{aligned} \quad (2.19)$$

ii. Turunan  $L(\sigma, \beta)$  terhadap  $\beta$

Sebelum menurunkan  $L(\sigma, \beta)$  terhadap  $\beta$ , persamaan (2.6) diubah terlebih dahulu ke dalam bentuk sebagai berikut.

$$\begin{aligned} \ln L(\sigma, \beta) &= -\frac{1}{2} \sum_{i=1}^n \frac{y_i}{(\mathbf{X}' \beta)^2 \sigma^2} + \sum_{i=1}^n \frac{1}{(\mathbf{X}' \beta) \sigma^2} - \frac{1}{2} \sum_{i=1}^n \frac{1}{y_i \sigma^2} - \frac{n}{2} \ln(2\pi y_i^3) - \frac{n}{2} \ln \sigma^2 \\ \ln L(\sigma, \beta) &= -\frac{1}{2} \sum_{i=1}^n \frac{y_i \beta^{-2}}{\mathbf{X}' \sigma^2} + \sum_{i=1}^n \frac{\beta^{-1}}{\mathbf{X}' \sigma^2} - \frac{1}{2} \sum_{i=1}^n \frac{1}{y_i \sigma^2} - \frac{n}{2} \ln(2\pi y_i^3) - \frac{n}{2} \ln \sigma^2 \end{aligned} \quad (2.20)$$

Turunan persamaan (2.18) terhadap  $\beta$  yang disamadengankan nol adalah

$$\begin{aligned} \frac{\partial \ln L(\sigma, \beta)}{\partial \beta} &= \frac{3}{2} \sum_{i=1}^n \frac{y_i \beta^{-3}}{\mathbf{X}' \sigma^2} - \sum_{i=1}^n \frac{\beta^{-2}}{\mathbf{X}' \sigma^2} = 0 \\ 0 &= \frac{3}{2} \sum_{i=1}^n \frac{y_i}{(\mathbf{X}' \beta^3) \sigma^2} - \sum_{i=1}^n \frac{1}{(\mathbf{X}' \beta^2) \sigma^2} \end{aligned} \quad (2.21)$$

Persamaan (2.7) dan (2.9) di atas tidak memberikan penyelesaian karena kedua persamaan di atas saling terkait satu sama lain. Misalnya pada persamaan (2.7) setelah fungsi *log likelihood* diturunkan terhadap ternyata turunannya masih

mempunyai parameter lain yaitu Begitu juga dengan persamaan (2.9) sehingga pengestimasi kedua parameter ini harus dilakukan secara bersamaan yaitu dengan menggunakan metode MLE yang diselesaikan dengan metode numerik iterasi yang disebut sebagai metode Newton-Raphson.

## 2.7 Regresi Gamma

Model regresi gamma adalah sebagai berikut:

$$\mu_i = E(Y_i) = \exp(\beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}) \quad (2.21)$$

dengan

$$\mathbf{x}_i = [1 \quad x_{i1} \quad x_{i2} \quad \dots \quad x_{ik}]_{1 \times (1+k)}^T; \quad \boldsymbol{\beta} = [\beta_0 \quad \beta_1 \quad \dots \quad \beta_k]_{1 \times (1+k)}^T$$

Berdasarkan persamaan (2.20) maka diperoleh:

$$\mu_i = \frac{\alpha}{\gamma_i} = \exp(\mathbf{x}_i^T \boldsymbol{\beta}) \text{ sehingga:}$$

$$\gamma_i = \frac{\alpha}{\exp(\mathbf{x}_i^T \boldsymbol{\beta})} \quad (2.22)$$

Dengan melakukan substitusi persamaan (2.22) ke persamaan (2.21), maka diperoleh fungsi kepadatan peluang dari model regresi gamma adalah sebagai berikut:

$$f(y_i | \alpha, \gamma) = \begin{cases} \frac{\left(\frac{\alpha}{\exp(\mathbf{x}_i^T \boldsymbol{\beta})}\right)^\alpha y_i^{\alpha-1} e^{-\left(\frac{\alpha}{\exp(\mathbf{x}_i^T \boldsymbol{\beta})}\right) y_i}}{\Gamma(\alpha)} & , y_i > 0, \alpha > 0, \gamma > 0 \\ 0 & , \text{ untuk } y \text{ yang lainnya} \end{cases} \quad (2.23)$$

Peubah respons dalam regresi gamma merupakan peubah kontinu non negatif  $Y_i > 0$  dengan ukuran kemiringan ke kanan yang ditentukan oleh besarnya  $v$  dengan nilai harapan sebagai berikut:

$$\mu_i = \frac{v}{\gamma_i} \text{ sehingga}$$

$$\gamma_i = \frac{v}{\mu_i} \quad (2.24)$$

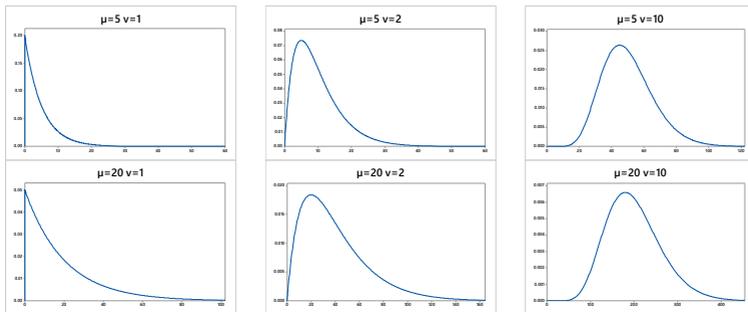
Dengan melakukan substitusi persamaan (2.24) ke persamaan (2.19), maka fungsi padat peluangnya sebagai berikut  $Y_i \sim \text{Gamma}(\mu_i, \nu)$ , (McCullagh & Nelder, 1989):

$$f(y_i) = \frac{1}{\Gamma(\nu)y_i} \left( \frac{\nu y_i}{\mu_i} \right)^\nu \exp\left( -\frac{\nu y_i}{\mu_i} \right)^i \quad (2.25)$$

Dimana  $y_i > 0, \nu > 0, \mu_i > 0, i = 0, 1, 2, \dots, n$

Berdasarkan fungsi padat peluang tersebut diperoleh  $E(Y_i) = \mu_i$  dan

$$\text{Var}(Y_i) = \mu_i^2 / \nu = \mu_i^2 \sigma^2.$$



**Gambar 2.1** Distribusi Gamma

Pada model regresi gamma, parameter  $\nu = \sigma^{-2}$  diasumsikan konstan untuk semua pengamatan, sehingga fungsi densitasnya mempunyai bentuk yang sama untuk semua pengamatan. Model regresi gamma merupakan suatu model yang tergolong dalam model linear terampat (*Generalized Linear Model*), dengan mengasumsikan bahwa peubah respons  $Y_i \sim \text{Gamma}(\mu_i, \nu)$ ,  $E(Y_i) = \mu_i = X\beta$  dengan fungsi hubung adalah  $g(\mu_i) = \mu_i^{-1}$ , kadang kala juga menggunakan fungsi hubung log natural (Czado, 2004).

## 2.8 Estimasi parameter

Penaksiran parameter dilakukan dengan metode *maximum likelihood estimator* (MLE). Metode MLE bertujuan untuk memaksimalkan nilai fungsi *lnlikelihood* (Casella dan Berge, 2002). Berdasarkan fungsi kepadatan peluang bersama dari Y pada persamaan (2.23), maka diperoleh fungsi *lnlikelihood* sebagai berikut:

$$\begin{aligned} \ln L(\alpha, \beta) &= \ln \left( \prod_{i=1}^n \left( \frac{\left( \frac{\alpha}{\exp(\mathbf{x}_i^T \boldsymbol{\beta})}\right)^\alpha}{\Gamma(\alpha)} y_i^{\alpha-1} \exp\left(-\frac{y_i \alpha}{\exp(\mathbf{x}_i^T \boldsymbol{\beta})}\right) \right) \right) \\ &= -n \ln \Gamma(\alpha) - \alpha \sum_{i=1}^n (\mathbf{x}_i^T \boldsymbol{\beta}) + \alpha n \ln \alpha + (\alpha - 1) \ln \sum_{i=1}^n y_i - \sum_{i=1}^n \left( \frac{y_i \alpha}{\exp(\mathbf{x}_i^T \boldsymbol{\beta})} \right) \quad (2.26) \end{aligned}$$

Penaksir parameter pada regresi gamma yaitu  $\hat{\alpha}$  dan  $\hat{\boldsymbol{\beta}}$  diperoleh dengan memaksimumkan fungsi  $L(\alpha, \beta)$ . Maksimum dari fungsi  $L(\alpha, \beta)$  dapat diperoleh dengan cara mencari turunan pertama fungsi *lnlikelihood* pada persamaan (2.26), untuk parameter  $\alpha$  dan  $\boldsymbol{\beta}$  kemudian disamadengankan nol.

Turunan parsial fungsi  $\ln L(\alpha, \beta)$  terhadap parameter  $\boldsymbol{\beta}$  :

$$\frac{\partial \ln L(\alpha, \beta)}{\partial \boldsymbol{\beta}} = -\alpha \sum_{i=1}^n \mathbf{x}_i^T + \alpha \sum_{i=1}^n \mathbf{x}_i^T y_i \left( \exp(\mathbf{x}_i^T \boldsymbol{\beta}) \right)^{-1} \quad (2.27)$$

Turunan parsial fungsi  $\ln L(\alpha, \beta)$  terhadap parameter  $\alpha$  :

$$\frac{\partial \ln L(\alpha, \beta)}{\partial \alpha} = -n\psi(\alpha) - \sum_{i=1}^n (\mathbf{x}_i^T \boldsymbol{\beta}) + n(\ln \alpha + 1) + \sum_{i=1}^n y_i - \sum_{i=1}^n \left( \frac{y_i}{\exp(\mathbf{x}_i^T \boldsymbol{\beta})} \right) \quad (2.28)$$

Persamaan (2.27) dan (2.28) menghasilkan penyelesaian yang tidak *close form*, sehingga untuk penyelesaiannya digunakan pendekatan numerik yaitu algoritma *Newton Raphson*.

## 2.9 Metode Newton-Raphson

Metode Newton-Raphson adalah metode numerik untuk menyelesaikan persamaan non-linear secara iteratif seperti persamaan likelihood yang memaksimumkan suatu fungsi. Dasar dari metode ini adalah pendekatan deret Taylor linear. Dalam metode Newton-Raphson dibutuhkan turunan pertama dan kedua dari fungsi log-likelihoodnya. Untuk mengestimasi  $\sigma$  dan  $\beta$  dengan metode Newton-Raphson diperlukan estimasi awal dari  $\sigma$  dan  $\beta$ . Karena model regresi inverse Gaussian menggunakan fungsi hubung log, dan setelah diperoleh nilai estimator parameter regresi inverse Gaussian. Maka persamaan regresi inverse

Gaussian dugaan dengan variabel respons berdistribusi inverse Gaussian adalah sebagai berikut:

$$\hat{\mu} = e^{\hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2} + \dots + \hat{\beta}_k X_{ik}} \quad (2.29)$$

Fungsi likelihood dengan parameter  $\beta$  dapat diselesaikan sehingga memperoleh nilai estimator  $\beta$  dengan menggunakan metode Newton Raphson. Rumus estimasi parameter  $\beta$  pada iterasi ke-(t+1) dalam proses iterasi ( $t = 0, 1, 2, \dots$ ) adalah sebagai berikut:

$$\hat{\theta}_{(t+1)} = \hat{\theta}_{(t)} - D(\theta_t)^{-1} d(\theta_t) \quad (2.30)$$

Dimana,

$\hat{\theta}_{(t+1)}$  : estimasi parameter  $\theta$  pada iterasi ke-(t+1)

$\hat{\theta}_{(t)}$  : estimasi parameter  $\theta$  pada iterasi ke-t

$D(\theta_t)^{-1}$  : matriks turunan pertama fungsi likelihood

$d(\theta_t)$  : matriks turunan kedua fungsi likelihood

Menurut Chapra & Canale (1988: 134) proses iterasi dengan menggunakan metode Newton-Raphson terus dilakukan hingga didapatkan nilai  $\theta$  yang konvergen, yaitu sampai konvergen, yaitu

sampai  $\left| \frac{\hat{\theta}_{(t+1)} - \hat{\theta}_{(t)}}{\hat{\theta}_{(t)}} \right| < \delta$ , dengan  $\delta$  bilangan yang kecil,  $\delta > 0$ .

## 2.10 Uji Goodness of Fit

Uji *goodness of fit* digunakan untuk mengetahui apakah model layak atau tidak layak digunakan. Langkah-langkah pengujian dengan uji *goodness of fit* (Jong & Heller, 2008:71):

1. Merumuskan hipotesis  
 $H_0$  : Model Layak Digunakan  
 $H_1$  : Model tidak layak dignakan
2. Memilih taraf signifikan :  $\alpha$
3. Menentukan statistik uji  
 Statistik yang digunakan adalah Deviance.
4. Kriteria keputusan :  
 Tolak  $H_0$  jika *Deviance*  $> X^2_{\alpha, (n-k)}$

Dengan  $n$  adalah banyaknya pengamatan dan  $k$  adalah banyaknya parameter.

## 5. Kesimpulan

**Tabel 2.3.** Deviance untuk respons distribusi Keluarga Eksponensia

| Distribusi       | Deviance   |
|------------------|--|
| Normal           | $\frac{1}{\sigma^2} \sum_{i=1}^n (y_i - \hat{\mu}_i)$  |
| Poisson          | $2 \sum_{i=1}^n \left\{ y_i \ln \left( \frac{y_i}{\hat{\mu}_i} \right) - (y_i - \hat{\mu}_i) \right\}$   |
| Binomial         | $2 \sum_{i=1}^n n_i \left\{ y_i \ln \left( \frac{y_i}{\hat{\mu}_i} \right) + (n_i - y_i) \ln \left( \frac{n_i - y_i}{1 - \hat{\mu}_i} \right) \right\}$                                    |
| Gamma            | $2v \sum_{i=1}^n \left\{ -\ln \left( \frac{y_i}{\hat{\mu}_i} \right) + \left( \frac{y_i - \hat{\mu}_i}{\hat{\mu}_i} \right) \right\}$  |
| Inverse gaussian | $\frac{1}{\sigma^2} \sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{\hat{\mu}_i^2 y_i}$  |
| Negatif Binomial | $2 \sum_{i=1}^n \left\{ y_i \ln \left( \frac{y_i}{\hat{\mu}_i} \right) - \left( y_i + \frac{1}{K} \right) \ln \left( \frac{Y_i + \frac{1}{K}}{\hat{\mu}_i + \frac{1}{K}} \right) \right\}$ |

### 2.11 Uji Signifikansi Parameter dengan Uji Wald

Setelah dilakukan uji dengan deviance dilakukan uji signifikansi dari variabel penjelas, dengan melihat apakah terdapat variabel penjelas yang tidak signifikan di dalam model. Jika terdapat variabel penjelas yang tidak signifikan, perlu dilakukan reduksi terhadap variabel penjelas tersebut. Statistik uji yang digunakan untuk menyelidiki tingkat signifikansi dari variabel penjelas yaitu dengan uji Wald. Uji Wald akan membandingkan estimator likelihood parameter  $\beta_j$  dengan estimator standar errornya. Variabel penjelas yang dikeluarkan terlebih dahulu yaitu variabel yang tidak signifikan yang memiliki nilai p-value paling besar, lalu dibuat model baru tanpa memasukkan variabel yang

direduksi tersebut. Langkah-langkah uji Wald adalah sebagai berikut (Agresti, 2007: 11):

1. Merumuskan Hipotesis

$$H_0 : \beta_j = 0$$

$$H_1 : \beta_j \neq 0$$

2. Memilih taraf signifikan :  $\alpha$

3. Menentukan statistik uji

Statistik yang digunakan adalah Uji Wald.

$$W = \left\{ \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \right\} \sim X_1^2, j = 0, 1, 2, \dots, k$$

4. Kriteria keputusan :

$$\text{Tolak } H_0 \text{ jika } W > X_{(1 - \alpha/2, 1)}^2 \text{ atau } W < X_{(\alpha/2, 1)}^2$$

$$\text{atau } H_0 \text{ ditolak jika } p\text{-value} < \alpha$$

5. Kesimpulan

## 2.12 Metode Rasional untuk Prediksi Banjir

Salah satu metode yang sering digunakan dalam menghitung atau memperkirakan besarnya debit rencana adalah metode Rasional. Metode ini digunakan untuk daerah yang luas pengalirannya kurang dari 300 ha (Goldman et.al., 1986, dalam Suripin, 2004). Metode Rasional dikembangkan berdasarkan asumsi bahwa curah hujan yang terjadi mempunyai intensitas seragam dan merata di seluruh daerah pengaliran selama paling sedikit sama dengan waktu konsentrasi ( $t_c$ ). Persamaan matematikanya adalah sebagai berikut :

$$Q = 0.278 \times C \times I \times A \quad (2.31)$$

Q : Debit ( $m^3/detik$ )

0.278 : Konstanta Jika satuan luas daerah menggunakan  $km^2$

C : Koefisien Aliran

I : Intensitas curah hujan ( $mm/jam$ )

A : Luas daerah aliran ( $km^2$ )

### 2.13 Metode Mononobe

Dalam perencanaan bangunan pengairan (misalnya drainase), debit rencana sangat diperlukan untuk mengetahui kapasitas yang seharusnya dapat ditampung oleh sebuah drainase, agar semua debit air dapat ditampung dan teralirkan. Metode mononobe adalah metode yang digunakan untuk menghitung intensitas curah hujan. Persamaannya adalah sebagai berikut.

$$I = \frac{R_{24}}{24} \left( \frac{24}{T} \right)^{\frac{2}{3}} \quad (2.32)$$

dimana :

- I : Intensitas curah hujan (mm/jam)
- t : Lamanya curah hujan / durasi curah hujan (jam)
- $R_{24}$  : Curah hujan rencana dalam suatu periode ulang, yang nilainya didapat dari tahapan sebelumnya (tahapan analisis frekuensi)

### 2.14 Uji Korelasi

Korelasi terjadi diantara dua variabel dimana suatu nilai variabel berhubungan dengan nilai variabel lainnya. Salah satu cara untuk melihat hubungan diantara dua variabel adalah dengan melihat nilai koefisien korelasi, yaitu suatu nilai yang mengukur kekuatan hubungan antara dua variabel. Koefisien korelasi biasanya mengacu pada koefisien korelasi momen-produk Pearson, sesuai nama orang yang mengembangkannya yaitu Pearson. Koefisien korelasi Pearson diperoleh dengan membagi kovarians kedua variabel dengan perkalian simpangan bakunya dengan persamaan sebagai berikut :

$$\rho_{x,y} = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X) \text{var}(Y)}} \quad (2.33)$$

Nilai koefisien korelasi Pearson berkisar diantara -1 sampai 1. Kedua variabel dikatakan memiliki hubungan yang erat bila nilai korelasi mendekati 1, baik itu positif maupun negatif. Namun bila nilai korelasi 0 menunjukkan bahwa kedua variabel tidak memiliki

hubungan erat. Nilai korelasi yang positif menunjukkan adanya hubungan berbanding lurus diantara kedua variabel tersebut, sebaliknya hubungan yang berbanding terbalik terlihat dari nilai korelasi yang negatif (Triola, 2010).

Pengujian hipotesisnya sebagai berikut:

$H_0$  : Tidak ada hubungan antara kedua variabel.

$H_1$  : Terdapat hubungan antar kedua variabel.

Statistik uji yang digunakan adalah

$$t = \frac{r_{x,y} \sqrt{n-2}}{\sqrt{1-r^2}} \quad (2.34)$$

Keputusan tolak  $H_0$  jika  $|t_{hit}| = t_{(\frac{\alpha}{2}, n-2)}$

## 2.15 Root Mean Square Error

RMSE adalah metode alternatif untuk mengevaluasi teknik peramalan yang digunakan untuk mengukur tingkat akurasi hasil prakiraan suatu model. RMSE merupakan nilai rata-rata dari jumlah kuadrat kesalahan, juga dapat menyatakan ukuran besarnya kesalahan yang dihasilkan oleh suatu model prakiraan. Nilai RMSE rendah menunjukkan bahwa variasi nilai yang dihasilkan oleh suatu model prakiraan mendekati variasi nilai obeservasinya.

$$RMSEP = \sqrt{\frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n}} \quad (2.35)$$

*RMSEP = Root Mean Square Error Prediction*

$n$  = jumlah sampel

$Y_i$  = nilai aktual ke- $i$

$\hat{Y}_i$  = nilai prediksi ke- $i$

## 2.16 Tinjauan Non Statistik

Tinjauan non statistik pada penelitian meliputi pemahaman yang meliputi curah hujan dan banjir.

### 2.10.1 Curah Hujan

Curah hujan adalah jumlah air yang jatuh di permukaan tanah datar selama periode tertentu yang diukur dengan satuan tinggi milimeter (mm) di atas permukaan horizontal. Dalam

penjelasan lain curah hujan juga dapat diartikan sebagai ketinggian air hujan yang terkumpul dalam tempat yang datar, tidak menguap, tidak meresap dan tidak mengalir. Indonesia merupakan negara yang memiliki angka curah hujan yang bervariasi dikarenakan daerahnya yang berada pada ketinggian yang berbeda-beda. Curah hujan 1 (satu) milimeter, artinya dalam luasan satu meter persegi pada tempat yang datar tertampung air setinggi satu milimeter tempat yang datar tertampung air setinggi satu milimeter atau tertampung air setinggi 1 liter.

Menurut Linsley (1996: 49), jenis-jenis hujan berdasarkan intensitas curah hujan, yaitu:

- 1) hujan ringan, kecepatan jatuh sampai 2,5 mm/jam;
- 2) hujan menengah, dari 2,5-7,6 mm/jam.
- 3) hujan lebat, lebih dari 7,6 mm/jam

### **2.10.2 Banjir Genangan**

Banjir dalam pengertian umum adalah debit aliran air sungai dalam jumlah yang tinggi, atau debit aliran air di sungai secara relatif lebih besar dari kondisi normal akibat hujan yang turun di hulu atau di suatu tempat tertentu terjadi secara terus menerus, sehingga air tersebut tidak dapat ditampung oleh alur sungai yang ada, maka air melimpah keluar dan menggenangi daerah sekitarnya (Peraturan Dirjen RLPS No.04 thn 2009). Banjir merupakan peristiwa dimana daratan yang biasanya kering (bukan daerah rawa) menjadi tergenang oleh air, hal ini disebabkan oleh curah hujan yang tinggi dan kondisi topografi wilayah berupa dataran rendah hingga cekung. Selain itu, terjadinya banjir juga dapat disebabkan oleh limpasan air permukaan (*runoff*) yang meluap dan volumenya melebihi kapasitas pengaliran sistem drainase atau sistem aliran sungai. Banjir dapat terjadi akibat naiknya permukaan air lantaran curah hujan yang diatas normal, perubahan suhu, tanggul/bendungan yang bobol, pencairan salju yang cepat, terhambatnya aliran air di tempat lain (Ligal, 2008).

Dibyo Saputro (1984) mengatakan penyebab banjir dan lamanya genangan bukan hanya disebabkan oleh meluapnya air sungai, melainkan oleh kelebihan curah hujan dan fluktuasi muka air laut khususnya dataran alluvial pantai, unit-unit geomorfologi seperti daerah rawa, rawa belakang, dataran banjir, pertemuan sungai dengan dataran alluvial merupakan tempat-tempat yang rentan banjir. Sebab-sebab banjir alami (Kodoatie & Sugiyanto, 2002) meliputi curah hujan, pengaruh fisiologi, erosi dan sedimentasi, kapasitas sungai, kapasitas drainase yang tidak memadai, pengaruh air pasang.

*(Halaman ini sengaja dikosongkan)*

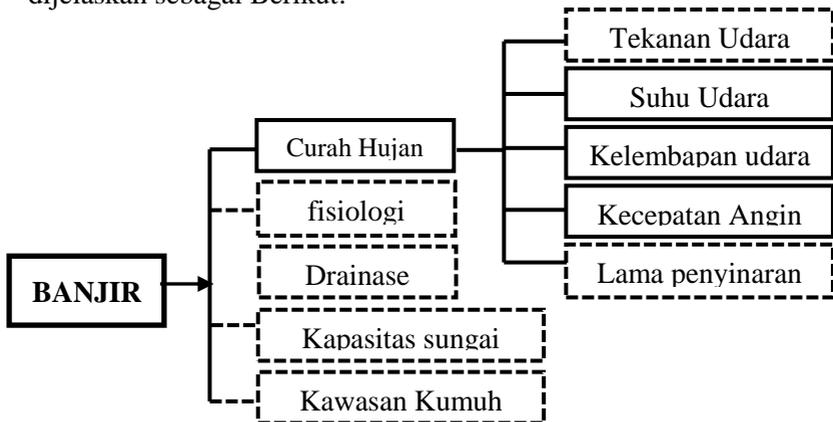
## BAB III METODOLOGI PENELITIAN

### 3.1 Sumber Data

Sumber data yang digunakan merupakan data sekunder yang di peroleh dari Dinas Pekerjaan Umum Bina Marga dan Pematusan dan data Online BMKG (<http://dataonline.bmkg.go.id>). Data yang diperoleh dari Dinas Pekerjaan Umum Bina Marga dan Pematusan merupakan data curah hujan harian mulai tahun 2010 sampai dengan 2018 di 6 stasiun Kota Surabaya yaitu Gunungsari, Wonokromo, Wonorejo, Keputih, Kedungcowek, dan Gubeng. Sedangkan untuk data yang diperoleh dari BMKG online merupakan data yang memengaruhi curah hujan seperti kelembaban rata rata, kecepatan angin rata rata, dan suhu udara (temperature) rata rata. Data *training* berjumlah 3652 yaitu pada 1 Januari 2009 sampai 30 November 2018. Sedangkan data *testing* berjumlah 31 data yaitu 1-31 Desember 2018.

### 3.2 Variabel Penelitian

Rincian variabel yang akan dianalisis pada penelitian ini dapat dijelaskan sebagai Berikut:



(Kodoatie & Sugivanto, 2002)

**Gambar 3.1** Rancangan variabel penelitian

Pada gambar 3.1 diketahui bahwa banjir disebabkan oleh banyak faktor antara lain curah hujan tinggi, pengaruh fisiologi, drainase lahan, pembuangan sampah sembarangan, infiltrasi tanah dan kawasan kumuh dan lain lain (Kodoatie & Sugiyanto, 2002). Namun dikarenakan keterbatasan waktu dan tidak tersedianya data pada variabel penyebab banjir yang diberi kotak garis putus putus, sehingga variabel yang digunakan dalam penelitian ini adalah sebagai berikut.

**Tabel 3.1** Variabel Penelitian

| No. | Variabel | Deskripsi        | Satuan | Jenis Data |
|-----|----------|------------------|--------|------------|
| 1.  | $X_1$    | Kelembaban Udara | %      | Rasio      |
| 2.  | $X_2$    | Suhu Udara       | °c     | Rasio      |
| 3.  | $X_3$    | Kecepatan Angin  | Knot   | Rasio      |
| 4.  | $Y_t$    | Curah Hujan      | Mm     | Rasio      |

### 3.3 Variabel dan Definisi Operasional

Variabel yang digunakan dalam penelitian ini terdiri dari satu variabel respon dan tiga variabel prediktor. Adapun variabel-variabel tersebut adalah sebagai berikut.

#### a. Variabel Respon

Penelitian ini menggunakan variabel respon curah hujan yang berasal dari Dinas Pekerjaan Umum Bina Marga dan Pematusan. Adapun definisi operasional dari variabel respon yang digunakan adalah jumlah air yang jatuh di permukaan tanah datar selama periode tertentu yang diukur dengan satuan tinggi milimeter (mm) di atas permukaan horizontal.

#### b. Variabel Prediktor

Peneiltian ini menggunakan tiga variabel yang diduga berpengaruh terhadap intensitas curah hujan yang terjadi. Pemilihan variabel prediktor tersebut didasarkan pada hubungan paling tinggi pengaruhnya terhadap curah hujan. Variabel prediktor atau penjelas merupakan variabel yang diukur di stasiun yang berbeda yaitu Stasiun Juanda, Perak 1 dan Perak 2. Adapun tiga variabel prediktor tersebut adalah sebagai berikut:

1. Temperatur rata rata ( $X_1$ )

Temperatur adalah suatu ukuran dingin atau panasnya keadaan atau sesuatu lainnya.

2. Kelembaban rata rata ( $X_2$ )

Kelembaban adalah tingkat kebasahan udara (jumlah air yang terkandung di udara) yang dinyatakan dengan persentase nisbi/relatif terhadap titik jenuhnya. Udara jenuh dengan kelembaban 100% jika di dalam 1 M<sup>3</sup> udara pada temperatur 30<sup>0</sup>C mengandung 30 gram uap air. Sedangkan pada suhu 20<sup>0</sup>C mengandung 17 gram uap air.

3. Kecepatan angin rata rata ( $X_3$ )

Kecepatan angin adalah sebuah kuantitas atmosferik fundamental. Kecepatan angin disebabkan oleh pergerakan angin dari tekanan tinggi ke tekanan rendah.

### 3.4 Struktur Data

Struktur data secara umum yang digunakan dalam penelitian ini terdapat pada Tabel 3.2. struktur data terdiri atas data pengukuran yang berasal dari setiap rayon dan data pengukuran dari BMKG.

**Tabel 3.2** Struktur Data dari Setiap Stasiun (univariat)

| Waktu<br>(H/B/T) | t | Y              | Data pengukuran stasiun BMKG |           |           |
|------------------|---|----------------|------------------------------|-----------|-----------|
|                  |   |                | $X_{1s}$                     | $X_{2s}$  | $X_{3s}$  |
| 01/01/2008       | 1 | Y <sub>1</sub> | $X_{1,1}$                    | $X_{2,1}$ | $X_{3,1}$ |
| 02/01/2008       | 2 | Y <sub>2</sub> | $X_{1,2}$                    | $X_{2,2}$ | $X_{3,2}$ |
| ⋮                | ⋮ | ⋮              | ⋮                            | ⋮         | ⋮         |
| t                | t | Y <sub>t</sub> | $X_{1,t}$                    | $X_{2,t}$ | $X_{3,t}$ |
| ⋮                | ⋮ | ⋮              | ⋮                            | ⋮         | ⋮         |
| 31/12/2018       | n | Y <sub>n</sub> | $X_{1,n}$                    | $X_{2,n}$ | $X_{3,n}$ |

\*Ket : Y<sub>t</sub> = curah hujan di level kecamatan,  
s = stasiun (1,2,3)

**Tabel 3.3** Struktur Data dari Setiap Stasiun (multivariat)

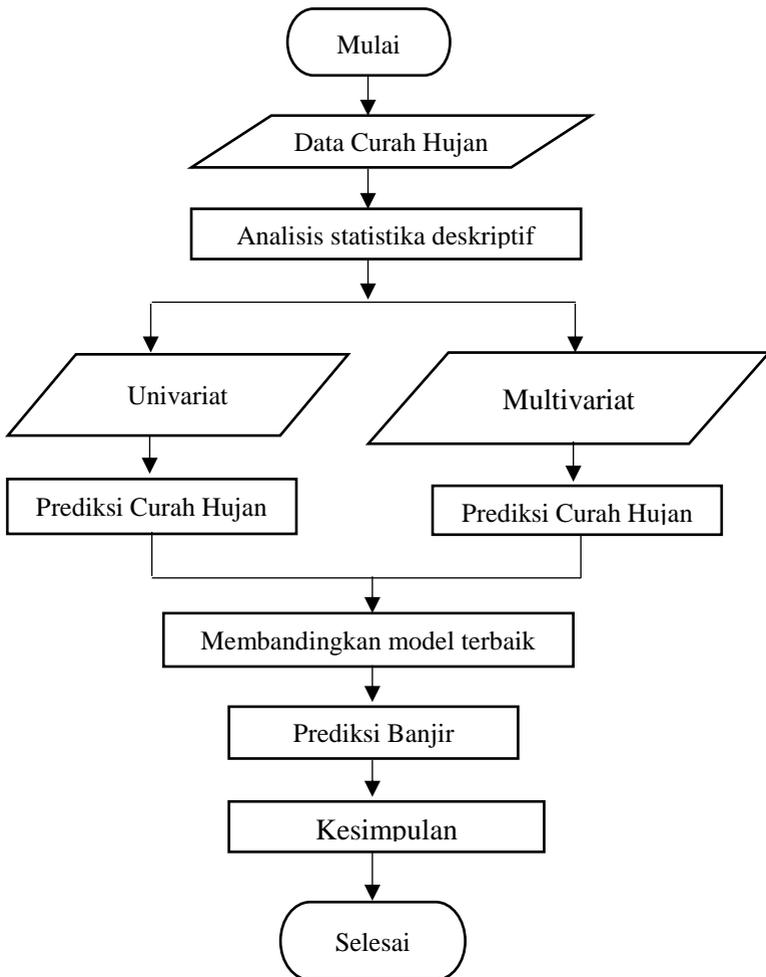
| Tanggal        | t | Curah Hujan | Data pengukuran stasiun BMKG |                            |                            |
|----------------|---|-------------|------------------------------|----------------------------|----------------------------|
|                |   |             | Temperatur                   | Kelembapan                 | Kecepatan Angin            |
|                |   |             | $X_{11} \cdots X_{13}$       | $X_{21} \cdots X_{23}$     | $X_{31} \cdots X_{33}$     |
| 01 / 01 / 2009 | 1 | $y_{r,1}$   | $X_{11,1} \cdots X_{13,1}$   | $X_{21,1} \cdots X_{23,1}$ | $X_{31,1} \cdots X_{33,1}$ |
| 02 / 01 / 2009 | 2 | $y_{r,2}$   | $X_{11,2} \cdots X_{13,2}$   | $X_{21,2} \cdots X_{23,2}$ | $X_{31,2} \cdots X_{33,2}$ |
| 03 / 01 / 2009 | 3 | $y_{r,3}$   | $X_{11,3} \cdots X_{13,3}$   | $X_{21,3} \cdots X_{23,3}$ | $X_{31,3} \cdots X_{33,3}$ |
| ⋮              | ⋮ | ⋮           | ⋮                            | ⋮                          | ⋮                          |
| t              | t | $y_r$       | $X_{11,r} \cdots X_{13,r}$   | $X_{21,r} \cdots X_{23,r}$ | $X_{31,t} \cdots X_{33,r}$ |
| ⋮              | ⋮ | ⋮           | ⋮                            | ⋮                          | ⋮                          |
| 31 / 12 / 2018 | n | $y_m$       | $X_{11,n} \cdots X_{13,n}$   | $X_{21,n} \cdots X_{23,n}$ | $X_{31,n} \cdots X_{33,n}$ |

### 3.5 Langkah Penelitian

Langkah-langkah analisis yang akan dilakukan berdasarkan dengan tujuan pada penelitian ini adalah sebagai berikut.

1. Melakukan eksplorasi terhadap variabel-variabel penelitian yang digunakan.
  - a. Melakukan analisis statistika deskriptif dengan menggunakan *time series plot* dan *boxplot*.
  - b. Melakukan analisis hubungan antara variabel respon dengan prediktor menggunakan *scatter plot* dan korelasi.
2. Melakukan pemodelan curah hujan dengan regresi inverse Gaussian
  - a. Pemodelan curah hujan setiap rayon dengan masing masing variabel BMKG.
  - b. Pemodelan curah hujan setiap rayon dengan hasil PCA variabel BMKG.
3. Membandingkan akurasi hasil prediksi dari curah hujan berdasarkan pemodelan secara univariate maupun dengan menggunakan PCA pada variabel prediktornya.
4. Memprediksi banjir dengan metode rasional berdasarkan hasil prediksi curah hujan terbaik.
  - a. Memprediksi variabel prediktor
  - b. Melakukkann perhitungan banjir dengan metode rasional
5. Menarik kesimpulan dan saran.

Langkah-langkah analisis secara umum digambarkan pada diagram alir pada Gambar 3.1.



**Gambar 3.2** Diagram Alir Langkah Analisis

*(Halaman ini sengaja dikosongkan)*

## BAB IV HASIL DAN PEMBAHASAN

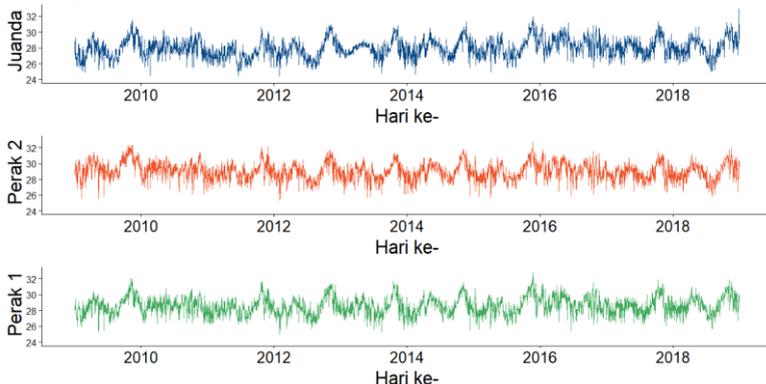
Pada bab ini akan diulas mengenai karakteristik variabel klimatologi BMKG dan variabel respon curah hujan yang kemudian akan dilanjutkan dengan pemodelan hasil prediksi banjir berdasarkan hasil prediksi curah hujan dengan menggunakan regresi inverse gaussian.

### 4.1 *Time Series Plot*

Time Series plot digunakan sebagai analisis trend untuk melihat kecenderungan pergerakan data pada periode-periode waktu sebelumnya. Time Series plot yaitu membuat diagram scatter berdasarkan data dengan waktu. Berikut merupakan hasil dari Time Series plot untuk variabel prediktor dan respon.

#### 4.1.1 *Time Series Plot Temperatur di Setiap Stasiun*

Deskripsi mengenai karakteristik variabel temperatur perlu dilakukan supaya didapatkan informasi umum tentang temperatur di masing masing stasiun selama periode 2009 sampai dengan tahun 2018. Pada gambar 4.1 telah disajikan perbedaan temperature dari tahun 2009 sampai dengan tahun 2018 di masing masing stasiun.



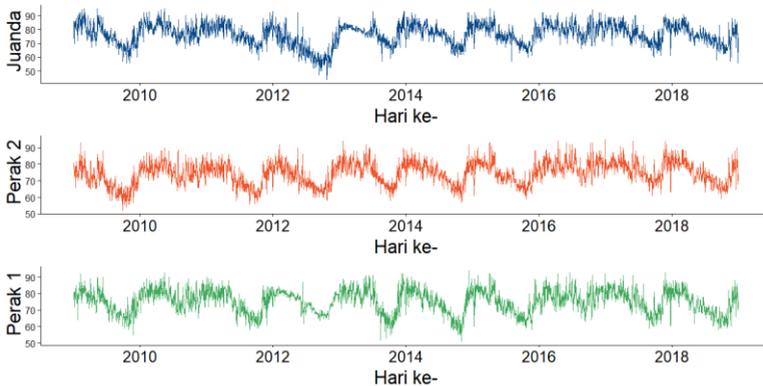
**Gambar 4.1** Time series plot temperature di setiap stasiun

Berdasarkan hasil plot variabel Temperature dengan seluruh stasiun yaitu menunjukkan pola yang hampir sama di masing masing stasiun. Hal ini mengindikasikan bahwa suhu atau

temperatur di Surabaya cenderung tetap meskipun diukur di berbagai tempat yang berbeda. Dalam kurun waktu sepuluh tahun tidak mengindikasikan adanya suhu yang ekstrim yaitu dapat dilihat pada plot yang masih menunjukkan nilai konstan.

#### 4.1.2 Time Series Plot Kelembapan di Setiap Stasiun

Deskripsi mengenai karakteristik variabel kelembapan perlu dilakukan supaya didapatkan informasi umum tentang temperature di masing masing stasiun selama periode 2009 sampai engan tahun 2018. Pada gambar 4.2 telah disajikan perbedaan kelembapan dari tahun 2009 sampai dengan tahun 2018 di masing masing stasiun.

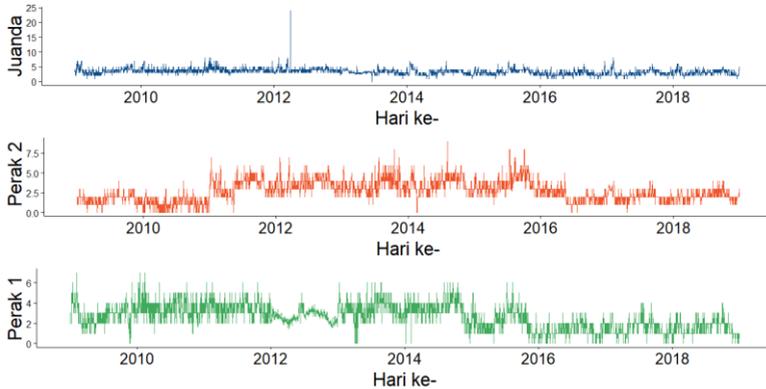


**Gambar 4.2** Time series plot kelembapan di setiap stasiun

Kelembapan yang terjadi di masing masing stasiun menunjukkan kemiripan. Artinya tidak ada perbedaan yang signifikan meskipun diukur di tempat yang berbeda pada wilayah Kota Surabaya.

#### 4.1.3 Time Series Plot Kecepatan Angin di Setiap Stasiun

Deskripsi mengenai karakteristik variabel kecepatan angin perlu dilakukan supaya didapatkan informasi umum tentang temperature di masing masing stasiun selama periode 2009 sampai engan tahun 2018. Pada gambar 4.3 telah disajikan perbedaan t kecepatan angin dari tahun 2009 sampai dengan tahun 2018 di masing masing stasiun.



**Gambar 4.3** Time series plot Kecepatan angin di setiap stasiun

Kecepatan angin yang terjadi di stasiun Juanda di tahun 2019 menunjukkan angka yang sangat tinggi pada bulan Januari. Pada bulan Januari Surabaya dan kota kota lain di Jawa Timur sedang dalam musibah terguyur angin kencang. Sedangkan pada di hari lainnya di stasiun Juanda kecepatan anginnya sangat rendah. Berbeda halnya dengan yang terjadi pada stasiun Perak 1 maupun perak 2 yang cenderung tinggi kecepatan anginnya. Secara keseluruhan jika dilihat pada plot dari tahun 2009 sampai dengan tahun 2018 kecepatan angina di tiga tempat pengukuran Surabaya cenderung mengalami penurunan.

