



TUGAS AKHIR - KS184822

**PREDIKSI BANJIR GENANGAN DI KOTA
SURABAYA BERDASARKAN PREDIKSI CURAH
HUJAN BERBASIS DATA BMKG DENGAN
PENDEKATAN *NU - SUPPORT VECTOR
REGRESSION* DAN *NEURAL NETWORK***

**ICHA TIRHISS FEBRIANA
NRP 062115 4000 0049**

Dosen Pembimbing

Dr. rer. pol. Dedy Dwi Prastyo, S.Si., M.Si.

Dr. Ardhasena Sopaheluwakan

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

(Halaman ini sengaja dikosongkan)



TUGAS AKHIR - KS184822

**PREDIKSI BANJIR GENANGAN DI KOTA
SURABAYA BERDASARKAN PREDIKSI CURAH
HUJAN BERBASIS DATA BMKG DENGAN
PENDEKATAN *NU - SUPPORT VECTOR
REGRESSION* DAN *NEURAL NETWORK***

**ICHA TIRHISS FEBRIANA
NRP 062115 4000 0049**

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

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

(Halaman ini sengaja dikosongkan)



FINAL PROJECT - KS184822

**FLOOD PREDICTION IN SURABAYA BASED ON
RAINFALL FORECASTING WITH PREDICTORS FROM
BMKG DATA BY USING NU - SUPPORT VECTOR
REGRESSION AND NEURAL NETWORK**

**ICHA TIRHISS FEBRIANA
NRP 062115 4000 0049**

Supervisors

Dr. rer. pol. Dedy Dwi Prastyo, S.Si., M.Si.

Dr. Ardhasena Sopaheluwakan

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

(Halaman ini sengaja dikosongkan)

LEMBAR PENGESAHAN

PREDIKSI BANJIR GENANGAN DI KOTA SURABAYA BERDASARKAN PREDIKSI CURAH HUJAN BERBASIS DATA BMKG DENGAN PENDEKATAN *NU – SUPPORT VECTOR REGRESSION* DAN *NEURAL NETWORK*

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:

Icha Tirhiss Febriana
NRP. 062115 4000 0049

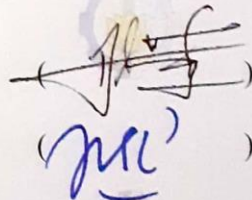
Disetujui oleh Pembimbing:

Dr. rer. pol. Dedy Dwi Prastyo, S.Si., M.Si.

NIP. 19831204 200812 1 002

Dr. Ardhasena Sopaheluwakan

NIP. 19760201 201012 1 001



Mengetahui,
Kepala Departemen Statistika



Dr. Suhartono

NIP. 19710929 199512 1 001

SURABAYA, JULI 2019

(Halaman ini sengaja dikosongkan)

PREDIKSI BANJIR GENANGAN DIKOTA SURABAYA BERDASARKAN PREDIKSI CURAH HUJAN BERBASIS DATA BMKG DENGAN PENDEKATAN *NU - SUPPORT VECTOR REGRESSION* DAN *NEURAL NETWORK*

Nama Mahasiswa : Icha Tirhiss Febriana
NRP : 062115 4000 0049
Departemen : Statistika-FMKSD-ITS
**Dosen Pembimbing : Dr.rer.pol. Dedy Dwi Prastyo, S.Si., M.Si.
Dr. Ardhasena Sopaheluwakan**

Abstrak

*Surabaya adalah kota metropolitan terbesar kedua di Indonesia setelah Jakarta, dimana pertumbuhan pembangunan, ekonomi dan perkembangan properti berkembang pesat di daerah ini. Selain itu Kota Surabaya seringkali memenangkan penghargaan internasional dalam bidang lingkungan dan tata wilayah kota. Namun meskipun demikian, bencana banjir kerap saja menjadi permasalahan yang belum dapat diatasi dengan tuntas. Banjir yang terjadi di Kota Surabaya disebabkan oleh berbagai hal, selain curah hujan yang ekstrem juga disebabkan oleh perubahan tata guna lahan akibat perkembangan pembangunan yang menyebabkan semakin berkurangnya daerah resapan banjir. Adapun dampak yang ditimbulkan dari bencana banjir di Kota Surabaya antara lain yaitu terhambatnya roda perekonomian di Jawa Timur serta menimbulkan kerugian yang besar bagi masyarakat Surabaya. Oleh karena itu penanganan permasalahan banjir di Kota Surabaya menjadi hal sangat penting untuk dilakukan sebagai salah satu representasi perkotaan di Indonesia. Pada penelitian ini akan dilakukan prediksi banjir di Kota Surabaya berdasarkan prediksi curah hujan berbasis data BMKG dengan pendekatan *v-SVR* dan *NN*. Kriteria kebaikan yang digunakan untuk membandingkan antara kedua pendekatan tersebut adalah nilai *RMSE*. Sedangkan prediksi banjir dilakukan menggunakan metode rasional. Berdasarkan hasil analisis diketahui bahwa metode terbaik untuk memodelkan curah hujan di Surabaya adalah metode *v-SVR*, sebab metode *v-SVR* dengan fungsi kernel radial menghasilkan nilai *RMSE* yang lebih kecil dibandingkan metode *NN* dengan fungsi aktivasi *tanh*. Adapun hasil dari prediksi banjir menunjukkan bahwa sebagian*

besar selisih antara debit hidrolika dan debit hidrologi bernilai positif dan hanya terdapat dua saluran yang memiliki selisih negatif. Hal tersebut berarti bahwa sepanjang Desember 2018 diprediksi akan terjadi banjir di Kota Surabaya, tepatnya di daerah Gubeng pada saluran Mojo dan saluran Sidoluhur. Hasil dari penelitian ini diharapkan dapat menjadi bahan pertimbangan bagi pemerintahan Kota Surabaya dalam perencanaan dan mitigasi banjir.