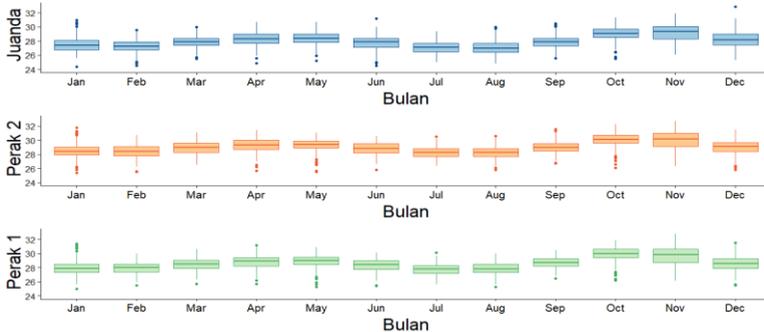
## 4.2 Boxplot

Boxplot adalah salah satu cara dalam statistik deskriptif untuk menggambarkan secara grafik dari data numeris melalui lima ukuran diantaranya adalah nilai maksimum dan minimum, kuartil I, kuartil II dan kuartil III. Selain lima ukuran diatas boxplot mampu mnunjukkan adanya outlier pada data serta mampu juga dalam menentukan distribusi dari suatu data.

### 4.2.1 Boxplot Temperatur di Setiap Stasiun

Temperature merupakan variabel yang sifatnya berubah ubah artinya menyesuaikan suhu udara di tempat tersebut. Keadaan temperature di daerah Surabaya menurut penelitian atau data BMKG adalah hamper sama dengan untuk setiap tahunnya. Artinya dalam setiap tahunnya ada pola musiman yang terdapat dalam pengukuran variabel temperature. Berikut merupakan hasil

boxplot untuk temperature yang dikukur mulai tahun 2009 sampai dengan tahun 2018.

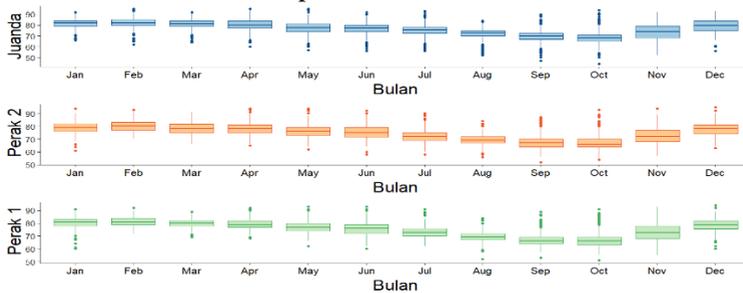


**Gambar 4.4** Boxplot Temperature di setiap stasiun

*Boxplot* variabel temperatur yang disajikan pada Gambar 4.4 memberikan informasi bahwa secara deskriptif tidak terjadi cuaca yang eksrim serta selisih rata rata yang cukup kecil. Begitu juga untuk nilai mediannya hampir sama di setiap stasiun selama kurun waktu sepuluh tahun.

#### 4.2.2 Boxplot Kelembapan di Setiap Stasiun

Kelembapan yang terjadi di kota Surabaya selama kurun waktu sepuluh tahun mengalami berbagai variasi yang cukup beragam. Kelembapan bisa dipicu dari adanya hujan dan juga berasal dari asap asap kendaraan. Kelembapan yang terjadi di kota Surabaya terus mengalami peningkatan setiap tahunnya. Berikut disajikan boxplot kelembapan dari setiap stasiun selama sepuluh tahun dari tahun 2009 sampai tahun 2018.



**Gambar 4.5** Boxplot Kelembapan di setiap Stasiun

**Gambar 4.5** menginformasikan bahwa pada stasiun Juanda memiliki nilai median yang hamper sama pada setiap tahunnya di

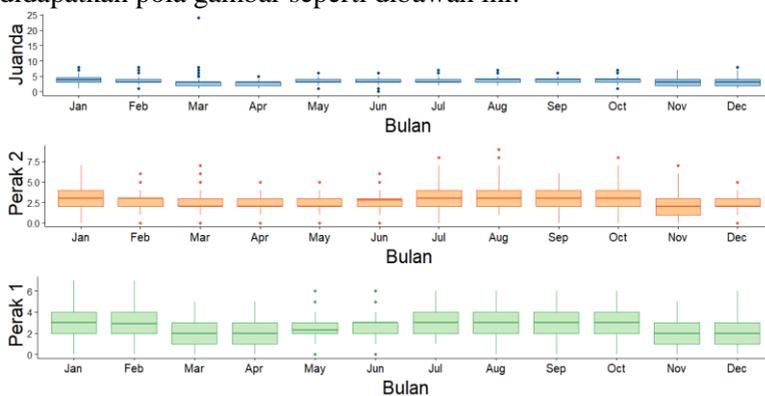
bulan yang sama. Tidak hanya itu kelembapan yang diukur yang di stasiun ini tergolong mengelompok secara rendah di semua bulan.

Pada stasiun Perak 1 terdapat perbedaan rata rata yang cukup signifikan berdasarkan tinggi rendahnya Boxplot pada gambar diatas. Namun secara keseluruhan nilai medianya hamper sama atau tidak jauh berbeda. Sedangkan pada stasiun Perak 2 menunjukkan perbedaan yang sangat jelas kelembapan yang terjadi pada bulan bulan tertentu. Perbedaan nilai tengah juga terlihat sangat signifikan. Hal ini menunjukkan bahwa terjadi variasi yang sangat tinggi pada pengukuran kelembapan di stasiun ini.

Secara keseluruhan terjadi perbedaan yang signifikan kadar kelembapan yang terjadi di pengukuran stasiun wilayah Surabaya.

#### 4.2.3 Boxplot Kecepatan Angin di Setiap Stasiun

Sampai dengan bulan Desember tahun 2018, kecepatan angin menurut masing masing stasiun BMKG sangat bervariasi. Setelah dilakukan pengukuran setiap hari untuk kecepatan angin didapatkan pola gambar seperti dibawah ini.



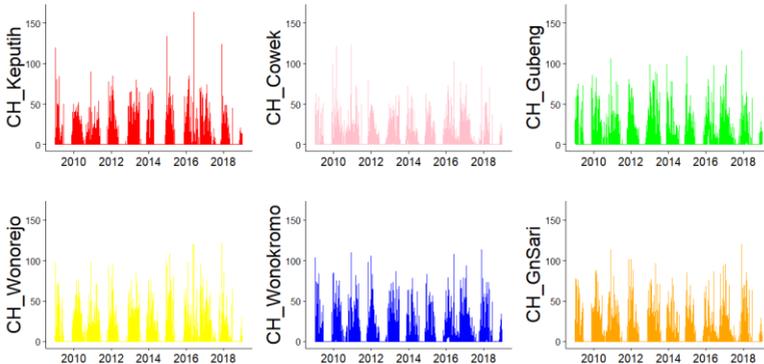
**Gambar 4.6** Boxplot Kecepatan Angin di setiap Stasiun

Kecepatan angin di Surabaya sangat bervariasi. Kecepatan angin yang diukur di stasiun Juanda memiliki nilai yang relative kecil jika dibandingkan dengan stasiun yang lain. Namun pada bulan Maret terjadi kecepatan angina yang sangat tinggi di Surabaya dan daerah Jawa Timur dan mengakibatkan kerusakan

kerusakan rumah warga. Sedangkan pada stasiun Perak 1 dan Perak 2 terjadi naik turun untuk kecepatan angin selama kurun waktu sepuluh tahun di Surabaya. Nilai median yang variatif menunjukkan bahwa kecepatan angin di daerah daerah Surabaya cukup relative berbeda beda.

### 4.3 Curah Hujan

Curah hujan adalah jumlah air yang jatuh di permukaan tanah datar selama periode tertentu yang diukur dengan satuan tinggi (mm) diatas permukaan horizontal bila tidak terjadi evaporasi, runoff dan infiltrasi. Berikut merupakan plot yang menggambarkan kondisi curah hujan di Suarabaya yang diukur di enam (6) stasiun Balai PSAWS Bukit Peketingan yaitu Keputih, Gubeng, Wonorejo, Wonokromo, Kedung Cowek dan Gunung sari disajikan pada **Gambar 4.7**

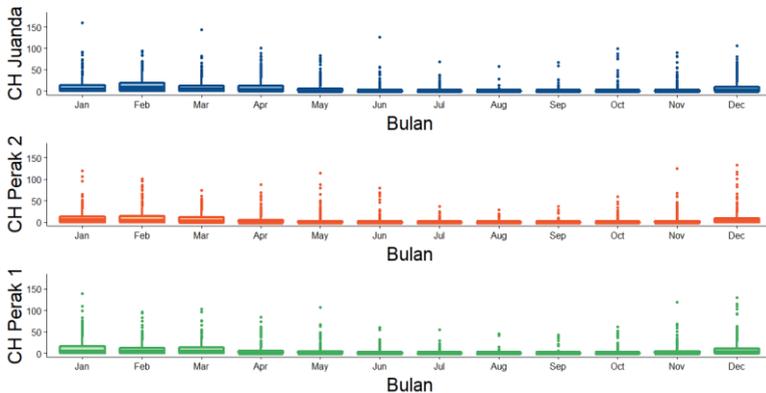


**Gambar 4.7** Time series Plot Curah Hujan Gabungan Balai Surabaya

Curah hujan di Surabaya sangat bervariasi, berdasarkan data pencatatan dari Balai PSAWS hari hujan tinggi terjadi pada bulan November sampai dengan April. Kondisi ini terjadi secara terus menerus selama sepuluh tahun. Pada bulan Mei sampai dengan Oktober terjadi hujan namun tidak sering saat bulan November-April. Hal ini terjadi karena adanya pergerakan angin Monsun Barat yang terjadi pada bulan Oktober sampai dengan April. Angin Monsun Barat merupakan angin yang bergerak dari benua Asia menuju ke benua Australia. Angin ini membawa banyak sekali uap air sehingga menyebabkan daerah yang dilaluinya mengalami musim penghujan termasuk kota Surabaya. Dampak yang

ditimbulkan dari adanya angin monsun dengan iklim yang kurang baik menyebabkan kondisi hujan yang sangat lebat sehingga mengakibatkan curah hujan sangat tinggi.

Untuk lebih jelasnya mengenai curah hujan yang terjadi di Kota Surabaya selama sepuluh tahun disajikan dalam Boxplot setiap bulan seperti dibawah ini. Jika dilihat dari median Boxplot kondisi curah hujan tertinggi terjadi pada November sampai April.



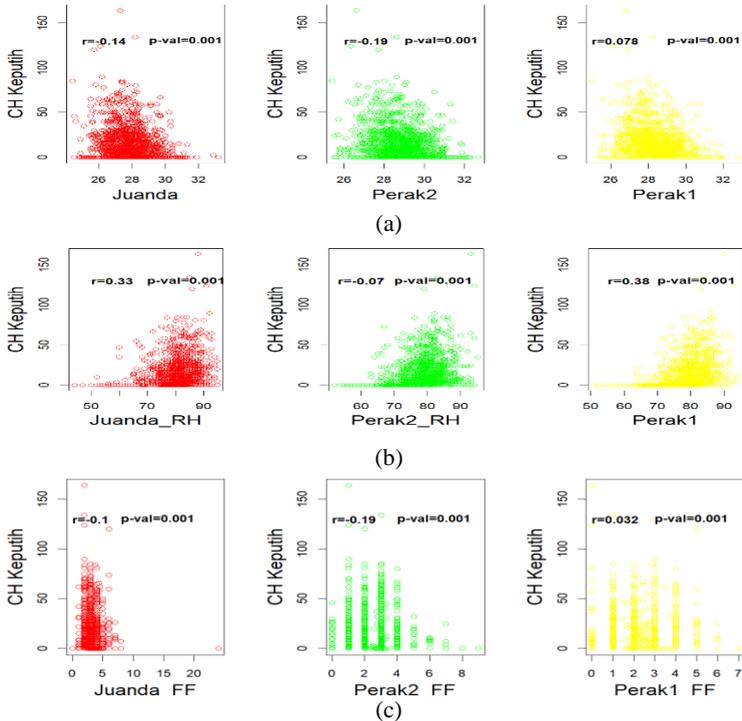
**Gambar 4.8** BoxPlot Curah Hujan Gabungan di Balai Surabaya

#### 4.4 Scatter Plot

Scatter plot atau yang biasa disebut dengan diagram tebar merupakan plot yang menggambarkan hubungan diantara variabel. Hubungan yang terjadi bisa negative ataupun positif. Berikut merupakan beberapa scatter plot dari variabel curah hujan di seluruh balai PSAWS dengan variabel BMKG.

##### 4.4.1. Scatter Plot Curah Hujan Keputih dengan Variabel Stasiun BMKG

Hubungan yang terjadi antara curah hujan dengan dengan variabel BMKG yang meliputi Temperature, Kelembapan dan Kecepatan angin yang memengaruhi curah hujan sangat menarik untuk di deskripsikan. Hubungan yang dimaksud adalah berdasarkan nilai korelasi untuk menentukan seberapa erat hubungan diantara kedua variabel. Selain itu apakah terjadi kelinearan terhadap hubungan yang terjadi menjadi salah satu parameternya. Berikut merupakan hasil *scatterplot*.



**Gambar 4.9** Scatter Plot Curah Hujan Keputih dengan (a) Temperatur, (b) Kelembapan, dan (c) Kecepatan Angin

**Gambar 4.9** menginformasikan bahwa pada plot curah hujan keputih dengan variabel Temperatur dari semua stasiun menunjukkan hubungan yang negative pada stasiun Juanda dan perak 1 yang masing masing berkorelasi sebesar -0.14 dan -0.19. Yang berarti adalah hubungan yang terjadi variabel curah hujan di Keputih dengan variabel Temeperatur di stasiun Juanda dan perak 1 memiliki hubungan yang berbanding terbalik dan tingkat hubungannya cukup rendah berdasarkan nilai korelasinya. Sedangkan hubungan yang terjadi pada variabel curah hujan dengan kelembapan di stasiun Perak 2 memiliki hubungan yang positif dan berbanding lurus dengan nilai korelasi sebesar 0.078, yang artinya adalah hubungannya sangat rendah.

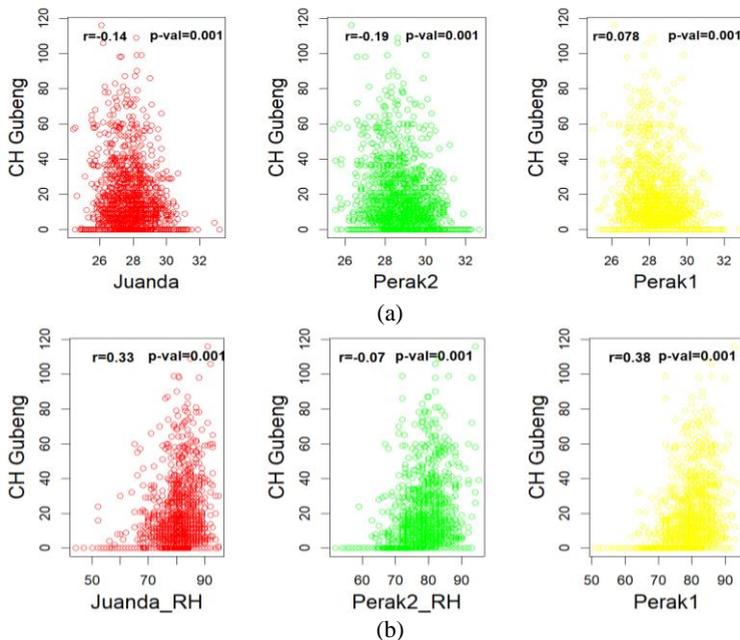
Hubungan yang terjadi antara curah hujan keputih dengan variabel kelembapan dari semua stasiun menunjukkan hubungan

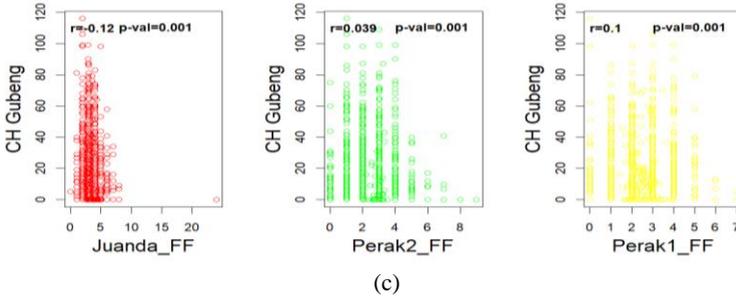
yang positif pada stasiun Juanda dan perak 1 yang masing masing berkorelasi sebesar 0.33 dan 0.38. Nilai korelasi tersebut cukup tinggi menunjukkan hubungan yang positif dan memiliki hubungan yang berbanding lurus dengan curah hujan. Sedangkan pada variabel kelembapan di stasiun Perak 2 menunjukkan hubungan yang negatif dengan korelasi yang sangat kecil yaitu sebesar -0.14.

Hubungan yang terjadi antara curah hujan keputih dengan variabel kecepatan angin dari semua stasiun menunjukkan hubungan yang negatif pada stasiun Juanda dan perak 1 yang masing masing berkorelasi sebesar -0.1 dan -0.081. Nilai korelasi tersebut sangat kecil hubungannya dan berbanding terbalik dengan curah hujan. Sedangkan hubungan lurus terjadi pada variabel kecepatan angin di stasiun Perak 2 dengan nilai korelasi sebesar 0.032.

#### 4.4.2. Scatter Plot Curah Hujan Gubeng dengan Variabel Stasiun BMKG

Berikut disajikan plot curah hujan di Gubeng dengan semua variabel BMKG.





**Gambar 4.10** Scatter Plot Curah Hujan Gubeng dengan (a) Temperatur, (b) Kelembapan, dan (c) Kecepatan Angin

Berdasarkan **Gambar 4.10** menginformasikan bahwa pada plot curah hujan Gubeng dengan variabel Temperatur dari semua stasiun menunjukkan hubungan yang negative pada stasiun Juanda dan perak 1 yang masing masing berkorelasi sebesar  $-0.11$  dan  $-0.17$ . Yang berarti adalah hubungan yang terjadi variabel curah hujan di Gubeng dengan variabel Temeperatur di stasiun Juanda dan perak1 memiliki hubungan yang berbanding terbalik dan tingkat hubungannya cukup rendah berdasarkan nilai korelasinya. Sedangkan hubungan yang terjadi pada variabel curah hujan dengan kelembapan di stasiun Perak 2 memiliki hubungan yang positif dan berbanding lurus dengan nilai korelasi sebesar  $0.066$ ., yang artinya adalah hubungannya sangat rendah. Korelasi tertinggi terjadi pada stasiun Perak 1.

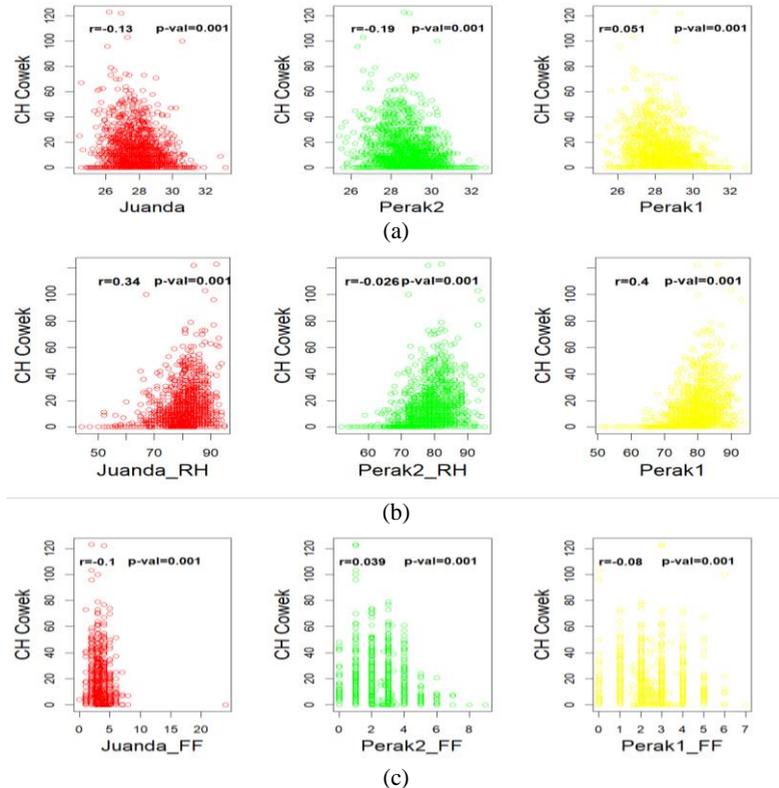
Hubungan yang terjadi antara curah hujan Gubeng dengan variabel kelembapan dari semua stasiun menunjukkan hubungan yang positif pada stasiun Juanda dan perak 1 yang masing masing berkorelasi sebesar  $0.35$  dan  $0.39$ . Nilai korelasi tersebut cukup tinggi menunjukkan hubungan yang positif dan memiliki hubungan yang berbanding lurus dengan curah hujan. Sedangkan pada variabel kelembpan di stasiun Perak 2 menunjukkan hubungan yang negatif dengan korelasi yang sangat kecil yaitu sebesar  $-0.62$ . Korelasi tertinggi terjadi pada stasiun Perak 1.

Hubungan yang terjadi antara curah hujan Gubeng dengan variabel kecepatan angin dari semua stasiun menunjukkan hubungan yang negatif pada stasiun Juanda dan perak 1 yang masing masing berkorelasi sebesar  $-0.12$  dan  $-0.11$ . Nilai korelasi tersebut sangat kecil hubungannya dan berbanding terbalik dengan

curah hujan. Sedangkan hubungan lurus terjadi pada variabel kecepatan angin di stasiun Perak 2 dengan nilai korelasi sebesar 0.039. Korelasi tertinggi terjadi pada stasiun Juanda.

#### 4.4.3. Scatter Plot Curah Hujan Kedung Cowek dengan Variabel Stasiun BMKG

Berikut disajikan plot curah hujan di Kedung Cowek dengan semua variabel BMKG.



**Gambar 4.11** Scatter Plot Curah Hujan Kedung Cowek dengan (a) Temperatur, (b) Kelembaban, dan (c) Kecepatan Angin

**Gambar 4.11** menginformasikan bahwa pada plot curah hujan Gubeng dengan variabel Temperatur dari semua stasiun menunjukkan hubungan yang negative pada stasiun Juanda dan perak 1 yang masing masing berkorelasi sebesar -0.13 dan -0.19. Yang berarti adalah hubungan yang terjadi variabel curah hujan di

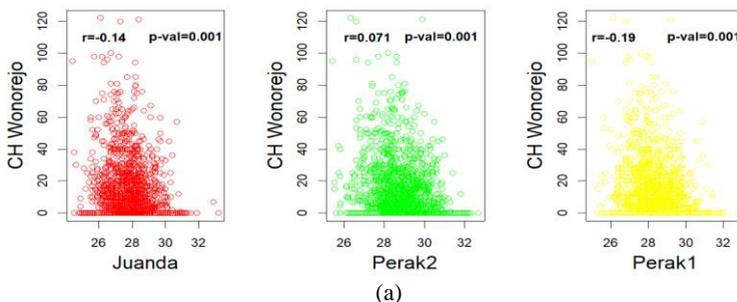
Gubeng dengan variabel Temperatur di stasiun Juanda dan perak 1 memiliki hubungan yang berbanding terbalik dan tingkat hubungannya cukup rendah berdasarkan nilai korelasinya. Sedangkan hubungan yang terjadi pada variabel curah hujan dengan kelembapan di stasiun Perak 2 memiliki hubungan yang positif dan berbanding lurus dengan nilai korelasi sebesar 0.051., yang artinya adalah hubungannya sangat rendah. Korelasi tertinggi terjadi pada stasiun Perak 1.

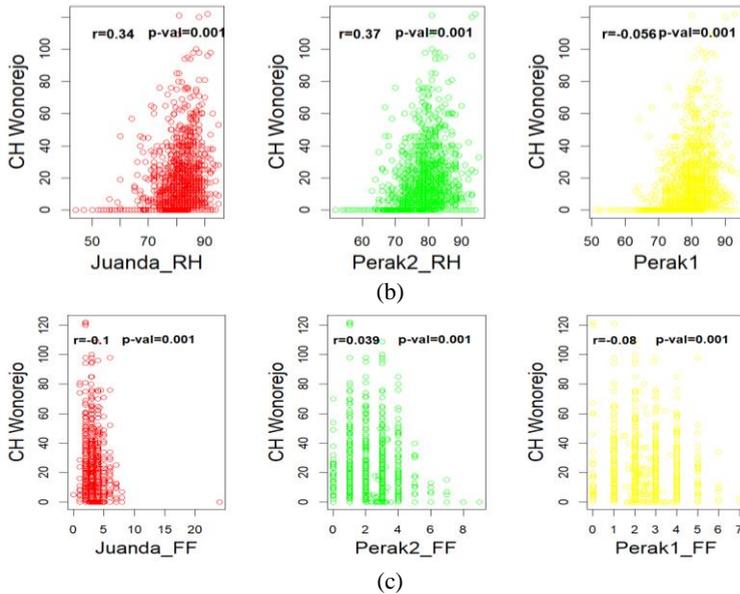
Hubungan yang terjadi antara curah hujan Gubeng dengan variabel kelembapan dari semua stasiun menunjukkan hubungan yang positif pada stasiun Juanda dan perak 1 yang masing masing berkorelasi sebesar 0.34 dan 0.4. Nilai korelasi tersebut cukup tinggi menunjukkan hubungan yang positif dan memiliki hubungan yang berbanding lurus dengan curah hujan. Sedangkan pada variabel kelembapan di stasiun Perak 2 menunjukkan hubungan yang negatif dengan korelasi yang sangat kecil yaitu sebesar -0.03. Korelasi tertinggi terjadi pada stasiun Perak 1.

Hubungan yang terjadi antara curah hujan Gubeng dengan variabel kecepatan angin dari semua stasiun menunjukkan hubungan yang negatif pada stasiun Juanda dan perak 1 yang masing masing berkorelasi sebesar -0.11 dan -0.08. Nilai korelasi tersebut sangat kecil hubungannya dan berbanding terbalik dengan curah hujan.

#### 4.4.4. Scatter Plot Curah Hujan Wonorejo dengan Variabel Stasiun BMKG

Hubungan antara dua variabel yang berasal dari curah hujan Wonorejo dengan variabel stasiun BMKG dapat diketahui dengan menggunakan *scatter plot*. Berikut hasil *scatter plot*nya.





**Gambar 4.12** Scatter Plot Curah Wonorejo dengan (a) Temperatur, (b) Kelembaban, dan (c) Kecepatan Angin

Berdasarkan **Gambar 4.12** menginformasikan bahwa pada plot curah hujan Gubeng dengan variabel Temperatur dari semua stasiun menunjukkan hubungan yang negative pada stasiun Juanda dan perak 1 yang masing masing berkorelasi sebesar -0.14 dan -0.19. Yang berarti adalah hubungan yang terjadi variabel curah hujan di Gubeng dengan variabel Temeperatur di stasiun Juanda dan perak 1 memiliki hubungan yang berbanding terbalik dan tingkat hubungannya cukup rendah berdasarkan nilai korelasinya. Sedangkan hubungan yang terjadi pada variabel curah hujan dengan kelembaban di stasiun Perak 2 memiliki hubungan yang positif dan berbanding lurus dengan nilai korelasi sebesar 0.071., yang artinya adalah hubungannya sangat rendah. Korelasi tertinggi terjadi pada stasiun Perak 1.

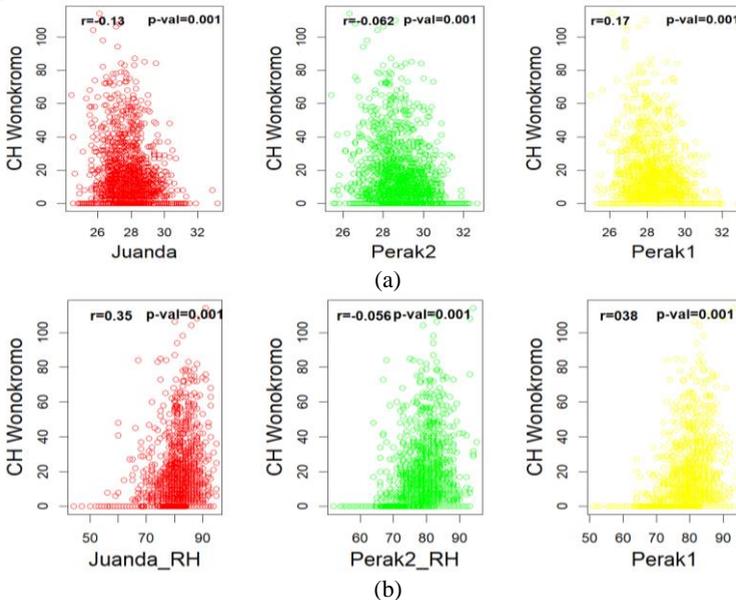
Hubungan yang terjadi antara curah hujan Gubeng dengan variabel kelembaban dari semua stasiun menunjukkan hubungan yang positif pada stasiun Juanda dan perak 1 yang masing masing berkorelasi sebesar 0.34 dan 0.37. Nilai korelasi tersebut cukup tinggi menunjukkan hubungan yang positif dan memiliki hubungan

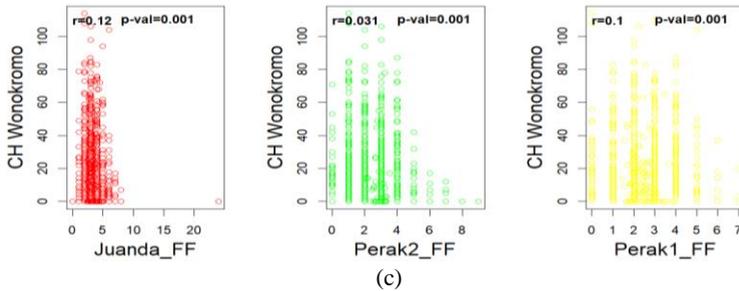
yang berbanding lurus dengan curah hujan. Sedangkan pada variabel kelembapan di stasiun Perak 2 menunjukkan hubungan yang negatif dengan korelasi yang sangat kecil yaitu sebesar  $-0.06$ . Korelasi tertinggi terjadi pada stasiun Perak 1.

Hubungan yang terjadi antara curah hujan Gubeng dengan variabel kecepatan angin dari semua stasiun menunjukkan hubungan yang negatif pada stasiun Juanda dan perak 1 yang masing masing berkorelasi sebesar  $-0.11$  dan  $-0.095$ . Nilai korelasi tersebut sangat kecil hubungannya dan berbanding terbalik dengan curah hujan. Sedangkan hubungan lurus terjadi pada variabel kecepatan angin di stasiun Perak 2 dengan nilai korelasi sebesar  $0.029$ . Korelasi tertinggi terjadi pada stasiun Juanda.

#### 4.4.5. Scatter Plot Curah Hujan Wonokromo dengan Variabel Stasiun BMKG

Berikut disajikan plot curah hujan di Wonokromo dengan semua variabel BMKG dan semua Stasiun yaitu stasiun Juanda, perak1 dan Perak 2.





**Gambar 4.13** Scatter Plot Curah Hujan Wonokromo dengan (a) Temperatur, (b) Kelembaban, dan (c) Kecepatan Angin)

Berdasarkan **Gambar 4.13** menginformasikan bahwa pada plot curah hujan Gubeng dengan variabel Temperatur dari semua stasiun menunjukkan hubungan yang negative pada stasiun Juanda dan perak 1 yang masing masing berkorelasi sebesar -0.14 dan -0.19. Yang berarti adalah hubungan yang terjadi variabel curah hujan di Gubeng dengan variabel Temeperatur di stasiun Juanda dan perak 1 memiliki hubungan yang berbanding terbalik dan tingkat hubungannya cukup rendah berdasarkan nilai korelasinya. Sedangkan hubungan yang terjadi pada variabel curah hujan dengan kelembapan di stasiun Perak 2 memiliki hubungan yang positif dan berbanding lurus dengan nilai korelasi sebesar 0.082. Yang artinya adalah hubungannya sangat rendah. Korelasi tertinggi terjadi pada stasiun Perak 2.

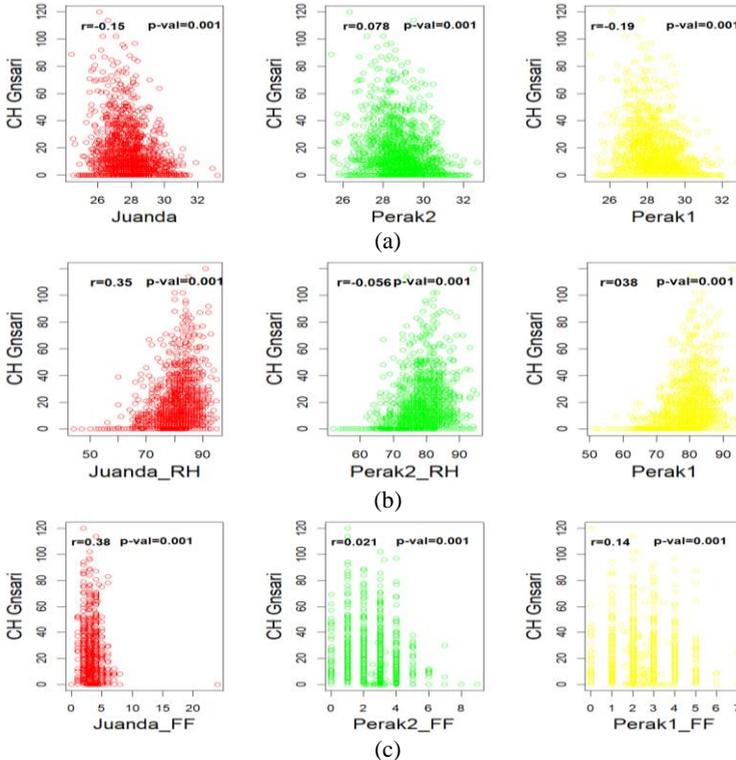
Hubungan yang terjadi antara curah hujan Gubeng dengan variabel kelembapan dari semua stasiun menunjukkan hubungan yang positif pada stasiun Juanda dan perak 1 yang masing masing berkorelasi sebesar 0.35 dan 0.38. Nilai korelasi tersebut cukup tinggi menunjukkan hubungan yang positif dan memiliki hubungan yang berbanding lurus dengan curah hujan. Sedangkan pada variabel kelembpan di stasiun Perak 2 menunjukkan hubungan yang negatif dengan korelasi yang sangat kecil yaitu sebesar -0.07. Korelasi tertinggi terjadi pada stasiun Perak 1.

Hubungan yang terjadi antara curah hujan Gubeng dengan variabel kecepatan angin dari semua stasiun menunjukkan hubungan yang negatif pada stasiun Juanda dan perak 1 yang masing masing berkorelasi sebesar -0.12 dan -0.1. Nilai korelasi tersebut sangat kecil hubungannya dan berbanding terbalik dengan

curah hujan. Sedangkan hubungan lurus terjadi pada variabel kecepatan angin di stasiun Perak 2 dengan nilai korelasi sebesar 0.031. Korelasi tertinggi terjadi pada stasiun Juanda.

#### 4.4.6. Scatter Plot Curah Hujan Gunung Sari dengan Variabel Stasiun BMKG

Berikut disajikan plot curah hujan di Wonokromo dengan semua variabel BMKG dan semua Stasiun.



**Gambar 4.14** Scatter Plot Curah Hujan Gunung Sari dengan (a) Temperatur, (b) Kelembaban, dan (c) Kecepatan Angin

**Gambar 4.14** menginformasikan bahwa pada plot curah hujan Gubeng dengan variabel Temperatur dari semua stasiun menunjukkan hubungan yang negative pada stasiun Juanda dan perak 1 yang masing masing berkorelasi sebesar -0.15 dan -0.19. Yang berarti adalah hubungan yang terjadi variabel curah hujan di Gubeng dengan variabel Temeperatur di stasiun Juanda dan perak

1 memiliki hubungan yang berbanding terbalik dan tingkat hubungannya cukup rendah berdasarkan nilai korelasinya. Sedangkan hubungan yang terjadi pada variabel curah hujan dengan kelembapan di stasiun Perak 2 memiliki hubungan yang positif dan berbanding lurus dengan nilai korelasi sebesar 0.08. Yang artinya adalah hubungannya sangat rendah. Korelasi tertinggi terjadi pada stasiun Perak 2.

Hubungan yang terjadi antara curah hujan Gubeng dengan variabel kelembapan dari semua stasiun menunjukkan hubungan yang positif pada stasiun Juanda dan Perak 1 yang masing-masing berkorelasi sebesar 0.37 dan 0.39. Nilai korelasi tersebut cukup tinggi menunjukkan hubungan yang positif dan memiliki hubungan yang berbanding lurus dengan curah hujan. Sedangkan pada variabel kelembapan di stasiun Perak 2 menunjukkan hubungan yang negatif dengan korelasi yang sangat kecil yaitu sebesar -0.08. Korelasi tertinggi terjadi pada stasiun Perak 1.

Hubungan yang terjadi antara curah hujan Gubeng dengan variabel kecepatan angin dari semua stasiun menunjukkan hubungan yang negatif yang masing-masing berkorelasi sebesar -0.12, -0.094 dan -0.0021. Nilai korelasi tersebut sangat kecil hubungannya dan berbanding terbalik dengan curah hujan. Korelasi tertinggi terjadi pada stasiun Juanda.

#### **4.5 Pemodelan Univariate Curah Hujan di Surabaya dengan Variabel di Setiap Stasiun**

Pemodelan dilakukan dengan menggunakan metode regresi Inverse Gaussian dan regresi Gamma yang termasuk dalam keluarga distribusi eksponensial. Data curah hujan meliputi data masing-masing kecamatan yang terdiri dari curah hujan Keputih, Gubeng, Wonorejo, Wonokromo, Kedung Cowek dan Gunung Sari. Sedangkan variabel BMKG yang dimaksudkan adalah Temperatur, Kelembapan dan Kecepatan Angin. Tujuan akhir dari pemodelan secara univariate adalah untuk mendapatkan model terbaik dengan berdasarkan nilai error yang terkecil. Berikut merupakan hasil pemodelannya.

##### **4.5.1 Pemodelan Curah Hujan Keputih**

Pemodelan yang dilakukan yaitu dengan meregresikan variabel curah hujan Keputih sebagai respon dengan variabel

stasiun BMKG sebagai prediktor. Sebagai prediktor tambahan yaitu variabel *dummy* merupakan variabel yang berdasarkan dari statistika deskriptif didapatkan untuk hari hujan yang dikategorikan dengan yaitu pada bulan November sampai dengan April sedangkan pada bulan Mei sampai dengan Oktober merupakan hari hari yang jarang turun hujan dikategorikan dengan angka 0. Serta variabel lag yang merupakan variabel  $Y_{t-k}$  yang digunakan untuk meningkatkan model.

### I. Pemodelan Curah Hujan Keputih dengan Stasiun Juanda

Berikut adalah hasil pemodelan dari curah hujan keputih yang diawali dengan meregresikan variabel respon dengan masing masing variabel prediktor dengan regresi inverse Gaussian (link="1/ $\mu^2$ "), Gamma (link=inverse), Gamma (link=inverse) dan Gamma dengan fungsi "sqrt".

**Tabel 4.1** Summary CH Keputih dengan Temperatur

|        | Gaussian |       | inverse |       | identity |       | sqrt   |       |
|--------|----------|-------|---------|-------|----------|-------|--------|-------|
|        | Est      | p-val | est     | p-val | est      | p-val | est    | p-val |
| Interc | -0.273   | 0.002 | -1.191  | 0.002 | 33.87    | 0.002 | 14.705 | 0.002 |
| Temp   | 0.012    | 0.002 | 0.049   | 0.002 | -1.02    | 0.002 | -0.443 | 0.002 |

**Tabel 4.2** Summary CH Keputih dengan Kelembapan

|        | Gaussian |       | inverse |       | identity |       | sqrt   |       |
|--------|----------|-------|---------|-------|----------|-------|--------|-------|
|        | Est      | p-val | est     | p-val | est      | p-val | est    | p-val |
| Interc | 0.347    | 0.002 | 1.351   | 0.002 | -6.52    | 0.002 | -4.078 | 0.002 |
| RH     | -0.004   | 0.002 | -0.014  | 0.002 | 0.148    | 0.002 | 0.081  | 0.002 |

**Tabel 4.3** Summary CH Keputih dengan Kecepatan Angin

|        | Gaussian |       | inverse |       | identity |       | sqrt   |       |
|--------|----------|-------|---------|-------|----------|-------|--------|-------|
|        | Est      | p-val | est     | p-val | est      | p-val | est    | p-val |
| Interc | 0.001    | 0.982 | 0.031   | 0.167 | 6.372    | 0.002 | -4.078 | 0.002 |
| Angin  | 0.011    | 0.002 | 0.049   | 0.002 | -0.27    | 0.002 | 0.081  | 0.002 |

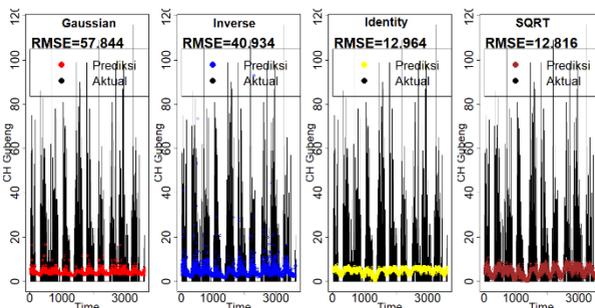
Berdasarkan ketiga Gambar di atas didapatkan informasi bahwa pemodelan terhadap masing masing variabel Temperature, Kelembapan dan kecepatan angin menghasilkan hasil yang signifikan atau berpengaruh terhadap model berdasarkan tiga kriteria yaitu inverse Gaussian , Gamma (link=inverse), Gamma (link=inverse ) dan Gamma dengan fungsi "sqrt". Hal ini dapat

dilihat dari nilai *p-value* yang kurang dari 0.05. Namun hal berbeda ditunjukkan saat dilakukan pengujian secara bersamaan.