Kata Kunci : Banjir, Curah Hujan, NN, Prediksi, SVR

FLOOD PREDICTION IN SURABAYA BASED ON RAINFALL FORECASTING WITH PREDICTORS FROM BMKG DATA BY USING NU – SUPPORT VECTOR REGRESSION AND NEURAL NETWORK

Name : Icha Tirhiss Febriana
Student Number : 062115 4000 0049
Department : Statistics
Supervisors : Dr.rer.pol. Dedy Dwi Prastyo, S.Si., M.Si.
Dr. Ardhasena Sopaheluwakan

Abstract

Surabaya is second largest metropolitan city in Indonesia after Jakarta, whereas development, economic and properties are growing rapidly in this area. Moreover, Surabaya often wins international awards in the field of environment and urban planning governance. But, unfortunately Surabaya still having problem with floods. Floods often occur in Surabaya caused by land use changes due to development growth that have impacted a reduction in flood absorption areas. The impact of the floods in the Surabaya city among others, was hampering of the economy in East Java and causing huge losses for Surabaya's society. Therefore, the handling of flood problems in Surabaya city is very important to do as one of the urban representations in Indonesia. In this study, flood prediction will be carried out based on rainfall prediction which refers to BMKG data using v-SVR and NN approach. The goodness criterion used to compare the two approaches is the RMSE value. While flood prediction is calculated by using a rational method. The results of analysis informs that the best method for modelling rainfall in Surabaya city is the v-SVR method. The v-SVR with radial kernel functions produces a smaller RMSE than the NN with the activation function of tanh. The flood prediction shows that most of the difference between hydraulic discharge and hydrological discharge is positive and there are only two channels that have a negative difference. This means that throughout December 2018 it is predicted that there will be flooding in Surabaya city, precisely in the Gubeng Area, i.e. Mojo channel and Sidoluhur channel. The results of this study are expected to be considered by Surabaya city government in flood planning and mitigation.

Keywords : Flood, NN, Prediction, Rainfall, SVR

(Halaman ini sengaja dikosongkan)

KATA PENGANTAR

Puji Syukur penulis panjatkan atas rahmat dan hidayah yang telah diberikan Allah SWT sehingga penulis dapat menyelesaikan Laporan Tugas Akhir yang berjudul “**Prediksi Banjir Genangan di Kota Surabaya Berdasarkan Prediksi Curah Hujan Berbasis Data BMKG dengan Pendekatan *Nu – Support Vector Regression dan Neural Network***” dengan tepat waktu.

Penulis menyadari dalam penyusunan Tugas Akhir ini tidak terlepas dari bantuan maupun dukungan dari berbagai pihak. Pada kesempatan ini penulis menyampaikan terima kasih kepada:

1. Dr. rer. pol. Dedy Dwi Prastyo, S.Si., M.Si. selaku dosen pembimbing utama yang telah meluangkan waktu memberikan bimbingan, saran, serta motivasi selama penyusunan Tugas Akhir berlangsung.
2. Dr. Ardhasena Sopaheluwakan selaku co-pembimbing Tugas Akhir dan Kepala Bidang Klimatologi dan Kualitas Udara BMKG Jakarta Pusat, yang telah meluangkan waktu dan dengan sangat sabar memberikan bimbingan, saran, dukungan serta motivasi selama penyusunan Tugas Akhir.
3. Dr. Ir. Setiawan, MS. dan Dr. R. Moh Atok, S.Si., M.Si. selaku dosen penguji yang telah memberikan masukan dan bantuan dalam menyelesaikan Tugas Akhir.
4. Dr. Santi Wulan Purnami, S.Si., M.Si. selaku Ketua Program Studi Sarjana yang telah memberikan fasilitas, sarana, dan prasarana.
5. Prof. Dr. Drs. I Nyoman Budiantara, M.Si. selaku dosen wali yang telah banyak memberikan saran dan arahan dalam proses belajar di Departemen Statistika.
6. Orang tua, saudara, serta keluarga besar penulis yang selalu memberikan doa, kasih sayang, serta dukungan sehingga menjadi semangat dalam menghadapi kesulitan selama penyusunan Tugas Akhir.
7. Teman-teman seperjuangan *stefa*, Afifah, Henidar, Devita, Risda dan Angga yang selalu memberikan bantuan, dukungan, dan semangat dalam penyusunan Tugas Akhir.

8. Sahabat-sahabat saya, Rosi, Indah, Shindy, Fonda, Habib, Master Hendri, keluarga besar Vivacious serta teman-teman lain yang selalu memberikan semangat kepada penulis dalam penyusunan Tugas Akhir.
9. Seluruh pihak yang turut membantu dalam penyelesaian laporan Tugas Akhir ini baik secara langsung maupun tidak langsung.

Penulis menyadari masih banyak kekurangan dalam pembuatan laporan Tugas Akhir ini. Penulis berharap semoga laporan Tugas Akhir ini bermanfaat dan menambah wawasan bagi pembaca. Kritik dan saran sangat diperlukan untuk perbaikan di masa yang akan datang.

Surabaya, Juni 2019

Penulis

DAFTAR ISI

	Halaman
HALAMAN JUDUL	i
LEMBAR PENGESAHAN.....	v
ABSTRAK	vii
KATA PENGANTAR.....	xi
DAFTAR ISI.....	xiii
DAFTAR GAMBAR.....	xv
DAFTAR TABEL	xix
DAFTAR LAMPIRAN	xxi
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	8
1.3 Tujuan.....	8
1.4 Manfaat Penelitian.....	9
1.5 Batasan Masalah.....	10
BAB II TINJAUAN PUSTAKA	11
2.1 Statistika Deskriptif	11
2.2 <i>Pre-Processing</i> Data.....	11
2.3 Definisi Peramalan (<i>Forecasting</i>)	16
2.4 Analisis <i>Time Series</i>	16
2.5 <i>Support Vector Regression</i> (SVR)	19
2.6 <i>Artificial Neural Network</i> (ANN).....	23
2.7 Validasi Model dan Pemilihan Metode Terbaik.....	26
2.8 Perhitungan Matematis Banjir.....	28
2.9 Curah Hujan	32
2.10 Banjir	33
2.11 Dinas Pekerjaan Umum Bina Marga dan Pematusan Kota Surabaya	33
2.12 Balai PSAWS Buntung Peketingan Surabaya.....	34
2.13 Badan Meteorologi, Klimatologi, dan Geofisika Kota Surabaya	34
BAB III METODOLOGI PENELITIAN.....	35
3.1 Sumber Data	35
3.2 Daerah Penelitian	35