**Tabel 4.4** Summary CH Keputih dengan Tiga Variabel

|        | Gaussian |       | inverse |       | identity |       | sqrt   |       |
|--------|----------|-------|---------|-------|----------|-------|--------|-------|
|        | est      | p-val | est     | p-val | est      | p-val | est    | p-val |
| Interc | 0.345    | 0.002 | 1.512   | 0.002 | -19.1    | 0.002 | -8.275 | 0.002 |
| Temp   | 0.001    | 0.002 | -0.004  | 0.002 | 0.445    | 0.002 | 0.069  | 0.002 |
| RH     | 0.004    | 0.965 | -0.015  | 0.432 | 0.16     | 0.002 | 0.194  | 0.002 |
| Angin  | 0.001    | 0.977 | -0.002  | 0.711 | -0.19    | 0.002 | -0.086 | 0.002 |

Hasil pemodelan dengan regresi inverse Gaussian dan Gamma dengan link “inverse” menunjukkan hasil yang signifikan hanya variabel Temperatur dan intercept. Yaitu dengan nilai *p-value* sebesar 0.002 dan kurang dari 0.05. Sedangkan variabel kelembapan dan kecepatan angin tidak berpengaruh signifikan terhadap model karena nilai *p-value* lebih besar daripada nilai  $\alpha$  (0.05). Pada metode Gamma dengan fungsi link “identity” dan “sqrt” didapatkan hasil yang signifikan untuk semua variabel.



**Gambar 4.15** Prediksi Curah Hujan Keputih dengan tiga Variabel

Hasil prediksi diatas dengan menggunakan regresi Gamma dengan fungsi link “sqrt” didapatkan nilai *Root Mean Square Error* (RMSE) terkecil yaitu sebesar 12.816. Sedangkan RMSE Untuk ketiga metode lainnya masing masing sebesar 57.844 (inverse Gaussian), 40.984 (Gamma “inverse”) dan 12.964 (Gamma “identity”). Nilai AIC dari yang terbesar yaitu 33221 (inverse Gaussian), 30068 (gamma “sqrt”), 30026 (inverse), dan 29956 (identity). Dengan memasukkan tiga variabel kedalam model ternyata tidak menaikkan hasil prediksi karena variabel tersebut kurang memberikan kontribusi yang besar kedalam model.

Dibandingkan dengan menggunakan metode lainnya Gamma “sqrt” memiliki hasil yang lebih bagus dari hasil prediksinya dan dari segi visualnya. Dugaan adanya kasus multikolinieritas dapat dilihat dengan melihat nilai VIF (*Variance Inflation Factor*), jika nilai VIF lebih dari 10 mengindikasikan adanya multikolinieritas.

Tabel 4.5 VIF untuk Semua Model (Juanda)

|                  | VIF ( <i>Variance Inflation Factor</i> ) |            |             |
|------------------|--|------------|-------------|
|                  | Temperatur                               | Kelembapan | Kecep_Angin |
| Inverse Gaussian | 1.0007                                   | 2.727      | 2.727       |
| Gamma(inverse)   | 1.642                                    | 2.1733     | 2.116       |
| Gamma(identity)  | 318766.5                                 | 1847666.2  | 1406916.2   |
| Gamma(sqrt)      | 72.757009                                | 3.683      | 65.945      |

Nilai VIF yang cenderung berada dibawah nilai 10 mengindikasikan tidak terjadi kasus multikolinieritas antara variabel prediktor. Sehingga variabel yang masuk kedalam model adalah variabel intersep dan Temperature. Berbeda halnya dengan metode regresi gamma dengan fungsi link “identity” dan “sqrt “ semua variabel saling berhubungan erat dan terjadi kasus multikolinieritas karena nilai VIF lebih dari 10. Sehingga dilakukan penanganan pada kasus ini dengan melakukan analisis komponen utama (PCA) yaitu dengan membentuk kelompok yang memuat ketiga variabel tersebut. Berikut hasil analisis komponen utama variabel prediktor.

Tabel 4.6 Analisis Komponen Utama (Juanda)

|                        | PC1    | PC2     | PC3     |
|------------------------|--------|---------|---------|
| Standard deviation     | 7.5487 | 1.15813 | 0.97523 |
| Proportion of Variance | 0.9613 | 0.02263 | 0.01605 |
| Cumulative Proportion  | 0.9613 | 0.98395 | 1       |

Berdasarkan tabel diatas didapatkan bahwa PC1 sudah cukup untuk mewakili kelompok dari variabel prediktor dimana pada PC1 sudah memuat semua informasi pada variabel prediktor dengan nilai *cumulative proportion* sebesar 0.9613 atau 96.13%.

Pada data curah hujan, banyak data yang terdapat angka 0 atau tidak hujan pada bulan Mei sampai dengan Oktober. Agar

prediksi yang dihasilkan dari regresi inverse Gaussian dan Gamma tidak mendapatkan nilai error yang tinggi pada bulan bulan tidak hujan tersebut, maka ditambahkan variabel *dummy*. Berikut merupakan hasil pemodelan yang dihasilkan.

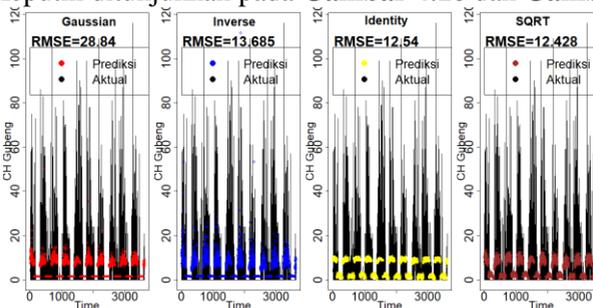
Tabel 4.7. Pemodelan Curah Hujan dengan Temperatur dan *dummy*

|        | Gaussian |       | inverse |       | identity |       | Sqrt   |       |
|--------|----------|-------|---------|-------|----------|-------|--------|-------|
|        | est      | p-val | est     | p-val | est      | p-val | Est    | p-val |
| Interc | 0.272    | 0.001 | -0.291  | 0.01  | 16.09    | 0.001 | 7.589  | 0.002 |
| Temp   | 0.004    | 0.004 | 0.033   | 0.001 | -0.51    | 0.001 | -0.225 | 0.002 |
| dumy   | -0.369   | 0.001 | -0.505  | 0.001 | 7.424    | 0.001 | 1.682  | 0.002 |

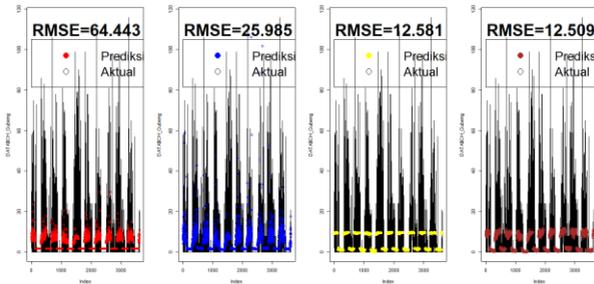
Tabel 4.8. Pemodelan Curah Hujan dengan ketiga variabel dan *dummy*

|        | Gaussian |       | inverse |       | identity |       | sqrt   |       |
|--------|----------|-------|---------|-------|----------|-------|--------|-------|
|        | est      | p-val | est     | p-val | est      | p-val | Est    | p-val |
| Interc | 0.539    | 1E-04 | 1.936   | 0.004 | -5.61    | 0.002 | -4.417 | 0.002 |
| Temp   | -0.002   | 0.014 | -0.012  | 0.008 | 0.045    | 0.002 | 0.036  | 0.002 |
| RH     | -0.002   | 0.45  | -0.015  | 0.08  | 0.133    | 0.002 | 0.109  | 0.002 |
| Angin  | -0.001   | 0.732 | -0.002  | 0.719 | -0.06    | 0.002 | -0.064 | 0.002 |
| dumy   | -0.317   | 3E-04 | -0.403  | 0.002 | 7.917    | 0.002 | 1.819  | 0.002 |

Pemodelan yang dihasilkan dengan meregresikan variabel curah hujan dengan Temperatur dan *dummy* didapatkan hasil yang signifikan pada kedua variabel tersebut. Hal ini ditunjukkan nilai pvalue yang sangat kecil pada regresi gamma inverse dan inverse Gaussian dan kurang dari 0.05. Sedangkan dengan pemodelan yang dilakukan pada **Tabel 4.8** dihasilkan variabel Temperatur dan *dummy*. Hasil ketepatan prediksi model dengan nilai aktual curah hujan Keputih ditunjukkan pada **Gambar 4.16** dan **Gambar 4.17**.



**Gambar 4.16** Prediksi Curah Hujan dengan Temperature dan *Dummy*



**Gambar 4.17** Prediksi CH Keputih dengan ketiga Variabel dan *dummy*

Prediksi curah hujan yang terjadi dengan metode inverse Gaussian menunjukkan hasil yang cukup jauh nilai real dari curah hujan dengan nilai RMSE sebesar 28.64. Pada metode regresi gamma dengan fungsi link Inverse didapatkan prediksi dengan nilai RMSE sebesar 13.826. Sedangkan dengan menggunakan gamma dengan fungsi link “identity” dan “sqrt” didapatkan nilai error yang rendah yaitu sebesar 12.54 dan 12.428. Berdasarkan keempat metode tersebut dengan memasukkan variabel *dummy* dan Temperature metode regresi gamma dengan fungsi link “sqrt” menjadi yang terbaik. Pada prediksi dengan menggunakan ketiga variabel BMKG didapatkan hasil yang tidak lebih baik daripada dengan menggunakan variabel temperature .

Prediksi yang didapatkan dengan menambah variabel *dummy* masih tidak bisa untuk di seajajarkan dengan nilai actual curah hujan. Oleh karena itu diperlukan variabel tambahan yang bisa meningkatkan hasil prediksi dari model dinamakan dengan variabel lag. Alasan utama ditambahkan variabel lag adalah curah hujan yang terjadi di waktu tertentu dipengaruhi oleh curah hujan yang sebelumnya. Berikut merupakan hasil pemodelannya.

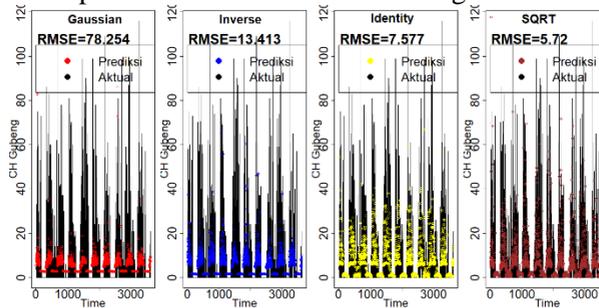
Tabel 4.9 Pemodelan Curah Hujan dengan ketiga variabel, *Dummy* dan lag1

|        | Gaussian |       | inverse |       | Identity |       | Sqrt   |       |
|--------|----------|-------|---------|-------|----------|-------|--------|-------|
|        | est      | p-val | est     | p-val | est      | p-val | Est    | p-val |
| Interc | 0.252    | 0.002 | 1.6     | 0.001 | -3.59    | 0.002 | -3.639 | 0.001 |
| Temp   | 0.001    | 0.84  | -0.008  | 0.38  | 0.086    | 0.002 | 0.089  | 0.001 |
| RH     | -0.002   | 0.001 | -0.011  | 0.001 | 0.03     | 0.002 | 0.03   | 0.001 |
| Angin  | 0.001    | 0.988 | 0.001   | 0.973 | -0.04    | 0.002 | -0.052 | 0.001 |
| Dumy   | -0.136   | 0.001 | -0.408  | 0.001 | 5.595    | 0.002 | 1.505  | 0.001 |
| lag1   | -0.001   | 0.001 | -0.001  | 0.029 | 0.394    | 0.001 | 0.068  | 0.001 |

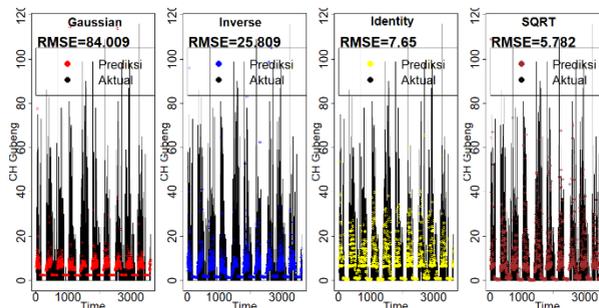
Tabel 4.10 Pemodelan Curah Hujan dengan Temperatur, *Dummy* dan lag1

|        | Gaussian |       | inverse |       | identity |       | Sqrt   |       |
|--------|----------|-------|---------|-------|----------|-------|--------|-------|
|        | est      | p-val | est     | p-val | est      | p-val | Est    | p-val |
| Interc | 0.272    | 0.001 | -0.262  | 0.023 | 6.968    | 0.001 | 6.968  | 0.001 |
| Temp   | 0.004    | 0.004 | 0.032   | 0.001 | -0.21    | 0.001 | -0.211 | 0.001 |
| dumy   | -0.367   | 0.001 | -0.494  | 0.001 | 1.38     | 0.001 | 1.38   | 0.001 |
| lag1   | -0.001   | 0.004 | -0.001  | 0.001 | 0.068    | 0.001 | 0.068  | 0.001 |

Semua Variabel berpengaruh signifikan dalam pemodelan. Hal ini dikarenakan nilai pvalue lebih kecil daripada nilai 0.05. Sedangkan pada regresi gamma dengan fungsi link “identity” dan “sqrt” semua variabel berpengaruh signifikan terhadap model. Prediksi dari pemodelan tersebut adalah sebagai berikut.



**Gambar 4.18.** Prediksi Curah Hujan Keputih dengan Variabel Temperatur, *dummy* dan lag1



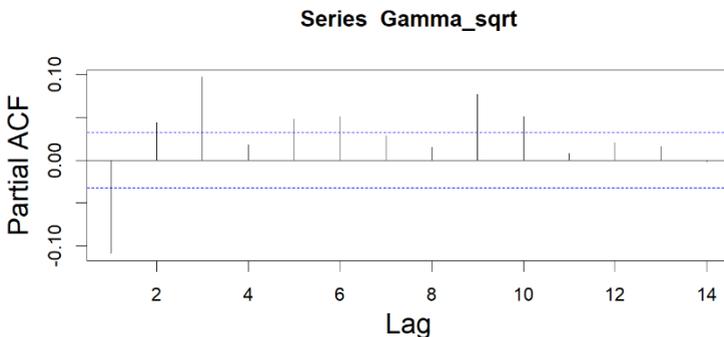
**Gambar 4.19.** Prediksi Curah Hujan Keputih dengan ketiga variabel Juanda, *Dummy* dan lag1

Prediksi yang terjadi dengan memodelkan variabel temperature dengan dummy dan lag 1 menggunakan regresi inverse Gaussian memiliki nilai RMSE yang paling besar yaitu

sebesar 84.009, disusul regresi Gamma dengan fungsi link “inverse”, “identity” dan “sqrt” yaitu 25.809, 7.65 dan 5.72. Secara visual dapat dilihat bahwa dengan metode regresi Gamma dengan fungsi link identity dan “sqrt” lebih mendekati nilai actual dan mampu menangkap prediksi yang lebih tinggi namun hanya beberapa prediksi saja. Sedangkan regresi Gamma dengan fungsi link “inverse” secara visual lebih baik daripada inverse Gaussian karena prediksinya lebih naik dan nilai RMSE nya juga lebih kecil.

Prediksi yang terjadi dengan memodelkan variabel temperatur, *Dummy* dan lag 1 menghasilkan nilai RMSE yang lebih kecil jika dibandingkan dengan hasil prediksi memodelkan ketiga variabel, *dummy* dengan lag1. Berdasarkan dari segi visual hasilnya lebih menyebar dan mampu menangkap prediksi yang jauh terhadap curah hujan. Perbedaan regresi gamma fungsi link “Identity” dan “sqrt” mampu menangkap prediksi yang ekstrim.

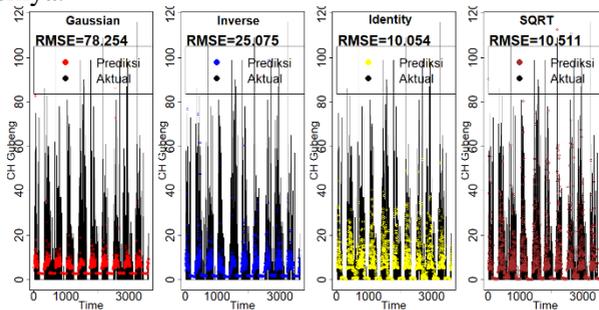
Penambahan jumlah lag diperlukan dalam pemodelan ini karena curah hujan tidak hanya dipengaruhi curah hujan satu hari saja (lag1). Untuk mengetahui curah hujan yang berpengaruh terhadap curah hujan saat ini dapat dilihat dari nilai PACF dari residual yang terbentuk dari model sebelumnya. Berikut merupakan hasil dari plot PACF residual dengan maksimal lag sebesar 14 dari pemodelan diatas untuk menentukan lag berapa saja yang signifikan.



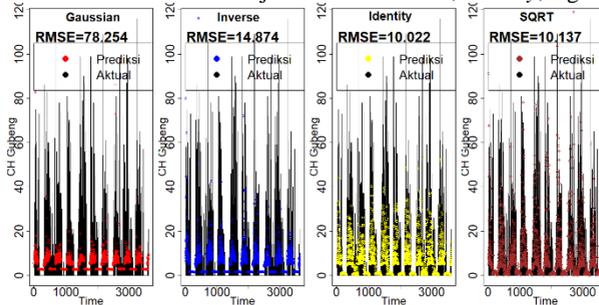
**Gambar 4.20** PACF Residual Gamma “sqrt”

Berdasarkan jumlah lag maksimal sebanyak 14 didapatkan hasil bahwa dengan regresi Gamma dengan fungsi link identity dan sqrt

hanya lag 1,2 dan lag 3 yang signifikan.. Berikut merupakan hasil prediksinya.



**Gambar 4.21.** Prediksi Curah Hujan Semua Variabel, *Dummy*, lag 1 dan lag2



**Gambar 4.22.** Prediksi Curah Hujan Temperatur, *Dummy*, lag 1 dan lag2

Secara visual terlihat bahwa setelah penambahan lag2 pada model ternyata menghasilkan prediksi yang turun. Jika dilihat berdasarkan nilai RMSE juga mengalami penurunan bila dibandingkan dengan pemodelan sebelumnya. Sehingga berdasarkan hasil prediksi hingga sampai dengan lag2 didapatkan dengan menggunakan variabel Temperatur, *dummy*, dan lag 1, metode Gamma dengan fungsi link “sqrt” menjadi yang terbaik diantara ketiga metode lainnya dengan nilai RMSE sebesar 5.72. Berikut merupakan hasil *summary* dari pemilihan model terbaik.

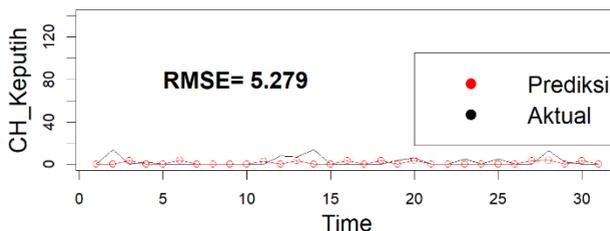
**Tabel 4.12** Pemodelan Curah Hujan Terbaik (Ketiga Variabel Juanda)

|                 | estimasi | Std.eror | pvalue |
|-----------------|----------|----------|--------|
| Intersep        | -3.639   | 0.185695 | 0.001  |
| Temperatur      | 0.089    | 0.005192 | 0.001  |
| Kelembapan      | 0.03     | 0.00151  | 0.001  |
| Kecapatan Angin | -0.052   | 0.011031 | 0.001  |
| <i>Dummy</i>    | 1.505    | 0.151092 | 0.001  |
| lag1            | 0.068    | 0.012304 | 0.001  |

Pemodelan secara individu menghasilkan keputusan bahwa semua variabel berpengaruh signifikan terhadap model. Hal ini dikarenakan nilai dari pvalue yang kurang dari alpha ( $\alpha = 0.05$ ). Hasil model yang terbentuk adalah sebagai berikut.

$$Y_{1,t} = (-3.639 + 0.089x_{1,t} + 0.03x_{2,t} - 0.052x_{3,t} + 1.505Dummy + 0.068Y_{1,t-1})^2$$

Pemodelan terbaik yaitu dalam pengujian data testing. Pemodelan dilakukan dengan membuat data *testing dummy* (1 dan 0) harian yang didapatkan dari jika pada setiap tanggal 1 Desember 2009 sampai dengan 1 Desember 2018 mengalami hujan minimal lima kali maka dikategorikan dengan 1 dan dilakukukan rata rata dari curah hujan. Sedangkan jika pada setiap tanggal 1 Desember 2009 sampai 1 Desember 2018 terdapat kurang dari 5 kali Hujan dikategorikan dengan 0. Berikut merupakan hasilnya.



**Gambar 4.23.** Prediksi Data Testing Juanda

Secara visual hasil prediksi yang dihasilkan tidak sesuai dengan pola data testing. Terlihat bahwa pola data testing lebih fluktuasi sedangkan pada data hasil prediksi lebih konstan. Berdasarkan hasil pengujian data testing yang berasal dari pemodelan data training yang memiliki RMSE sebesar 5.72 didapatkan nilai RMSE yang lebih kecil yaitu sebesar 5.279. Hal ini berarti model yang terbentuk dari data training sudah sangat baik yaitu dengan menggunakan regresi Gamma fungsi link “sqrt”. Secara visual hasil prediksi terlihat konstan namun jika dilihat dari nilai RMSE sangat kecil menunjukkan hasil model yang baik.

## II. Pemodelan Keputih dengan Stasiun Perak 1

Pemodelan yang dilakukan diawali dengan meregresikan variabel curah hujan dengan variabel dari stasiun perak 1 yaitu Temperatur, Kelembapan dan Kecepatan Angin. Metode yang digunakan adalah regresi inverse Gaussian, regresi gamma dengan

fungsi link “inverse”, regresi gamma dengan fungsi link “identity” dan “sqrt”. Pemodelan yang lainnya yaitu menambahkan variabel *dummy* dan variabel lag untuk hasil prediksi terbaik dengan kriteria kebaikan model menggunakan RMSE.

**Tabel 4.13** Summary CH Keputih 1 dengan Kelembapan (Perak1)

|        | Gaussian |       | inverse |       | identity |       | sqrt   |       |
|--------|----------|-------|---------|-------|----------|-------|--------|-------|
|        | est      | p-val | est     | p-val | est      | p-val | Est    | p-val |
| Interc | 0.357    | 0.002 | 1.362   | 0.002 | -9.48    | 0.002 | -4.584 | 0.002 |
| RH     | -0.004   | 0.002 | -0.015  | 0.002 | 0.186    | 0.002 | 0.089  | 0.002 |
| AIC    | -33225   |       | -30055  |       | -29953   |       | -30142 |       |

**Tabel 4.14** Summary CH Keputih dengan Kecepatan Angin (Perak1)

|        | Gaussian |       | inverse |       | identity |       | sqrt   |       |
|--------|----------|-------|---------|-------|----------|-------|--------|-------|
|        | est      | p-val | est     | p-val | est      | p-val | est    | p-val |
| Interc | 0.009    | 0.074 | 0.107   | 0.001 | 7.762    | 0.002 | 2.829  | 0.002 |
| FF     | 0.011    | 0.001 | 0.032   | 0.001 | -0.853   | 0.001 | -0.187 | 0.001 |
| AIC    | -33225   |       | -29845  |       | -29844   |       | -29844 |       |

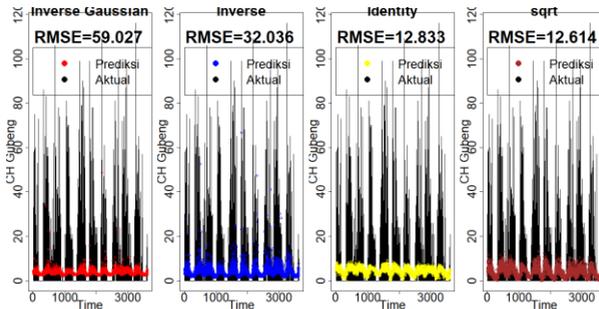
Berdasarkan hasil *summary* secara individu pemodelan antara curah hujan Perak 1 dengan masing masing variabel Perak 1 didapatkan bahwa secara individu variabel Temperature, kelembapan dan kecepatan angin berpengaruh signifikan terhadap model. Hal ini ditandai dengan nilai dari *pvalue* yang kurang dari 0.05 dengan menggunakan keempat metode tersebut. Pengaruh secara bersama akan dilakukan di pemodelan yang ditampilkan dalam **Tabel 4.15**.

**Tabel 4.15** Summary CH Keputih dengan variabel Stasiun Perak 1

|        | Gaussian |       | inverse |       | identity |       | sqrt   |       |
|--------|----------|-------|---------|-------|----------|-------|--------|-------|
|        | Est      | p-val | est     | p-val | est      | p-val | est    | p-val |
| Interc | 0.791    | 0.002 | 2.793   | 0.001 | -14.5    | 0.002 | -5.666 | 0.002 |
| Temp   | -0.01    | 0.002 | -0.033  | 0.001 | 0.059    | 0.002 | 0.024  | 0.002 |
| RH     | -0.006   | 0.002 | -0.021  | 0.001 | 0.217    | 0.002 | 0.089  | 0.002 |
| FF     | 0.001    | 0.002 | 0.003   | 0.005 | 0.415    | 0.002 | 0.103  | 0.002 |
| AIC    | -33221   |       | -30082  |       | -30020   |       | -30179 |       |

Hasil pemodelan dengan regresi inverse Gaussian, Gamma dengan link “inverse”, regresi Gamma dengan link “Identity”, dan regresi Gamma dengan “sqrt” menunjukkan hasil yang signifikan untuk semua variabel. Hal ini dikarenakan nilai p-value kurang dari 0.05. Artinya ketiga variabel dari stasiun Perak 1 benar benar berpengaruh signifikan terhadap model yang terbentuk. Hasil

estimasi dari curah hujan dan nilai kuadrat error ditampilkan dalam **Gambar 4.24**.



**Gambar 4.24** Prediksi Curah Hujan Keputih dengan tiga Variabel

Berdasarkan plot hasil prediksi diatas dengan menggunakan regresi inverse Gaussian didapatkan nilai *Root Mean Square Error* (RMSE) terbesar yaitu sebesar 59.62. Untuk regresi gamma dengan fungsi link “inverse”, “identity” dan “sqrt” dihasilkan nilai RMSE masing masing sebesar 32.8, 12.833 dan 12.614. Jadi dengan menggunakan ketiga variabel metode regresi gamma dengan “sqrt” menjadi yang terbaik. Secara visual terlihat dengan memasukkan tiga variabel kedalam model ternyata tidak menaikkan hasil prediksi. Prediksi yang dihasilkan masih jauh dari harapan atau curah hujan aktual. Dugaan adanya kasus multikolinieritas dapat dilihat dengan melihat nilai VIF (*Variance Inflation Factor*), jika nilai VIF lebih dari 10 mengindikasikan adanya multikolinieritas. Berikut merupakan hasil pengujian multikolinieritas.

**Tabel 4.16** VIF untuk Semua Model (Perak 1)

|                  | VIF ( <i>Variance Inflation Factor</i> ) |            |             |
|------------------|--|------------|-------------|
|                  | Temperatur                               | Kelembapan | Kecep_Angin |
| Inverse Gaussian | 4035.16                                  | 4035.74    | 46.17       |
| Gamma(inverse)   | 3.07                                     | 3.10       | 1.02        |
| Gamma(identity)  | 74512.76                                 | 1120411.06 | 1269263.64  |
| Gamma(sqrt)      | 1.350668                                 | 2.602062   | 2.238671    |

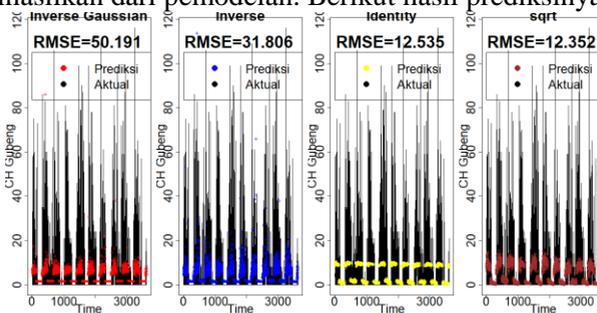
Berdasarkan **Tabel 4.16** dengan menggunakan metode inverse Gaussian dan regresi gamma dengan fungsi link identity mengindikasikan adanya kasus multikolinieritas. Hal ini dapat di dibuktikan dengan nilai VIF yang melebihi 10 pada semua

variabel. Sedangkan VIF untuk semua variabel pada regresi gamma dengan fungsi link “inverse” dan “sqrt” kurang dari 10 dan tidak terjadi kasus multikolinieritas. Adanya hubungan yang erat diantara variabel yang menyebabkan hail prediksi menjadi tidak baik dan juga model yang dihasilkan kurang bagus. Sehingga pada kasus multikolinieritas tersebut dilakukan penanganan dengan melakukan analisis komponen utama (PCA) dan mengelompokkannya. Berikut merupakan hasil dari PCA.

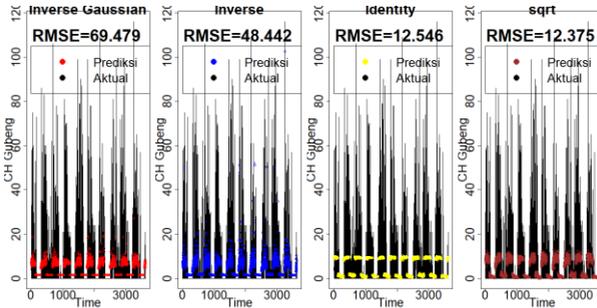
**Tabel 4.17** Analisis Komponen Utama (Perak 1)

|                        | PC1    | PC2    | PC3     |
|------------------------|--------|--------|---------|
| Standard deviation     | 7.1239 | 1.2149 | 0.95914 |
| Proportion of Variance | 0.9549 | 0.0278 | 0.01731 |
| Cumulative Proportion  | 0.9549 | 0.9827 | 1       |

Tabel diatas menginformasikan bahwa PC1 sudah cukup untuk mewakili kelompok dari variabel prediktor dimana pada PC1 sudah memuat semua informasi pada variabel prediktor dengan nilai cumulative proportion sebesar 0.9549 atau 95.49%. Pemodelan yang dilakukan ada dua yaitu curah hujan dengan variabel hasil analisis komponen utama untuk regresi inverse Gaussian dan regresi gamma dengan fungsi link identity serta pemodelan curah hujan dengan semua variabel untuk regresi gamma dengan fungsi link Inverse. Karena ada kondisi dimana terdapat bulan bulan hujan dan tidak maka digunakan variabel *dummy* dalam pemodelan dengan tujuan untuk mengururangi error yang dihasilkan dari pemodelan. Berikut hasil prediksinya.



**Gambar 4.25** Prediksi Curah Hujan Keputih dengan tiga Variabel dan *dummy*



**Gambar 4.26** Prediksi Curah Hujan Keputih PCI (Perak1) dan *Dummy*

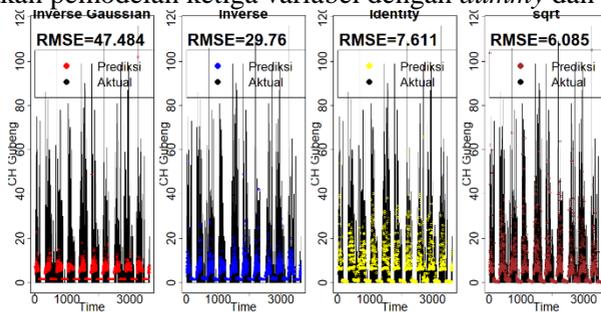
Perbandingan hasil prediksi kedua gambar diatas Berdasarkan nilai RMSE dengan menggunakan ketiga variabel perak 1 nilai RMSE lebih kecil dan hasil prediksi lebih meningkat atau sesuai dengan nilai dari observasi data actual curah hujan. Pada penggunaan metode regresi inverse Gaussian dan regresi Gamma dengan inverse tidak bisa menaikkan hasil prediksi. Berbeda halnya dengan kedua metode lainnya yaitu regresi Gamma “Identity” dan Gamma “sqrt” yang prediksinya yaitu bisa menangkap prediksi yang ekstrim. Berikut merupakan hasil summary dari pemodelan menggunakan ketiga variabel Perak 1.

**Tabel 4.18** Summary Curah Hujan dengan Ketiga Variabel dan *dummy*

|        | Gaussian |       | inverse |       | identity |       | sqrt   |       |
|--------|----------|-------|---------|-------|----------|-------|--------|-------|
|        | est      | p-val | est     | p-val | est      | p-val | est    | p-val |
| Interc | 0.64     | 0.001 | 2.181   | 0.001 | -4.894   | 0.002 | -3.915 | 0.002 |
| Temp   | -0.004   | 0.003 | -0.02   | 0.001 | 0.019    | 0.002 | 0.015  | 0.002 |
| RH     | -0.003   | 0.002 | -0.014  | 0.001 | 0.069    | 0.002 | 0.056  | 0.002 |
| Angi   | 0.001    | 0.603 | -0.002  | 0.709 | 0.203    | 0.002 | 0.163  | 0.002 |
| dumy   | -0.344   | 0.001 | -0.381  | 0.001 | 7.722    | 0.002 | 1.715  | 0.002 |
| AIC    | -33219   |       | -30171  |       | -30246   |       | -30149 |       |

Pemodelan yang dihasilkan setelah menambahkan variabel *dummy* ternyata mengakibatkan variabel kecepatan angin tidak signifikan terhadap model dengan menggunakan metode Inverse Gaussian dan regresi gamma dengan fungsi link Inverse. Hal ini dikarenakan pvalue yang dihasilkan lebih dari 0.05. Sedangkan jika menggunakan regresi Gamma dengan fungsi link “identity” dan “sqrt” semua variabel berpengaruh signifikan terhadap model.

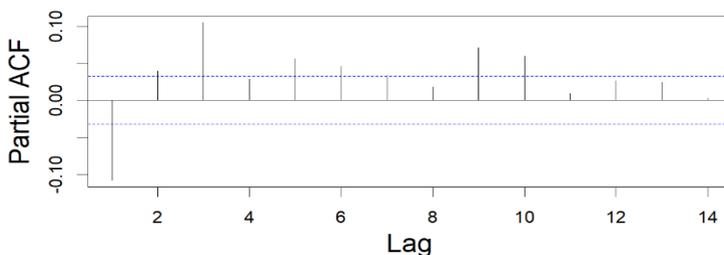
Prediksi yang didapatkan dari pemodelan tersebut masih tidak bisa untuk menangkap nilai aktual yang tinggi. Oleh karena itu diperlukan variabel tambahan yang bisa meningkatkan model dinamakan dengan variabel lag. Alasan utama ditambahkan variabel lag adalah curah hujan yang terjadi di waktu tertentu dipengaruhi oleh curah hujan yang sebelumnya. Berikut merupakan pemodelan ketiga variabel dengan *dummy* dan lag1.



**Gambar 4.27** Prediksi Curah Hujan Keputih dengan ketiga variabel (Perak1), lag1 dan *Dummy*

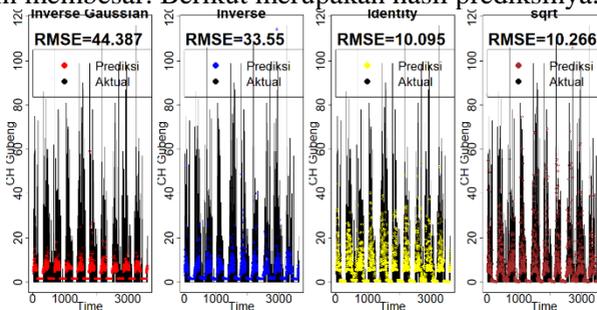
Penambahan variabel lag mampu menangkap prediksi yang *extreme*. Secara visual dengan menambah variabel lag1 membuat hasil prediksi menjadi lebih baik dari pada pemodelan tanpa lag. Jika dilihat dari RMSE dengan menggunakan regresi Gamma dengan fungsi link “sqrt” mengalami penurunan yaitu yang awalnya 12.352 menjadi 6.085. Untuk mengetahui curah hujan yang berpengaruh terhadap curah hujan dapat dilihat nilai PACF dari residual yang terbentuk dari model. Berikut hasil plot PACF residual dengan maksimal lag sebesar 14.

**Series Gamma\_sqrt**



**Gambar 4.28** PACF Residual Gamma “sqrt”

Berdasarkan jumlah lag maksimal sebanyak 14 didapatkan hasil bahwa dengan metode inverse Gaussian dan regresi Gamma dengan fungsi link inverse menunjukkan lag yang signifikan berpengaruh terhadap model dan hasil prediksi adalah lag 1, lag 2, lag 3, lag 4, lag 5 dan lag 6. Sedangkan pada regresi Gamma dengan fungsi link “identity” dan “sqrt” lag1,2, 3 dan lag 5. Variabel lag tidak berpengaruh signifikan jika dari hasil pemodelan didapatkan nilai RMSE yang tidak berubah atau semakin membesar. Berikut merupakan hasil prediksinya.



**Gambar 4.29** Prediksi Curah Hujan Keputih dengan ketiga variabel (Perak1), lag1, lag2 dan *Dummy*

Secara visual terlihat bahwa setelah penambahan lag2 pada model ternyata menghasilkan prediksi yang turun. Pemodelan dengan lag1 memiliki RMSE 6.085 sedangkan pemodelan sampai dengan lag2 memiliki RMSE sebesar 10.266. Sehingga berdasarkan hasil prediksi hingga sampai dengan lag2 didapatkan bahwa dengan menggunakan ketiga variabel, *dummy*, dan lag1, metode Regresi Gamma dengan fungsi link “sqrt” menjadi yang terbaik diantara ketiga metode lainnya dengan nilai RMSE sebesar 6.085. Berikut merupakan hasil summary dari model terbaik.

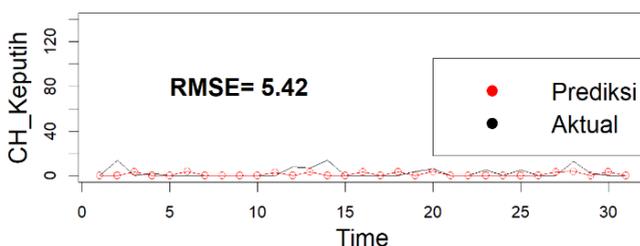
**Tabel 4.19** Summary Curah Hujan dengan Ketiga Variabel *dummy* dan lag1

|                 | estimasi | Std.error | pvalue |
|-----------------|----------|-----------|--------|
| Intersep        | -2.961   | 0.289     | 0.001  |
| Temperatur      | 0.012    | 0.005     | 0.01   |
| Kelembapan      | 0.044    | 0.003     | 0.001  |
| Kecapatan Angin | 0.094    | 0.024     | 0.001  |
| <i>Dummy</i>    | 1.415    | 0.147     | 0.001  |
| lag1            | 0.064    | 0.013     | 0.001  |

Hasil *summary* berdasarkan model terbaik dihasilkan semua variabel berpengaruh signifikan terhadap model. Hal ini dikarenakan hasil nilai signifikansi uji (*pvalue*) kurang dari 0.05. berikut merupakan hasil pemodelannya.

$$Y_{1,t} = (-2.961 + 0.012x_{1,t} + 0.044x_{2,t} - 0.094x_{3,t} + 1.405Dummy + 0.064Y_{1,t-1})^2$$

Hasil akhir dari pemodelan terbaik yaitu dalam pengujian dari data testing. Data testing berjumlah 31 data yaitu merupakan tanggal 1 Desember 2018 sampai dengan 31 Desember 2018. Berikut merupakan hasil pengujian dari data *testing*.



**Gambar 4.30** Prediksi Data Testing (Perak 1)

Hasil pengujian data testing yang berasal dari pemodelan data training yang memiliki RMSE sebesar 6.085 didapatkan nilai RMSE yang lebih kecil yaitu sebesar 5.42. Hal ini berarti model yang terbentuk dari data training sudah sangat baik yaitu dengan menggunakan regresi Gamma fungsi link “sqrt”. Secara visual hasil prediksi terlihat konstan namun jika dilihat dari nilai RMSE sangat kecil menunjukkan hasil model yang baik.

### III. Pemodelan Keputih dengan Stasiun Perak 2

Pemodelan yang dilakukan diawali dengan meregresikan variabel curah hujan dengan variabel dari stasiun perak 2 yaitu Temperatur rata rata, Kelembapan rata rata dan Kecepatan Angin rata rata. Berikut merupakan hasil pemodelannya.