3.3	Variabel Penelitian	38
3.4	Definisi Operasional.....	40
3.5	Struktur Data.....	41
3.6	Langkah Penelitian.....	43
BAB IV ANALISIS DAN PEMBAHASAN		51
4.1	<i>Pre-Processing</i> Data Penelitian	51
4.2	Gambaran Umum Curah Hujan dan Variabel Klimatologi di Kota Surabaya.....	55
4.3	Peramalan <i>Univariate</i> Variabel <i>Predictor</i> dengan <i>Holt-Winters</i>	75
4.4	<i>Principal Component Analysis</i>	77
4.5	Penentuan <i>Input Lag</i> yang Signifikan	78
4.6	Analisis Curah Hujan di Balai Keputih	87
4.7	Analisis Curah Hujan di Balai Kedung Cowek.....	113
4.8	Analisis Curah Hujan di Balai Gubeng	114
4.9	Analisis Curah Hujan di Balai Wonorejo	116
4.10	Analisis Curah Hujan di Balai Wonokromo	117
4.11	Analisis Curah Hujan di Balai Gunung Sari.....	119
4.12	Prediksi Banjir di Kota Surabaya	120
BAB V KESIMPULAN DAN SARAN		129
5.1	Kesimpulan	129
5.2	Saran.....	130
DAFTAR PUSTAKA.....		131
LAMPIRAN.....		137
BIODATA PENULIS.....		175

DAFTAR GAMBAR

	Halaman
Gambar 2.1	<i>Loss Function</i> (a) <i>Quadratic</i> , (b) <i>Least Modulus</i> , (c) <i>Huber</i> , dan (d) <i>ϵ-Insensitive</i> 19
Gambar 2.2	Ilustrasi Perhitungan <i>Error</i> pada SVR..... 20
Gambar 2.3	Arsitektur <i>Neural Network</i> 24
Gambar 2.4	Macam-macam Bentuk Saluran 31
Gambar 3.1	Letak Administratif Kota Surabaya..... 36
Gambar 3.2	Pola Hujan Surabaya Berdasarkan Jenisnya 36
Gambar 3.3	Wilayah Pengamatan Balai PSAWS Buntung Peketingan dan BMKG Kota Surabaya 37
Gambar 3.4	Kerangka Berfikir atau Kerangka Penelitian ... 38
Gambar 3.5	Diagram Alir Penelitian Keseluruhan..... 47
Gambar 3.6	Diagram Alir <i>Nu-Support Vector Regression</i> 48
Gambar 3.7	Diagram Alir <i>Neural Network</i> 49
Gambar 4.1	<i>Missing Value</i> (a) Data Balai PSAWS dan (b) Data BMKG..... 52
Gambar 4.2	<i>Time Series Plot</i> Curah Hujan di 6 Balai PSAWS Surabaya 56
Gambar 4.3	Boxplot CH di 6 Balai PSAWS Surabaya 56
Gambar 4.4	<i>Time Series Plot</i> (Atas) dan <i>Boxplot</i> (Bawah) Variabel (a) Temperatur, (b) Kelembapan Udara, dan (c) Kecepatan Angin 60
Gambar 4.5	<i>Scatterplot</i> Temperatur di 3 Stasiun BMKG dan Curah Hujan (a) Keputih, (b) Kedung Cowek, (c) Gubeng, (d) Wonorejo, (e) Wonokromo, dan (f) Gunung Sari 63
Gambar 4.6	<i>Scatterplot</i> Kelembapan Udara di 3 Stasiun BMKG dan Curah Hujan (a) Keputih, (b) Kedung Cowek, (c) Gubeng, (d) Wonorejo, (e) Wonokromo, dan (f) Gunung Sari..... 66
Gambar 4.7	<i>Scatterplot</i> Kecepatan Angin di 3 Stasiun BMKG dan Curah Hujan (a) Keputih, (b) Kedung Cowek, (c) Gubeng, (d) Wonorejo, (e) Wonokromo, dan (f) Gunung Sari 69

Gambar 4.8	ACF (Kiri) dan PACF (Kanan) Curah Hujan Balai (a) Keputih, (b) Kedung Cowek, (c) Gubeng, (d) Wonorejo, (e) Wonokromo, dan (f) GnSari.....	71
Gambar 4.9	<i>Scatterplot</i> antara CH dengan CH lag 1, lag 2, lag 3, lag 4, lag 5, dan lag 6 di Balai (a) Keputih, (b) Kedung Cowek, (c) Gubeng, (d) Wonorejo, (e) Wonokromo, dan (f) Gunung Sari.....	74
Gambar 4.10	Peramalan <i>Univariate</i> Menggunakan <i>Holt-Winters</i> Variabel <i>Predictor</i> (a) Temperatur Juanda (RMSE = 2,926509), (b) Kelembapan Udara Juanda (RMSE = 16,87781), (c) Kecepatan Angin Juanda (RMSE = 2,431392), (d) Temperatur Perak 2 (RMSE = 2,760564), (e) Kelembapan Udara Perak 2 (RMSE = 15,01662), (f) Kecepatan Angin Perak 2 (2,174094), (g) Temperatur Perak 1 (RMSE = 2,829435), (h) Kelembapan Udara Perak 1 (RMSE = 15,04157), (i) Kecepatan Angin Perak 1 (RMSE = 1,84213).....	77
Gambar 4.11	Plot PACF Curah Hujan di Balai (a) Keputih, (b) Kedung Cowek, (c) Gubeng, (d) Wonorejo, (e) Wonokromo, dan (f) Gunung Sari.....	79
Gambar 4.12	Plot ACF (Kiri) dan PACF (Kanan) Curah Hujan di Balai (a) Keputih, (b) Kedung Cowek, (c) Gubeng, (d) Wonorejo, (e) Wonokromo, dan (f) Gunung Sari setelah <i>Differencing</i> Musiman Lag 365.....	82
Gambar 4.13	Plot (a) Model 1, (b) Model 2, (c) Model 3, (d) Model 4, (e) Model 5, (f) Model 6, (g) Model 7, (h) Model 8	90
Gambar 4.14	Pengaruh Penambahan Lag pada RMSE (a) $Y_1, t = f(Y_1, t - 1 + Y_1, t - 2 + Y_1, t - 3 + Y_1, t - 4 + Y_1, t - 5)$, (b) $Y_1, t = f(Y_1, t - 1 + Y_1, t - 2 + Y_1, t - 3 + \dots + Y_1, t - 10)$, (c) $Y_1, t = f(Y_1, t - 1 + Y_1, t - 2 + Y_1, t - 3 + \dots + Y_1, t - 15)$	93

Gambar 4.15	Plot <i>Training</i> Pemodelan CH Keputih dengan <i>Input</i> lag PACF yang Signifikan	94
Gambar 4.16	Plot <i>Training</i> Pemodelan CH Keputih dengan <i>Input</i> lag ARIMA yang Signifikan	95
Gambar 4. 17	Plot (a) Model 1, (b) Model 2, (c) Model 3, (d) Model 4, (e) Model 5, (f) Model 6, (g) Model 7, (h) Model 8	97
Gambar 4.18	Plot <i>Training</i> Pemodelan CH Keputih dengan Kombinasi <i>Input</i> Prediktor dan lag PACF (a) Model 1, (b) Model 3, (c) Model 5, (d) Model 2, (e) Model 4, dan (f) Model 6.....	99
Gambar 4. 19	Plot <i>Training</i> Pemodelan CH Keputih dengan Kombinasi <i>Input</i> Prediktor dan lag ARIMA (a) Model 7, (b) Model 9, (c) Model 11, (d) Model 8, (e) Model 10, dan (f) Model 12.....	100
Gambar 4.20	Plot <i>Training</i> Pemodelan CH Keputih dengan Kombinasi <i>Input</i> Prediktor dan lag ARIMA (a) Model 13, (b) Model 14, (c) Model 15, dan (d) Model 16	102
Gambar 4.21	Model Terbaik untuk Prediksi Curah Hujan Keputih (a) <i>Training</i> dan (b) <i>Testing</i>	103
Gambar 4.22	Karakteristik Variabel Prediktor pada <i>Training</i> (Kiri) dan <i>Testing</i> (Kanan) di Stasiun (a) Juanda, (b) Perak 2, dan (c) Perak 1	104
Gambar 4.23	Prediksi Model Terbaik dengan NA untuk Curah Hujan Keputih (a) <i>Training</i> dan (b) <i>Testing</i>	105
Gambar 4.24	<i>Boxplot</i> CH Bulan Desember 2009-2017	106
Gambar 4.25	Model Terbaik untuk Prediksi Curah Hujan Keputih menggunakan Informasi <i>Prior Knowledge</i> (a) <i>Training</i> dan (b) <i>Testing</i>	107
Gambar 4.26	Model Terbaik CH Keputih dengan NN (a) <i>Training</i> dan (b) <i>Testing</i>	110
Gambar 4.27	Arsitektur Jaringan NN dari Model Terbaik Prediksi Curah Hujan Keputih	111

Gambar 4.28	Model Terbaik Prediksi Curah Hujan Kedung Cowek (a) <i>Training SVR</i> , (b) <i>Testing SVR</i> , (c) <i>Training NN</i> , dan (d) <i>Testing NN</i>	114
Gambar 4.29	Model Terbaik Prediksi Curah Hujan Gubeng (a) <i>Training SVR</i> , (b) <i>Testing SVR</i> , (c) <i>Training NN</i> , dan (d) <i>Testing NN</i>	115
Gambar 4.30	Model Terbaik Prediksi Curah Hujan Wonorejo (a) <i>Training SVR</i> , (b) <i>Testing SVR</i> , (c) <i>Training NN</i> , dan (d) <i>Testing NN</i>	116
Gambar 4.31	Model Terbaik Prediksi Curah Hujan Wonokromo (a) <i>Training SVR</i> , (b) <i>Testing SVR</i> , (c) <i>Training NN</i> , dan (d) <i>Testing NN</i>	118
Gambar 4.32	Model Terbaik Prediksi Curah Hujan Gunung Sari (a) <i>Training SVR</i> , (b) <i>Testing SVR</i> , (c) <i>Training NN</i> , dan (d) <i>Testing NN</i>	119
Gambar 4.33	Rayon PU Binamarga & Pematusan Kota Surabaya dan Balai PSAWS	120

DAFTAR TABEL

	Halaman
Tabel 2.1 Macam-macam Fungsi Kernel.....	23
Tabel 2.2 Tabel Koefisien <i>Run-off</i> (C)	29
Tabel 2.3 Tabel Kekasaran <i>Manning</i> untuk Saluran	32
Tabel 3.1 Kebutuhan Data Penelitian	39
Tabel 3.2 Struktur Data Penelitian Keseluruhan	41
Tabel 3.3 Struktur Data Penelitian untuk PCA	42
Tabel 4.1 Jumlah Data <i>Missing Value</i> pada Data Penelitian...	52
Tabel 4.2 Statistika Deskriptif Data Hasil Imputasi	54
Tabel 4.3 Hasil Kumulatif Proporsi.....	78
Tabel 4.4 <i>Input</i> Lag yang Signifikan dari PACF.....	80
Tabel 4.5 Model ARIMA untuk CH di 6 Balai PSAWS	83
Tabel 4.6 <i>Input</i> Lag yang Signifikan dari ARIMA	87
Tabel 4.7 Skenario Pemodelan ν -SVR CH Keputih dengan <i>Input</i> Prediktor	88
Tabel 4.8 Skenario Pemodelan ν -SVR CH Keputih dengan <i>Input</i> PCA.....	96
Tabel 4.9 Skenario Pemodelan ν -SVR CH Keputih dengan Kombinasi Keempat Macam <i>Input</i>	98
Tabel 4.10 Hasil Analisis <i>Neural Network</i>	109
Tabel 4.11 Perbandingan Metode ν -SVR dan NN pada data <i>Testing</i> Curah Hujan Keputih	112
Tabel 4.12 Sinkronisasi Enam Balai Hujan terhadap 5 Rayon Pematuan.....	121
Tabel 4.13 Hasil Prediksi Curah Hujan Maksimum pada 6 Balai PSAWS	122
Tabel 4.14 Perhitungan Banjir di Keputih	122
Tabel 4.15 Perhitungan Banjir di Kedung Cowek.....	123
Tabel 4.16 Perhitungan Banjir di Gubeng	124
Tabel 4.17 Perhitungan Banjir di Wonorejo	125
Tabel 4.18 Perhitungan Banjir di Wonokromo	126
Tabel 4.19 Perhitungan Banjir di Gunung Sari.....	127

(Halaman ini sengaja dikosongkan)