**Tabel 4.20** Summary CH Keputih dengan Temperatur (Perak2)

|        | Gaussian |       | inverse |       | identity |       | sqrt   |       |
|--------|----------|-------|---------|-------|----------|-------|--------|-------|
|        | est      | p-val | est     | p-val | est      | p-val | est    | p-val |
| Interc | -0.3     | 0.002 | -1.427  | 0.002 | 33.87    | 0.002 | 15.638 | 0.002 |
| Temp   | 0.012    | 0.002 | 0.057   | 0.002 | -1.02    | 0.002 | -0.461 | 0.002 |

**Tabel 4.21** Summary CH Keputih dengan Kelembapan (Perak2)

|        | Gaussian |       | inverse |       | identity |       | sqrt   |       |
|--------|----------|-------|---------|-------|----------|-------|--------|-------|
|        | est      | p-val | est     | p-val | est      | p-val | est    | p-val |
| Interc | 0.284    | 0.002 | 1.161   | 0.002 | -10.57   | 0.002 | -4.584 | 0.002 |
| RH     | -0.029   | 0.002 | -0.012  | 0.002 | 0.203    | 0.002 | 0.089  | 0.002 |
| AIC    | -33225   |       | -30011  |       | -30002   |       | -30142 |       |

**Tabel 4.22** Summary CH Keputih dengan Kecepatan Angin (Perak2)

|        | Gaussian |       | inverse |       | identity |       | sqrt   |       |
|--------|----------|-------|---------|-------|----------|-------|--------|-------|
|        | est      | p-val | est     | p-val | est      | p-val | est    | p-val |
| Interc | 0.003    | 0.334 | 0.074   | 0.009 | 7.695    | 0.002 | 3.333  | 0.002 |
| Angin  | 0.014    | 0.002 | 0.046   | 0.002 | -0.855   | 0.002 | -0.37  | 0.002 |
| AIC    | -33225   |       | -29866  |       | -29872   |       | -29884 |       |

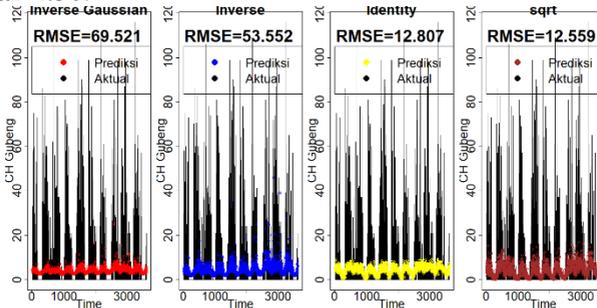
Berdasarkan hasil *summary* secara individu pemodelan antara curah hujan Perak 1 dengan masing masing variabel Perak 1 didapatkan bahwa secara individu variabel Temperature, kelembapan dan kecepatan angin berpengaruh signifikan terhadap model. Hal ini ditandai dengan nilai dari pvalue yang kurang dari 0.05 dengan menggunakan ketiga metode tersebut. Pengaruh secara bersama akan dilakukan di pemodelan yang ditampilkan dalam **Tabel 4.23**.

**Tabel 4.23** Summary CH Keputih dengan variabel Stasiun Perak 2

|        | Gaussian |       | inverse |       | identity |       | sqrt   |       |
|--------|----------|-------|---------|-------|----------|-------|--------|-------|
|        | Est      | p-val | est     | p-val | est      | p-val | Est    | p-val |
| Interc | 1.092    | 0.002 | 3.333   | 0.002 | -11.44   | 0.002 | -5.016 | 0.002 |
| Temp   | -0.02    | 0.002 | -0.054  | 0.002 | 0.035    | 0.002 | 0.016  | 0.002 |
| RH     | -0.007   | 0.002 | -0.02   | 0.002 | 0.214    | 0.002 | 0.094  | 0.002 |
| Angin  | -0.004   | 0.002 | -0.01   | 0.09  | -0.371   | 0.002 | -0.163 | 0.002 |
| AIC    | -33221   |       | -30023  |       | -30047   |       | -32022 |       |

Hasil pemodelan dengan regresi inverse Gaussian dan regresi Gamma dengan link “Identity” dan “sqrt” menunjukkan hasil yang signifikan untuk semua variabel. Hal ini dikarenakan nilai p-value kurang dari 0.05. Sedangkan dengan menggunakan Gamma dengan link “inverse” didapatkan bahwa variabel Kecepatan angin tidak berpengaruh signifikan terhadap model. Artinya kedua variabel dari stasiun Perak 2 benar benar berpengaruh signifikan terhadap model yang terbentuk. Hasil

estimasi dari curah hujan dan nilai kuadrat error ditampilkan dalam **Gambar 4.30**.



**Gambar 4.30** Prediksi Curah Hujan Keputih dengan tiga Variabel

**Gambar 4.30** menunjukkan plot hasil prediksi dengan menggunakan regresi inverse Gaussian didapatkan nilai *Root Mean Square Error* (RMSE) terbesar yaitu sebesar 69.521. Untuk regresi gamma dengan fungsi link inverse, identity dan sqrt dihasilkan nilai RMSE masing masing sebesar 53.552, 12.807 dan 12.559. Secara visual terlihat dengan memasukkan tiga variabel kedalam model ternyata tidak menaikkan hasil prediksi. Prediksi yang terjadi tidak mampu menangkap curah hujan yang tinggi. Dugaan adanya kasus multikolinieritas dapat dilihat dengan melihat nilai VIF (*Variance Inflation Factor*), jika nilai VIF lebih dari 10 mengindikasikan adanya multikolinieritas.

**Tabel 4.24** VIF untuk Semua Model (Perak 2)

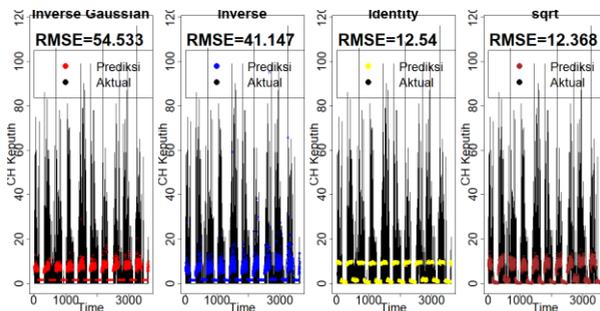
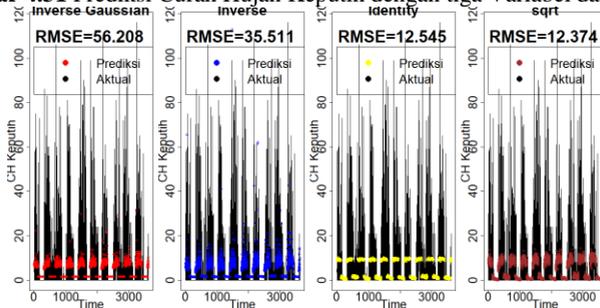
|                  | VIF ( <i>Variance Inflation Factor</i> ) |            |             |
|------------------|--|------------|-------------|
|                  | Temperatur                               | Kelembapan | Kecep_Angin |
| Inverse Gaussian | 2804.03769                               | 3178.52722 | 20.82436    |
| Gamma(inverse)   | 9.125438                                 | 11.340604  | 2.078728    |
| Gamma(identity)  | 21282.34                                 | 3107150.3  | 3029683     |
| Gamma(sqrt)      | 1.080729                                 | 8.789274   | 8.616122    |

Berdasarkan Tabel 4.24 didapatkan hasil bahwa dengan menggunakan ketiga metode yaitu inverse Gaussian, regresi Gamma dengan fungsi link “inverse” dan “identity” terdapat indikasi adanya multikolinieritas. Hal terjadi karena nilai VIF dari semua variabel lebih dari 10. Sehingga dilakukan pengelompokan dengan menggunakan PCA untuk menangani kasus multikolinieritas. Berikut merupakan hasil dari PCA.

**Tabel 4.25** Analisis Komponen Utama (Perak 2)

|                        | PC1    | PC2     | PC3     |
|------------------------|--------|---------|---------|
| Standard deviation     | 6.8376 | 1.27106 | 0.93771 |
| Proportion of Variance | 0.9493 | 0.03281 | 0.01785 |
| Cumulative Proportion  | 0.9493 | 0.98215 | 1       |

Berdasarkan tabel diatas didapatkan bahwa PC1 sudah cukup untuk mewakili kelompok dari variabel prediktor dimana pada PC1 sudah memuat semua informasi pada variabel prediktor dengan nilai cumulative proportion sebesar 0.9493 atau 94.93%. Pemodelan yang dilakukan ada dua yaitu curah hujan dengan variabel PC1 perak dan curah hujan dengan melibatkan seluruh variabel prediktor. Dikarenakan secara teori ketiga variabel tersebut memengaruhi curah hujan, maka variabel tersebut tetap dimasukkan kedalam pemodelan. Karena ada kondisi dimana terdapat bulan bulan hujan dan tidak maka digunakan variabel *dummy* dalam pemodelan dengan tujuan untuk mengurugi error yang dihasilkan dari pemodelan. Berikut hasil prediksinya.

**Gambar 4.31** Prediksi Curah Hujan Keputih dengan tiga Variabel dan *dummy*

**Gambar 4.32** Prediksi Curah Hujan Keputih PC1 dan *Dummy*

Berdasarkan perbandingan nilai RMSE didapatkan bahwa dengan menggunakan semua variabel perak 2 memiliki kebaikan model yang baik dengan RMSE yang lebih kecil yaitu sebesar 12.368 sedangkan dengan menggunakan variabel PC dan *dummy* memiliki RMSE sebesar 12.374 yaitu pada regresi Gamma dengan menggunakan fungsi link “sqrt”. Secara visual hasil prediksi tidak dapat menjangkau nilai actual yang sangat tinggi pada data curah hujan. Karena hasil dari RMSE tidak berbeda jauh antara kedua model maka kedua model tersebut digunakan dalam pemodelan bersama lag. Berikut merupakan hasil summary dari pemodelan menggunakan ketiga variabel Perak 2 dan PC dengan *Dummy*.

**Tabel 4.26** Pemodelan curah Hujan dengan PC (Perak 2)

|       | Gaussian |       | inverse |       | identity |       | sqrt   |       |
|-------|----------|-------|---------|-------|----------|-------|--------|-------|
|       | est      | p-val | est     | p-val | est      | p-val | est    | p-val |
| Inter | 0.379    | 0.001 | 0.6     | 0.002 | 1.289    | 0.002 | 1.039  | 0.002 |
| PC    | 0.002    | 0.005 | 0.008   | 0.002 | -0.058   | 0.002 | -0.047 | 0.002 |
| dumy  | -0.358   | 0.001 | -0.439  | 0.002 | 7.847    | 0.002 | 1.781  | 0.002 |
| AIC   | -33223   |       | -30014  |       | -30236   |       | -30399 |       |

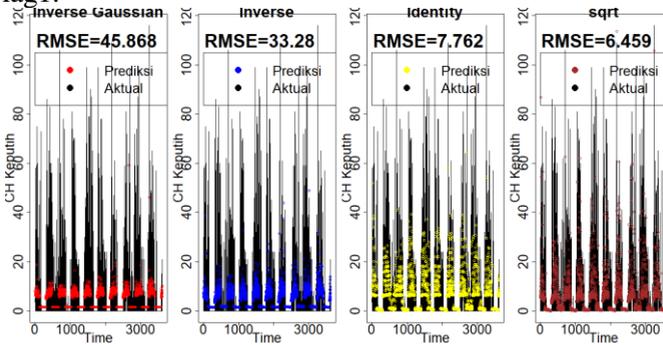
Tabel diatas menginformasikan bahwa variabel hasil analisis komponen utama (PC) berpengaruh signifikan terhadap model begitu juga dengan variabel *dummy*. Hal ini dikarenakan nilai p-vlaue kurang dari 0.05.

**Tabel 4.27** Pemodelan curah Hujan dengan Ketiga Variabel (Perak 2)

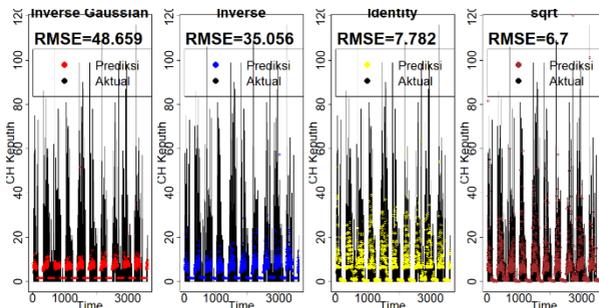
|        | Gaussian |       | inverse |       | identity |       | sqrt   |       |
|--------|----------|-------|---------|-------|----------|-------|--------|-------|
|        | est      | p-val | est     | p-val | est      | p-val | est    | p-val |
| Interc | 0.582    | 0.095 | 1.807   | 0.001 | -3.133   | 0.002 | -2.549 | 0.002 |
| Temp   | -0.003   | 0.727 | -0.016  | 0.223 | 0.01     | 0.002 | 0.008  | 0.224 |
| RH     | -0.002   | 0.253 | -0.011  | 0.001 | 0.059    | 0.002 | 0.048  | 0.002 |
| FF     | -0.001   | 0.485 | -0.001  | 0.914 | -0.102   | 0.002 | -0.083 | 0.002 |
| dumy   | -0.355   | 0.001 | -0.427  | 0.002 | 7.891    | 0.002 | 1.804  | 0.002 |

Perbandingan yang ditujukan antara pemodelan curah hujan dengan seluruh variabel perak dengan *dummy* menghasilkan semua variabel tidak signifikan kecuali *dummy* pada metode inverse Gaussian. Sedangkan dengan metode Gamma dengan fungsi link Inverse hanya variabel Kelembapan dan *dummy* yang signifikan. Pemodelan dengan menggunakan regresi Gamma fungsi link “identity” menunjukkan hasil yang signifikan untuk semua variabel. Sedangkan dengan “sqrt” variabel tidak

berpengaruh signifikan. Hal ini menunjukkan bahwa variabel perak 2 kurang memberikan kontribusi terhadap model pada regresi inverse Gaussian dan gamma “inverse”. Sehingga pemodelan dengan menggunakan PC dan *dummy* dipakai dalam pemodelan selanjutnya dengan lag. Berikut pemodelan PC *dummy* dan lag1.

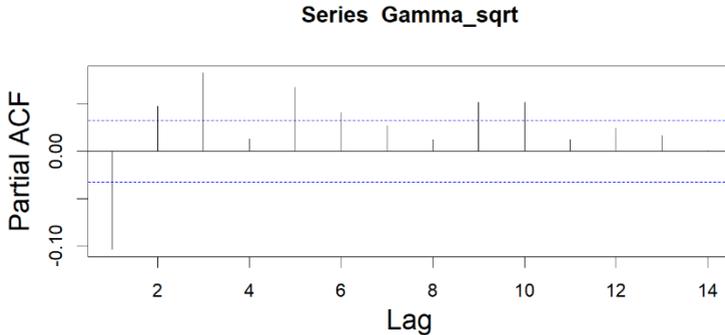


**Gambar 4.33** Prediksi Curah Hujan Keputih, PC1, lag1 dan Dummy



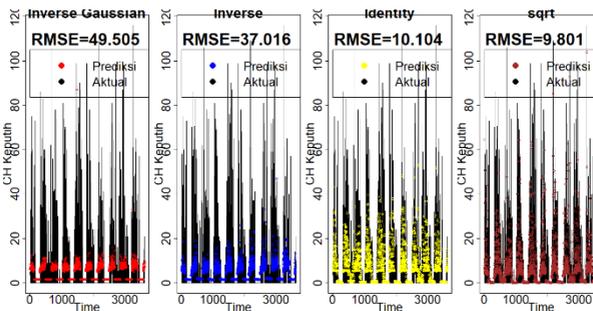
**Gambar 4.34** Prediksi Curah Hujan Keputih variabel Perak 2, lag1 dan *Dummy*

Penambahan lag1 pada metode regresi Gamma “sqrt” memberikan hasil yang bagus bagi pemodelan daripada tanpa adanya lag1. Karena dengan menambahkan lag mampu menangkap prediksi yang sifatnya ekstrim. Pada perbandingan dengan variabel PC dan lag1 dengan tiga variabel (perak 2) memberikan hasil dengan menggunakan PC ternyata memberikan hasil lebih baik daripada dengan menggunakan ketiga variabel. Hal ini dibuktikan dengan nilai RMSE sebesar 6.459.



**Gambar 4.35** PACF Residual Gamma “sqrt”

Berdasarkan jumlah lag maksimal sebanyak 14 didapatkan hasil bahwa dengan metode inverse Gaussian dan regresi Gamma dengan fungsi link inverse menunjukkan lag yang signifikan berpengaruh terhadap model dan hasil prediksi adalah lag 1, lag 2, lag 3, lag 4, lag 5. Sedangkan pada regresi Gamma dengan fungsi link “identity” dan “sqrt” hanya lag 3 dan lag 5. Variabel lag tidak berpengaruh signifikan jika hasil pemodelan didapatkan nilai RMSE yang tidak semakin membesar. Berikut merupakan hasil prediksinya.



**Gambar 4.36** Prediksi Curah Hujan variabel Perak 2, lag1, lag2 dan *Dummy*

Secara visual terlihat bahwa setelah penambahan lag2 pada model ternyata menghasilkan prediksi yang turun. Jika dilihat berdasarkan nilai RMSE juga mengalami penurunan bila dibandingkan dengan pemodelan sebelumnya. Pemodelan dengan lag1 memiliki RMSE 6.459 sedangkan pemodelan sampai dengan lag2 memiliki RMSE sebesar 9.801. Sehingga berdasarkan hasil

prediksi yang telah dilakukan hingga sampai dengan lag2 didapatkan bahwa dengan menggunakan ketiga variabel, *dummy*, dan lag1, metode Regresi Gamma dengan fungsi link “sqrt” menjadi yang terbaik diantara ketiga metode lainnya dengan nilai RMSE sebesar 6.459. Berikut merupakan hasil summary dari model terbaik yang dipilih.

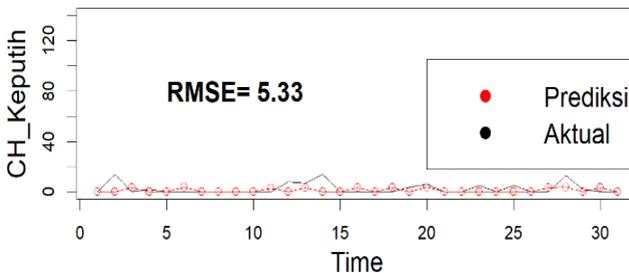
**Tabel 4.28** Summary Model Terbaik (Perak2)

|              | estimasi | Std.eror | pvalue |
|--------------|----------|----------|--------|
| Intersep     | 0.88174  | 0.0486   | 0.002  |
| PC1          | -0.03794 | 0.004314 | 0.002  |
| <i>dummy</i> | 1.482376 | 0.151093 | 0.002  |
| lag1         | 0.060604 | 0.012016 | 0.002  |

Hasil *summary* berdasarkan model terbaik dihasilkan semua variabel berpengaruh signifikan terhadap model. Hal ini dikarenakan hasil nilai signifikansi uji (*pvalue*) kurang dari 0.05. Berikut merupakan hasil pemodelannya.

$$Y_{1,t} = (-0.082 - 0.037PC_{1,t} + 1.48Dummy + 0.06Y_{1,t-1})^2$$

Hasil akhir dari pemodelan terbaik yaitu dalam pengujian dari data testing. Data testing berjumlah 31 data yaitu merupakan tanggal 1 Desember 2018 sampai dengan 31 Desember 2018. Berikut merupakan hasil pengujian dari data *testing*.



**Gambar 4.37** Pemodelan Data Testing (perak 2)

Pengujian data testing yang berasal dari pemodelan data training yang memiliki RMSE sebesar 6.459 didapatkan nilai RMSE yang lebih kecil yaitu sebesar 4.334. Hal ini berarti model yang terbentuk dari data training sudah sangat baik yaitu dengan menggunakan regresi Gamma fungsi link “sqrt”. Secara visual

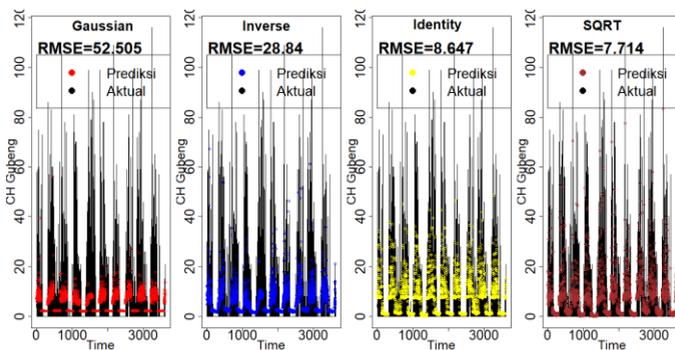
hasil prediksi terlihat konstan namun jika dilihat dari nilai RMSE sangat kecil menunjukkan hasil model yang baik.

#### 4.5.2 Pemodelan Curah Hujan Gubeng

Variabel curah hujan Gubeng sebagai respon diregresikan dengan variabel stasiun BMKG sebagai prediktor. Sebagai prediktor tambahan yaitu variabel *dummy* merupakan variabel yang berdasarkan dari statistika deskriptif didapatkan untuk hari hujan yang dikategorikan dengan yaitu pada bulan November sampai dengan April sedangkan pada bulan Mei sampai dengan Oktober merupakan hari hari yang jarang turun hujan dikategorikan dengan angka 0. Serta variabel lag yang merupakan variabel  $Y_{t-k}$  yang digunakan untuk meningkatkan model.

#### I. Pemodelan Curah Hujan Gubeng dengan Stasiun Juanda

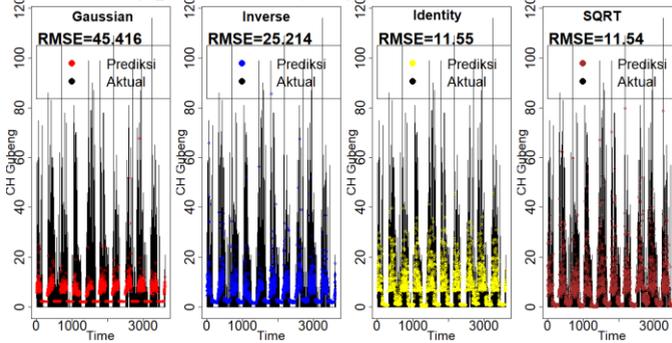
Pemodelan dari curah hujan Gubeng yang diawali dengan meregresikan variabel respon dengan masing masing variabel prediktor dengan regresi inverse Gaussian (link="1/ $\mu^2$ "), Gamma (link=inverse), Gamma (link=identity) dan Gamma dengan fungsi "sqrt". Pemodelan bertahap pada stasiun Gubeng dengan menambahkan variabel *dummy*, analisis komponen utama dan variabel lag terdapat pada **lampiran 12**. Berikut merupakan pemodelan terbaik.



**Gambar 4.38** Estimasi Curah Hujan Gubeng dengan variabel *dummy* dan *lag* 1

Pemodelan dengan penambahan variabel *lag* 1 berpengaruh signifikan terhadap peningkatan hasil prediksi seperti terlihat pada gambar diatas. Nilai RMSE juga semakin turun yaitu sebesar 7.714 pada metode regresi Gamma "sqrt". Penambahan *lag* 1 mampu

meningkatkan prediksi yang cukup tinggi bila dibandingkan dengan pemodelan tanpa lag1. Berikut merupakan pemodelan dari penambahan lag pada model yang terbentuk diatas.



**Gambar 4.39** Estimasi Curah Hujan Gubeng dengan *dummy*, *lag1* dan *lag2*

Penambahan *lag2* pada model ternyata menghasilkan prediksi yang turun. Jika dilihat berdasarkan nilai RMSE juga mengalami penurunan bila dibandingkan dengan pemodelan sebelumnya. Pemodelan dengan *lag1* memiliki RMSE 7.144 sedangkan pemodelan sampai dengan *lag2* memiliki RMSE sebesar 11.54. Sehingga berdasarkan hasil prediksi yang telah dilakukan hingga sampai dengan *lag2* didapatkan bahwa dengan menggunakan ketiga variabel, *dummy*, dan *lag1*, metode Regresi Gamma dengan fungsi link “sqrt” menjadi yang terbaik diantara ketiga metode lainnya dengan nilai RMSE sebesar 7.144. Berikut merupakan hasil summary dari model terbaik.

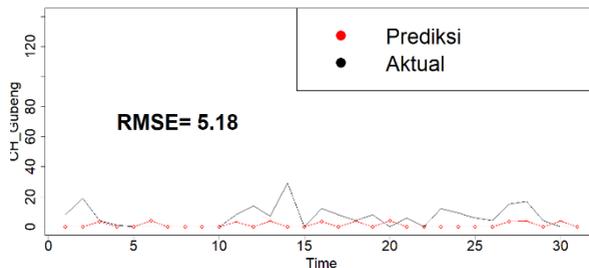
**Tabel 4.29** Summary Model Terbaik (Juanda)

|                 | Estimasi | Std.error | pvalue |
|-----------------|----------|-----------|--------|
| Intersep        | -5.669   | 0.336     | 0.001  |
| Temperatur      | 0.172    | 0.01      | 0.001  |
| Kelembapan      | 0.032    | 0.002     | 0.001  |
| Kecapatan Angin | -0.12    | 0.007     | 0.001  |
| Dummy           | 1.455    | 0.135     | 0.001  |
| lag1            | 0.054    | 0.01      | 0.001  |

Hasil *summary* berdasarkan model terbaik dihasilkan semua variabel berpengaruh signifikan terhadap model. Hal ini dikarenakan hasil nilai signifikansi uji (*p-value*) kurang dari 0.05. berikut merupakan hasil pemodelannya.

$$Y_{1,t} = (-5.669 + 0.172x_{1,t} + 0.032x_{2,t} - 0.012x_{3,t} + 1.455Dummy + 0.054Y_{1,t-1})^2$$

Pemodelan terbaik dari data training digunakann untuk memprediksi data testing berjumlah 31 data yaitu merupakan tanggal 1 Desember 2018 sampai dengan 31 Desember 2018. Berikut merupakan hasil pengujian dari data *testing*.

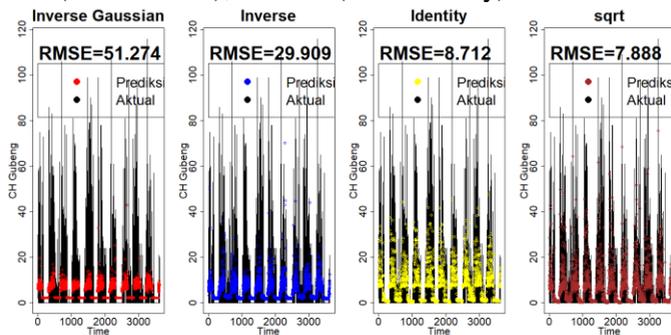


**Gambar 4.40** Prediksi Data Testing (Juanda)

Secara visual hasil prediksi terlihat konstan dan pola yang dihasilkan antara data testing dan prediksinya kurang menghasilkan ketepatan. Hasil pengujian data testing memiliki RMSE sebesar 5.18 lebih kecil dibandingkan dengan RMSE data training yaitu sebesar 7.144. Artinya model yang terbentuk dari data training sudah sangat baik yaitu dengan menggunakan regresi Gamma fungsi link “sqrt”.

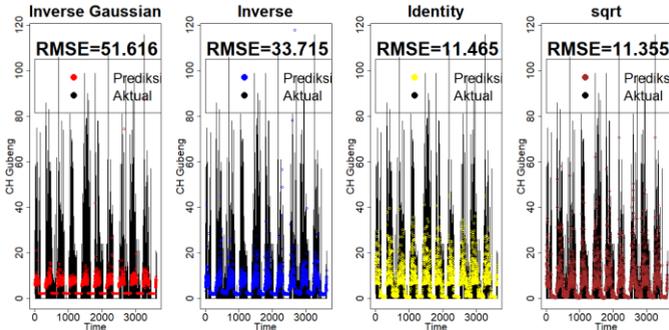
## II. Pemodelan Curah Hujan Gubeng dengan Perak 1

Pemodelan dari curah hujan Gubeng yang diawali dengan meregresikan variabel respon dengan masing masing variabel prediktor dengan regresi inverse Gaussian (link= $1/\mu^2$ ), Gamma (link=inverse), Gamma (link=identity) dan Gamma “sqrt”.



**Gambar 4.41** Estimasi CH Gubeng dengan ketiga variabel, *dummy* dan *lag1*

Penambahan *lag1* mampu meningkatkan prediksi yang cukup tinggi bila dibandingkan dengan pemodelan tanpa *lag1*. Nilai RMSE juga semakin turun yaitu sebesar 7.888 pada metode regresi Gamma “sqrt”. Berikut merupakan pemodelan dari penambahan *lag2* pada model yang terbentuk diatas.



**Gambar 4.42** Estimasi CH dengan Ketiga Variabel, *dummy*, *lag1* dan *lag2*

Penambahan *lag2* pada model menghasilkan prediksi yang semakin baik. Pemodelan dengan *lag1* pada Gamma “sqrt” memiliki RMSE sebesar 7.888. Pemodelan sampai dengan *lag2* Gamma “sqrt” memiliki RMSE sebesar 11.355. Sehingga berdasarkan hasil prediksi hingga sampai dengan *lag2* didapatkan bahwa dengan menggunakan ketiga variabel, *dummy*, dan *lag1*, metode Regresi Gamma dengan fungsi link “sqrt” menjadi yang terbaik diantara ketiga metode lainnya dengan nilai RMSE sebesar 7.888. Berikut merupakan hasil summary dari model terbaik.

**Tabel 4.30** Summary Model Terbaik Gamma ”sqrt” (Perak1)

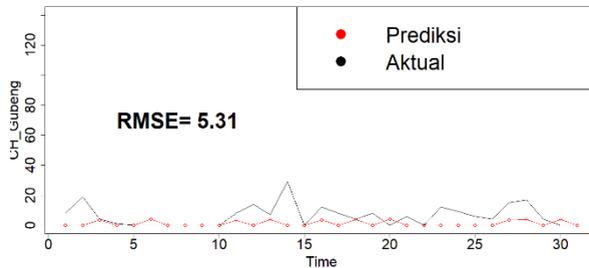
|                 | estimasi | Std.error | pvalue |
|-----------------|----------|-----------|--------|
| Intersep        | -2.8014  | 0.4879    | 0.0001 |
| Temperatur      | 0.0126   | 0.0015    | 0.0001 |
| Kelembapan      | 0.0466   | 0.0053    | 0.0001 |
| Kecapatan Angin | 0.0083   | 0.0557    | 0.8817 |
| Dummy           | 5.5102   | 0.7661    | 0.0001 |
| <i>lag1</i>     | 0.2987   | 0.0675    | 0.0001 |

Hasil *summary* berdasarkan model terbaik dihasilkan semua variabel berpengaruh signifikan terhadap model. Hal ini

$$Y_{1,t} = (-2.801 + 0.013x_{1,t} + 0.046x_{2,t} - 0.008x_{3,t} + 5.51Dummy + 0.2987Y_{1,t-1})^2$$

dikarenakan hasil nilai signifikansi uji (*pvalue*) kurang dari 0.05. berikut merupakan pemodelannya.

Hasil pemodelan terbaik dari data *training* digunakan dalam pengujian dari data testing. Data testing berjumlah 31 data yaitu merupakan tanggal 1 Desember 2018 sampai dengan 31 Desember 2018. Berikut merupakan hasil pengujian data *testing*.



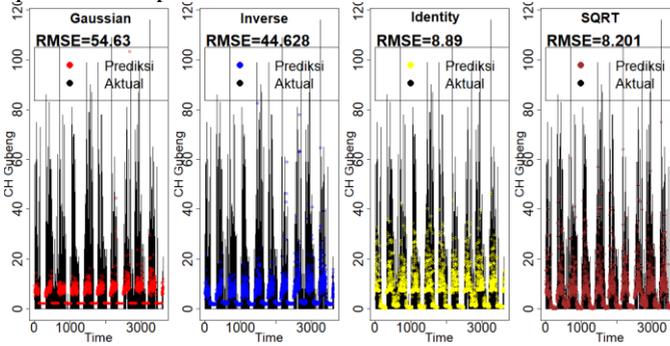
**Gambar 4.43** Prediksi Data Testing (Perak 1)

Secara visual terlihat bahwa pola yang terbentuk antara prediksi data testing dengan actual data testing tidak sesuai. Hasil prediksi terlihat konstan. Namun nilai RMSE yang dihasilkan lebih kecil jika dibandingkan dengan RMSE data *training* yaitu sebesar 5.31. Hal ini berarti model yang terbentuk dari data training sudah baik yaitu dengan menggunakan regresi Gamma fungsi link “sqrt”.

### III. Pemodelan Curah Hujan Gubeng dengan Perak 2

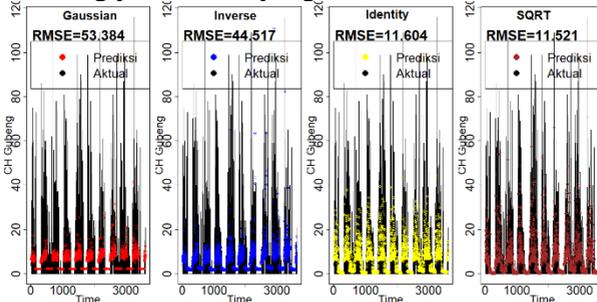
Pemodelan dari curah hujan Gubeng yang diawali dengan meregresikan variabel respon dengan masing masing variabel prediktor. Kemudian pemodelan tersebut dilanjutkan dengan menambahkan variabel *dummy* dengan kategori 1 untuk bulan hujan yaitu November sampai dengan April (bulan bulan basah) dan kategori 0 (bulan bulan kering) bulan Mei sampai Oktober. Penambahan variabel lag didasarkan pada curah hujan di waktu

sekarang dipengaruhi oleh curah hujan sebelumnya (*lag*) . Berikut merupakan hasil pemodelan terbaik.



**Gambar 4.44** Estimasi CH Gubeng dengan Ketiga Variabel, *dummy* dan *lag* 1

Pemodelan dengan variabel *lag* 1 berpengaruh signifikan terhadap peningkatan hasil prediksi seperti terlihat pada gambar diatas. Nilai RMSE juga semakin turun yaitu sebesar 8.201 pada metode regresi Gamma “sqrt”. Penambahan *lag* 1 mampu meningkatkan prediksi yang cukup tinggi bila dibandingkan dengan pemodelan tanpa *lag* 1. Berikut merupakan pemodelan dari penambahan *lag* pada model yang terbentuk diatas.



**Gambar 4.45** Estimasi CH dengan Ketiga Variabel *dummy lag* 1 dan *lag* 2

Secara visual terlihat bahwa setelah penambahan *lag* 2 pada model menghasilkan prediksi yang semakin baik. Pemodelan dengan *lag* 1 pada Gamma “sqrt” memiliki RMSE sebesar 8.201. Sedangkan pemodelan sampai dengan *lag* 2 Gamma “sqrt” memiliki RMSE sebesar 11.521. Sehingga berdasarkan hasil prediksi hingga sampai dengan *lag* 2 didapatkan bahwa dengan menggunakan ketiga variabel, *dummy*, dan *lag* 1, metode Regresi Gamma dengan fungsi link “sqrt” menjadi yang terbaik diantara

ketiga metode lainnya dengan nilai RMSE sebesar 8.201. Berikut merupakan hasil summary dari model terbaik.

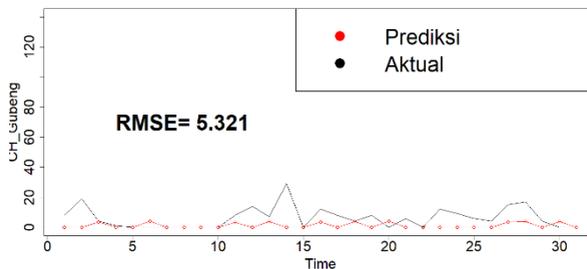
**Tabel 4.31** Summary Model Terbaik (Juanda)

|                 | Estimasi | Std.error | pvalue |
|-----------------|----------|-----------|--------|
| Intersep        | -5.47    | 0.618     | 0.001  |
| Temperatur      | 0.077    | 0.016     | 0.001  |
| Kelembapan      | 0.059    | 0.003     | 0.001  |
| Kecapatan Angin | 0.007    | 0.025     | 0.774  |
| Dummy           | 1.389    | 0.164     | 0.001  |
| lag1            | 0.048    | 0.011     | 0.001  |

Hasil *summary* berdasarkan model terbaik dihasilkan semua variabel berpengaruh signifikan terhadap model. Hal ini dikarenakan hasil nilai signifikansi uji (*pvalue*) kurang dari 0.05.

$$Y_{2,t} = (-5.47 + 0.077x_{1,t} + 0.059x_{2,t} + 0.007x_{3,t} + 1.389Dummy + 0.048Y_{1,t-1})^2$$

Pemodelan terbaik data *training* digunakan dalam pengujian dari data *testing*. Data *testing* berjumlah 31 data yaitu merupakan tanggal 1 Desember 2018 sampai dengan 31 Desember 2018. Berikut merupakan hasil pengujian dari data *testing*.



**Gambar 4.46** Prediksi Data Testing

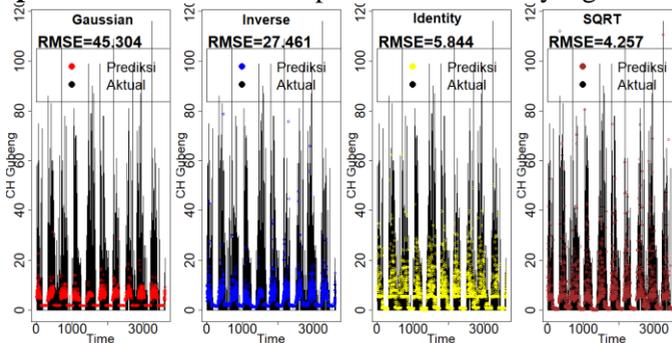
Hasil prediksi yang dihasilkan cenderung konstan dan tidak membentuk pola actual data testing. Hal ini dikarenakan banyaknya observasi data *testing* yang mengandung nilai 0. Berdasarkan hasil pengujian data testing yang berasal dari pemodelan data training yang memiliki RMSE sebesar 8.201 didapatkan nilai RMSE yang lebih kecil yaitu sebesar 5.321. Hal ini berarti model yang terbentuk dari data training sudah baik yaitu dengan menggunakan regresi Gamma fungsi link “sqrt”.

### 4.5.3 Pemodelan Curah Hujan Kedung Cowek

Regresi variabel curah hujan Kedung Cowek sebagai respon dengan variabel stasiun BMKG sebagai prediktor. Sebagai prediktor tambahan yaitu variabel *dummy* merupakan variabel yang berdasarkan dari statistika deskriptif didapatkan untuk hari hujan yang dikategorikan dengan yaitu pada bulan November sampai dengan April sedangkan pada bulan Mei sampai dengan Oktober merupakan hari hari yang jarang turun hujan dikategorikan dengan angka 0. Serta variabel lag yang merupakan variabel  $Y_{t-k}$  yang digunakan untuk meningkatkan model

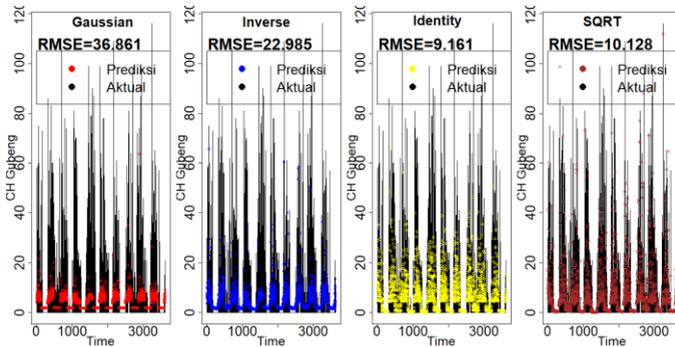
#### I. Pemodelan Curah Hujan Kd. Cowek dengan Juanda

Pemodelan dari curah hujan Kedung Cowek yang diawali dengan meregresikan variabel respon dengan masing masing variabel prediktor. Hasil pemodelan secara bertahap berada di **Lampiran 13**. Berikut adalah pemodelan terbaik yang terbentuk



**Gambar 4.47** Estimasi CH dengan Ketiga Variabel, *dummy* dan *lag1*

Pemodelan dengan variabel *lag1* meningkatkan hasil prediksi seperti terlihat pada gambar diatas. Nilai RMSE juga semakin turun yaitu sebesar 4.257 pada metode regresi Gamma “sqrt”. Penambahan *lag1* mampu meningkatkan prediksi yang cukup tinggi bila dibandingkan dengan pemodelan tanpa *lag1* (**lampiran 13**). Berikut merupakan pemodelan dari penambahan lag pada model yang terbentuk diatas.



**Gambar 4.48** Estimasi CH dengan Ketiga Variabel, dummy, lag1 dan lag2

Secara visual terlihat bahwa setelah penambahan lag2 pada model ternyata menghasilkan prediksi yang turun. Pemodelan dengan lag1 Gamma “sqrt” memiliki RMSE sebesar 4.267. Sedangkan pemodelan sampai dengan lag2 Gamma “sqrt” yang memiliki RMSE sebesar 10.128. Sehingga berdasarkan hasil prediksi hingga sampai dengan lag2 didapatkan bahwa dengan menggunakan ketiga variabel, *dummy*, dan lag1, metode Regresi Gamma dengan fungsi link “sqrt” menjadi yang terbaik diantara ketiga metode lainnya dengan nilai RMSE sebesar 4.267. Berikut merupakan hasil summary dari model terbaik.

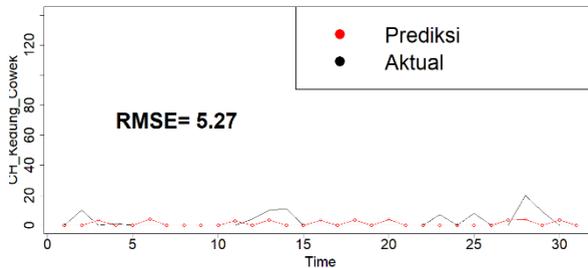
**Tabel 4.32** Summary Model Terbaik (Juanda)

|                 | Estimasi | Std.error | pvalue |
|-----------------|----------|-----------|--------|
| Intersep        | -3.881   | 0.309     | 0.001  |
| Temperatur      | 0.096    | 0.007     | 0.001  |
| Kelembapan      | 0.031    | 0.002     | 0.001  |
| Kecapatan Angin | -0.056   | 0.02      | 0.004  |
| Dummy           | 1.232    | 0.135     | 0.001  |
| lag1            | 0.084    | 0.013     | 0.001  |

Hasil *summary* berdasarkan model terbaik dihasilkan semua variabel berpengaruh signifikan terhadap model. Hal ini dikarenakan hasil nilai signifikansi uji (*pvalue*) kurang dari 0.05. berikut merupakan hasil pemodelannya.

$$Y_{1,t} = (-3.881 + 0.096x_{1,t} + 0.031x_{2,t} - 0.056x_{3,t} + 1.232Dummy + 0.084Y_{1,t-1})^2$$

Hasil akhir dari pemodelan terbaik yaitu dalam pengujian dari data testing. Berikut merupakan hasil pengujian data *testing*.