DAFTAR LAMPIRAN

	Halaman
Lampiran 1. <i>Scatterplot</i> antara Temperatur Lag 1-5 di 3 Stasiun BMKG dengan CH di 6 Balai PSAWS	137
Lampiran 2. <i>Scatterplot</i> antara Kelembapan Udara Lag 1-5 di 3 Stasiun BMKG dengan Curah Hujan di 6 Balai PSAWS	139
Lampiran 3. <i>Scatterplot</i> antara Kecepatan Angin Lag 1-5 di 3 Stasiun BMKG dengan Curah Hujan di 6 Balai PSAWS	141
Lampiran 4. Hasil Prediksi Variabel <i>Predictor</i> Menggunakan <i>Holt-Winters</i>	144
Lampiran 5. Pemodelan ν -SVR CH Keputih dengan <i>Input</i> Prediktor BMKG yang telah di Optimasi Parameter dengan 400 Kombinasi.....	145
Lampiran 6. Plot <i>Training</i> Pemodelan ν -SVR CH Keputih dengan Prediktor Klimatologi Perak 2	146
Lampiran 7. Plot <i>Training</i> Pemodelan ν -SVR CH Keputih dengan Prediktor Klimatologi Perak 1	147
Lampiran 8. Pemodelan ν -SVR CH Keputih dengan Prediktor PCA	149
Lampiran 9. Plot Hasil Prediksi dan Aktual Menggunakan <i>Neural Network</i> untuk Prediksi CH Keputih	149
Lampiran 10. Plot 6 Percobaan Terbaik Prediksi Curah Hujan di Kedung Cowek.....	155
Lampiran 11. Plot 6 Percobaan Terbaik Prediksi Curah Hujan di Gubeng	155
Lampiran 12. Plot 6 Percobaan Terbaik Prediksi Curah Hujan di Wonorejo	156
Lampiran 13. Plot 6 Percobaan Terbaik Prediksi Curah Hujan di Wonokromo	156
Lampiran 14. Plot 6 Percobaan Terbaik Prediksi Curah Hujan di Gunung Sari	157
Lampiran 15. <i>Syntax Pre-Processing</i> Data	157

Lampiran 16. <i>Syntax</i> Eksplorasi Data, <i>Missing Value</i> , <i>Boxplot</i> Desember dan <i>Matrix Plot</i> Prediktor.....	161
Lampiran 17. <i>Syntax Holt-Winters</i>	165
Lampiran 18. <i>Syntax</i> PCA.....	165
Lampiran 19. <i>Syntax</i> ν -SVR.....	166
Lampiran 20. <i>Syntax</i> NN.....	169
Lampiran 21. Alternatif Solusi untuk Mengatasi Prediksi Konstan.....	171
Lampiran 22. Surat Pernyataan Data	172
Lampiran 23. Surat Pernyataan Data (Lanjutan)	173

BAB I PENDAHULUAN

1.1 Latar Belakang

Indonesia adalah sebuah negara kepulauan yang secara geografis terletak 6° LU – 11° LS dan 95° BT – 11° BT, antara lautan Pasifik dan Lautan Hindi, antara Benua Asia dan Benua Australia, dan pada pertemuan dua rangkaian pegunungan yaitu Sirkum Pasifik dan Sirkum Mediterranean (Salowong, 2008). Indonesia juga terletak pada zona cincin api (*Ring of Fire*) dunia yaitu daerah patahan dengan potensi bencana yang sangat besar. Berdasarkan letaknya tersebut, menjadikan Indonesia sebagai salah satu negara yang rawan terhadap terjadinya bencana alam dengan karakteristik yang berbeda (BMKG, 2010). Bencana alam merupakan peristiwa atau rangkaian peristiwa yang mengancam dan mengganggu kehidupan masyarakat yang disebabkan oleh alam yang mengakibatkan timbulnya korban jiwa manusia, kerusakan lingkungan, kerugian harta benda, dan dampak psikologis. Bencana alam tersebut diantaranya gempa bumi, tsunami, gunung meletus, banjir, kekeringan, dan tanah longsor (BNPB, 2018). Suatu badan atau instansi yang menangani masalah kebencanaan di Indonesia yaitu Badan Nasional Penanggulangan Bencana (BNPB). BNPB bertugas mencatat seluruh kejadian kebencanaan di Indonesia yang meliputi jumlah kejadian bencana, jumlah korban dan dampak yang ditimbulkannya.

Secara garis besar, pada tahun 2018 kejadian bencana banjir menduduki peringkat pertama bencana yang paling banyak terjadi di Indonesia sekaligus merupakan bencana yang menimbulkan kerugian paling besar pula (Gupta, 2003). Permasalahan banjir sampai saat ini belum dapat diselesaikan secara tuntas, bahkan dari tahun ke tahun kejadian banjir terus meningkat baik dalam hal frekuensi, intensitas, maupun sebarannya (BNPB, 2018). Menurut Cahyadi (2015) terdapat banyak faktor yang dapat menyebabkan bencana banjir terjadi, selain dipengaruhi oleh perubahan iklim seperti curah hujan ekstrem, banjir juga dipengaruhi oleh perubahan tata guna lahan, infrastruktur, akumulasi aliran air,

jumlah drainase yang tersedia, kepadatan bangunan, jenis tanah, pendangkalan sungai, maupun kerusakan DAS. Kondisi demikian sering terjadi di kota-kota besar di Indonesia seperti halnya Surabaya.

Surabaya merupakan Ibukota Provinsi Jawa Timur yang rawan terhadap bencana banjir. Selain dikenal dengan Kota Pahlawan, Surabaya juga merupakan kota metropolitan terbesar kedua setelah Jakarta. Dimana pertumbuhan pembangunan, ekonomi dan perkembangan properti berkembang sangat cepat di kota ini. Dikutip dari berita harian Kompas (2019), Kota Surabaya telah menerima tiga penghargaan sekaligus, yaitu Adipura Kencana, Kinerja Pengurangan Sampah, dan Nirwasita Tantra dari Kementerian Lingkungan Hidup dan Kehutanan (KLHK). Bahkan Surabaya pernah menjadi kota favorit di ajang penghargaan internasional bergengsi dunia "*The 4th Guangzhou International Award for Urban Innovation*" yang digelar di Guangzhou China pada tanggal 7 Desember 2018 (Damanik, 2019). Hal ini berarti Surabaya telah mendapatkan penghargaan lebih banyak di bandingkan Jakarta dalam bidang lingkungan dan tata perencanaan kota. Namun ironisnya, banjir masih saja menjadi permasalahan yang sangat sulit diatasi di Kota Surabaya. Dinas Pekerjaan Umum Bina Marga dan Pematusan Kota Surabaya mengatakan bahwa hampir 50 persen dari total wilayah Surabaya terendam banjir saat musim hujan tiba dengan ketinggian yang bervariasi mulai dari 10 –70 cm dan waktu genangan paling lama sekitar enam jam. Selain itu kejadian banjir yang kerap terjadi secara rutin setiap musim penghujan tiba di Kota Surabaya tentunya dapat menghambat dan berdampak luas bagi roda perekonomian yang ada sebab Surabaya sebagai kota terbesar kedua memiliki peran yang besar dalam menggerakkan roda perekonomian di Provinsi Jawa Timur. Oleh karena itu, penanganan permasalahan banjir di Kota Surabaya menjadi hal yang sangat penting untuk dilakukan sebagai representasi perkotaan di Indonesia.