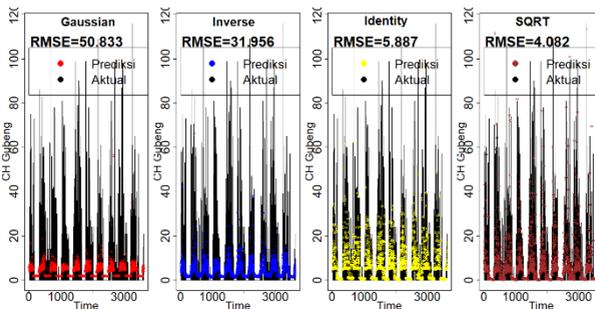


**Gambar 4.49** Prediksi Data Testing

Jika dilihat dari pola yang terbentuk kurang sesuai antara data testing dengan data prediksi. Berdasarkan hasil pengujian data testing yang berasal dari pemodelan data training yang memiliki RMSE sebesar 4.267 didapatkan nilai RMSE yang lebih besar yaitu sebesar 5.27. Hal ini berarti model yang terbentuk dari data training sudah baik yaitu dengan regresi Gamma fungsi link “sqrt”.

## II. Pemodelan Curah Hujan Kd. Cowek dengan Perak 1

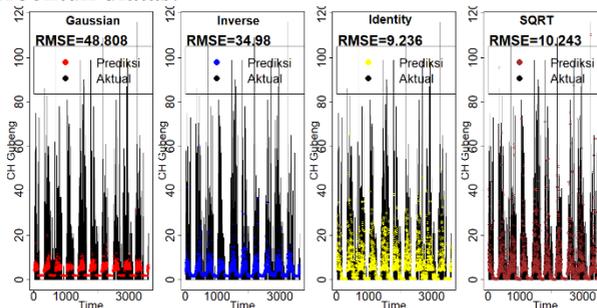
Hasil pemodelan terbaik dari curah hujan Kedung Cowek yang diawali dengan meregresikan variabel respon dengan masing-masing variabel prediktor. Pemodelan secara bertahap dari awal berada di **Lampiran 13**. Berikut pemodelan terbaik



**Gambar 4.50** Estimasi CH dengan Ketiga Variabel, dummy dan lag1

Pemodelan dengan penambahan variabel *lag1* berpengaruh signifikan terhadap peningkatan hasil prediksi. Nilai RMSE juga semakin turun yaitu sebesar 4.082 pada metode regresi Gamma “sqrt”. Penambahan *lag1* mampu meningkatkan prediksi yang cukup tinggi bila dibandingkan dengan pemodelan tanpa *lag1*.

Berikut merupakan pemodelan dari penambahan lag pada model yang terbentuk diatas.



**Gambar 4.51** Estimasi CH dengan Ketiga Variabel *dummy lag1* dan *lag2*

Pemodelan dengan lag1 Gamma “sqrt” memiliki RMSE sebesar 4.062. Sedangkan pemodelan sampai dengan lag2 Gamma “sqrt” yang memiliki RMSE sebesar 10.243. Berikut merupakan hasil summary dari model terbaik.

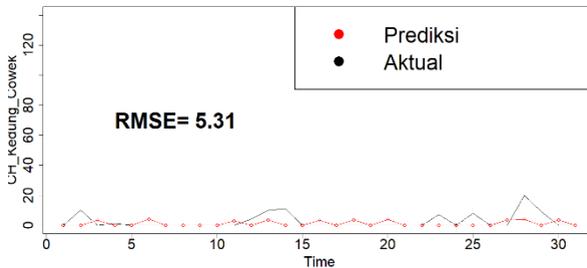
**Tabel 4.33** Summary Model Terbaik Gamma “sqrt” (Perak1)

|                 | Estimasi | Std.error | pvalue |
|-----------------|----------|-----------|--------|
| Intersep        | 0.679    | 9.703     | 0.944  |
| Temperatur      | -0.058   | 0.068     | 0.396  |
| Kelembapan      | 0.026    | 0.106     | 0.809  |
| Kecapatan Angin | -0.048   | 1.164     | 0.967  |
| Dummy           | 1.289    | 6.441     | 0.841  |
| lag1            | 0.087    | 0.642     | 0.892  |

Hasil *summary* berdasarkan model terbaik dihasilkan semua variabel berpengaruh signifikan terhadap model kecuali variabel kecepatan angin. Hal ini dikarenakan hasil nilai signifikansi uji (*pvalue*) kurang dari 0.05. Berikut merupakan hasil pemodelannya.

$$Y_{1,t} = (0.679 - 0.058x_{1,t} + 0.026x_{2,t} - 0.048x_{3,t} + 1.289Dummy + 0.087Y_{1,t-1})^2$$

Hasil akhir dari pemodelan terbaik yaitu dalam pengujian dari data testing. Data testing berjumlah 31 data yaitu merupakan tanggal 1 Desember 2018 sampai dengan 31 Desember 2018. Berikut merupakan hasil pengujian dari data *testing*.

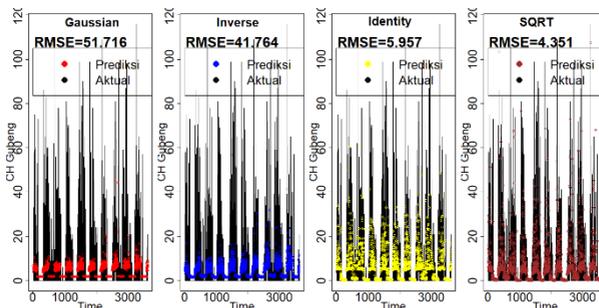


**Gambar 4.52** Prediksi Data Testing (Perak 1)

Pola yang terbentuk antara data testing dengan data prediksi menunjukkan hasil yang kurang sesuai dan prediksi dihasilkan nilai yang konstan. Hal ini dikarenakan banyaknya observasi dari data *testing* yang bernilai 0. Hasil pengujian data testing yang berasal dari pemodelan data training yang memiliki RMSE sebesar 4.061 didapatkan nilai RMSE yang lebih besar yaitu sebesar 5.321. Hal ini berarti model yang terbentuk dari data training sudah baik yaitu dengan menggunakan regresi Gamma fungsi link “sqrt”.

### III. Pemodelan Curah Hujan Kd. Cowek dengan Perak 2

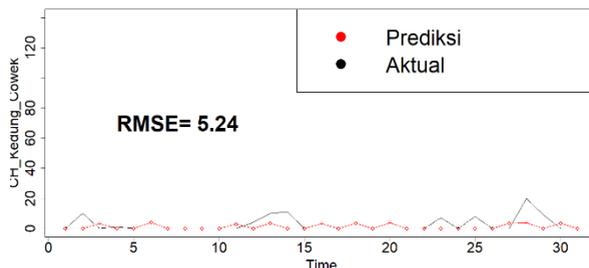
Pemodelan dari curah hujan Kedung Cowek diawali dengan meregresikan variabel respon dengan masing masing variabel prediktor. Selanjutnya dilakukan pemodelan dengan menambahkan variabel *dummy* dan variabel lag seperti terlampir pada **lampiran 13**. Berikut merupakan hasil pemodelan terbaik.



**Gambar 4.53** Estimasi CH Kd. Cowek dengan Ketiga Variabel, *dummy* dan *lag1*

Penambahan variabel *lag1* berpengaruh signifikan terhadap peningkatan hasil prediksi seperti terlihat pada gambar diatas. Nilai

RMSE juga semakin turun yaitu sebesar 9.419 pada metode regresi Gamma “identity”. Berikut merupakan hasil pengujian data *testing*.



**Gambar 4.54** Prediksi Data Testing

Pola data testing dengan data prediksi didapatkan hasil yang kurang tepat. Secara visual hasil prediksi terlihat konstan. Alasan mengapa hasil prediksi konstan adalah karena data testing mengandung banyak angka 0. Hasil pengujian data testing yang berasal dari pemodelan data training yang memiliki RMSE sebesar 4.357 didapatkan nilai RMSE yang lebih besar yaitu sebesar 5.24. Hal ini berarti model yang terbentuk dari data training sudah baik yaitu dengan menggunakan regresi Gamma fungsi link “sqrt”.

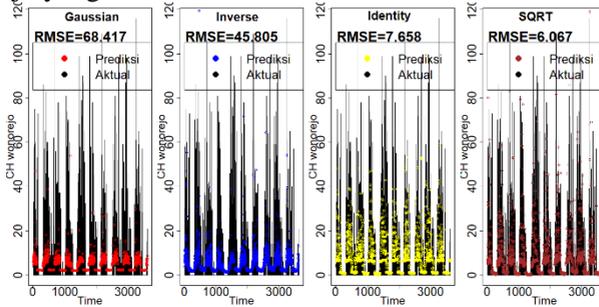
#### 4.5.4 Pemodelan Curah Hujan Wonorejo

Pemodelan yang dilakukan yaitu dengan meregresikan variabel curah hujan wonorejo sebagai respon dengan variabel stasiun BMKG sebagai prediktor. Sebagai prediktor tambahan yaitu variabel *dummy* merupakan variabel yang berdasarkan dari statistika deskriptif didapatkan untuk hari hujan yang dikategorikan dengan yaitu pada bulan November sampai dengan April sedangkan pada bulan Mei sampai dengan Oktober merupakan hari hari yang jarang turun hujan dikategorikan dengan angka 0. Serta variabel lag yang merupakan variabel  $Y_{t-k}$  yang digunakan untuk meningkatkan model

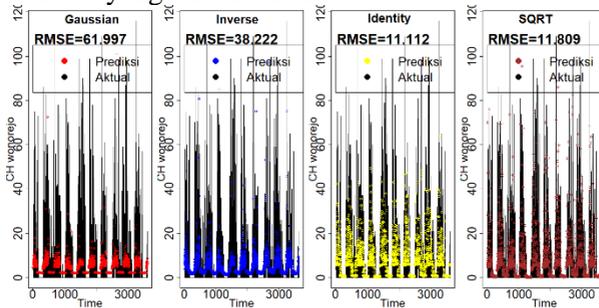
##### I. Pemodelan Curah Hujan Wonorejo dengan Juanda

Pemodelan secara bertahap yang diawali dengan meregresikan variabel respon dengan masing masing variabel prediktor dengan regresi inverse Gaussian (link= $1/\mu^2$ ), Gamma (link=inverse), Gamma (link=identity) dan Gamma dengan fungsi “sqrt sampai pemodelan dengan variabel *dummy*

menghasilkan prediksi yang masih kurang sesuai jika dibandingkan dengan data actual curah hujan Wonorejo seperti pada **lampiran 14**. Berikut hasil pemodelan dari curah hujan Wonorejo yang terbaik.



**Gambar 4.55** Estimasi Curah Hujan Wonorejo dengan variabel *dummy* dan *lag1*. Penambahan variabel *lag1* berpengaruh signifikan terhadap peningkatan hasil prediksi seperti terlihat pada gambar diatas. Nilai RMSE juga semakin turun yaitu sebesar 6.067 pada metode regresi Gamma “sqrt”. Berikut merupakan pemodelan dari penambahan lag pada model yang terbentuk diatas.



**Gambar 4.56.** Estimasi CH dengan Ketiga Variabel, *dummy*, *lag1* dan *lag2*

Secara visual terlihat bahwa setelah penambahan *lag2* pada model ternyata menghasilkan prediksi yang turun. Pemodelan dengan *lag1* memiliki RMSE 6.067 sedangkan pemodelan sampai dengan *lag2* memiliki RMSE sebesar 11.809. Sehingga model terbaik yang dihasilkan yaitu dengan ketiga variabel, *dummy*, dan *lag1*, metode Regresi Gamma dengan fungsi link “sqrt” menjadi yang terbaik diantara ketiga metode lainnya dengan nilai RMSE

sebesar 6.067. Berikut merupakan hasil summary dari model terbaik.

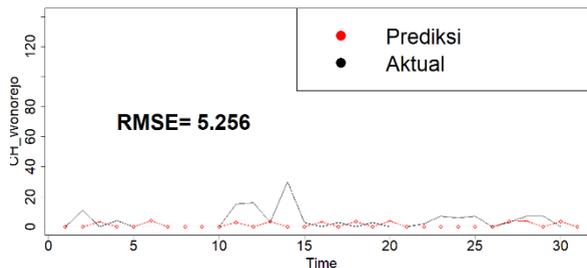
**Tabel 4.35** Summary Model Terbaik Gamma “sqrt” (Juanda)

|                 | Estimasi | Std.error | pvalue |
|-----------------|----------|-----------|--------|
| Intersep        | -4.195   | 0.192     | 0.001  |
| Temperatur      | 0.104    | 0.005     | 0.001  |
| Kelembapan      | 0.034    | 0.002     | 0.001  |
| Kecapatan Angin | -0.062   | 0.007     | 0.001  |
| Dummy           | 1.381    | 0.135     | 0.001  |
| lag1            | 0.068    | 0.011     | 0.001  |

Hasil *summary* berdasarkan model terbaik dihasilkan semua variabel berpengaruh signifikan terhadap model. Hal ini dikarenakan hasil nilai signifikansi uji (*pvalue*) kurang dari 0.05. berikut merupakan hasil pemodelannya.

$$Y_{1,t} = (-4.195 + 0.104x_{1,t} + 0.034x_{2,t} - 0.062x_{3,t} + 1.381Dummy + 0.068Y_{1,t-1})^2$$

Hasil akhir dari pemodelan terbaik yaitu dalam pengujian dari data testing. Data testing berjumlah 31 data yaitu merupakan tanggal 1 Desember 2018 sampai dengan 31 Desember 2018. Berikut merupakan hasil pengujian dari data *testing*.

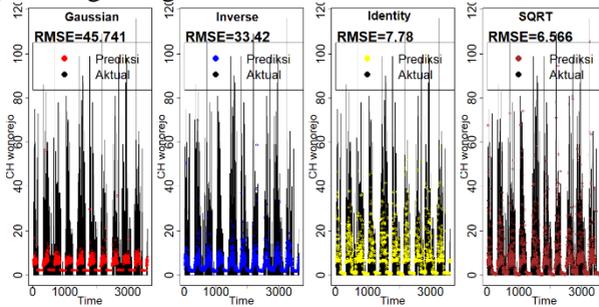


**Gambar 4.57** Pemodelan Data Testing

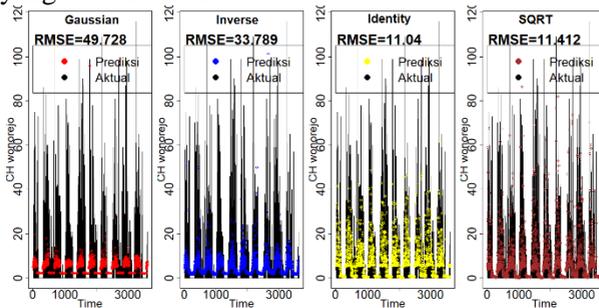
Secara visual hasil prediksi terlihat cukup konstan. Pola data testing dengan data prediksi didapatkan hasil yang kurang tepat. Hasil pengujian data testing yang berasal dari pemodelan data training yang memiliki RMSE sebesar 6.067 didapatkan nilai RMSE yang lebih kecil yaitu sebesar 5.256. Hal ini berarti model yang terbentuk dari data training sudah sangat baik.

## II. Pemodelan Curah Hujan Wonorejo dengan Perak 1

Berikut hasil pemodelan dari curah hujan Wonorejo yang diawali dengan meregresikan variabel respon dengan masing masing variabel prediktor sampai dengan menambahkan variabel dummy dan lag sebagai model terbaik.



**Gambar 4.58** Estimasi CH Wonorejo dengan Ketiga variabel, *dummy* dan *lag1* Pemodelan dengan menambahkan variabel *lag1* berpengaruh signifikan terhadap peningkatan hasil prediksi. Nilai RMSE juga semakin baik yaitu sebesar 6.566 pada metode regresi Gamma “sqrt”. Penambahan *lag1* mampu meningkatkan prediksi yang cukup tinggi bila dibandingkan dengan pemodelan tanpa *lag1*. Berikut merupakan pemodelan dari penambahan lag pada model yang terbentuk diatas.



**Gambar 4.59** Estimasi CH dengan Ketiga variabel, *dummy*, *lag1* dan *lag2*

Secara visual terlihat bahwa setelah penambahan *lag2* pada model ternyata menghasilkan prediksi yang turun. Jika dilihat berdasarkan nilai RMSE juga mengalami penurunan bila dibandingkan dengan pemodelan sebelumnya. Pemodelan dengan *lag1* memiliki RMSE 6.566 sedangkan pemodelan sampai dengan *lag2* memiliki RMSE sebesar 11.412. Sehingga berdasarkan hasil

prediksi yang telah dilakukan hingga sampai dengan lag2 didapatkan bahwa dengan menggunakan ketiga variabel, *dummy*, dan lag1, metode Regresi Gamma dengan fungsi link “sqrt” menjadi yang terbaik diantara ketiga metode lainnya dengan nilai RMSE sebesar 6.566. Berikut merupakan hasil summary dari model terbaik.

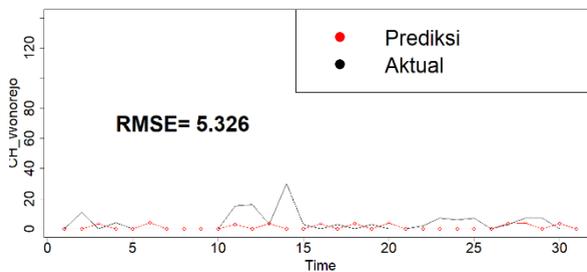
**Tabel 4.36.** Summary Model Terbaik Gamma “sqrt” (Perak1)

|                 | Estimasi | Std.eror | pvalue |
|-----------------|----------|----------|--------|
| Intersep        | -1.956   | 0.169    | 0.001  |
| Temperatur      | 0.011    | 0.004    | 0.009  |
| Kelembapan      | 0.04     | 0.002    | 0.001  |
| Kecapatan Angin | -0.096   | 0.005    | 0.001  |
| Dummy           | 1.266    | 0.131    | 0.001  |
| lag1            | 0.063    | 0.011    | 0.001  |

Hasil *summary* berdasarkan model terbaik dihasilkan semua variabel berpengaruh signifikan terhadap model. Hal ini dikarenakan hasil nilai signifikansi uji (*pvalue*) kurang dari 0.05.

$$Y_{1,t} = (-1.956 + 0.011x_{1,t} + 0.04x_{2,t} - 0.096x_{3,t} + 1.266Dummy + 0.063Y_{1,t-1})^2$$

Hasil akhir dari pemodelan terbaik yaitu dalam pengujian dari data testing. Data testing berjumlah 31 data yaitu merupakan tanggal 1 Desember 2018 sampai dengan 31 Desember 2018. Berikut merupakan hasil pengujian dari data *testing*.



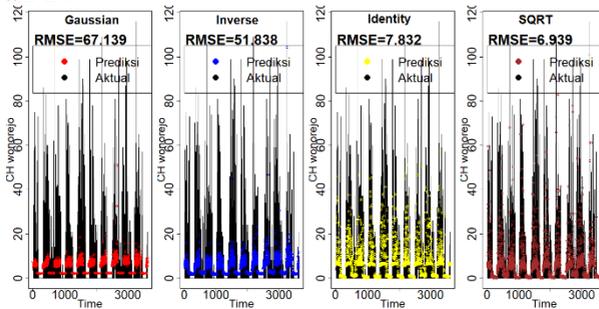
**Gambar 4.60** Prediksi Data Testing

Pemodelan menggunakan regresi Gamma fungsi link “sqrt” menjadi yang terbaik. Secara visual hasil prediksi terlihat konstan. Pola data testing dengan data prediksi didapatkan hasil yang kurang tepat. Hasil pengujian data testing yang berasal dari pemodelan data training yang memiliki RMSE sebesar 6.566

didapatkan nilai RMSE yang lebih kecil yaitu sebesar 5.326. Hal ini berarti model yang terbentuk dari data training sudah baik.

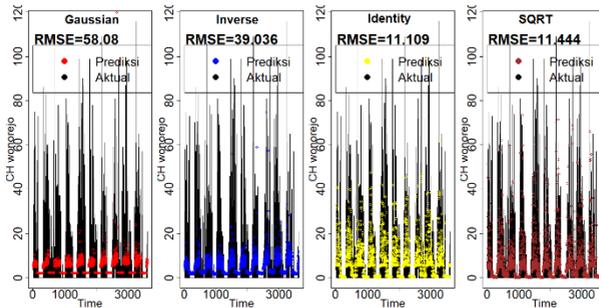
### III. Pemodelan Curah Hujan Wonorejo dengan Perak 2

Berikut hasil pemodelan dari curah hujan Wonorejo yang diawali dengan meregresikan variabel respon dengan masing masing variabel prediktor dengan regresi inverse Gaussian ( $\text{link}="1/\mu^2"$ ), Gamma ( $\text{link}=\text{inverse}$ ), Gamma ( $\text{link}=\text{identity}$ ) dan Gamma dengan fungsi “sqrt” sampai dengan penambahan variabel *dummy* dan lag1. Pemodelan secara bertahap seperti pada lampiran 14.



**Gambar 4.61** Estimasi CH Wonorejo dengan Ketiga Variabel, *dummy* dan lag1

Penambahan variabel *lag1* berpengaruh signifikan terhadap peningkatan hasil prediksi. Nilai RMSE juga semakin turun yaitu sebesar 6.939 pada metode regresi Gamma “sqrt”. Berikut merupakan pemodelan dari penambahan lag pada model yang terbentuk diatas.



**Gambar 4.62** Estimasi CH dengan Ketiga Variabel, *dummy*, lag1 dan lag2

Secara visual terlihat bahwa setelah penambahan lag2 pada model ternyata menghasilkan prediksi yang turun. Jika dilihat

berdasarkan nilai RMSE juga mengalami penurunan bila dibandingkan dengan pemodelan sebelumnya. Pemodelan dengan lag1 memiliki RMSE 6.939 sedangkan pemodelan sampai dengan lag2 memiliki RMSE sebesar 11.444. Sehingga berdasarkan hasil prediksi yang telah dilakukan hingga sampai dengan lag2 didapatkan bahwa dengan menggunakan ketiga variabel, *dummy*, dan lag1, metode Regresi Gamma dengan fungsi link “sqrt” menjadi yang terbaik diantara ketiga metode lainnya dengan nilai RMSE sebesar 6.939. Berikut merupakan hasil summary dari model terbaik.

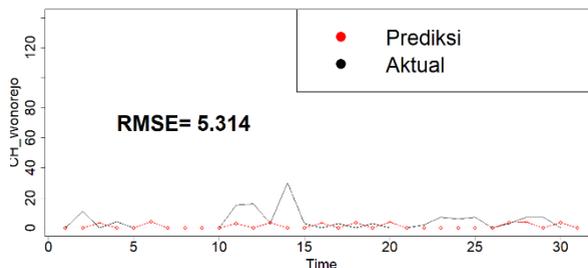
**Tabel 4.37** Summary Model Terbaik Gamma “sqrt” (Perak2)

|                 | Estimasi | Std.eror | Pvalue |
|-----------------|----------|----------|--------|
| Intersep        | -2.563   | 0.1205   | 0.0001 |
| Temperatur      | 0.0065   | 0.0013   | 0.0001 |
| Kelembapan      | 0.0488   | 0.0022   | 0.0001 |
| Kecapatan Angin | -0.0849  | 0.0041   | 0.0001 |
| Dummy           | 1.2965   | 0.1303   | 0.0001 |
| lag1            | 0.0587   | 0.0102   | 0.0001 |

Hasil *summary* berdasarkan model terbaik dihasilkan semua variabel berpengaruh signifikan terhadap model. Hal ini dikarenakan hasil nilai signifikansi uji (*pvalue*) kurang dari 0.05.

$$Y_{1,t} = (-2.56 + 0.0065x_{1,t} + 0.049x_{2,t} - 0.085x_{3,t} + 1.2965Dummy + 0.058Y_{1,t-1})^2$$

Pengecekan pemodelan terbaik dari hasil data training yaitu dalam pengujian dari data testing. Data testing berjumlah 31 data yaitu merupakan tanggal 1 Desember 2018 sampai dengan 31 Desember 2018. Berikut merupakan hasil pengujian data *testing*.



**Gambar 4.63** Prediksi Data Testing

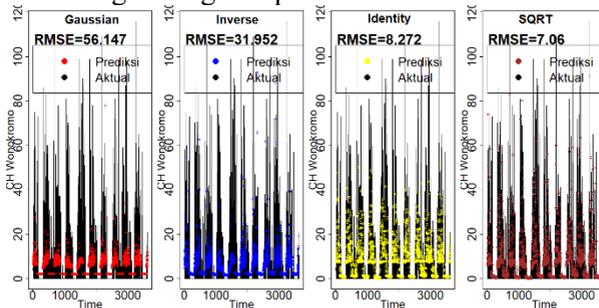
Secara visual hasil prediksi terlihat konstan. Pola data testing dengan data prediksi didapatkan hasil yang kurang tepat. Hasil pengujian data testing yang berasal dari pemodelan data training yang memiliki RMSE sebesar 6.929 didapatkan nilai RMSE yang lebih kecil yaitu sebesar 5.314. Hal ini berarti model yang terbentuk dari data training sudah sangat baik.

#### 4.5.5 Pemodelan Curah Hujan Wonokromo

Pemodelan yang dilakukan yaitu dengan meregresikan variabel curah hujan wonokromo` sebagai respon dengan variabel stasiun BMKG sebagai prediktor. Sebagai prediktor tambahan yaitu variabel *dummy* merupakan variabel yang berdasarkan dari statistika deskriptif didapatkan untuk hari hujan yang dikategorikan dengan yaitu pada bulan November sampai dengan April sedangkan pada bulan Mei sampai dengan Oktober merupakan hari hari yang jarang turun hujan dikategorikan dengan angka 0. Serta variabel lag yang merupakan variabel  $Y_{t-k}$  (berdasarkan curah hujan sekarang dipengaruhi oleh curah hujan sebelumnya atau masa lampau) yang digunakan untuk meningkatkan model

##### I. Pemodelan Curah Hujan Wonokromo dengan Juanda

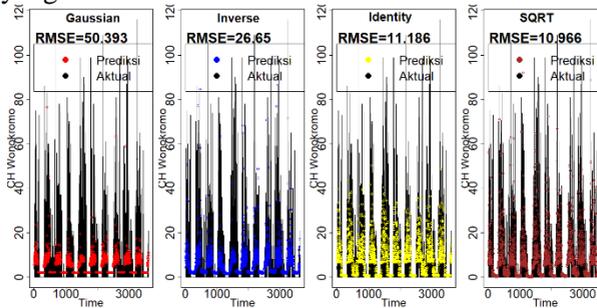
Berikut hasil pemodelan dari curah hujan Wonokromo yang diawali dengan meregresikan variabel respon dengan masing masing variabel prediktor dengan regresi inverse Gaussian (link="1/ $\mu^2$ "), Gamma (link=inverse), Gamma (link=identity) dan Gamma dengan fungsi "sqrt".



**Gambar 4.64** Estimasi CH dengan Ketiga Variabel, *dummy* dan *lag*1

Penambahan variabel *lag*1 berpengaruh signifikan terhadap peningkatan hasil prediksi seperti terlihat pada gambar diatas. Nilai

RMSE juga semakin turun yaitu sebesar 7.06 pada metode regresi Gamma “sqrt”. Penambahan *lag1* mampu meningkatkan prediksi yang cukup tinggi bila dibandingkan dengan pemodelan tanpa *lag1*. Berikut merupakan pemodelan dari penambahan lag pada model yang terbentuk diatas.



**Gambar 4.65** Estimasi CH Ketiga Variabel, *dummy*, *lag1* dan *lag2*

Secara visual terlihat bahwa setelah penambahan *lag2* pada model ternyata menghasilkan prediksi yang turun. Pemodelan dengan *lag1* memiliki RMSE 7.06 sedangkan pemodelan sampai dengan *lag2* memiliki RMSE sebesar 10.966. Sehingga berdasarkan hasil prediksi hingga sampai dengan *lag2* didapatkan bahwa dengan menggunakan ketiga variabel, *dummy*, dan *lag1*, metode Regresi Gamma dengan fungsi link “sqrt” menjadi yang terbaik diantara ketiga metode lainnya dengan nilai RMSE sebesar 7.06. Berikut merupakan hasil summarynya.

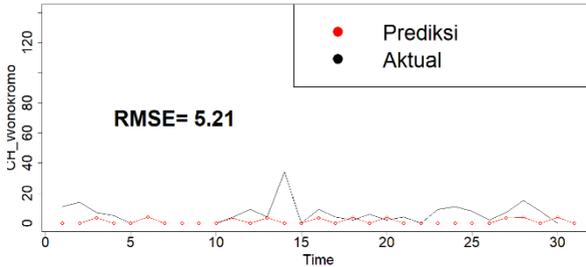
**Tabel 4.38** Summary Model Terbaik Gamma “sqrt” (Juanda)

|                 | Estimasi | Std.eror | pvalue |
|-----------------|----------|----------|--------|
| Intersep        | -4.3496  | 0.1897   | 0.0001 |
| Temperatur      | 0.1067   | 0.0048   | 0.0001 |
| Kelembapan      | 0.0348   | 0.0015   | 0.0001 |
| Kecapatan Angin | -0.0626  | 0.0079   | 0.0001 |
| Dummy           | 1.5417   | 0.1351   | 0.0001 |
| <i>lag1</i>     | 0.0579   | 0.0092   | 0.0001 |

Hasil *summary* berdasarkan model terbaik dengan Gamma “sqrt” dihasilkan semua variabel berpengaruh signifikan terhadap model. Hal ini dikarenakan hasil nilai signifikansi uji (*pvalue*) kurang dari 0.05. berikut merupakan hasil pemodelannya.

$$Y_{1,t} = (-4.349 + 0.1067x_{1,t} + 0.035x_{2,t} - 0.0626x_{3,t} + 1.542Dummy + 0.0579Y_{1,t-1})^2$$

Hasil akhir dari pemodelan terbaik yaitu dalam pengujian dari data testing. Data testing berjumlah 31 data yaitu merupakan tanggal 1 Desember 2018 sampai dengan 31 Desember 2018. Berikut merupakan hasil pengujian dari data *testing*.

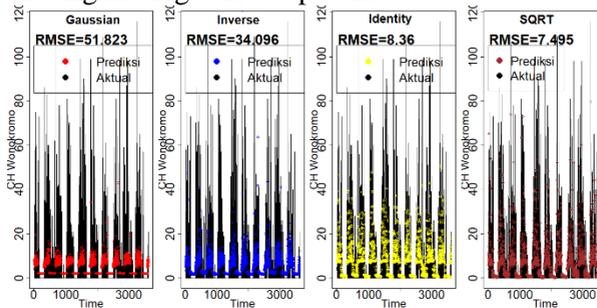


Gambar 4.66 Prediksi Data Testing

Secara visual hasil prediksi terlihat konstan. Pola data testing dengan data prediksi didapatkan hasil yang kurang tepat. Hal ini dikarenakan banyaknya observasi data testing yang bernilai 0. Hasil pengujian data testing yang berasal dari pemodelan data training yang memiliki RMSE sebesar 7.06 didapatkan nilai RMSE yang lebih kecil yaitu sebesar 5.21. Hal ini berarti model yang terbentuk dari data training sudah sangat baik.

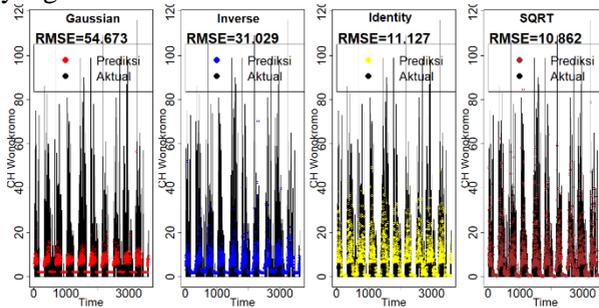
## II. Pemodelan Curah Hujan Wonokromo dengan Perak 1

Berikut hasil pemodelan terbaik dari curah hujan Wonokromo yang diawali dengan meregresikan variabel respon dengan masing masing variabel prediktor.



**Gambar 4.67** Estimasi CH dengan Ketiga Variabel, *dummy* dan *lag1*

Penambahan variabel *lag1* berpengaruh signifikan terhadap peningkatan hasil prediksi seperti terlihat pada gambar diatas. Nilai RMSE juga semakin turun yaitu sebesar 7.495 pada metode regresi Gamma “sqrt”. Penambahan *lag1* mampu meningkatkan prediksi yang cukup tinggi bila dibandingkan dengan pemodelan tanpa *lag1*. Berikut merupakan pemodelan dari penambahan lag pada model yang terbentuk diatas.



**Gambar 4.68** Estimasi CH dengan Ketiga Variabel, *dummy*, *lag1* dan *lag2*

Secara visual terlihat bahwa setelah penambahan *lag2* pada model ternyata menghasilkan prediksi yang turun. Jika dilihat berdasarkan nilai RMSE juga mengalami penurunan bila dibandingkan dengan pemodelan sebelumnya. Pemodelan dengan *lag1* memiliki RMSE 7.495, sedangkan pemodelan sampai dengan *lag2* memiliki RMSE sebesar 10.862. Sehingga berdasarkan hasil prediksi yang telah dilakukan hingga sampai dengan *lag2* didapatkan bahwa dengan menggunakan ketiga variabel, *dummy*, dan *lag1*, metode Regresi Gamma dengan fungsi link “sqrt” menjadi yang terbaik diantara ketiga metode lainnya dengan nilai RMSE sebesar 9.785. Berikut merupakan hasil summary dari model terbaik.

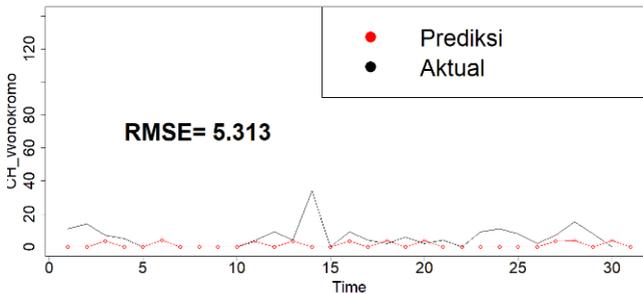
**Tabel 4.39** Summary Model Terbaik Gamma “sqrt” (perak 1)

|                 | estimasi | Std.eror | pvalue |
|-----------------|----------|----------|--------|
| Intersep        | -1.956   | 0.169    | 0.001  |
| Temperatur      | 0.011    | 0.004    | 0.009  |
| Kelembapan      | 0.04     | 0.002    | 0.001  |
| Kecapatan Angin | -0.096   | 0.005    | 0.001  |
| Dummy           | 1.266    | 0.131    | 0.001  |
| <i>lag1</i>     | 0.063    | 0.011    | 0.001  |

Hasil *summary* berdasarkan model terbaik dihasilkan semua variabel berpengaruh signifikan terhadap model. Hal ini dikarenakan hasil nilai signifikansi uji (*pvalue*) kurang dari 0.05. Berikut merupakan hasil pemodelannya.

$$Y_{1,t} = (-1.956 + 0.011x_{1,t} + 0.04x_{2,t} - 0.096x_{3,t} + 1.266Dummy + 0.063Y_{1,t-1})^2$$

Hasil akhir dari pemodelan terbaik yaitu dalam pengujian dari data testing. Data testing berjumlah 31 data yaitu merupakan tanggal 1 Desember 2018 sampai dengan 31 Desember 2018. Pemodelan yang dilakukan yaitu dengan membuat variabel *dummy* harian pada setiap bulan Desember di tahun 2009 sampai 2018. Jika pada setiap tanggal 1 Desember 2009 sampai tahun 2018 terdapat lima kali hujan dikategorikan dengan 1 dan dilakukan rata-rata curah hujan. Namun jika hujan kurang dari 5 dikategorikan dengan 0. Berikut merupakan hasil pengujian dari data *testing*.



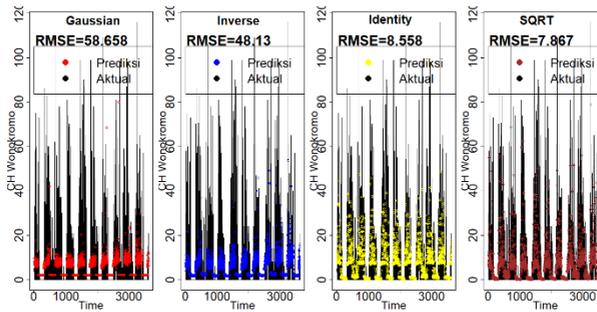
**Gambar 4.69** Prediksi Data Testing

Secara visual hasil prediksi terlihat konstan. Pola data testing dengan data prediksi didapatkan hasil yang kurang tepat. Hasil pengujian data testing yang berasal dari pemodelan data training yang memiliki RMSE sebesar 9.785 didapatkan nilai RMSE yang lebih kecil yaitu sebesar 5.313. Hal ini berarti model yang terbentuk dari data training sudah sangat baik.

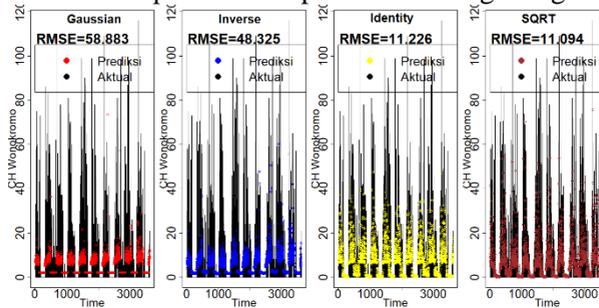
### III. Pemodelan Curah Hujan Wonokromo dengan Perak 2

Berikut hasil pemodelan dari curah hujan Wonokromo yang diawali dengan meregresikan variabel respon dengan masing

masing variabel prediktor). Pemodelan bertahap sampai didapatkan model terbaik ditunjukkan pada **Lampiran 15**.



**Gambar 4.70** Estimasi Curah Hujan Wonokromo dengan *dummy* dan *lag1*  
 Penambahan *lag1* mampu meningkatkan prediksi yang cukup tinggi bila dibandingkan dengan pemodelan tanpa *lag1*. Nilai RMSE juga semakin turun yaitu sebesar 7.967 pada metode regresi Gamma “sqrt”. Berikut pemodelan dengan *lag2*.



**Gambar 4.71** Estimasi Curah Hujan Wonokromo dengan *dummy*, *lag1* dan *lag2*

Secara visual terlihat bahwa setelah penambahan *lag2* pada model ternyata menghasilkan prediksi yang turun. Jika dilihat berdasarkan nilai RMSE juga mengalami penurunan bila dibandingkan dengan pemodelan sebelumnya. Pemodelan dengan *lag1* memiliki RMSE 7.867 sedangkan pemodelan sampai dengan *lag2* memiliki RMSE sebesar 11.084. Berdasarkan hasil prediksi yang telah dilakukan hingga sampai dengan *lag2* didapatkan bahwa dengan menggunakan ketiga variabel, *dummy*, dan *lag1*, metode Regresi Gamma dengan fungsi link “sqrt” menjadi yang terbaik diantara ketiga metode lainnya dengan nilai RMSE sebesar 6.939. Berikut merupakan hasil summary dari model terbaik.

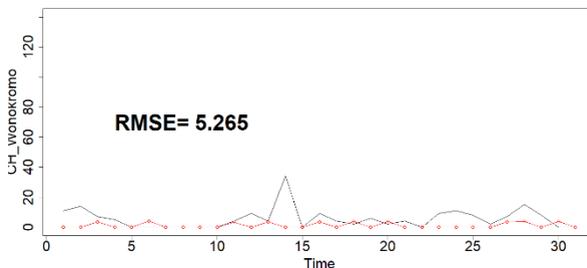
**Tabel 4.40** Summary Model Terbaik Gamma “sqrt” (Perak 2)

|                 | estimasi | Std.eror | Pvalue |
|-----------------|----------|----------|--------|
| Intersep        | -2.563   | 0.121    | 0.001  |
| Temperatur      | 0.007    | 0.002    | 0.001  |
| Kelembapan      | 0.049    | 0.003    | 0.001  |
| Kecapatan Angin | -0.085   | 0.005    | 0.001  |
| Dummy           | 1.297    | 0.131    | 0.001  |
| lag1            | 0.059    | 0.011    | 0.001  |

Hasil *summary* berdasarkan model terbaik dihasilkan semua variabel berpengaruh signifikan. Hal ini dikarenakan hasil nilai signifikansi uji (*pvalue*) kurang dari 0.05. Berikut modelnya.

$$Y_{1,t} = (-2.563 + 0.007x_{1,t} + 0.049x_{2,t} - 0.085x_{3,t} + 1.297Dummy + 0.059Y_{1,t-1})^2$$

Hasil akhir dari pemodelan terbaik yaitu dalam pengujian dari data testing. Data testing berjumlah 31 data yaitu merupakan tanggal 1 Desember 2018 sampai dengan 31 Desember 2018. Berikut merupakan hasil pengujian dari data *testing*.

**Gambar 4.72** Pemodelan Data Testing

Secara visual hasil prediksi terlihat konstan. Pola data testing dengan data prediksi didapatkan hasil yang kurang tepat. Hal ini dikarenakan banyaknya observasi dari data testing yang bernilai 0. Hasil pengujian data testing yang berasal dari pemodelan data training yang memiliki RMSE sebesar 6.939 didapatkan nilai RMSE yang lebih kecil yaitu sebesar 5.265. Hal ini berarti model yang terbentuk dari data training sudah baik.

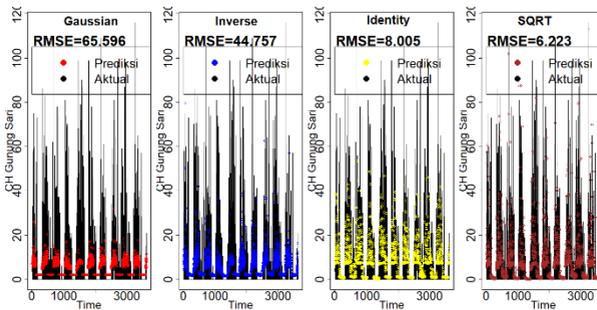
#### 4.5.6 Pemodelan Curah Hujan Gunung Sari

Pemodelan yang dilakukan yaitu dengan meregresikan variabel curah hujan Gunung Sari` sebagai respon dengan variabel

stasiun BMKG sebagai prediktor. Sebagai prediktor tambahan yaitu variabel *dummy* merupakan variabel yang berdasarkan dari statistika deskriptif didapatkan untuk hari hujan yang dikategorikan dengan yaitu pada bulan November sampai dengan April sedangkan pada bulan Mei sampai dengan Oktober merupakan hari hari yang jarang turun hujan dikategorikan dengan angka 0. Serta variabel lag yang merupakan variabel  $Y_{t-k}$  yang digunakan untuk meningkatkan model.

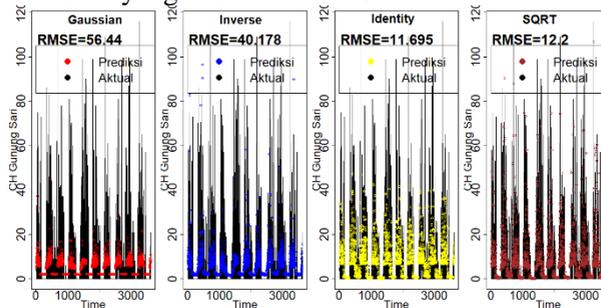
### I. Pemodelan Curah Hujan Gunung Sari dengan Juanda

Berikut hasil pemodelan terbaik dari curah hujan Gunung Sari yang diawali dengan meregresikan variabel respon.



**Gambar 4.73** Estimasi CH dengan Ketiga Variabel, *dummy* dan *lag1*

Penambahan variabel *lag1* berpengaruh signifikan terhadap peningkatan hasil prediksi seperti terlihat pada gambar diatas. Nilai RMSE juga semakin turun yaitu sebesar 6.223 pada metode regresi Gamma “sqrt”. Berikut merupakan pemodelan dari penambahan lag pada model yang terbentuk diatas.



**Gambar 4.74** Estimasi CH dengan Ketiga Variabel, *dummy lag1* dan *lag2*

Secara visual terlihat bahwa setelah penambahan lag2 pada model ternyata menghasilkan prediksi yang turun. Jika dilihat berdasarkan nilai RMSE juga mengalami penurunan bila dibandingkan dengan pemodelan sebelumnya. Pemodelan dengan lag1 memiliki RMSE 6.223 sedangkan pemodelan sampai dengan lag2 memiliki RMSE sebesar 12.2. Sehingga berdasarkan hasil prediksi hingga sampai dengan lag2 didapatkan bahwa dengan menggunakan ketiga variabel, *dummy*, dan lag1, metode Regresi Gamma dengan fungsi link “sqrt” menjadi yang terbaik diantara ketiga metode lainnya dengan nilai RMSE sebesar 6.223. Berikut merupakan hasil summary dari model terbaik.

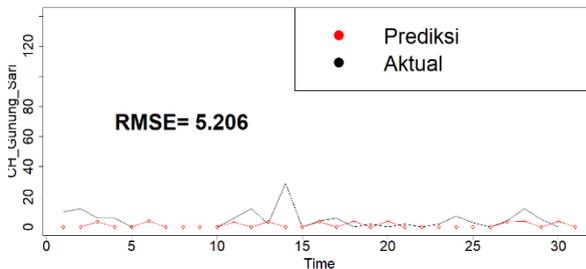
**Tabel 4.41** Summary Model Terbaik Gamma “sqrt” (Juanda)

|                 | Estimasi | Std.error | Pvalue |
|-----------------|----------|-----------|--------|
| Intersep        | -4.166   | 0.186     | 0.001  |
| Temperatur      | 0.103    | 0.005     | 0.001  |
| Kelembapan      | 0.034    | 0.002     | 0.001  |
| Kecapatan Angin | -0.059   | 0.008     | 0.001  |
| Dummy           | 1.529    | 0.134     | 0.001  |
| lag1            | 0.065    | 0.01      | 0.001  |

Hasil *summary* berdasarkan model terbaik dihasilkan semua variabel berpengaruh signifikan terhadap model. Hal ini dikarenakan hasil nilai signifikansi uji (*pvalue*) kurang dari 0.05. berikut merupakan hasil pemodelannya.

$$Y_{1,t} = (-4.166 + 0.103x_{1,t} + 0.034x_{2,t} - 0.059x_{3,t} + 1.529Dummy + 0.065Y_{1,t-1})^2$$

Akhir dari pemodelan terbaik yaitu dalam pengujian dari data testing. Data testing berjumlah 31 data yaitu merupakan tanggal 1 Desember 2018 sampai dengan 31 Desember 2018. Berikut merupakan hasil pengujian dari data *testing*.

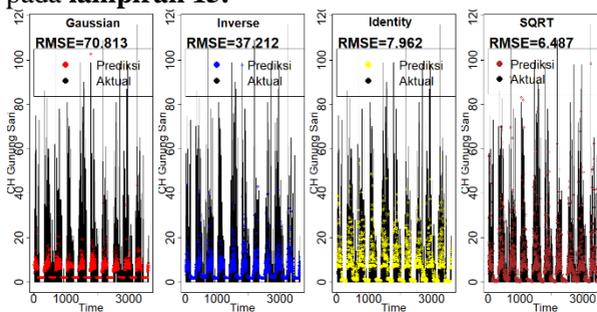


**Gambar 4.75** Pemodelan Data Testing

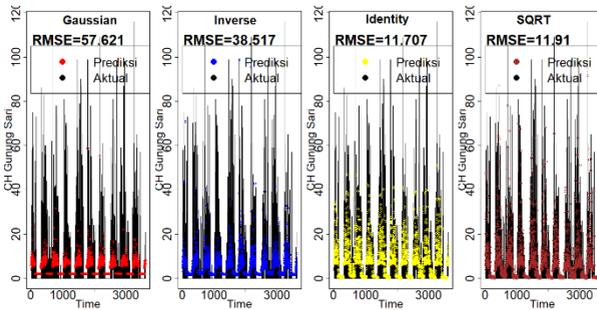
Secara visual hasil prediksi terlihat konstan. Pola data testing dengan data prediksi didapatkan hasil yang kurang tepat. Hal ini dikarenakan banyaknya observasi dari data testing yang bernilai 0. Hasil pengujian data testing yang berasal dari pemodelan data training yang memiliki RMSE sebesar 6.223 didapatkan nilai RMSE yang lebih kecil yaitu sebesar 5.206. Hal ini berarti model yang terbentuk dari data training sudah baik.

## II. Pemodelan Curah Hujan Gunung Sari dengan Perak 1

Pemodelan dari curah hujan Gunung Sari diawali dengan meregresikan variabel respon dengan masing masing variabel prediktor dengan regresi inverse Gaussian ( $\text{link}="1/\mu^2"$ ), Gamma ( $\text{link}=\text{inverse}$ ), Gamma ( $\text{link}=\text{identity}$ ) dan Gamma dengan fungsi "sqrt". Pemodelan selanjutnya adalah dengan menambahkan variabel dummy, hingga menambahkan variabel lag seperti pada **lampiran 15**.

**Gambar 4.76** Estimasi Curah Hujan Gunung Sari dengan *dummy* dan *lag*1

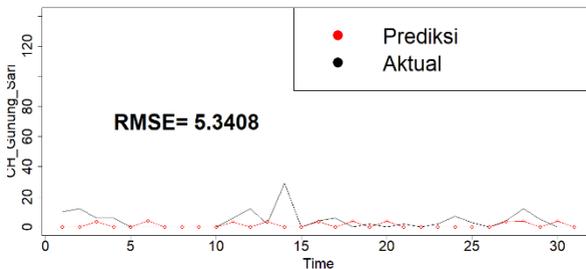
Penambahan variabel *lag*1 berpengaruh signifikan terhadap peningkatan hasil prediksi seperti terlihat pada gambar diatas. Nilai RMSE juga semakin turun yaitu sebesar 6.487 pada metode regresi Gamma "sqrt". Penambahan *lag*1 mampu meningkatkan prediksi yang cukup tinggi bila dibandingkan dengan pemodelan tanpa *lag*1. Berikut merupakan pemodelan dari penambahan *lag* pada model yang terbentuk diatas.



**Gambar 4.77** Estimasi Curah Hujan Gunung Sari dengan *dummy*, *lag1* dan *lag2*

Secara visual terlihat bahwa setelah penambahan *lag2* pada model ternyata menghasilkan prediksi yang turun. Jika dilihat berdasarkan nilai RMSE juga mengalami penurunan bila dibandingkan dengan pemodelan sebelumnya. Pemodelan dengan *lag1* memiliki RMSE 6.487 sedangkan pemodelan sampai dengan *lag2* memiliki RMSE sebesar 11.91. Berdasarkan hasil prediksi sampai dengan *lag2* didapatkan bahwa dengan ketiga variabel, *dummy*, dan *lag1*, metode Regresi Gamma dengan fungsi link “*sqrt*” menjadi yang terbaik dengan nilai RMSE sebesar 6.487.

Pemodelan terbaik data training digunakan dalam pengujian data testing. Berikut merupakan hasil prediksi data *testing*.



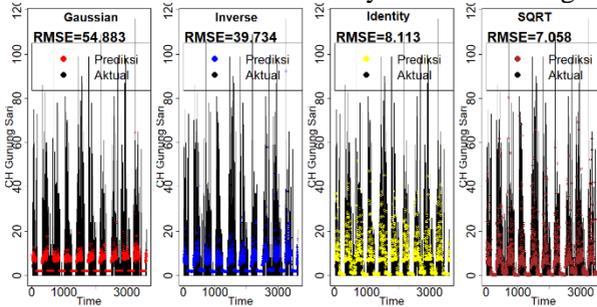
**Gambar 4.78** Pemodelan Data Testing

Secara visual hasil prediksi terlihat konstan. Pola data testing dengan data prediksi didapatkan hasil yang kurang tepat. Hasil pengujian data testing yang berasal dari pemodelan data training yang memiliki RMSE sebesar 6.487 didapatkan nilai

RMSE yang lebih kecil yaitu sebesar 5.3408. Hal ini berarti model yang terbentuk dari data training sudah sangat baik.

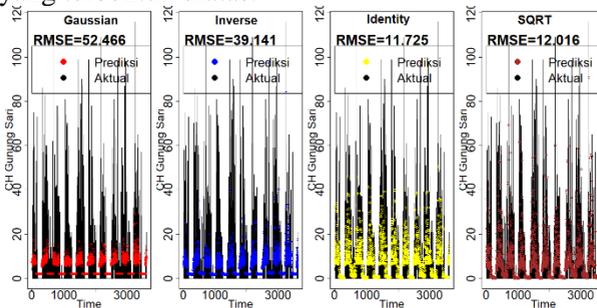
### III. Pemodelan Curah Hujan Gn. Sari dengan Perak 2.

Pemodelan yang dilakukan yaitu dengan meregresikan variabel curah hujan Gunung sari dengan variabel stasiun Perak 2 dengan menambahkan variabel dummy dan variabel lag.



**Gambar 4.79** Estimasi Curah Hujan Gunung Sari dengan *dummy* dan *lag*1

Penambahan variabel *lag*1 berpengaruh signifikan terhadap peningkatan hasil prediksi seperti terlihat pada gambar diatas. Nilai RMSE juga semakin turun yaitu sebesar 7.058 pada metode regresi Gamma “sqrt”. Penambahan *lag*1 mampu meningkatkan prediksi yang cukup tinggi bila dibandingkan dengan pemodelan tanpa *lag*1. Berikut merupakan pemodelan dari penambahan lag pada model yang terbentuk diatas.



**Gambar 4.80** Estimasi Curah Hujan Gunung Sari dengan *dummy lag*1 dan *lag*2

Secara visual terlihat bahwa setelah penambahan *lag*2 pada model ternyata menghasilkan prediksi yang turun. Jika dilihat berdasarkan nilai RMSE juga mengalami penurunan bila dibandingkan dengan pemodelan sebelumnya. Pemodelan dengan

lag1 memiliki RMSE 7.058 sedangkan pemodelan sampai dengan lag2 memiliki RMSE sebesar 12.016. Sehingga berdasarkan hasil prediksi yang telah dilakukan hingga sampai dengan lag2 didapatkan bahwa dengan menggunakan ketiga variabel, *dummy*, dan lag1, metode Regresi Gamma dengan fungsi link “sqrt” menjadi yang terbaik diantara ketiga metode lainnya dengan nilai RMSE sebesar 10.534. Berikut merupakan hasil summary dari model terbaik.

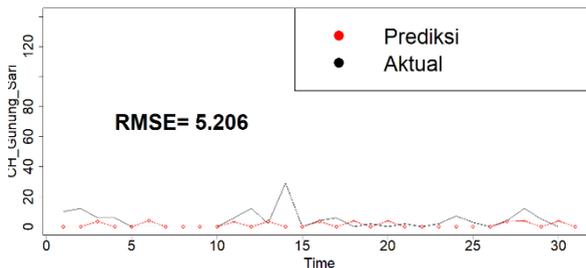
**Tabel 4.42** Summary Model Terbaik Gamma “identity” (Perak 2)

|                 | estimasi | Std.error | pvalue |
|-----------------|----------|-----------|--------|
| Intersep        | -2.683   | 0.153     | 0.001  |
| Temperatur      | 0.009    | 0.003     | 0.002  |
| Kelembapan      | 0.05     | 0.003     | 0.001  |
| Kecapatan Angin | -0.087   | 0.005     | 0.001  |
| Dummy           | 1.425    | 0.14      | 0.001  |
| lag1            | 0.057    | 0.01      | 0.001  |

Hasil *summary* berdasarkan model terbaik dihasilkan semua variabel berpengaruh signifikan terhadap model. Hal ini dikarenakan hasil nilai signifikansi uji (*pvalue*) kurang dari 0.05. berikut merupakan hasil pemodelannya.

$$Y_{1,t} = (-2.683 + 0.009x_{1,t} + 0.05x_{2,t} - 0.087x_{3,t} + 1.425Dummy + 0.057Y_{1,t-1})^2$$

Pengujian kebaikan model terbaik dari data *training* digunakan dalam data *testing*. Data testing berjumlah 31 data yaitu merupakan tanggal 1 Desember 2018 sampai dengan 31 Desember 2018. Berikut merupakan hasil pengujian dari data *testing*.



**Gambar 4.81.** Prediksi Data Testing

Secara visual hasil prediksi terlihat konstan. Pola data testing dengan data prediksi didapatkan hasil yang kurang tepat.

Hasil pengujian data testing yang berasal dari pemodelan data training yang memiliki RMSE sebesar 6.223 didapatkan nilai RMSE yang lebih kecil yaitu sebesar 5.206. Hal ini berarti model yang terbentuk dari data training sudah sangat baik.

#### **4.6 Pemodelan Multivariat Curah Hujan dengan Variabel Stasiun BMKG**

Pemodelan dilakukan dengan menggunakan metode regresi Inverse Gaussian dan regresi Gamma yang termasuk dalam keluarga distribusi eksponensial. Data curah hujan meliputi data masing masing kecamatan yang terdiri dari curah hujan Keputh, Gubeng, Wonorejo, Wonokromo, kedung Cowek dan Gunung Sari. Sedangkan variabel BMKG yang dimaksudkan adalah Temperature, Kelembapan dan Kecepatan Angin. Maksud dari pemodelan secara multivariate yaitu melempokkan variabel yang sama dari ketiga stasiun dengan menggunakan analisis komponen utama. Tujuan yang dimaksud adalah untuk mengetahui bagaimana hasil pemodelan jika dilakukan secara bersama. Tujuan akhir dari pemodelan secara multivariat adalah untuk mendapatkan model terbaik dengan berdasarkan nilai error yang terkecil dan membandingkan hasil dengan pemodelan secara univariate. Berikut merupakan hasil pemodelannya.

##### **4.6.1 Pemodelan Multivariat Curah Hujan Keputih dengan variabel Stasiun BMKG**

Pemodelan yang dilakukan yaitu dengan meregresikan variabel curah hujan Keputih sebagai respon dengan variabel stasiun BMKG sebagai prediktor. Dimana variabel prediktornya sudah dilakukan pengeleompokan dengan menggunakan analisis komponen utama (PCA). Sebagai prediktor tambahan yaitu variabel *dummy* dan variabel lag yang merupakan variabel  $Y_{t-k}$  yang digunakan untuk meningkatkan model.

Langkah awal yang dilakukan adalah melakukan pengelompokan variabel berdasarkan variabel yang sama dengan menggunakan PCA. *Threshold cumulative Proportion* yang digunakan adalah sebesar 80% karena dengan persentase tersebut sudah bisa mewakili hasil pengelompokan dari variabel yang sama

dan berasal dari stasiun yang berbeda. Berikut merupakan hasil dari analisis komponen utama.

**Tabel 4.43.** Summary PCA Temperature dari Ketiga Stasiun

|                        | PC1    | PC2     | PC3     |
|------------------------|--------|---------|---------|
| Standard deviation     | 1.8685 | 0.54603 | 0.33202 |
| Proportion of Variance | 0.8953 | 0.07646 | 0.02827 |
| Cumulative Proportion  | 0.8953 | 0.97173 | 1       |

**Tabel 4.44** Summary PCA Kelembapan dari Ketiga Stasiun

|                        | PC1     | PC2     | PC3     |
|------------------------|---------|---------|---------|
| Standard deviation     | 11.5507 | 3.7645  | 2.42588 |
| Proportion of Variance | 0.8693  | 0.09234 | 0.03834 |
| Cumulative Proportion  | 0.8693  | 0.96166 | 1       |

**Tabel 4.45** Summary PCA Kecepatan Angin dari Ketiga Stasiun

|                        | PC1    | PC2    | PC3    |
|------------------------|--------|--------|--------|
| Standard deviation     | 1.6106 | 1.004  | 0.8316 |
| Proportion of Variance | 0.6041 | 0.2348 | 0.161  |
| Cumulative Proportion  | 0.6041 | 0.8389 | 1      |

Berdasarkan hasil analisis komponen utama yang disajikan dalam ketiga tabel diatas dengan menggunakan *Threshold cumulative Proportion* sebesar 80% didapatkan hasil bahwa jumlah PC yang terbentuk pada analisis PCA variabel Temperatur dan Kelembapan adalah satu yaitu dengan nilai *cumulative Proportion* masing masing sebesar 0.8953 dan 0.8693. Sedangkan untuk variabel kecepatan angin jumlah PC yang terbentuk adalah sebanyak dua yaitu dengan *cumulative Proportion* sebesar 0.8389. Hasil dari PC tersebut akan dimodelkan secara univariate dengan menggunakan regresi inverse Gaussian, regresi gamma dengan fungsi link Inverse, regresi Gamma dengan fungsi link identity dan regresi Gamma dengan fungsi link sqrt. Berdasarkan dari masing masing PC yang terbentuk. Berikut merupakan hasil pemodelannya

**Tabel 4.46.** Pemodelan Curah Hujan Keputih dengan PC1 (Temperature)

|        | Gaussian |       | inverse |       | identity |       | sqrt   |       |
|--------|----------|-------|---------|-------|----------|-------|--------|-------|
|        | est      | p-val | est     | p-val | est      | p-val | est    | p-val |
| Interc | 0.043    | 0.002 | 0.208   | 0.002 | 5.245    | 0.002 | 2.285  | 0.002 |
| PC1    | 0.007    | 0.002 | 0.034   | 0.002 | -0.766   | 0.002 | -0.29  | 0.002 |
| AIC    | -33223   |       | -30054  |       | -30066   |       | -29913 |       |

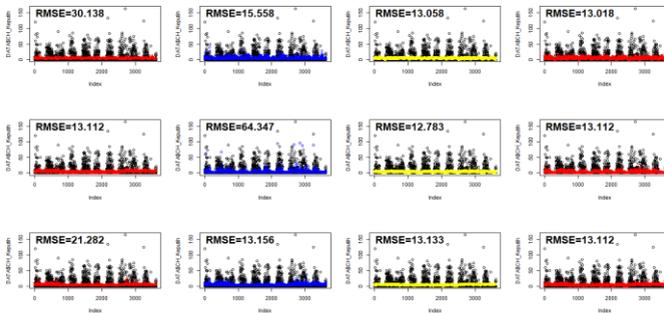
**Tabel 4.47.** Pemodelan Curah Hujan Keputih dengan PC1 (Kelembapan)

|        | Gaussian |       | inverse |       | identity |       | sqrt   |       |
|--------|----------|-------|---------|-------|----------|-------|--------|-------|
|        | Est      | p-val | est     | p-val | est      | p-val | est    | p-val |
| Interc | 0.072    | 0.002 | 0.273   | 0.002 | 4.338    | 0.002 | 1.884  | 0.002 |
| PC1    | 0.003    | 0.002 | 0.009   | 0.002 | -0.135   | 0.002 | -0.057 | 0.002 |
| AIC    | -33223   |       | -30055  |       | -29900   |       | -30232 |       |

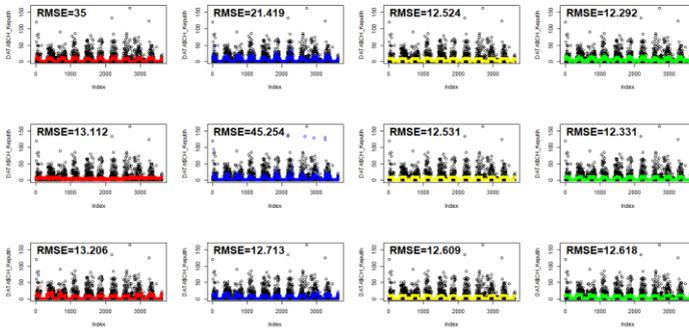
**Tabel 4.48.** Pemodelan Curah Hujan Keputih dengan PC1 (Kecepatan angin)

|        | Gaussian |       | inverse |       | identity |       | sqrt   |         |
|--------|----------|-------|---------|-------|----------|-------|--------|---------|
|        | Est      | p-val | est     | p-val | est      | p-val | est    | p-val   |
| Interc | 0.042    | 0.002 | 0.204   | 0.002 | 5.333    | 0.002 | 2.301  | 0.002   |
| PC1    | 0.01     | 0.002 | 0.041   | 0.002 | -0.659   | 0.002 | -0.274 | 0.002   |
| PC2    | 0.007    | 0.012 | 0.017   | 0.021 | -0.119   | 0.002 | -0.106 | 0.00473 |
| AIC    | -33221   |       | -29872  |       | -29872   |       | -29892 |         |

Hasil pemodelan multivariate yang ditunjukkan pada ketiga tabel diatas dengan menggunakan keempat metode didapatkan hasil variabel PC yang signifikan secara univariate. Hal ini dikarenakan nilai p-value yang kurang dari nilai alpha (0.05). Hasil prediksi dari masing masing metode dapat dilihat pada **Gambar 4.33**

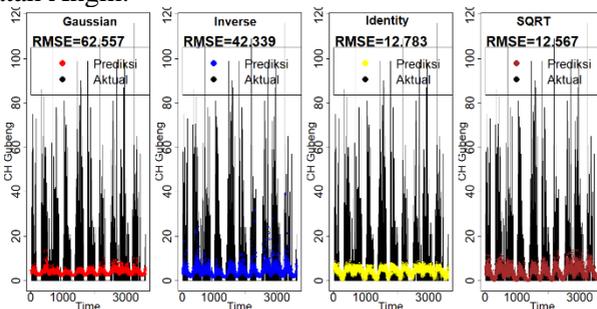
**Gambar 4.82** Prediksi Curah Hujan Keputih dengan variabel PC

Secara visual terlihat bahwa pemodelan yang dilakukan secara individu terhadap pengelompokan yang sama masih terlihat jauh dari nilai actual pada hasil prediksinya. Nilai RMSE juga masih berkisar cukup besar pada metode inverse Gaussian dan regresi Gamma dengan fungsi link inverse. Sedangkan dengan menggunakan regresi gamma dengan fungsi link identity dan sqrt nilai RMSE cenderung lebih kecil. Berikut pemodelan dengan menggunakan masing masing PC dengan menambahkan *dummy*.



**Gambar 4.83** Prediksi Curah Hujan Keputih dengan variabel PC dan *dummy*

Perbandingan yang dihasilkan diantara pemodelan dengan dan tanpa variabel *dummy* adalah dengan menggunakan variabel *dummy* menjadikan nilai RMSE menjadi turun. Pada metode regresi Gamma “sqrt” pemodelan dengan menggunakan variabel PC1\_Temperature dengan *dummy* memberikan nilai RMSE terkecil yaitu sebesar 12.292. secara umum dnegan pemodelan masing masing variabel PC dengan *dummy* memberikan hasil bahwa dengan menggunakan regresi gamma “sqrt” memiliki nilai RMSE yang terkecil jikan dbandingkan dengan metode lain. Dengan adanya variabel *dummy* mampu menangkap hasil prediksi yang 0 sehingga pada pemodelan sebelumnya tanpa *dummy* memiliki RMSE lebih besar. Berikut merupakan hasil pemodelan secara bersama dari PC Temperatur, PC Kelembapan dan PC Kecepatan Angin.



**Gambar 4.84** Prediksi Curah Hujan Keputih dengan ketiga variabel PC

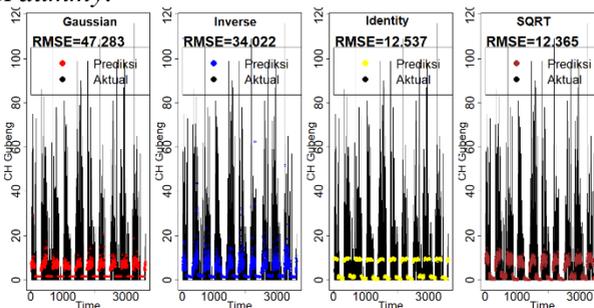
Hasil pemodelan dengan ketiga variabel PC menghasilkan nilai prediksi yang lebih naik jika dibandingkan dengan pemodelan

masing masing PC. Berikut merupakan summary dari pemodelan ketiga variabel PC dengan curah hujan Keputih.

**Tabel 4.49.** Pemodelan Curah Hujan Keputih dengan semua PC

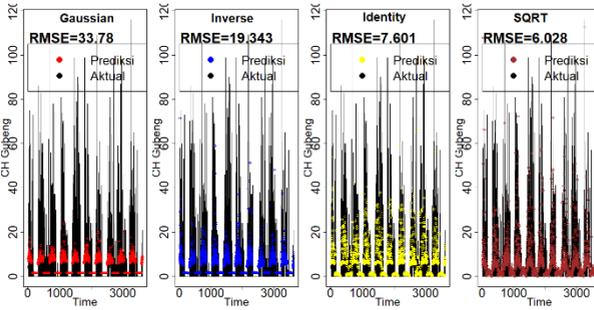
|        | Gaussian |       | Inverse |       | identity |       | Sqrt   |       |
|--------|----------|-------|---------|-------|----------|-------|--------|-------|
|        | Est      | p-val | est     | p-val | est      | p-val | est    | p-val |
| Interc | 0.083    | 0.002 | 0.292   | 0.001 | 4.324    | 0.002 | 1.878  | 0.002 |
| PC1    | -0.012   | 0.002 | -0.028  | 0.001 | 0.064    | 0.002 | 0.042  | 0.002 |
| PC2    | 0.004    | 0.002 | 0.013   | 0.001 | -0.149   | 0.002 | -0.057 | 0.002 |
| Pca    | 0.001    | 0.002 | 0.002   | 0.233 | 0.149    | 0.002 | -0.09  | 0.002 |
| PCb    | 0.002    | 0.002 | 0.003   | 0.002 | -0.088   | 0.002 | 0.001  | 0.98  |
| AIC    | -33219   |       | 30070   |       | -30065   |       | -30270 |       |

Hasil dari pemodelan dengan menggunakan semua PC pada regresi Gamma “sqrt” tidak mampu menaikkan hasil prediksi adalah karena variabel PC<sub>2</sub> Kecepatan angin tidak berpengaruh signifikan terhadap model. Hal ini karena nilai pvalue lebih dari 0.05. Dengan metode regresi inverse Gaussian dan gamma “identity” semua variabel berpengaruh signifikan. Sehingga jika dibandingkan dengan prediksi secara individu mengalami kenaikan. Sedangkan dengan Gamma “inverse” pada variabel PC<sub>1</sub> kecepatan angin tidak berpengaruh signifikan, karena pvalue (0.232) lebih dari 0.05. Berikut hasil prediksi semua PC dengan variabel *dummy*.

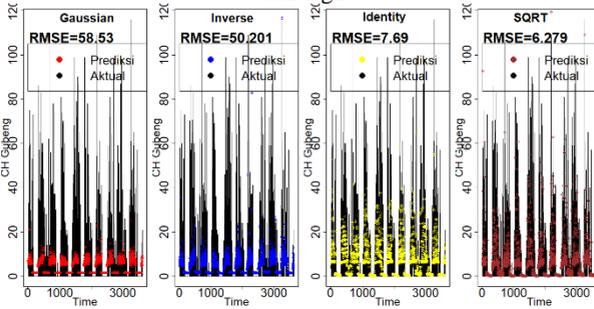


**Gambar 4.85.** Prediksi Curah Hujan Keputih dengan ketiga PC dan *dummy*

Penambahan variabel *dummy* mampu menurunkan nilai RMSE. Pada metode regresi Gamma “sqrt” memiliki nilai RMSE terkecil yaitu 12.365. Pada pemodelan dengan variabel PC<sub>1</sub> temperature memiliki RMSE sebesar 12.567. Sehingga akan dibandingkan kedua pemodelan tersebut dengan menambahkan lag1. Berikut hasil pemodelannya.



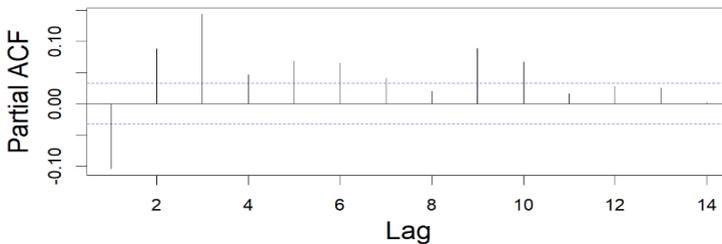
**Gambar 4.86.** Prediksi CH Keputih dengan variabel PC Temperatur, *dummy* dan lag1



**Gambar 4.87.** Prediksi CH Keputih dengan ketiga PC, *dummy* dan lag1

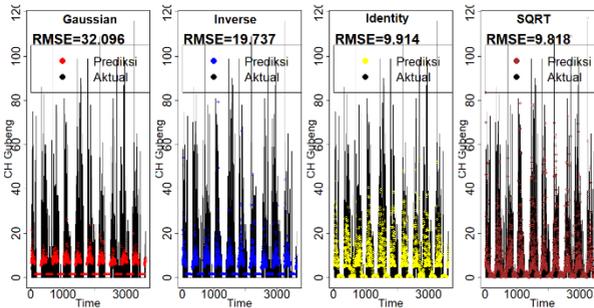
Kedua gambar diatas dapat dibandingkan bahwa dengan pemodelan curah hujan dengan variabel PC Temperatur memiliki kebaikan model yang lebih bagus daripada ketiga variabel hasil PC untuk semua metode. Model terbaik yaitu dengan menggunakan regresi gamma”sqrt” dengan nilai RMSE sebesar 6.028. Untuk mengetahui curah hujan yang berpengaruh terhadap curah hujan saat ini dapat dilihat dari nilai PACF dari residual yang terbentuk dari model sebelumnya. Berikut merupakan hasil dari plot PACF.

**Series Gamma\_sqrt**



**Gambar 4.88.** PACF Residual Regresi Gamma “sqrt”

Berdasarkan jumlah lag maksimal sebanyak 14 didapatkan hasil bahwa dengan metode regresi Gamma “sqrt” lag yang berpengaruh signifikan terhadap model yang terbentuk adalah yang keluar dari garis biru putus putus yaitu pada lag ke 1,2 dan lag3. Berikut merupakan pemodelan dari penambahan lag pada model yang terbentuk diatas.

**Gambar 4.89.** Prediksi Curah Hujan Keputih dengan PC Temperatur, *dummy*, lag 1 dan lag 2

Perbedaan yang signifikan terlihat saat penambahan lag2 yaitu hasil prediksi menjadi lebih besar. Meskipun dengan menambah lag mampu meningkatkan hasil prediksi semakin ke atas namun nilai RMSE yang terjadi menunjukkan kenaikan yang signifikan pula yaitu 9.818. Sedangkan model awal yang tanpa adanya lag2 memiliki RMSE sebesar 6.028. Kelebihan dari metode regresi Gamma “identity” dan “sqrt” yaitu mampu menangkap prediksi yang tinggi atau bernilai ekstrim. Karena setelah penambahan lag RMSE semakin tinggi maka pemodelan terbaik cukup sampai dengan lag1 yaitu variabel PC Temperatur, *dummy* dan lag1. Berikut merupakan hasil *summary* model terbaik.

**Tabel 4.50.** Summary Model Terbaik

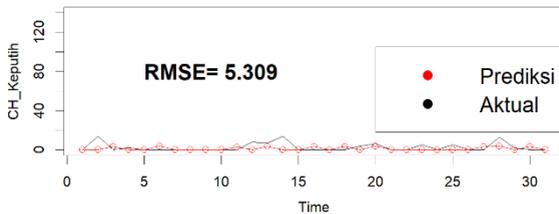
|                | Gamma(sqrt) |           |        |
|----------------|-------------|-----------|--------|
|                | estimate    | Std.Error | pvalue |
| Intercept      | 1.07667     | 0.04963   | 0.002  |
| PC_Temperature | -0.21501    | 0.01001   | 0.002  |
| <i>dummy</i>   | 1.36921     | 0.12943   | 0.002  |
| lag1           | 0.06463     | 0.01143   | 0.002  |

Berdasarkan hasil *summary* didapatkan bahwa dengan pengujian secara individu semua variabel berpengaruh signifikan

terhadap model. Hal ini dikarenakan nilai dari pvalue kurang dari nilai alpha yaitu 0.05. Nilai RMSE yang kecil menunjukkan model yang baik. Berikut merupakan model yang terbentuk.

$$Y_{1,t} = (-1.076 - 0.215PCX_{1,t} + 1.369Dummy + 0.057Y_{1,t-1})^2$$

Hasil akhir dari pemodelan terbaik yaitu dalam pengujian dari data testing. Data *testing* berjumlah 31 data yaitu merupakan tanggal 1 Desember 2018 sampai dengan 31 Desember 2018. Berikut merupakan hasil pengujian dari data *testing*. Pengujian data testing dikatakan baik jika nilai RMSE kecil dan prediksinya tidak jauh berbeda dengan data *training*. Pemodelan yang dilakukan yaitu dengan membuat variabel *dummy* harian pada setiap bulan Desember di tahun 2009 sampai 2018. Jika pada setiap tanggal 1 Desember 2009 sampai tahun 2018 terdapat lima kali hujan dikategorikan dengan 1 dan dilakukan rata rata curah hujan. Namun jika hujan kurang dari 5 dikategorikan dengan 0. Berikut hasil prediksinya.



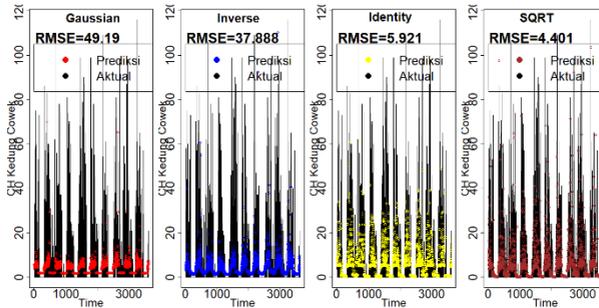
**Gambar 4.90** Prediksi Data Testing

Berdasarkan hasil pengujian data testing yang berasal dari pemodelan data training yang memiliki RMSE sebesar 6.379 didapatkan nilai RMSE yang lebih kecil yaitu sebesar 5.309. Hal ini berarti model yang terbentuk dari data training sudah sangat baik yaitu dengan menggunakan regresi Gamma fungsi link “sqrt”. Secara visual hasil prediksi terlihat konstan namun jika dilihat dari nilai RMSE sangat kecil menunjukkan hasil model yang baik.

#### 4.6.2 Pemodelan Multivariat Curah Hujan Kedung Cowek dengan variabel Stasiun BMKG

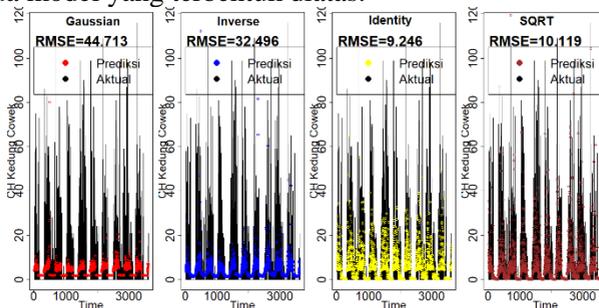
Pemodelan yang dilakukan yaitu dengan meregresikan variabel curah hujan Kedung Cowek sebagai respon dengan variabel stasiun BMKG sebagai prediktor. Dimana variabel

prediktornya sudah dilakukan pengelempokan dengan menggunakan analisis komponen utama (PCA). Sebagai prediktor tambahan yaitu variabel *dummy* dan variabel lag yang merupakan variabel  $Y_{t-k}$  yang digunakan untuk meningkatkan model. Berikut merupakan model terbaik yang terbentuk.



**Gambar 4.91** Prediksi Curah Hujan Kd. Coweikd dengan ketiga variabel PC, *dummy* dan lag1

Penambahan variabel *lag1* berpengaruh signifikan terhadap peningkatan hasil prediksi seperti terlihat pada gambar diatas. Nilai RMSE juga semakin turun yaitu sebesar 4.401 pada metode regresi Gamma “sqrt”. Pemodelan dengan menggunakan regresi Gamma “identity” dan “sart” mampu memberikan prediksi curah hujan yang baik dibandingkan dengan regresi Inverse Gaussian dan gamma “inverse”. Berikut merupakan pemodelan dari penambahan lag pada model yang terbentuk diatas.



**Gambar 4.92.** Prediksi CH dengan ketiga variabel PC, *dummy*, *lag1* dan *lag2*

Perbedaan yang signifikan terlihat saat penambahan *lag2* yaitu hasil prediksi menjadi lebih besar dengan RMSE sebesar 10.119. Kelebihan dari metode regresi Gamma “identity” dan

“sqrt” yaitu mampu menangkap prediksi yang tinggi atau bernilai ekstrim. Pengaruh dari penambahan *lag* yaitu mampu meningkatkan hasil prediksi. Karena setelah penambahan lag RMSE semakin meninggi maka pemodelan terbaik cukup sampai dengan lag1 yaitu variabel PC Temperatur, PC Kelembapan, PC1\_Angin, *dummy* dan lag1. Berikut merupakan hasil summary dari model terbaik.

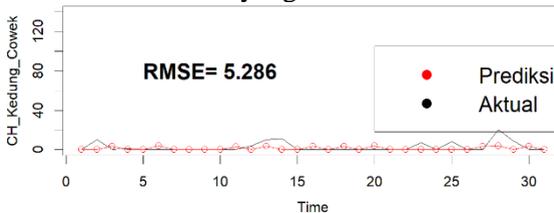
**Tabel 4.51.** Summary Model Terbaik

|                | estimasi | Std. error | Pvalue |
|----------------|----------|------------|--------|
| Intersep       | 0.9121   | 0.045      | 0.0001 |
| PC_Temperature | 0.0378   | 0.0028     | 0.0001 |
| PC_Kelembapan  | -0.0256  | 0.0015     | 0.0001 |
| PC1_Angin      | -0.0684  | 0.0061     | 0.0001 |
| dummy1         | 1.1796   | 0.1356     | 0.0001 |
| lag1           | 0.0793   | 0.0132     | 0.0001 |

Berdasarkan hasil *summary* didapatkan bahwa dengan pengujian secara individu semua variabel berpengaruh signifikan terhadap model. Hal ini dikarenakan nilai dari pvalue kurang dari nilai alpha yaitu 0.05. Nilai RMSE yang kecil menunjukkan model yang baik. Berikut merupakan model yang terbentuk.

$$Y_{1,t} = (0.912 - 0.038PCX_{1,t} - 0.026PCX_{2,t} + PC_1X_{3,t} + 1.179Dummy + 0.079Y_{1,t-1})^2$$

Setelah terbentuk model terbaik dari data training akan diuji dengan menggunakan data testing untuk mengetahui apakah data training sudah baik hasil prediksinya. Cara mengeceknya dengan melihat nilai RMSE yang dihasilkan oleh data testing.



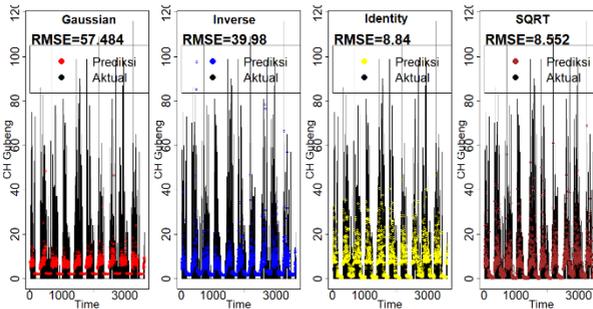
**Gambar 4.93** Pemodelan Data Testing

Secara visual hasil prediksi terlihat konstan. Pola data testing dengan data prediksi didapatkan hasil yang kurang tepat. Hasil pengujian data testing yang berasal dari pemodelan data training yang memiliki RMSE sebesar 9.246 didapatkan nilai

RMSE yang lebih kecil yaitu sebesar 5.286. Hal ini berarti model yang terbentuk dari data training sudah sangat baik.

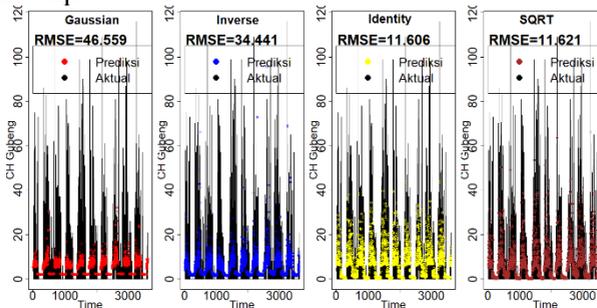
#### 4.6.3 Pemodelan Multivariat Curah Hujan Gubeng dengan variabel Stasiun BMKG

Pemodelan yang dilakukan yaitu dengan meregresikan variabel curah hujan Gubeng sebagai respon dengan variabel stasiun BMKG sebagai prediktor. Dimana variabel prediktornya sudah dilakukan pengelempokan dengan menggunakan analisis komponen utama (PCA). Sebagai prediktor tambahan yaitu variabel *dummy* dan variabel lag yang merupakan variabel  $Y_{t-k}$  yang digunakan untuk meningkatkan model. Berdasarkan dari masing masing PC yang terbentuk. Berikut merupakan hasil pemodelannya.