Penyebab banjir atau air menggenang di Surabaya banyak disebabkan oleh tumpukan sampah dan intensitas curah hujan yang tinggi (Suhartono, 2017). Dimana kesadaran masyarakat Surabaya akan lingkungan dan sampah masih sangat kurang, sehingga air

hujan yang turun tidak dapat mengalir dengan baik akibat tersumbat oleh sampah. Disisi lain terjadinya curah hujan ekstrem atau intensitas curah hujan yang sangat tinggi tidak diimbangi dengan sistem drainase yang baik, artinya debit saluran air tidak sesuai dengan intensitas curah hujan yang turun atau dengan kata lain intensitas curah hujan lebih tinggi daripada debit air pada saluran drainase yang ada sehingga menyebabkan air hujan yang turun tidak langsung terserap oleh tanah atau mengalir ke saluran drainase melainkan air hujan tersebut akan meluap ke permukaan dan menyebabkan genangan atau banjir. Adapun penyebab kecilnya debit air pada saluran drainase yaitu dikarenakan adanya tumpukan sampah yang menghambat saluran air.

Banyak penelitian yang telah dilakukan untuk menganalisis faktor-faktor yang mempengaruhi banjir maupun pemodelan banjir untuk prediksi. Pirasteh, *et al.* (2011) melakukan pemodelan dan simulasi daerah rawan banjir di bagian selatan Semenanjung Malaysia. Adapun metode yang digunakan yaitu model Jaringan Saraf Tiruan yang dikembangkan pada MATLAB menggunakan tujuh faktor penyebab banjir. Hasil yang diperoleh berdasarkan SIG yaitu faktor-faktor penyebab banjir diantaranya curah hujan, kemiringan, ketinggian, akumulasi aliran, tanah, penggunaan lahan, dan geologi. Kemudian Tehrany, Pradhan, & Ahmad (2015) dengan penelitiannya tentang pemodelan banjir menggunakan SVM didapatkan hasil bahwa semua faktor pengkondisian banjir yang meliputi curah hujan, ketinggian, kemiringan, kelengkungan, indeks daya aliran (SPI), indeks basah topografi (TWI), jarak dari sungai, geologi, penggunaan / tutupan lahan (LULC), tanah, dan permukaan limpasan memiliki pengaruh positif yang cukup terhadap analisis banjir. Selain itu pula disimpulkan bahwa teknik SVM adalah alat yang efisien dan andal dalam penilaian kerentanan banjir. Adapun Khosravi, Nohani, & Pourghasemi (2016) juga melakukan pemodelan banjir dengan sepuluh faktor pengkondisi banjir diantaranya curah hujan, sudut kemiringan, kelengkungan rencana, ketinggian, indeks basah topografi, indeks daya aliran, jarak dari sungai, geologi, penggunaan lahan, dan indeks vegetasi. Hasil yang didapatkan dari penelitian tersebut

yaitu model rasio frekuensi memiliki AUC tertinggi dibandingkan dengan model lainnya.

Berdasarkan penelitian-penelitian terdahulu yang telah dilakukan dapat diketahui bahwa penyebab utama banjir dari segi iklim yaitu disebabkan oleh curah hujan, sedangkan faktor non-iklim lainnya yaitu dipengaruhi oleh banyaknya sampah, jumlah drainase, tata guna lahan dan pendangkalan sungai. Namun data tentang jumlah drainase, perubahan tata guna lahan dan mengkuantifikasi sampah sulit untuk didapatkan atau diamati. Sehingga dalam penelitian ini variabel-variabel tersebut untuk sementara akan dibatasi (diabaikan) atau dianggap konstan dan diasumsikan jumlah drainase serta kebiasaan masyarakat masih sama atau tidak terjadi perubahan perilaku. Oleh karena itu variabel yang akan diperhatikan dalam penelitian ini untuk memprediksi banjir hanyalah variabel curah hujan. Akan tetapi curah hujan yang terjadi bukan hanya dipengaruhi oleh curah hujan pada periode sebelumnya melainkan juga disebabkan oleh faktor-faktor lainnya. Sehingga perlu dilakukan prediksi curah hujan berdasarkan faktor-faktor yang mempengaruhi curah hujan terlebih dahulu sebelum memprediksi banjir genangan.

Beberapa penelitian terdahulu yang membahas mengenai curah hujan beserta faktor-faktor yang mempengaruhinya adalah Navianti, Usadha, & Widjajati (2012) dengan penelitiannya tentang prediksi curah hujan di Kota Surabaya dan didapatkan kesimpulan bahwa prediksi curah hujan berdasarkan variabel yang mempengaruhi terjadinya hujan seperti suhu udara, kelembaban relatif, kecepatan angin, tekanan udara, total lapisan awan, dan lama penyinaran matahari diperoleh keakuratan prediksi curah hujan sebesar 77,68% dari sebelas eksperimen yang dilakukan. Sehingga pada penelitian ini akan dilakukan prediksi curah hujan pada tujuh balai PSAWS Buntung Peking Surabaya berdasarkan variabel-variabel yang diduga berpengaruh diantaranya temperatur, kelembapan udara, dan kecepatan angin yang diperoleh dari tiga stasiun BMKG di Surabaya. Dikarenakan sumber skala data antara variabel yang mempengaruhi dan dipengaruhi berbeda, maka perlu digunakan metode pemodelan

yang mengadopsi konsep pemodelan *statistical downscaling* untuk menaksir curah hujan di setiap balai.