**Gambar 4.94** Prediksi CH Gubeng dengan ketiga PC, *dummy* dan lag1

Penambahan variabel *lag1* berpengaruh signifikan terhadap peningkatan hasil prediksi seperti terlihat pada gambar di atas. Nilai RMSE juga semakin turun yaitu sebesar 8.552 pada metode regresi Gamma “sqrt”.



**Gambar 4.95** Prediksi Curah Hujan Gubeng dengan PC, *dummy*, lag1 dan lag2

Perbedaan yang signifikan terlihat saat penambahan lag2 yaitu hasil prediksi menjadi lebih turun. Meskipun dengan menambah lag mampu meningkatkan hasil prediksi semakin ke atas namun nilai RMSE yang terjadi menunjukkan kenaikan yang signifikan pula yaitu 11.621. Sedangkan model awal yang tanpa adanya lag2 memiliki RMSE sebesar 8.552 pada metode Gamma "sqrt". Karena setelah penambahan lag RMSE semakin tinggi maka pemodelan terbaik cukup sampai dengan lag1 yaitu variabel PC Temperatur, PC Kelembaban, PC1\_Angin, *dummy* dan lag1. Berikut merupakan hasil summary dari model terbaik.

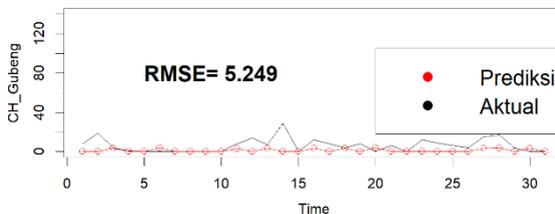
**Tabel 4.52.** Summary Model Terbaik Gamma "sqrt"

|                | estimasi | Std.eror | Pvalue |
|----------------|----------|----------|--------|
| Intersep       | 1.1776   | 0.0657   | 0.0001 |
| PC_Temperature | 0.0572   | 0.0095   | 0.0001 |
| PC_Kelembaban  | -0.0328  | 0.0024   | 0.0001 |
| PC1_Angin      | -0.0783  | 0.0186   | 0.0001 |
| <i>dummy</i> 1 | 1.3284   | 0.177    | 0.0001 |
| lag1           | 0.0449   | 0.0113   | 0.0001 |

Berdasarkan hasil *summary* didapatkan bahwa dengan pengujian secara individu semua variabel berpengaruh signifikan terhadap model. Hal ini dikarenakan nilai dari pvalue kurang dari nilai alpha yaitu 0.05. Nilai RMSE yang kecil menunjukkan model yang baik. Berikut merupakan model yang terbentuk.

$$Y_{1,t} = (1.177 - 0.057PCX_{1,t} - 0.032PCX_{2,t} - 0.078PC_1X_{3,t} + 1.328Dummy + 0.045Y_{1,t-1})^2$$

Setelah terbentuk model terbaik dari data training akan diuji dengan menggunakan data testing untuk mengetahui apakah data training sudah baik hasil prediksinya. Cara mengeceknya dengan melihat nilai RMSE yang dihasilkan oleh data testing.

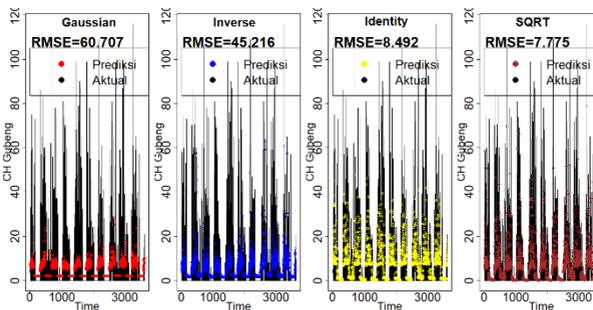


**Gambar 4.96** Pemodelan Data Testing (Gubeng)

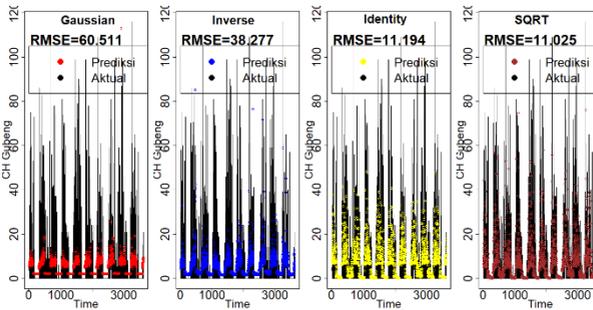
Secara visual hasil prediksi terlihat konstan. Pola data testing dengan data prediksi didapatkan hasil yang kurang tepat. Hasil pengujian data testing yang berasal dari pemodelan data training yang memiliki RMSE sebesar 6.223 didapatkan nilai RMSE yang lebih kecil yaitu sebesar 8.552. Hal ini berarti model yang terbentuk dari data training sudah sangat baik.

#### 4.6.4 Pemodelan Multivariat Curah Hujan Wonokromo dengan Variabel Stasiun BMKG

Pemodelan yang dilakukan yaitu dengan meregresikan variabel curah hujan Wonokromo sebagai respon dengan variabel stasiun BMKG sebagai prediktor. Dimana variabel prediktornya sudah dilakukan pengelempokan dengan menggunakan analisis komponen utama (PCA). Sebagai prediktor tambahan yaitu variabel *dummy* dan variabel lag yang merupakan variabel  $Y_{t-k}$  yang digunakan untuk meningkatkan model. Berdasarkan dari masing masing PC yang terbentuk. Berikut merupakan hasil pemodelannya.

**Gambar 4.97** Prediksi CH Wonokromo dengan ketiga PC, *dummy* dan lag1

Penambahan variabel *lag1* berpengaruh signifikan terhadap peningkatan hasil prediksi seperti terlihat pada gambar diatas. Nilai RMSE juga semakin turun yaitu sebesar 7.755 pada metode regresi Gamma “sqrt”. Berikut merupakan pemodelan dari penambahan lag pada model yang terbentuk diatas.



**Gambar 4.98.** Prediksi CH dengan semua PC, *dummy*, lag1 dan lag2

Perbedaan yang signifikan terlihat saat penambahan lag2 yaitu hasil prediksi menjadi lebih turun. Meskipun dengan menambah lag mampu meningkatkan hasil prediksi semakin ke atas namun nilai RMSE yang terjadi menunjukkan kenaikan yang signifikan pula yaitu 11.025. Sedangkan model awal yang tanpa adanya lag2 memiliki RMSE sebesar 7.755 pada metode Gamma "sqrt". Pengaruh dari penambahan lag yaitu mampu meningkatkan hasil prediksi. Setelah penambahan lag didapatkan RMSE semakin tinggi maka pemodelan terbaik cukup sampai dengan lag1 yaitu variabel PC Temperatur, PC Kelembaban, PC1\_Angin, *dummy* dan lag1. Berikut merupakan hasil summary dari model terbaik.

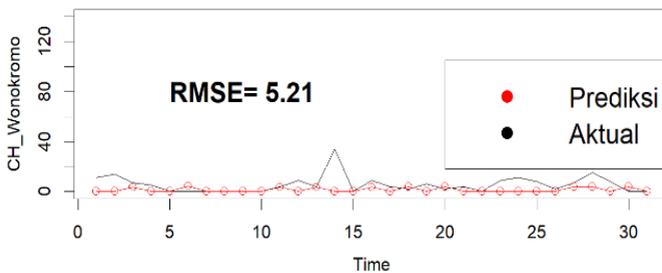
**Tabel 4.53.** Summary Model Terbaik Gamma "sqrt"

|                | estimasi | Std.eror | pvalue |
|----------------|----------|----------|--------|
| Intersep       | 1.0573   | 0.0429   | 0.0001 |
| PC_Temperature | 0.0338   | 0.0025   | 0.0001 |
| PC_Kelembapan  | -0.0312  | 0.0013   | 0.0001 |
| PC1_Angin      | -0.058   | 0.0043   | 0.0001 |
| PC2_Angin      | -0.0561  | 0.0088   | 0.0001 |
| <i>dummy</i> 1 | 1.4171   | 0.13     | 0.0001 |
| lag1           | 0.0516   | 0.0088   | 0.0001 |

Berdasarkan hasil *summary* didapatkan bahwa dengan pengujian secara individu semua variabel berpengaruh signifikan terhadap model. Hal ini dikarenakan nilai dari pvalue kurang dari nilai alpha yaitu 0.05. Nilai RMSE yang kecil menunjukkan model yang baik. Berikut merupakan model yang terbentuk.

$$Y_{1,t} = (1.057 - 0.034PCX_{1,t} - 0.0312PCX_{2,t} - 0.058PCX_{1,3,t} - 0.05PCX_{2,3,t} - 1.417Dummy + 0.05Y_{1,t-1})^2$$

Setelah terbentuk model terbaik dari data training akan diuji dengan menggunakan data testing untuk mengetahui apakah data training sudah baik hasil prediksinya. Cara mengeceknya dengan melihat nilai RMSE yang dihasilkan oleh data testing. Pemodelan yang dilakukan yaitu dengan membuat variabel *dummy* harian pada setiap bulan Desember di tahun 2009 sampai 2018. Jika pada setiap tanggal 1 Desember 2009 sampai tahun 2018 terdapat lima kali hujan dikategorikan dengan 1 dan dilakukan rata rata curah hujan. Namun jika hujan kurang dari 5 dikategorikan dengan 0. Berikut merupakan hasil pemodelannya.



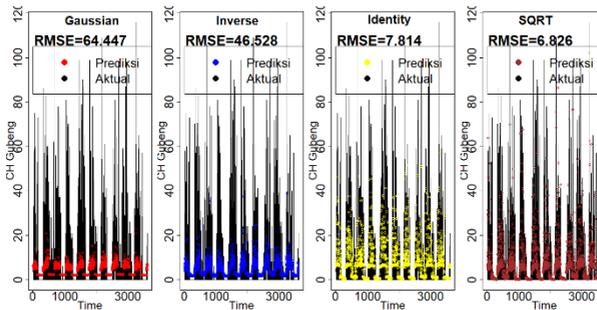
**Gambar 4.99.** Pemodelan Data Testing

Secara visual hasil prediksi terlihat konstan. Pola data testing dengan data prediksi didapatkan hasil yang kurang tepat. Hasil pengujian data testing yang berasal dari pemodelan data training yang memiliki RMSE sebesar 7.755 didapatkan nilai RMSE yang lebih kecil yaitu sebesar 5.21. Hal ini berarti model yang terbentuk dari data training sudah sangat baik.

#### 4.6.5 Pemodelan Multivariat Curah Hujan Wonorejo dengan variabel Stasiun BMKG

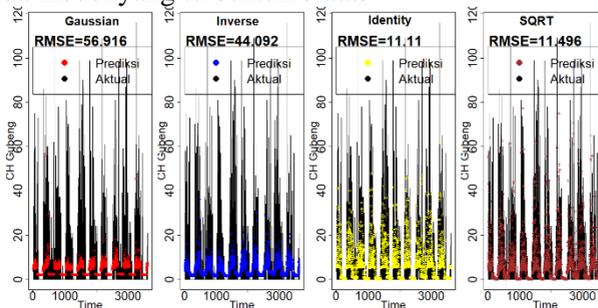
Pemodelan yang dilakukan yaitu dengan meregresikan variabel curah hujan Wonorejo sebagai respon dengan variabel stasiun BMKG sebagai prediktor. Dimana variabel prediktornya sudah dilakukan pengelempokan dengan menggunakan analisis komponen utama (PCA). Sebagai prediktor tambahan yaitu variabel *dummy* dan variabel lag yang merupakan variabel  $Y_{t-k}$

yang digunakan untuk meningkatkan model. Berikut merupakan hasil pemodelannya.



**Gambar 4.100** Prediksi CH dengan ketiga variabel PC, *dummy* dan lag1

Penambahan variabel *lag1* berpengaruh signifikan terhadap peningkatan hasil prediksi seperti terlihat pada gambar diatas. Nilai RMSE juga semakin turun yaitu sebesar 6.826 pada metode regresi Gamma “sqrt”. Berikut merupakan pemodelan dari penambahan lag pada model yang terbentuk diatas.



**Gambar 4.101.** Prediksi CH dengan Ketiga PC, *dummy*, lag 1 dan lag 2

Perbedaan yang signifikan terlihat saat penambahan lag2 yaitu hasil prediksi menjadi lebih turun. Meskipun dengan menambah lag mampu meningkatkan hasil prediksi semakin ke atas namun nilai RMSE yang terjadi menunjukkan kenaikan yang signifikan pula yaitu 11.496. Sedangkan model awal yang tanpa adanya lag2 memiliki RMSE sebesar 6.852 pada metode Gamma ”sqrt”. Pengaruh dari penambahan lag yaitu mampu meningkatkan hasil prediksi. Karena setelah penambahan lag RMSE samakin meninggi maka pemodelan terbaik cukup sampai dengan lag1 yaitu

variabel PC Temperatur, PC Kelembaban, PC1\_Angin, *dummy* dan lag1. Berikut merupakan hasil summary dari model terbaik.

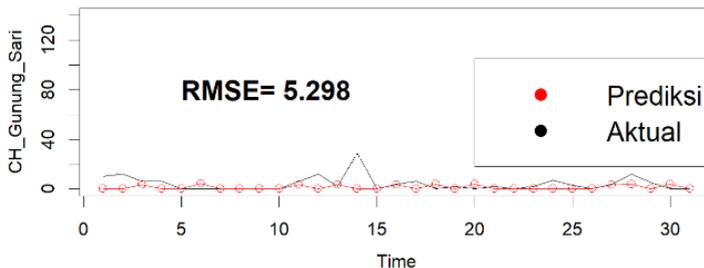
**Tabel 4.54.** Summary Model Terbaik Gamma "sqrt"

|                | estimasi | Std.eror | pvalue |
|----------------|----------|----------|--------|
| Intersep       | 1.0081   | 0.0458   | 0.0001 |
| PC_Temperature | 0.0326   | 0.0029   | 0.0001 |
| PC_Kelembaban  | -0.0303  | 0.0015   | 0.0001 |
| PC1_Angin      | -0.0567  | 0.0052   | 0.0001 |
| dummy1         | 1.2815   | 0.1352   | 0.0001 |
| lag1           | 0.0598   | 0.0105   | 0.0001 |

Berdasarkan hasil *summary* didapatkan bahwa dengan pengujian secara individu semua variabel berpengaruh signifikan terhadap model. Hal ini dikarenakan nilai dari pvalue kurang dari nilai alpha yaitu 0.05. Nilai RMSE yang kecil menunjukkan model yang baik. Berikut merupakan model yang terbentuk.

$$Y_{1,t} = (1.008 - 0.033PCX_{1,t} - 0.0303PCX_{2,t} - 0.0567PC_1X_{3,t} + 1.2815Dummy + 0.059Y_{1,t-1})^2$$

Setelah terbentuk model terbaik dari data training akan diuji dengan menggunakan data testing untuk mengetahui apakah data training sudah baik hasil prediksinya. Cara mengeceknya dengan melihat nilai RMSE yang dihasilkan oleh data testing.

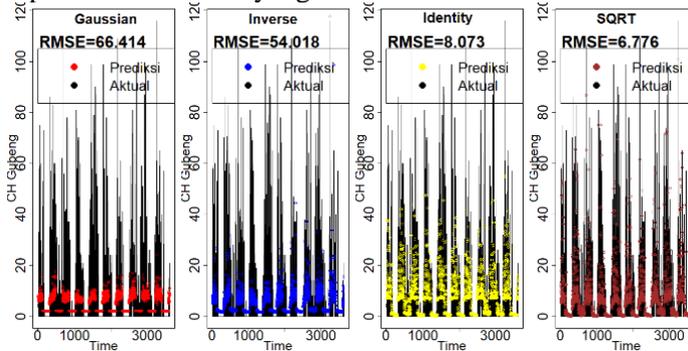


**Gambar 4.102** Prediksi Data Testing

Secara visual hasil prediksi terlihat konstan. Pola data testing dengan data prediksi didapatkan hasil yang kurang tepat. Hasil pengujian data testing yang berasal dari pemodelan data training yang memiliki RMSE sebesar 6.852 didapatkan nilai RMSE yang lebih kecil yaitu sebesar 5.298. Hal ini berarti model yang terbentuk dari data training sudah sangat baik

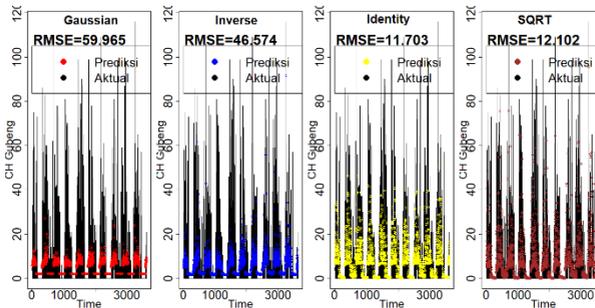
#### 4.6.6 Pemodelan Multivariat Curah Hujan Gunung Sari dengan variabel Stasiun BMKG

Pemodelan yang dilakukan yaitu dengan meregresikan variabel curah hujan Gunung Sari sebagai respon dengan variabel stasiun BMKG sebagai prediktor. Dimana variabel prediktornya sudah dilakukan pengelompokan dengan menggunakan analisis komponen utama (PCA). Sebagai prediktor tambahan yaitu variabel *dummy* dan variabel lag yang merupakan variabel  $Y_{t-k}$  yang digunakan untuk meningkatkan model. Berikut merupakan hasil pemodelan terbaik yang dihasilkan.



**Gambar 4.103** Prediksi Curah Hujan Gunung Sari dengan ketiga variabel PC, *dummy* dan lag1

Penambahan variabel *lag1* berpengaruh signifikan terhadap peningkatan hasil prediksi seperti terlihat pada gambar diatas. Nilai RMSE juga semakin turun yaitu sebesar 6.776 pada metode regresi Gamma “sqrt”. Penambahan *lag1* mampu meningkatkan prediksi yang cukup tinggi bila dibandingkan dengan pemodelan tanpa *lag1*. Pada penggunaan metode regresi inverse Gaussian dan regresi Gamma dengan inverse tidak bisa menaikkan hasil prediksi. Berbeda halnya dengan kedua metode lainnya yaitu regresi Gamma “Identity” dan Gamma “sqrt” yang bagus dalam hal prediksinya yaitu dapat menangkap prediksi yang sifatnya ekstrim. Berikut merupakan pemodelan dari penambahan lag pada model yang terbentuk diatas.



**Gambar 4.104.** Prediksi Curah Hujan Gunung Sari dengan Ketiga PC, *dummy*, *lag1* dan *lag2*

Perbedaan yang signifikan terlihat saat penambahan *lag2* yaitu hasil prediksi menjadi lebih turun. Meskipun dengan menambah lag mampu meningkatkan hasil prediksi semakin ke atas namun nilai RMSE yang terjadi menunjukkan kenaikan yang signifikan pula yaitu 12.102. Sedangkan model awal yang tanpa adanya *lag2* memiliki RMSE sebesar 6.776 pada metode Gamma “*sqrt*”. Kelebihan dari metode regresi Gamma “*identity*” dan “*sqrt*” yaitu mampu menangkap prediksi yang tinggi atau bernilai ekstrim. Pengaruh dari penambahan lag yaitu mampu meningkatkan hasil prediksi. Karena setelah penambahan lag RMSE semakin tinggi maka pemodelan terbaik cukup sampai dengan *lag1* yaitu variabel PC Temperatur, PC Kelembaban, PC1\_Angin, *dummy* dan *lag1*. Berikut merupakan hasil summary dari model terbaik.

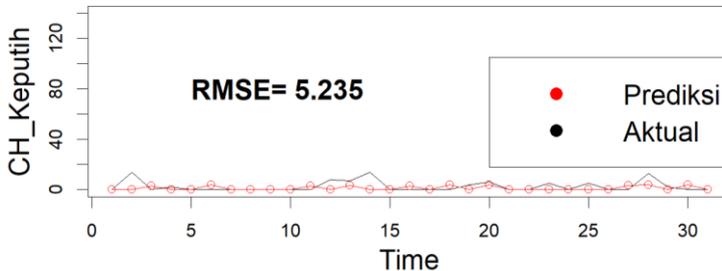
**Tabel 4.55.** Summary Model Terbaik Gamma “*sqrt*”

|                | Estimasi | Std.error | pvalue |
|----------------|----------|-----------|--------|
| Intersep       | 0.9884   | 0.0404    | 0.0001 |
| PC_Temperature | 0.0321   | 0.0019    | 0.0001 |
| PC_Kelembapan  | -0.0296  | 0.0013    | 0.0001 |
| PC1_Angin      | -0.0564  | 0.0032    | 0.0001 |
| <i>dummy1</i>  | 1.4405   | 0.1284    | 0.0001 |
| <i>lag1</i>    | 0.0592   | 0.0093    | 0.0001 |

Berdasarkan hasil *summary* didapatkan bahwa dengan pengujian secara individu semua variabel berpengaruh signifikan terhadap model. Hal ini dikarenakan nilai dari pvalue kurang dari nilai alpha yaitu 0.05. Nilai RMSE yang kecil menunjukkan model yang baik. Berikut merupakan model yang terbentuk.

$$Y_{1,t} = (0.9884 - 0.032PCX_{1,t} - 0.0296PCX_{2,t} - 0.0564PC_1X_{3,t} + 1.44Dummy + 0.059Y_{1,t-1})^2$$

Setelah terbentuk model terbaik dari data training akan diuji dengan menggunakan data testing untuk mengetahui apakah data training sudah baik hasil prediksinya. Cara mengeceknya dengan melihat nilai RMSE yang dihasilkan oleh data testing.



**Gambar 4.105.** Pemodelan Data *Testing*

Secara visual hasil prediksi terlihat konstan. Pola data testing dengan data prediksi didapatkan hasil yang kurang tepat. Hal ini dikarenakan banyaknya observasi data testing yang mengandung nilai 0. Hasil pengujian data testing yang berasal dari pemodelan data training yang memiliki RMSE sebesar 6.776 didapatkan nilai RMSE yang lebih kecil yaitu sebesar 5.235. Hal ini berarti model yang terbentuk dari data training sudah baik

#### 4.7. Pemilihan Model Terbaik

Pemilihan model terbaik berdasarkan nilai RMSE yang paling kecil atau minimum. Hasil dari ukuran kebaikan model ditampilkan dalam tabel dibawah ini yang melibatkan semua variabel prediktor dengan menambahkan dummy dan lag1.

**Tabel 4.56.** Summary RMSE Testing Setiap stasiun

| Balai     | Juanda | Perak1 | Perak2 | Multivariat |
|-----------|--------|--------|--------|-------------|
| Keputih   | 5.279  | 5.42   | 5.33   | 5.309       |
| Cowek     | 5.27   | 5.31   | 5.24   | 5.286       |
| Gubeng    | 5.18   | 5.31   | 5.321  | 5.249       |
| Wonorejo  | 5.256  | 5.326  | 5.314  | 5.298       |
| Wonokromo | 5.21   | 5.313  | 5.265  | 5.21        |
| Gn. Sari  | 5.206  | 5.3408 | 6.7    | 5.235       |

Berdasarkan tabel diatas dapat ditarik kesimpulan bahwa pemodelan secara univariat untuk masing masing stasiun di Juanda menghasilkan nilai RMSE yang terkecil. Sehingga pemilihan model terbaiknya adalah pada stasiun Juanda dengan Balai Keputih, Gubeng, Wonorejo, wonokromo, dan Gunung Sari. Sedangkan pada stasiun Perak 2 adalah pada balai Kd. Cowek.

#### 4.8 Prediksi Curah Hujan dan Banjir di Kota Surabaya

Pemodelan curah hujan terbaik didapatkan dengan menggunakan seluruh variabel prediktor yang berasal dari masing masing stasiun BMKG dengan pemodelan secara univariate. Berikut hasil curah hujan maksimum berdasarkan hasil prediksi model terbaik.

**Tabel 4.57.** Curah Hujan Maksimum

| Curah Hujan | Model Terbaik  | CH Prediksi Maksimum |
|-------------|----------------|----------------------|
| Keputih     | Stasiun Juanda | 3.971                |
| Kd. Cowek   | Stasiun Perak2 | 4.067                |
| Gubeng      | Stasiun Juanda | 4.161                |
| Wonorejo    | Stasiun Juanda | 4.020                |
| Wonokromo   | Stasiun Juanda | 4.055                |
| Gn. Sari    | Stasiun Juanda | 3.982                |

Curah hujan maksimum hasil prediksi tertinggi berada di balai Gubeng sebesar 4.161. Sedangkan Curah hujan terendah berada di Keputih sebesar 3.971. Berdasarkan curah hujan maksimum yang didapatkan dari hasil prediksi digunakan dalam prediksi banjir. Hidrolika dan hidrologi akan dilihat berdasarkan setiap saluran di setiap rayon. Berikut merupakan hasil prediksi banjir di masing masing balai.

##### 4.8.1 Prediksi Banjir di Keputih

Berdasarkan hasil prediksi Curah hujan maksimum yang telah diperoleh dari setiap balai.berikut merupakan hasil prediksi banjir dari masing masing saluran.

**Tabel 4.59.** Prediksi Banjir Balai Keputih

| Saluran                   | Debit (Hidrolika) | Debit (Hidrologi) | Keterangan |
|---------------------------|-------------------|-------------------|------------|
| Saluran HKSJN / Kalijudan | 0.865             | 0.054             | OK         |
| Saluran Kalijudan         | 8.251             | 0.035             | OK         |

|                          |         |       |    |
|--------------------------|---------|-------|----|
| Saluran Kalisari Timur   | 8.251   | 0.058 | OK |
| Saluran laban Sari       | 21.426  | 0.061 | OK |
| Saluran Larangan         | 63.549  | 0.067 | OK |
| Saluran Manyar Kertoadi  | 8.251   | 0.051 | OK |
| Saluran Manyar Sabrangan | 8.251   | 0.056 | OK |
| Saluran Mulyorejo        | 119.168 | 0.036 | OK |
| Saluran Ring Road ITS    | 35.023  | 0.059 | OK |
| Saluran Sutorejo         | 3.413   | 0.049 | OK |

Berdasarkan perhitungan matematis rasional, pada keadaan  $Q$  (debit) Hidrolika (saluran) lebih besar dari  $Q$  Hidrologi (curah hujan), maka diprediksi tidak berpotensi banjir. Sehingga dapat disimpulkan bahwa pada Desember 2018 saluran saluran di Keputih tidak berpotensi banjir.

#### 4.8.2 Prediksi Banjir di Gubeng

Berdasarkan hasil prediksi Curah hujan maksimum yang telah diperoleh dari setiap balai.berikut merupakan hasil prediksi banjir dari masing masing saluran.

**Tabel 4.58.** Prediksi Banjir Balai Gubeng

| Saluran                   | Debit (Hidrolika) | Debit (Hidrologi) | Keterangan |
|---------------------------|-------------------|-------------------|------------|
| Saluran Jeblokan          | 89.007            | 0.040             | OK         |
| Saluran Juwingan          | 16.247            | 0.057             | OK         |
| Saluran Kalibokor         | 284.374           | 0.035             | OK         |
| Saluran Kalidami          | 1281.770          | 0.058             | OK         |
| Saluran Kalisari THR      | 5.503             | 0.063             | OK         |
| Saluran Kamboja           | 5.503             | 0.060             | OK         |
| Saluran Kampung Seng      | 11.088            | 0.056             | OK         |
| Saluran Kedung Tarukan    | 176.320           | 0.065             | OK         |
| Saluran Kertajaya Viaduck | 21.426            | 0.080             | OK         |
| Saluran Menur             | 16.247            | 0.060             | OK         |
| Saluran Mojo              | 0.300             | 0.050             | OK         |
| Saluran Pacar             | 5.503             | 0.076             | OK         |
| Saluran Pucang Adi        | 3.413             | 0.063             | OK         |
| Saluran Pucang Jajar      | 11.932            | 0.054             | OK         |
| Saluran Pucang Rinenggo   | 1.842             | 0.057             | OK         |

|                              |        |       |    |
|------------------------------|--------|-------|----|
| Saluran Simokerto / Donorejo | 16.247 | 0.067 | OK |
| Saluran Simolawang           | 8.251  | 0.071 | OK |
| Saluran Slamet               | 3.413  | 0.056 | OK |
| Saluran Srikana              | 8.251  | 0.055 | OK |
| Saluran Sawah Pulo           | 5.503  | 0.066 | OK |
| Saluran Sidoluhur dsk        | 0.370  | 0.051 | OK |
| Saluran Sidotopo Lor         | 53.032 | 0.064 | OK |
| Saluran Tenggumung karya     | 0.865  | 0.034 | OK |
| Saluran Wonosari Lor         | 4.381  | 0.056 | OK |

Berdasarkan perhitungan matematis rasional, pada keadaan  $Q$  (debit) Hidrolika (saluran) lebih besar dari  $Q$  Hidrologi (curah hujan), maka diprediksi tidak berpotensi banjir. Sehingga dapat disimpulkan bahwa pada Desember 2018 saluran saluran di Gubeng tidak berpotensi banjir.

#### 4.8.3 Prediksi Banjir di Kd. Cowek

Berdasarkan hasil prediksi Curah hujan maksimum yang telah diperoleh dari setiap balai.berikut merupakan hasil prediksi banjir dari masing masing saluran.

**Tabel 4.59.** Prediksi Banjir Balai Kd. Cowek

| Saluran                     | Debit (Hidrolika) | Debit (Hidrologi) | Keterangan |
|-----------------------------|-------------------|-------------------|------------|
| Saluran Gading              | 89.007            | 0.054             | OK         |
| Saluran Jeblokan            | 103.401           | 0.047             | OK         |
| Saluran Kenjeran AL         | 103.401           | 0.064             | OK         |
| Saluran Kenjeran Lama       | 176.320           | 0.064             | OK         |
| Saluran Kenjeran Pantai Ria | 53.032            | 0.062             | OK         |
| Saluran Pegirian            | 307.968           | 0.039             | OK         |
| Saluran Kapas Madya         | 103.401           | 0.073             | OK         |
| Saluran Kedung Sroko        | 21.426            | 0.076             | OK         |
| Saluran Putra Agung         | 63.549            | 0.069             | OK         |
| Saluran Tambaksari          | 5.503             | 0.052             | OK         |
| Saluran Tambak Segaran      | 43.214            | 0.070             | OK         |
| Saluran Sidotopo Wetan      | 7.739             | 0.038             | OK         |

Berdasarkan perhitungan matematis rasional, pada keadaan  $Q$  (debit) Hidrolika (saluran) lebih besar dari  $Q$  Hidrologi (curah hujan), maka diprediksi tidak berpotensi banjir. Sehingga dapat disimpulkan bahwa pada Desember 2018 saluran saluran di Kd. Cowek tidak berpotensi banjir.

#### 4.8.4 Prediksi Banjir di Wonorejo

Berdasarkan hasil prediksi Curah hujan maksimum yang telah diperoleh dari setiap balai.berikut merupakan hasil prediksi banjir dari masing masing saluran.

**Tabel 4.60.** Prediksi Banjir Balai Wonorejo

| Saluran                     | Debit (Hidrolika) | Debit (Hidrologi) | Keterangan |
|-----------------------------|-------------------|-------------------|------------|
| Saluran Dupak Ps.Turi       | 1.773             | 0.042             | OK         |
| Saluran Kutilang PA.Pesapen | 3.413             | 0.073             | OK         |
| Saluran Ronggolawe          | 0.370             | 0.041             | OK         |
| Saluran Mojopahit           | 5.503             | 0.054             | OK         |
| Saluran Bubutan             | 17.315            | 0.073             | OK         |
| Saluran Embong Malang       | 21.426            | 0.073             | OK         |
| Saluran Komeri              | 5666.167          | 0.093             | OK         |
| Saluran Petemon Kuburan     | 1.042             | 0.041             | OK         |
| Saluran Pregolan Bunder     | 4.381             | 0.058             | OK         |
| Saluran Krembangan Jaya Sel | 4.615             | 0.063             | OK         |
| Saluran Ikan Mungsing       | 12.933            | 0.076             | OK         |
| Saluran Juwono              | 304463.214        | 0.109             | OK         |
| Saluran Embong Kemiri       | 37311.287         | 0.106             | OK         |
| Saluran Keputran            | 48.755            | 0.087             | OK         |
| Saluran Cempaka             | 7.498             | 0.063             | OK         |
| Saluran Ciliwung            | 3396.223          | 0.095             | OK         |
| Saluran Ikan Cucut          | 6.790             | 0.054             | OK         |
| Saluran Anwari              | 1586.754          | 0.082             | OK         |
| Saluran Sriwijaya           | 0.669             | 0.050             | OK         |
| Saluran Brawijaya           | 1.296             | 0.059             | OK         |
| Saluran Serayu              | 4.381             | 0.090             | OK         |

|                             |          |       |    |
|-----------------------------|----------|-------|----|
| Saluran Greges              | 3.413    | 0.075 | OK |
| Saluran Kali Mas            | 1.842    | 0.051 | OK |
| Saluran Krembangan Jaya Sel | 4.381    | 0.073 | OK |
| Saluran Ikan Mungsing       | 4.381    | 0.062 | OK |
| Saluran Juwono              | 8.251    | 0.077 | OK |
| Saluran Embong Kemiri       | 1.842    | 0.051 | OK |
| Saluran Keputran            | 4.381    | 0.090 | OK |
| Saluran Cempaka             | 4.381    | 0.085 | OK |
| Saluran Ciliwung            | 5.503    | 0.071 | OK |
| Saluran Ikan Cucut          | 4.381    | 0.061 | OK |
| Saluran Anwari              | 8.251    | 0.068 | OK |
| Saluran Sriwijaya           | 121.513  | 0.107 | OK |
| Saluran Brawijaya           | 3.722    | 0.061 | OK |
| Saluran Serayu              | 156.053  | 0.101 | OK |
| Saluran Greges              | 307.968  | 0.052 | OK |
| Saluran Kali Mas            | 1737.545 | 0.045 | OK |

Berdasarkan perhitungan matematis rasional, pada keadaan  $Q$  (debit) Hidrolika (saluran) lebih besar dari  $Q$  Hidrologi (curah hujan), maka diprediksi tidak berpotensi banjir. Sehingga dapat disimpulkan bahwa pada Desember 2018 saluran saluran di Wonorejo tidak berpotensi banjir.

#### 4.8.5 Prediksi Banjir di Wonokromo

Berdasarkan hasil prediksi Curah hujan maksimum yang telah diperoleh dari setiap balai.berikut merupakan hasil prediksi banjir dari masing masing saluran.

**Tabel 4.61.** Prediksi Banjir Balai Wonokromo

| Saluran                     | Debit (Hidrolika) | Debit (Hidrologi) | Keterangan |
|-----------------------------|-------------------|-------------------|------------|
| Saluran Margorejo           | 176.320           | 0.035             | OK         |
| Saluran Bendul Merisi Timur | 156.053           | 0.047             | OK         |
| Saluran Bendul Merisi Besar | 35.023            | 0.051             | OK         |
| Saluran Tepi Jl. A. Yani    | 16.247            | 0.035             | OK         |
| Saluran Jemursari Prapen    | 16.247            | 0.030             | OK         |

|                             |        |       |    |
|-----------------------------|--------|-------|----|
| Saluran Jemur Wonosari      | 18.725 | 0.055 | OK |
| Saluran Kutisari            | 24.682 | 0.045 | OK |
| Saluran Sarono Jiwo         | 1.842  | 0.041 | OK |
| Saluran Kali Rejo           | 1.842  | 0.043 | OK |
| Saluran Kali Rungkut        | 16.247 | 0.044 | OK |
| Saluran Tenggilis Mejoyo    | 16.247 | 0.043 | OK |
| Saluran Rungkut Kidul       | 2.592  | 0.053 | OK |
| Saluran Rungkut Asri Lor    | 24.682 | 0.051 | OK |
| Saluran Kedung Asem         | 4.381  | 0.037 | OK |
| Saluran Penjaringan Sari    | 4.381  | 0.033 | OK |
| Saluran Medokan Semampir    | 18.725 | 0.028 | OK |
| Saluran Semolowaru          | 8.251  | 0.034 | OK |
| Saluran Manyar              | 35.023 | 0.049 | OK |
| Saluran Manyar Rejo         | 4.381  | 0.048 | OK |
| Saluran Ngagel Wasono       | 21.426 | 0.053 | OK |
| Saluran Kalisumo            | 35.023 | 0.052 | OK |
| Saluran Ngagel Jaya Selatan | 1.842  | 0.045 | OK |
| Saluran Krukah              | 6.790  | 0.038 | OK |

Berdasarkan perhitungan matematis rasional, pada keadaan  $Q$  (debit) Hidrolika (saluran) lebih besar dari  $Q$  Hidrologi (curah hujan), maka diprediksi tidak berpotensi banjir. Sehingga dapat disimpulkan bahwa pada Desember 2018 saluran saluran di Wonokromo tidak berpotensi banjir.

#### 4.8.6 Prediksi Banjir di Gunung Sari

Berdasarkan hasil prediksi Curah hujan maksimum yang telah diperoleh dari setiap balai.berikut merupakan hasil prediksi banjir dari masing masing saluran.

**Tabel 4.62.** Prediksi Banjir Balai Gunung Sari

| Saluran             | Debit (Hidrolika) | Debit (Hidrologi) | Keterangan |
|---------------------|-------------------|-------------------|------------|
| Lidah wetan - Kulon | 89.007            | 0.064             | OK         |
| Jeruk               | 16.247            | 0.055             | OK         |
| Babadan Indah       | 1.842             | 0.044             | OK         |

|                       |            |       |    |
|-----------------------|------------|-------|----|
| Babadan Karang        | 11.932     | 0.051 | OK |
| Wiyung                | 27.879     | 0.051 | OK |
| Menganti              | 3.413      | 0.040 | OK |
| Gemol Kedurus         | 156.053    | 0.070 | OK |
| Gogor                 | 43.214     | 0.070 | OK |
| jajar Tunggal         | 31.324     | 0.059 | OK |
| Gunung Sari           | 53.032     | 0.051 | OK |
| Bogangin I            | 5.503      | 0.048 | OK |
| Bogangin Baru         | 2.592      | 0.038 | OK |
| Kebraon               | 6.790      | 0.044 | OK |
| Kebraon Manis         | 6.790      | 0.049 | OK |
| Prima Kebraon         | 11.932     | 0.053 | OK |
| Pondok Maritim        | 6.790      | 0.053 | OK |
| Karang Klumprik Barat | 11.932     | 0.051 | OK |
| Balas Klumprik        | 406711.775 | 0.088 | OK |

Berdasarkan perhitungan matematis rasional, pada keadaan  $Q$  (debit) Hidrolika (saluran) lebih besar dari  $Q$  Hidrologi (curah hujan), maka diprediksi tidak berpotensi banjir. Sehingga dapat disimpulkan bahwa pada Desember 2018 saluran saluran di Gunung Sari tidak berpotensi banjir.

*(Halaman ini sengaja dikosongkan)*

## **BAB V**

### **KESIMPULAN DAN SARAN**

#### **5.1 Kesimpulan**

Berdasarkan hasil analisis dan pembahasan maka dapat disimpulkan sebagai berikut: Karakteristik Curah Hujan selama sepuluh tahun mengindikasikan bahwa di bulan Mei sampai dengan Oktober menjadi bulan yang jarang terjadi hujan. Sedangkan pada Bulan November menjadi bulan yang sering turun hujan.

1. Karakteristik Curah Hujan selama sepuluh tahun mengindikasikan bahwa di bulan Mei sampai dengan Oktober menjadi bulan yang jarang terjadi hujan. Sedangkan pada Bulan November menjadi bulan yang sering turun hujan.
2. Pemodelan curah hujan Keputih dengan Stasiun Juanda didapatkan hasil terbaik yaitu dengan menggunakan metode regresi Gamma dengan fungsi link “sqrt” memiliki nilai RMSE 5.279 yaitu pemodelan antara curah hujan Keputih dengan semua variabel stasiun Juanda, dummy dan lag1.
3. Pemodelan curah hujan Kd.Cowek dengan Stasiun Perak2 didapatkan hasil terbaik yaitu dengan menggunakan metode regresi Gamma dengan fungsi link “sqrt” memiliki nilai RMSE 5.24 yaitu pemodelan antara curah hujan Keputih dengan semua variabel stasiun Juanda, dummy dan lag1.
4. Pemodelan curah hujan Gubeng dengan Stasiun Juanda didapatkan hasil terbaik yaitu dengan menggunakan metode regresi Gamma dengan fungsi link “sqrt” memiliki nilai RMSE 5.18 yaitu pemodelan antara curah hujan Keputih dengan semua variabel stasiun Juanda, dummy dan lag1.
5. Pemodelan curah hujan Wonorejo dengan Stasiun Juanda didapatkan hasil terbaik yaitu dengan menggunakan metode regresi Gamma dengan fungsi link “sqrt” memiliki nilai RMSE 5.256 yaitu pemodelan antara curah hujan Keputih dengan semua variabel stasiun Juanda, dummy dan lag1.
6. Pemodelan curah hujan Wonokromo dengan Stasiun Juanda didapatkan hasil terbaik yaitu dengan menggunakan metode regresi Gamma dengan fungsi link “sqrt” memiliki nilai

RMSE 5.21 yaitu pemodelan antara curah hujan Keputih dengan semua variabel stasiun Juanda, dummy dan lag1.