Menurut Benestad (2004), konsep dari *statistical downscaling* (SD) yaitu suatu teknik yang menggunakan model statistika untuk melihat hubungan antara suatu data yang berskala besar dengan data yang berskala lokal. Setelah mengetahui hubungan kedua gugus data tersebut, data yang berskala global digunakan untuk memprediksi data peubah iklim berskala lokal (balai pengamatan hujan). Namun dikarenakan data global yang biasanya digunakan pada metode *statistical downscaling* adalah model dari GCM bukan data pengamatan stasiun BMKG, sehingga konsep yang diadopsi dari *statistical downscaling* pada penelitian ini adalah konsep dalam memodelkan setiap stasiun BMKG untuk memprediksi curah hujan di setiap balai PSAWS Buntung Peketingan Surabaya.

Adapun penelitian-penelitian terdahulu mengenai prediksi curah hujan khususnya ARIMA sudah banyak dilakukan. Akan tetapi penggunaan metode ARIMA untuk peramalan curah hujan seringkali memiliki beberapa kelemahan dan menghasilkan akurasi yang belum optimal. Hal tersebut dikarenakan metode ARIMA yang didapatkan tidak mampu memenuhi asumsi residual yang *white noise* yaitu tidak ada *autocorrelation* dan berdistribusi normal (Karomah & Hendikawati, 2014). Zhai (2005) menyatakan bahwa ARIMA hanya mampu mengakomodasi data *time series* yang linier dan lemah terhadap data *time series* yang non-linier. Metode ARIMA sangat baik ketepatannya untuk prediksi jangka pendek dengan pola data yang linier, sebab ARIMA memprediksi berdasarkan lag-lag yang berpengaruh. Sedangkan untuk prediksi jangka panjang dengan pola data yang non-linier ketepatannya akan berkurang dan terus menurun (Zhang, 2003). Curah hujan sendiri memiliki keragaman yang besar baik dalam skala ruang maupun skala waktu serta pola data dari curah hujan seringkali berbentuk non-linier, sehingga dibutuhkan metode lain selain ARIMA yang dapat mengakomodasi pola data yang non-linier tersebut. Salah satu metode yang dapat mengatasi non-linieritas pada data yaitu *Support Vector Regression* (SVR) dan *Artificial Neural Network* (ANN).

Support Vector Regression (SVR) merupakan salah satu metode pengembangan dari *Support Vector Machine* (SVM) khusus untuk prediksi yang bebas dari asumsi, mampu mengatasi data curah hujan yang non-linier, serta dapat memberikan hasil yang lebih baik meskipun dengan data yang sedikit (Zhao, Tao, & Zio, 2015). Gun (1998) menyatakan bahwa SVR mengatasi kasus *overfitting* dengan cara mendapatkan parameter *learning* yang tepat dan mengatasi non-linieritas data dengan cara menambahkan fungsi kernel. Salah satu fungsi kernel yang sering digunakan dalam penelitian-penelitian terdahulu dan menghasilkan nilai akurasi prediksi yang lebih baik yaitu fungsi kernel *Radial Basis Function* (RBF). Harafani (2015) dengan penelitiannya yang berjudul optimasi parameter pada metode *Support Vector Machine* berbasis Algoritma Genetika menghasilkan kesimpulan bahwa fungsi kernel RBF memiliki nilai RMSE yang paling kecil dibandingkan fungsi kernel lainnya. Jika ditinjau dari jenisnya, SVR terdiri dari 2 jenis yaitu *Epsilon-Support Vector Regression* (ϵ -SVR) serta *Nu-Support Vector Regression* (ν -SVR). Namun pada penelitian ini hanya akan digunakan ν -SVR untuk prediksi curah hujan di Kota Surabaya. Metode ν -SVR merupakan salah satu metode peramalan dengan teknik regresi khusus untuk data non-linier dengan melibatkan ϵ -*insensitive loss function* dalam analisisnya serta menggunakan kernel tunggal dalam pemulus datanya. Dalam ν -SVR terdapat parameter ν dan C dimana parameter ν dapat mengontrol nilai ϵ -*insensitive loss function* optimal secara mudah dan parameter C berperan sebagai parameter *penalty* untuk data yang melebihi ϵ .

Adapun penelitian terdahulu yang berkaitan dengan penggunaan metode SVR dalam hal prediksi diantaranya yaitu Maharesi (2013) yang melakukan penelitian dengan judul penggunaan SVR pada prediksi *return* saham syariah BEI. Hasil yang diperoleh dari penelitian tersebut yaitu metode SVR lebih baik dibandingkan dengan model Regresi Linear Berganda berbasis *Ordinary Least Squares* (RLB-OLS), sebab SVR memiliki nilai MSE dan R^2 lebih kecil daripada RLB-OLS. Purwidhiati (2015) dengan penelitiannya mengenai prediksi pergerakan harga saham menggunakan ϵ -SVR dan ν -SVR

menghasilkan kesimpulan bahwa dari hasil perbandingan MSE diketahui metode ν -SVR dengan menggunakan Kernel *Laplacian* lebih baik digunakan untuk memprediksi daripada metode ϵ -SVR. Suci (2017) dalam penelitiannya yang berjudul peramalan curah hujan menggunakan metode ARIMA dan SVR menyatakan bahwa SVR merupakan metode terbaik karena menghasilkan prediksi dengan nilai RMSE dan sMAPE yang lebih rendah daripada ARIMA. Hal tersebut juga didukung oleh penelitian Wijaya (2017) mengenai penggunaan SVR sebagai prediksi yang menghasilkan kesimpulan bahwa SVR merupakan metode peramalan yang lebih akurat dengan nilai RMSE dan sMAPE yang lebih kecil.

Metode selain SVR yang dapat mengatasi masalah non-linieritas data dan tidak membutuhkan asumsi yaitu *Artificial Neural Network*. *Artificial Neural Network* atau yang sering disebut dengan *Neural Network* (NN) merupakan sebuah model yang mengadopsi cara kerja neuron pada otak manusia. Kelebihan lainnya dari metode NN yaitu adaptif terhadap perubahan parameter dan dapat mengklasifikasi pola *input* dan *output* (Kusumadewi, 2004). Penelitian sebelumnya yang berkenaan dengan penggunaan NN untuk prediksi cuaca pernah dilakukan oleh Baboo dan Shereef (2010). Dimana dalam penelitiannya tersebut, variabel yang digunakan antara lain curah hujan, temperatur, kelembapan udara, kecepatan angin, arah angin, dan tekanan atmosfer. Adapun kesimpulan yang diperoleh yaitu bahwa penggunaan metode *Artificial Neural Network* untuk meramalkan cuaca menghasilkan nilai kebaikan model yang bagus. Penelitian lainnya juga dilakukan oleh Patmawati (2013) mengenai prediksi curah hujan menggunakan ANN dan mendapatkan hasil prediksi curah hujan dengan tingkat akurasi mencapai 76,23%.