7. Pemodelan curah hujan Gunung Sari dengan Stasiun Juanda didapatkan hasil terbaik yaitu dengan menggunakan metode regresi Gamma dengan fungsi link “sqrt” memiliki nilai RMSE 5.206 yaitu pemodelan antara curah hujan Keputih dengan semua variabel stasiun Juanda, dummy dan lag1.
8. Berdasarkan curah hujan maksimum didapatkan hasil bahwa pada bulan Desember 2018, saluran saluran di kota Surabaya tidak berpotensi hujan.

## 5.2 Saran

Berdasarkan hasil analisis dan pembahasan yang telah dilakukan, maka saran yang dapat diberikan sebagai berikut.

1. Melakukan prediksi curah hujan dengan menggunakan metode yang berbasis *machine learning* karena pada metode ini mampu memprediksi hasil yang lebih akurat serta bisa menangkap hasil prediksi yang bersifat ekstrim.
2. Menambah Jumlah data *testing* untuk menguji Kebaikan dari model terbaik yang terbentuk.

## DAFTAR PUSTAKA

- Agresti, Alan. (2007). *An Introduction to Categorical Data Analysis Second Edition*. Florida: Departement of Statistics University of Florida.
- Chandler, R.E., Wheater, H.S., (2002). Analysis of rainfall variability using generalized linear models: a case study from the west of Ireland. *Water Resour. Res.* 38, 1192, <http://dx.doi.org/10.1029/2001WR000906>.
- Detiknews. (2019). Dipetik 14 Februari 2019, dari detiknews.com: <https://news.detik.com/berita-jawa-timur/4410154/banjir-semalam-masih-genangi-permukiman-warga-di-surabaya>.
- Detiknews. (2019). Dipetik 14 Februari 2019, dari detiknews.com: <https://news.detik.com/berita-jawa-timur/d-4410720/bpb-dan-linmas-surabaya-waspada-bencana-di-kota-pahlawan>
- Detiknews. (2019). Dipetik 14 Februari 2019, dari detiknews.com: <https://news.detik.com/berita-jawa-timur/3914854/hujan-tak-deras-sememi-terendam-air-gara-gara-box-culvert>.
- Dibiyosaputro, Suprpto. (1984). Flood Susceptibility And Hazard Survey Of The Kudus Pranata Welahan Area, Sentral Java. *Indonesia Tesis*, ITC, Enschede, Nadherlands.
- Draper, N. R., & Smith, H. (1992). *Analisis Regresi Terapan (2nd Edition)*. Alin Bahasa: I. B. Sumantri. Analisis Regresi Terapan (Edisi ke-2). Jakarta: PT Gramedia Pustaka Utama.
- Ewemoje, T. A., dan ewemooje, O. S. (2011). Best Distribution and Plotting Positions of Daily Maximum Flood Estimation at Ona River in Ogun-Oshun River Basin. *Agricultural Engineering International*, 13, 1-13.
- Garthwaite, P. H. (1994). An interpretation of partial least squares. *Journal of the American Statistical Association*, 89, 122-127.
- Hinrich Schütze; Christopher D. Manning. (1999). *Foundations of Statistical Natural Language Processing*. Cambridge, Mass: MIT Press. p. 304. ISBN 978-0-262-13360-9.
- Hogg, R. V., McKean, J. W., dan Craig, A. T. (2013). Introduction to Matematical Statistics 7 th ed. *Pearson Education*, Inc.

- Johnson, R.A. and Wichern, D.W. (2002). *Applied Multivariate Statistical Analysis*, 5th Edn. Prentice Hall. New Jersey.
- Jolliffe. (2002). *Principal Component Analysis*. Second Edition, Springer-Verlag.
- Jong, P. D. & Heller, G. Z. (2008). *Generalized Linear Models for Insurance Data*. Cambridge: Cambridge University Press.
- Kodoatie, R. J., & Sugiyanto. (2002). *Banjir, Beberapa Penyebab dan Metode Pengendaliannya dalam Prespektif Lingkungan*. Yogyakarta: Pustaka Pelajar.
- McCarthy, J., dan Maine, D. (1992). A Framework for Analyzing the Determinants of Maternal Mortality. *Studies in family Planning*, 69, 23-33.
- McCullagh, P. & Nelder, J. A. (1989). *Generalized Linear Models Edition*. London: Chapman & Hall.
- Nadarajah, S., dan Gupta, A. K. (2006). Some Bivariate Gamma Distributions. *Applied Mathematics* (19), 767-774.
- Putri, D. E., Purhadi, dan Prastyo, D. D. (2017). Parameter estimation and hypothesis testing on geographically weighted gamma regression. *Journal of Physics: Conference Series*, 1-7.
- Susanti, P. D., Miardini, A., dan Harjadi, B. (2017). Analisis kerentanan tanah longsor sebagai dasar mitigasi di Kabupaten Banjarnegara. *Jurnal Penelitian Pengelolaan Daerah Aliran Sungai*, 1, 49-59.
- Susilowati. (2010). Analisa Karakteristik Curah Hujan dan Kurva Intensitas Durasi Frekuensi (IDF) di Provinsi Lampung. *Rekayasa*, 14(1).
- Tripathi R C, Gupta C R, and Pair K P. (1993). Statistical Test Involving Several Independent Gamma Distribution *Ann. Inst. Statist. Math.* 45, 773-786.
- Upomo, T. C. (2016). Pemilihan Distribusi Probabilitas pada Analisa Hujan dengan Metode *Goodness of Fit Test*. *Teknik Sipil dan Perencanaan*, 18(2), 139-148.
- Wilks, D.S., Wilby, R.L., (1999). *The weather generation game: a review of stochastic weather models*. *Prog. Phys. Geogr.* 23, 329-357.
- Walpole, R. (2007). *Pengantar Statistika Edisi ke-3 Terjemahan Bambang Sumantri*. Jakarta: Gramedia.

## LAMPIRAN

### Lampiran 1. Syntax Penggabungan Data Stasiun Juanda, Perak 1 dan Perak2

```
setwd('E:/Juanda All')
files1 = list.files()

Juanda_R = NA
for(x in files1){
  BMKG1 = read_excel(x)
  Juanda_R = rbind(Juanda_R, BMKG1[9:39,])
}

Juanda_R      =      Juanda_R[rowSums(is.na(Juanda_R))      !=
ncol(Juanda_R), ]
colnames(Juanda_R)      =
c("Tanggal", "Tn", "Tx", "Tavg", "RH_avg", "RR", "ss", "ff_x", "ddd_
x", "ff_avg", "ddd_car")

Juanda_R = transform(Juanda_R, Tanggal = as.Date(Tanggal, "%d-
%m-%Y"),
  Tn = as.numeric(Tn),
  Tx = as.numeric(Tx),
  Tavg = as.numeric(Tavg),
  RH_avg = as.numeric(RH_avg),
  RR = as.numeric(RR),
  ss = as.numeric(ss),
  ff_x = as.numeric(ff_x),
  ff_avg = as.numeric(ff_avg),
  ddd_x = as.numeric(ddd_x))
```

### Lampiran 2. Syntax Time Series Plot Temperatur, Kelembapan dan Kecepatan Angin

```
Juanda_Temp = ggplot(data = DATA) + aes(y = Juanda_Tavg, x =
Tanggal) + ylim(24, 33) + labs(x = "Hari ke-", y = "Temperatur
Juanda") + geom_line(color = "#0c4c8a") + theme_minimal()
Perak2_Temp = ggplot(data = DATA) +
```

```

Perak2_Temp = ggplot(data = DATA) +
  aes(y = Perak2_Tavg, x = Tanggal) + ylim(24, 33) +
  labs(x = "Hari ke-", y = "Temperatur Perak 2") +
  geom_line(color = "#ef562d") +
  theme_minimal()
Perak1_Temp = ggplot(data = DATA) +
  aes(y = Perak1_Tavg, x = Tanggal) + ylim(24, 33) +
  labs(x = "Hari ke-", y = "Temperatur Perak 1") +
  geom_line(color = "#41ab5d") + theme_minimal()
win.graph()
plot_grid(Juanda_Temp,Perak2_Temp,Perak1_Temp, labels =
  NULL, nrow = 3, ncol = 1)          ddd_x =
as.numeric(ddd_x)

```

### Lampiran 3. *Syntax BoxPlot Curah Hujan*

```

Hari = strptime(DATA$Tanggal,"%d")
Bulan = strptime(DATA$Tanggal,"%b")
Tahun = strptime(DATA$Tanggal,"%Y")
month.abb =
c("Jan","Feb","Mar","Apr","May","Jun","Jul","Aug","Sep","
Oct","Nov","Dec")

# Curah Hujan Balai PSAWS
win.graph()
plot_grid(
  ggplot(data = DATA) +
    aes(x = factor(Bulan ,levels = month.abb), y =
DATA$CH_Keputih) +
  labs(x = "Bulan", y = "CH_Keputih") +
  ylim(0.00000,165)+
  geom_boxplot(color = "#0c4c8a", fill = "#9ecae1") +
  theme_minimal(),

```

```

ggplot(data = DATA) +
  aes(x = factor(Bulan ,levels = month.abb), y =
DATA$CH_KdCowek) +
  labs(x = "Bulan", y = "CH_KdCowek") +
  ylim(0.00000,165)+
  geom_boxplot(color = "#ef562d", fill = "#fec98d") +
  theme_minimal(),
ggplot(data = DATA) +
  aes(x = factor(Bulan ,levels = month.abb), y =
DATA$CH_Gubeng) +
  labs(x = "Bulan", y = "CH_Gubeng") +
  ylim(0.00000,165)+
  geom_boxplot(color = "#41ab5d", fill = "#c7e9c0") +
  theme_minimal(),
ggplot(data = DATA) +
  aes(x = factor(Bulan ,levels = month.abb), y =
DATA$CH_Wonorejo) +
  labs(x = "Bulan", y = "CH_Wonorejo") +
  ylim(0.00000,165)+
  geom_boxplot(color = "#0c4c8a", fill = "#9ecae1") +
  theme_minimal(),
ggplot(data = DATA) +
  aes(x = factor(Bulan ,levels = month.abb), y =
DATA$CH_Wonokromo) +
  labs(x = "Bulan", y = "CH_Wonokromo") +
  ylim(0.00000,165)+
  geom_boxplot(color = "#ef562d", fill = "#fec98d") +
  theme_minimal(),
ggplot(data = DATA) +
  aes(x = factor(Bulan ,levels = month.abb), y =
DATA$CH_GnSari) +
  labs(x = "Bulan", y = "CH_Gunungsari") +
  ylim(0.00000,165)+
  geom_boxplot(color = "#41ab5d", fill = "#c7e9c0") +
  theme_minimal(),
labels = NULL, nrow = 2, ncol = 3)

```

**Lampiran 4. Syntax ScatterPlot Curah Hujan dengan Prediktor**

```

win.graph()
a=ggscatter(DATA_APIK_FIX, y= "CH_Cowek__1", x =
"Juanda_T",
           color = "red", palette = c("#00AFBB"),shape = 21, size
= 1, # Points color, shape and size
           add = "reg.line", cor.method = "pearson",add.params =
list(color = "blue", fill = "green"),
           conf.int = TRUE, cor.coef.size = 4,ggtheme =
theme_pubr(),cor.coef = TRUE, cor.coeff.args = list(method =
"pearson", label.x = 3, label.sep = "\n"))
b=ggscatter(DATA_APIK_FIX, y= "CH_Cowek__1", x =
"Perak1_T",
           color = "green", palette = c("#00AFBB"),shape = 21,
size = 1, # Points color, shape and size
           add = "reg.line", cor.method = "pearson",add.params =
list(color = "blue", fill = "green"),
           conf.int = TRUE, cor.coef.size = 4,ggtheme =
theme_pubr(),cor.coef = TRUE, cor.coeff.args = list(method =
"pearson", label.x = 3, label.sep = "\n"))
c=ggscatter(DATA_APIK_FIX, y= "CH_Cowek__1", x =
"Perak2_T",
           color = "yellow", palette = c("#00AFBB"),shape = 21,
size = 1, # Points color, shape and size
           add = "reg.line", cor.method = "pearson",add.params =
list(color = "blue", fill = "green"),
           conf.int = TRUE, cor.coef.size = 4,ggtheme =
theme_pubr(),cor.coef = TRUE, cor.coeff.args = list(method =
"pearson", label.x = 3, label.sep = "\n"))
t= plot_grid(a,b,c,nrow=3)

```

### Lampiran 5. Syntax Pemodelan Curah Hujan dengan Juanda

```

DATA_d      <-      read_excel("C:/Users/Angga/Videos/tugas
akhir/DATA_d.xlsx")
DATA_d[14:20][DATA_d[14:20]== 0]<-0.00001
DATA=DATA_d
DATA=DATA[c(1:3621),]
Y1=DATA$CH_Keputih
#=====
# Y dengan Temperature
#=====
model=glm(Y1~DATA$Juanda_T,      inverse.gaussian(link      =
"1/mu^2"))
model1=glm(Y1~DATA$Juanda_T, Gamma(link="inverse"))
model2=glm(Y1~DATA$Juanda_T, Gamma(link="identity"))
model3=glm(Y1~DATA$Juanda_T, Gamma(link="sqrt"))
#=====
# Juanda Gabungan
#=====
model=glm(Y1~DATA$Juanda_RH+DATA$Juanda_T+DATA$Jua
nda_ff, inverse.gaussian(link = "1/mu^2"))
model1=glm(Y1~DATA$Juanda_RH+DATA$Juanda_T+DATA$Ju
anda_ff, Gamma(link="inverse"))
model2=glm(Y1~DATA$Juanda_RH+DATA$Juanda_T+DATA$Ju
anda_ff, Gamma(link="identity"))
model3=glm(Y1~DATA$Juanda_RH+DATA$Juanda_T+DATA$Ju
anda_ff, Gamma(link="sqrt"))
#=====
# Gabungan x1,x2,x3 +dummy
#=====
#Juanda (Temperature dan dummy)
mod=glm(Y1~DATA$Juanda_T+DATA$dummy,
Gamma(link="inverse"))
model=glm(Y1~DATA$Juanda_T+DATA$dummy,
inverse.gaussian(link = "1/mu^2"),start = coef(mod))
model1=glm(Y1~DATA$Juanda_T+DATA$dummy,
Gamma(link="inverse"))
model2=glm(Y1~DATA$Juanda_T+DATA$dummy,
Gamma(link="identity"))

```

```

#=====
# Dummy +lag1+ Juanda
#=====
Data1=cbind.data.frame(DATA$CH_Keputih,DATA$Juanda_T,DA
TA$Juanda_RH,DATA$Juanda_ff,DATA$dummy)
colnames(Data1)=c("CH_Keputih","Juanda_T","Juanda_RH","Juan
da_ff","dummy")
Data_lag=slide(Data1, Var = "CH_Keputih", slideBy = -c(1))
colnames(Data_lag)=c("Y1","Temperature","Kelembapan","kecep_
Angin","dummy","lag1")
dataa=Data_lag[2:3621,]
head(dataa)
model=glm(Y1~., inverse.gaussian(link = "1/mu^2"),dataa)
model1=glm(Y1~., Gamma(link="inverse"),dataa)
model2=glm(Y1~., Gamma(link="identity"),dataa)
model3=glm(Y1~., Gamma(link="sqrt"),dataa)
#=====
# Dummy +lag1+lag2 Juanda
#=====
Data1=cbind.data.frame(DATA$CH_Keputih,DATA$Juanda_T,DA
TA$Juanda_RH,DATA$Juanda_ff,DATA$dummy)
colnames(Data1)=c("CH_Keputih","Juanda_T","Juanda_RH","Juan
da_ff","dummy")
Data_lag=slide(Data1, Var = "CH_Keputih", slideBy = -c(1,2))
colnames(Data_lag)=c("Y1","Temperature","Kelembapan","kecep_
Angin","dummy","lag1","lag2")
dataa=Data_lag[3:3621,]
head(dataa)
model=glm(Y1~., inverse.gaussian(link = "1/mu^2"),dataa)
model1=glm(Y1~., Gamma(link="inverse"),dataa)
model2=glm(Y1~., Gamma(link="identity"),dataa)

```

### Lampiran 6. Syntax Pemodelan Curah Hujan dengan Perak1

```

#==== Perak 1
mm=glm(Y1~DATA$Perak1_T, Gamma(link="inverse"))
model=glm(Y1~DATA$Perak1_T, inverse.gaussian(link =
"1/mu^2"),start = coef(mm))
model1=glm(Y1~DATA$Perak1_T, Gamma(link="inverse"))
model2=glm(Y1~DATA$Perak1_T, Gamma(link="identity"))
model3=glm(Y1~DATA$Perak1_T,Gamma(link="sqrt"))
#=====
# Y dengan kelembapan
#=====
model=glm(Y1~DATA$Perak1_RH, inverse.gaussian(link =
"1/mu^2"))
model1=glm(Y1~DATA$Perak1_RH, Gamma(link="inverse"))
model2=glm(Y1~DATA$Perak1_RH, Gamma(link="identity"))
model3=glm(Y1~DATA$Perak1_RH, Gamma(link="sqrt"))
#Y~Kecepatan Angin
model=glm(Y1~DATA$Perak1_ff,inverse.gaussian(link=
"1/mu^2"))
model1=glm(Y1~DATA$Perak1_ff, Gamma(link="inverse"))
model2=glm(Y1~DATA$Perak1_ff, Gamma(link="identity"))
model3=glm(Y1~DATA$Perak1_ff, Gamma(link="sqrt"))
#=====
# Gabungan x1,x2,x3
#=====
library(tidyverse)
library(caret)
library(car)
#perak1
aw=glm(Y1~DATA$Perak1_T+DATA$Perak1_RH+DATA$Perak
1_ff, Gamma(link="inverse"))
model=glm(Y1~DATA$Perak1_T+DATA$Perak1_RH+DATA$Per
ak1_ff,start = coef(aw), inverse.gaussian(link = "1/mu^2"))
model1=glm(Y1~DATA$Perak1_T+DATA$Perak1_RH+DATA$P
erak1_ff, Gamma(link="inverse"))
model2=glm(Y1~DATA$Perak1_T+DATA$Perak1_RH+DATA$P
erak1_ff, Gamma(link="identity"))
model3=glm(Y1~DATA$Perak1_T+DATA$Perak1_RH+DATA$P
erak1_ff, Gamma(link="sqrt"))

```

```

#=====
# Gabungan x1,x2,x3 +dummy
#=====
aw=glm(Y1~DATA$Perak1_T+DATA$Perak1_RH+DATA$Perak1_ff+DATA$dummy, Gamma(link="inverse"))
model=glm(Y1~DATA$Perak1_T+DATA$Perak1_RH+DATA$Perak1_ff+DATA$dummy,start = coef(aw), inverse.gaussian(link = "1/mu^2"))
model1=glm(Y1~DATA$Perak1_T+DATA$Perak1_RH+DATA$Perak1_ff+DATA$dummy, Gamma(link="inverse"))
model2=glm(Y1~DATA$Perak1_T+DATA$Perak1_RH+DATA$Perak1_ff+DATA$dummy, Gamma(link="identity"))
model3=glm(Y1~DATA$Perak1_T+DATA$Perak1_RH+DATA$Perak1_ff+DATA$dummy, Gamma(link="sqrt"))
#=====
# Gabungan x1,x2,x3 +dummy+lag1
#=====
data1=cbind.data.frame(DATA$CH_Keputih,DATA$Perak1_T,DATA$Perak1_RH,DATA$Perak1_ff,DATA$dummy)
Data_lag=slide(data1, Var = "DATA$CH_Keputih", slideBy = -c(1))
colnames(Data_lag)=c("Y1", "Temperatur", "Kelembapan", "Kecepatan Angin", "dummy", "lag1")
dataa=Data_lag[2:3621,]
head(dataa)
mod=glm(Y1~., Gamma(link="inverse"),dataa)
model=glm(Y1~., inverse.gaussian(link = "1/mu^2"),dataa,start=coef(mod))
model1=glm(Y1~., Gamma(link="inverse"),dataa)
model2=glm(Y1~., Gamma(link="identity"),dataa)
model3=glm(Y1~., Gamma(link="sqrt"),dataa)

```

### Lampiran 7. Syntax Pemodelan Curah Hujan dengan Perak2

```

#==== Perak 2
mm=glm(Y1~DATA$Perak1_T, Gamma(link="inverse"))
model=glm(Y1~DATA$Perak2_T, inverse.gaussian(link = "1/mu^2"),start
= coef(mm))
model1=glm(Y1~DATA$Perak2_T, Gamma(link="inverse"))
model2=glm(Y1~DATA$Perak2_T, Gamma(link="identity"))
model3=glm(Y1~DATA$Perak2_T, Gamma(link="sqrt"))
# Perak 2
model=glm(Y1~DATA$Perak2_RH, inverse.gaussian(link = "1/mu^2"))
model1=glm(Y1~DATA$Perak2_RH, Gamma(link="inverse"))
model2=glm(Y1~DATA$Perak2_RH, Gamma(link="identity"))
model3=glm(Y1~DATA$Perak2_RH, Gamma(link="sqrt"))
# Perak 2
model=glm(Y1~DATA$Perak2_ff, inverse.gaussian(link = "1/mu^2"))
model1=glm(Y1~DATA$Perak2_ff, Gamma(link="inverse"))
model2=glm(Y1~DATA$Perak2_ff, Gamma(link="identity"))
model3=glm(Y1~DATA$Perak2_ff, Gamma(link="sqrt"))
#perak2
aw=glm(Y1~DATA$Perak2_T+DATA$Perak2_RH+DATA$Perak2_ff,
Gamma(link="inverse"))
model=glm(Y1~DATA$Perak2_T+DATA$Perak2_RH+DATA$Perak2_f
f,start = coef(aw), inverse.gaussian(link = "1/mu^2"))
model1=glm(Y1~DATA$Perak2_T+DATA$Perak2_RH+DATA$Perak2
_ff, Gamma(link="inverse"))
model2=glm(Y1~DATA$Perak2_T+DATA$Perak2_RH+DATA$Perak2
_ff, Gamma(link="identity"))
model3=glm(Y1~DATA$Perak2_T+DATA$Perak2_RH+DATA$Perak2
_ff, Gamma(link="sqrt"))
#Perak 2
aw=glm(Y1~DATA$Perak2_T+DATA$Perak2_RH+DATA$Perak2_ff+
DATA$dummy, Gamma(link="inverse"))
model=glm(Y1~DATA$Perak2_T+DATA$Perak2_RH+DATA$Perak2_f
f+DATA$dummy,start = coef(aw), inverse.gaussian(link = "1/mu^2"))
model1=glm(Y1~DATA$Perak2_T+DATA$Perak2_RH+DATA$Perak2
_ff+DATA$dummy, Gamma(link="inverse"))
model2=glm(Y1~DATA$Perak2_T+DATA$Perak2_RH+DATA$Perak2
_ff+DATA$dummy, Gamma(link="identity"))
model3=glm(Y1~DATA$Perak2_T+DATA$Perak2_RH+DATA$Perak2
_ff+DATA$dummy, Gamma(link="sqrt"))

```

```

#=====
#   PC Peerak 2
#=====
databaru_Temperatur1=as.data.frame(cbind(DATAS$Perak2_T,DAT
AS$Perak2_RH,DATAS$Perak2_ff))
data.pca_Juanda <- prcomp(databaru_Temperatur1[,1:3])
summary(data.pca_Juanda)
data.pca_Juanda
result_Perak2=as.data.frame(data.pca_Juanda$x)
PC1_Perak2= result_Perak2[,1]
#===== PC2_perak
model=glm(Y1~PC1_Perak2+DATA$dummy, inverse.gaussian(link
= "1/mu^2"))
model1=glm(Y1~PC1_Perak2+DATA$dummy,
Gamma(link="inverse"))
model2=glm(Y1~PC1_Perak2+DATA$dummy,
Gamma(link="identity"))
model3=glm(Y1~PC1_Perak2+DATA$dummy,
Gamma(link="sqrt"))
databaru_Angin1=as.data.frame(cbind(DATAS$Perak2_T,DATAS$Pe
rak2_RH,DATAS$Perak2_ff))
data.pca_Perak2 <- prcomp(databaru_Angin1[,1:3])
summary(data.pca_Perak2)
data.pca_Perak2
result_Perak2=as.data.frame(data.pca_Perak2$x)
PC1_Perak2= result_Perak2[,1]
data1=cbind.data.frame(DATAS$CH_Keputih,PC1_Perak2,DATAS$d
ummy)
Data_lag=slide(data1, Var = "DATAS$CH_Keputih", slideBy = -c(1))
colnames(Data_lag)=c("Y1","perak2","dummy","lag1")
dataa=Data_lag[2:3621,]
head(dataa)
model=glm(Y1~., inverse.gaussian(link = "1/mu^2"),dataa)
model1=glm(Y1~., Gamma(link="inverse"),dataa)
model2=glm(Y1~., Gamma(link="identity"),dataa)
model3=glm(Y1~., Gamma(link="sqrt"),dataa)

```

### Lampiran 8. *Syntax Pemodelan Curah Hujan PC*

```

##PC KELEMBAPAN
databaru_kelembapan=as.data.frame(cbind(DATA$Juanda_RH,DATA$Perak1_RH,DATA$Perak2_RH))
data.pca_RH <- prcomp(databaru_kelembapan[,1:3])
summary(data.pca_RH)
result1=as.data.frame(data.pca_RH$x)
PC1_RH= result1[,1]
model=glm(Y1~PC1_RH, inverse.gaussian(link = "1/mu^2"))
model1=glm(Y1~PC1_RH, Gamma(link="inverse"))
model2=glm(Y1~PC1_RH, Gamma(link="identity"))
model3=glm(Y1~PC1_RH, Gamma(link="sqrt"))
## Temperature
databaru_Temperatur=as.data.frame(cbind(DATA$Juanda_T,DATA$Perak2_T,DATA$Perak1_T))
data.pca_T <- prcomp(databaru_Temperatur[,1:3])
summary(data.pca_T)
result3=as.data.frame(data.pca_T$x)
PC1_T= result3[,1]
mop=glm(Y1~PC1_T, Gamma(link="inverse"))
model=glm(Y1~PC1_T, inverse.gaussian(link = "1/mu^2"),start = coef(mop))
model1=glm(Y1~PC1_T, Gamma(link="inverse"))
model2=glm(Y1~PC1_T, Gamma(link="identity"))
model3=glm(Y1~PC1_T, Gamma(link="sqrt"))
## Kecepatan angin
databaru_angin=as.data.frame(cbind(DATA$Juanda_ff,DATA$Perak2_ff,DATA$Perak1_ff))
data.pca_angin <- prcomp(databaru_angin[,1:3])
summary(data.pca_angin)
data.pca_angin
result5=as.data.frame(data.pca_angin$x)
PC1_a= result5[,1]
PC2_a= result5[,2]
model=glm(Y1~PC1_a+PC2_a, inverse.gaussian(link = "1/mu^2"),start = coef(mop))
model1=glm(Y1~PC1_a+PC2_a, Gamma(link="inverse"))
model2=glm(Y1~PC1_a+PC2_a, Gamma(link="identity"))
model3=glm(Y1~PC1_a+PC2_a, Gamma(link="sqrt"))

```

```

mopp=glm(Y1~PC1_T+PC1_RH+PC1_a+PC2_a,
Gamma(link="inverse"))
model=glm(Y1~PC1_T+PC1_RH+PC1_a+PC2_a,
inverse.gaussian(link = "1/mu^2"),start = coef(mopp))
model1=glm(Y1~PC1_T+PC1_RH+PC1_a+PC2_a,
Gamma(link="inverse"))
model2=glm(Y1~PC1_T+PC1_RH+PC1_a+PC2_a,
Gamma(link="identity"))
model3=glm(Y1~PC1_T+PC1_RH+PC1_a+PC2_a,
Gamma(link="sqrt"))
# gabungan +dummy
mopp=glm(Y1~PC1_T+PC1_RH+PC1_a+PC2_a+DATA$dummy,
Gamma(link="inverse"))
model=glm(Y1~PC1_T+PC1_RH+PC1_a+PC2_a+DATA$dummy
, inverse.gaussian(link = "1/mu^2"),start = coef(mopp))
model1=glm(Y1~PC1_T+PC1_RH+PC1_a+PC2_a+DATA$dumm
y, Gamma(link="inverse"))
model2=glm(Y1~PC1_T+PC1_RH+PC1_a+PC2_a+DATA$dumm
y, Gamma(link="identity"))
model3=glm(Y1~PC1_T+PC1_RH+PC1_a+PC2_a+DATA$dumm
y, Gamma(link="sqrt"))
# gabungan +dummy+lag
data1=cbind.data.frame(DATA$CH_Keputih,PC1_T,PC1_RH,PC1
_a,PC2_a,DATA$dummy)
Data_lag=slide(data1, Var = "DATA$CH_Keputih", slideBy = -
c(1))
colnames(Data_lag)=c("Y1","PC_Temperature","PC_Kelembapan
","PC1_Angin","PC2_Angin","dummy","lag1")
dataa=Data_lag[2:3621,]
head(dataa)
mopp=glm(Y1~., Gamma(link="inverse"),dataa)
model=glm(Y1~., inverse.gaussian(link = "1/mu^2"),start =
coef(mopp),dataa)
model1=glm(Y1~., Gamma(link="inverse"),dataa)
model2=glm(Y1~., Gamma(link="identity"),dataa)
model3=glm(Y1~.,
Gamma(link="sqrt"),dataa)model3=glm(Y1~PC1_a+PC2_a,
Gamma(link="sqrt"))

```

### Lampiran 9. Syntax Prediksi data Testing

```

dataaa=DATA_d[3622:3652,]
data=Prediksi_x
x15=data$P1_Temperature
x25=data$P1_Kelembapan
x35=data$P1_kecep_Angin
Y1=dataaa$CH_Keputih
dummy=dataaa$dummy

baru = cbind.data.frame(Y1,x15,x25,x35,dummy)
colnames(baru)=c("Y1","Temperatur","Kelembapan","Kecep_Angin",
"dummy")
datatest=cbind.data.frame(baru[1,],lag1)
colnames(datatest)=c("Y1","Temperatur","Kelembapan","Kecep_Angin",
"dummy","lag1")
head(datatest)
pred = matrix(ncol=1,nrow=31)
test = predict(model3, datatest)
pred[1,]= test
for (i in 2:31){
  Model = cbind(baru[i,],test)
  datatest = Model
  test = predict(model3, datatest)
  pred[i,]= test
}
data=cbind.data.frame(Y1,dummy,x15,x25,x35,pred)
colnames(data)=c("Y1","dummy","Temperatur","Kelembapan","Kecep_Angin",
"lag1")
test=predict(model3,data)
win.graph()
ts.plot(dataaa$CH_Keputih)
yhat=test
points(test,col="red")
text(15,12,font = 2, adj = 0, cex = 2,
paste0("RMSE=", round(RMSE(test, Y1),3)))

```

**Lampiran 10.** *Syntax Prediksi data Testing Multivariat PC*

```

dataaa=DATA_d[3622:3652,]
data=Prediksi_x
databaru_kelembapan=as.data.frame(cbind(data$P2_Temperatur
e,data$P1_Kelembapan,data$P2_kecep_Angin))
data.pca_RH <- prcomp(databaru_kelembapan[,1:3])
summary(data.pca_RH)
data.pca_RH
result1=as.data.frame(data.pca_RH$x)
PC1= result1[,1]

Y1=dataaa$CH_Keputih
dummy=dataaa$dummy

baru = cbind.data.frame(Y1,PC1,dummy)
colnames(baru)=c("Y1","perak2","dummy")
datatest=cbind.data.frame(baru[,1,],lag1)
colnames(datatest)=c("Y1","perak2","dummy","lag1")
head(datatest)
pred = matrix(ncol=1,nrow=31)
test = predict(model3, datatest)
pred[1,]= test

for (i in 2:31){
  Model = cbind(baru[i,],test)
  datatest = Model
  test = predict(model3, datatest)
  pred[i,]= test
}

data=cbind.data.frame(Y1,dummy,PC1,pred)
colnames(data)=c("Y1","dummy","perak2","lag1")
test=predict(model3,data)

win.graph()
ts.plot(dataaa$CH_Keputih)
yhat=test
points(test,col="red")
text(15,12,font = 2, adj = 0, cex = 2,
     paste0("RMSE=", round(RMSE(test,Y1),3)))

```

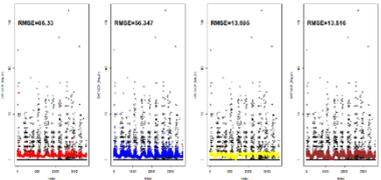
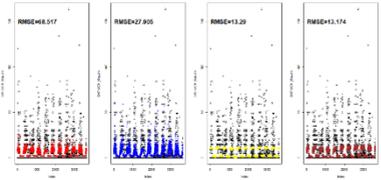
### Lampiran 11. Syntax output Pemodelan

```

library(DescTools)
win.graph()
par(mfrow=c(1,4))
#Model
plot(DATA$CH_Keputih,type = "p")
points(model$fitted.values,col="red")
text(30,150,font = 2, adj = 0, cex = 2,
     paste0("RMSE=", round(RMSE(model$fitted.values,Y1),3)))
#Model1
plot(DATA$CH_Keputih,type = "p")
points(model1$fitted.values,col="blue")
text(30,150,font = 2, adj = 0, cex = 2,
     paste0("RMSE=", round(RMSE(model1$fitted.values,Y1),3)))
#Model2
plot(DATA$CH_Keputih,type = "p")
points(model2$fitted.values,col="yellow")
text(30,150,font = 2, adj = 0, cex = 2,
     paste0("RMSE=", round(RMSE(model2$fitted.values,Y1),3)))
#Model3
plot(DATA$CH_Keputih,type = "p")
points(model3$fitted.values,col="red")
text(30,150,font = 2, adj = 0, cex = 2,
     paste0("RMSE=", round(RMSE(model3$fitted.values,Y1),3)))

```

### Lampiran 12. Prediksi Curah Hujan Gubeng

| Output  | Model                  | Stasiun |
|---|------------------------|---------|
|  | Semua Variabel         | Juanda  |
|  | Semua Variabel + dummy |         |

|  |                                      |                       |
|--|--------------------------------------|-----------------------|
|  | <p><b>Semua Variabel</b></p>         | <p><b>Perak 1</b></p> |
|  | <p><b>Semua Variabel + dummy</b></p> |                       |
|  | <p><b>Semua Variabel</b></p>         | <p><b>Perak 2</b></p> |
|  | <p><b>Semua Variabel + dummy</b></p> |                       |

**Lampiran 13. Prediksi Curah Hujan Kedung Cowek**

| <p><b>Output</b></p> | <p><b>Model</b></p>                  | <p><b>Stasiun</b></p> |
|----------------------|--------------------------------------|-----------------------|
|                      | <p><b>Semua Variabel</b></p>         | <p><b>Juanda</b></p>  |
|                      | <p><b>Semua Variabel + dummy</b></p> |                       |

|  |                                      |                       |
|--|--------------------------------------|-----------------------|
|  | <p><b>Semua Variabel</b></p>         | <p><b>Perak 1</b></p> |
|  | <p><b>Semua Variabel + dummy</b></p> |                       |
|  | <p><b>Semua Variabel</b></p>         | <p><b>Perak 2</b></p> |
|  | <p><b>Semua Variabel + dummy</b></p> |                       |

**Lampiran 14. Prediksi Curah Hujan Wonorejo**

| <p><b>Output</b></p> | <p><b>Model</b></p>          | <p><b>Stasiun</b></p> |
|----------------------|------------------------------|-----------------------|
|                      | <p><b>Semua Variabel</b></p> | <p><b>Juanda</b></p>  |

|  |                                      |                       |
|--|--------------------------------------|-----------------------|
|  | <p><b>Semua Variabel + dummy</b></p> |                       |
|  | <p><b>Semua Variabel</b></p>         | <p><b>Perak 1</b></p> |
|  | <p><b>Semua Variabel + dummy</b></p> |                       |
|  | <p><b>Semua Variabel</b></p>         | <p><b>Perak 2</b></p> |
|  | <p><b>Semua Variabel + dummy</b></p> |                       |

**Lampiran 15. Prediksi Curah Hujan Wonokromo**

| Output | Model                                | Stasiun               |
|--------|--------------------------------------|-----------------------|
|        | <p><b>Semua Variabel</b></p>         | <p><b>Juanda</b></p>  |
|        | <p><b>Semua Variabel + dummy</b></p> |                       |
|        | <p><b>Semua Variabel</b></p>         | <p><b>Perak 1</b></p> |
|        | <p><b>Semua Variabel + dummy</b></p> |                       |
|        | <p><b>Semua Variabel</b></p>         | <p><b>Perak 2</b></p> |

|  |                                      |  |
|--|--------------------------------------|--|
|  | <p><b>Semua Variabel + dummy</b></p> |  |
|--|--------------------------------------|--|

**Lampiran 16. Prediksi Curah Hujan Gunung Sari**

| Output | Model                                | Stasiun               |
|--------|--------------------------------------|-----------------------|
|        | <p><b>Semua Variabel</b></p>         | <p><b>Juanda</b></p>  |
|        | <p><b>Semua Variabel + dummy</b></p> |                       |
|        | <p><b>Semua Variabel</b></p>         | <p><b>Perak 1</b></p> |
|        | <p><b>Semua Variabel + dummy</b></p> |                       |

|  |                                      |                       |
|--|--------------------------------------|-----------------------|
|  | <p><b>Semua Variabel</b></p>         | <p><b>Perak 2</b></p> |
|  | <p><b>Semua Variabel + dummy</b></p> |                       |

**Lampiran 17.** Surat Pengambilan Data Badan Meteorologi  
Klimatologi dan Geofisika (BMKG)

**SURAT PERNYATAAN**

Saya yang bertanda tangan di bawah ini, mahasiswa Departemen  
Statistika FMKSD ITS:

Nama : Angga Mahendra

NRP : 06211540000035

menyatakan bahwa data yang digunakan dalam Tugas Akhir/ Thesis ini  
merupakan data sekunder yang diambil dari penelitian / buku/ Tugas  
Akhir/ Thesis/ publikasi / lainnya yaitu:

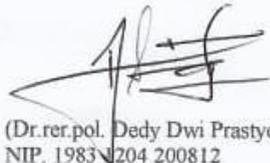
Sumber : <http://dataonline.bmkg.go.id>

Keterangan : Data iklim BMKG (Temperatur rata rata,  
Kelembaban rata rata, dan Kecepatan Angin rata  
rata) dari tahun 1 Januari 2009 sampai dengan 31  
Desember 2018 pada stasiun Juanda, Perak1 dan  
Perak 2 Surabaya.

Surat Pernyataan ini dibuat dengan sebenarnya. Apabila terdapat  
pemalsuan data maka saya siap menerima sanksi sesuai aturan yang  
berlaku.

Mengetahui  
Pembimbing Tugas Akhir

Surabaya,



(Dr.rer.pol. Dedy Dwi Prastyo, S.Si.,M.Si.)  
NIP. 19830204 200812

(Angga Mahendra)  
NRP. 06211540000035

\*(coret yang tidak perlu)

## Lampiran 18. Surat Pengambilan Data Dinas Pekerjaan Umum BinaMarga dan Pematusan

### SURAT KETERANGAN

Saya yang bertanda tangan di bawah ini menerangkan bahwa :

1. Mahasiswa Statistika FMKSD-ITS dengan identitas berikut :

Nama : Angga Mahendra

NRP : 06211540000035

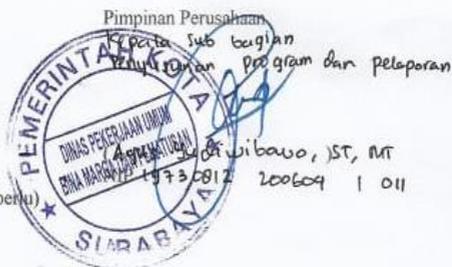
Telah mengambil data di instansi/perusahaan kami :

Nama Instansi : Dinas PU Bina Marga dan Pematusan

Divisi/ bagian : *Penyusunan Program dan Pelaporan*

sejak tanggal *1 Maret 2019* sampai dengan *30 April 2019* untuk keperluan Tugas Akhir/ ~~Thesis~~ Semester Gasal/Genap\* 20 / 20 .

2. Tidak Keberatan/Keberatan\* nama perusahaan dicantumkan dalam Tugas Akhir/ ~~Thesis~~-mahasiswa Statistika yang akan di simpan di Perpustakaan ITS dan dibaca di lingkungan ITS.
3. Tidak Keberatan/Keberatan\* bahwa hasil analisis data dari perusahaan dipublikasikan dalam E journal ITS yaitu Jurnal Sains dan Seni ITS.



\*(coret yang tidak perlu)

*(Halaman ini sengaja dikosongkan)*

## BIODATA PENULIS



Penulis yang bernama lengkap Angga Mahendra lahir di Kabupaten Lumajang, Jawa Timur pada tanggal 17 Juli 1996. Penulis merupakan anak pertama dari dua bersaudara. Pendidikan formal yang di-tempuh oleh penulis bermula pada tahun 2004-2010 yang ditempuh di SDN Kunir Kidul 04, pada tahun

2010-2013 penulis masuk ke SMPN 1 Tempeh dan dilanjutkan pada tahun 2013-2015 di SMAN 2 Lumajang. Penulis kemudian diterima sebagai mahasiswa di Departemen Statistika ITS melalui jalur SNMPTN pada tahun 2015. Selama menempuh pendidikan di bangku perguruan tinggi, penulis aktif mengikuti organisasi kemahasiswaan ITS mulai dari tingkat Departemen yaitu di FORSIS-ITS, PERS HIMASTA ITS yang berawal dengan menjadi staff kemudian berlanjut menjadi manajer di Divisi Analisis Data. Selain itu penulis juga aktif mengikuti kegiatan kemahasiswaan salah satunya yaitu menjadi panitia GERIGI ITS 2016/2017, kemudian penulis pernah mengikuti Organisasi Kemahasiswaan di tingkat Fakultas yaitu BEM FMKSD sebagai Ketua Divisi Kastrat dan Ketua DPM FMIPA 2017/2018. Berkaitan dengan adanya Tugas Akhir ini, apabila terdapat hal-hal yang dapat didiskusikan baik berupa kritik dan saran dapat menghubungi penulis melalui *e-mail* yaitu Mahendra.andra11@gmail.com.

*(Halaman ini sengaja dikosongkan)*