Berdasarkan uraian tersebut, maka peneliti ingin berkontribusi untuk mengembangkan metode guna memprediksi banjir genangan di Kota Surabaya. Adapun penelitian yang akan dilakukan yaitu mengenai prediksi curah hujan berdasarkan faktor-faktor yang mempengaruhi curah hujan menggunakan metode ν -SVR dan ANN. Kemudian hasil prediksi curah hujan tersebut akan dibandingkan menggunakan RMSE, metode yang menghasilkan RMSE terkecil akan digunakan untuk memprediksi curah hujan.

Hasil prediksi curah hujan selanjutnya digunakan untuk memprediksi kejadian banjir. Hasil dari penelitian ini diharapkan dapat membantu pemerintah Kota Surabaya dalam memahami tren banjir sehingga dapat berguna dalam memberikan rekomendasi terhadap kebijakan perencanaan dan pencegahan banjir (mitigasi bencana banjir). Selain itu dari hasil analisis pada penelitian ini juga diharapkan dapat mengurangi atau meminimalisir kerugian yang diakibatkan oleh bencana banjir.

1.2 Rumusan Masalah

Berdasarkan latar belakang diatas menunjukkan adanya masalah penelitian mengenai bagaimana prediksi curah hujan dan prediksi banjir genangan di Kota Surabaya. Terjadinya curah hujan dapat disebabkan oleh berbagai macam faktor diantaranya temperatur, kelembapan udara, dan kecepatan angin maupun curah hujan itu sendiri pada periode sebelumnya. Dimana ketersediaan data untuk faktor-faktor yang mempengaruhi curah hujan memiliki skala global (pengamatan stasiun), sedangkan data curah hujan sendiri memiliki skala lokal (pengamatan balai), sehingga diperlukan metode pemodelan pada data faktor-faktor yang mempengaruhi curah hujan dalam skala global (stasiun) guna memprediksi curah hujan dalam skala lokal (balai). Selain itu belum ada justifikasi mengenai metode pemodelan yang paling baik. Namun pada penelitian ini akan digunakan metode *Nu-Support Vector Regression* dan *Artificial Neural Network* sebagai pembanding untuk memprediksi curah hujan di Kota Surabaya. Masalah lainnya yang juga muncul dari latar belakang diatas yaitu bagaimana hasil prediksi banjir genangan di Kota Surabaya berdasarkan hasil prediksi curah hujan berbasis data BMKG menggunakan metode ν -SVR dan NN.

1.3 Tujuan

Berdasarkan rumusan masalah, adapun tujuan yang ingin dicapai pada penelitian ini adalah sebagai berikut :

1. Mendeskripsikan karakteristik curah hujan dan faktor-faktor yang diduga berpengaruh menggunakan statistika deskriptif.

2. Memprediksi curah hujan berbasis data BMKG Kota Surabaya menggunakan metode *Nu-Support Vector Regression* dan *Artificial Neural Network*.
3. Memprediksi banjir genangan di Kota Surabaya berdasarkan hasil prediksi curah hujan menggunakan perhitungan matematis dengan metode rasional.

1.4 Manfaat Penelitian

Manfaat yang diharapkan pada penelitian ini adalah sebagai berikut :

1. Bagi Keilmuan Statistika
Dapat menjadi referensi untuk penelitian selanjutnya mengenai mitigasi bencana banjir dalam rangka menjamin Surabaya bebas banjir di masa depan dengan menggunakan pendekatan secara statistika. Selain itu dapat memberikan wawasan keilmuan mengenai penggunaan metode *Nu-Support Vector Regression* dan *Artificial Neural Network* serta perhitungan matematis banjir menggunakan metode rasional.
2. Bagi Pemerintah Kota Surabaya terutama Dinas Pekerjaan Umum Bina Marga dan Pematuan serta Balai PSAWS Buntung Pekingting Surabaya
Dapat memberikan saran dan rekomendasi kepada pemerintah Kota Surabaya mengenai faktor-faktor yang perlu ditekan guna mengurangi kejadian banjir genangan di beberapa wilayah di Kota Surabaya. Selain itu dapat memberikan informasi mengenai daerah/kawasan mana saja di Kota Surabaya yang berpotensi tinggi terjadi banjir genangan berdasarkan hasil prediksi banjir. Sehingga dapat memberikan rekomendasi kepada pemerintah Kota Surabaya dalam mengambil kebijakan mengenai pembangunan infrastruktur pada daerah-daerah tertentu di Surabaya untuk mengurangi, mencegah, maupun menanggulangi kejadian banjir dikemudian hari. Diharapkan pula pemerintah Kota Surabaya dapat memberikan perhatian yang lebih terhadap daerah-daerah tersebut dengan memperbaiki maupun menambah infrastruktur - infrastruktur sesuai dengan kondisi

lingkungan setempat seperti rumah pompa, danau bozem dan perbaikan saluran drainase.

3. Bagi Badan Meteorologi, Klimatologi, dan Geofisika Kota Surabaya

Dapat menjadi bahan evaluasi bagi BMKG Juanda, Perak 1, maupun Perak 2 dalam memprediksi curah hujan kedepan menggunakan metode-metode statistika untuk kebutuhan praktis dengan tingkat kebaikan model yang lebih tinggi.

4. Bagi Masyarakat atau Pembaca

Dapat memberikan tambahan informasi bagi masyarakat guna langkah awal untuk mitigasi bencana banjir. Diharapkan masyarakat dapat turut serta membantu pemerintah Surabaya dalam menjaga dan merawat Kota Surabaya baik dalam hal infrastruktur, sampah dan lain-lainnya demi terwujudnya Surabaya yang bebas banjir.

1.5 Batasan Masalah

Batasan permasalahan dalam penelitian ini yaitu mengasumsikan tidak terjadi perubahan perilaku pada kebiasaan masyarakat Surabaya sehari-hari, selain itu diasumsikan *drainase* dan sistem salurannya tidak berubah atau tetap sama. Variabel - variabel lainnya diluar variabel penelitian dianggap konstan. Selain itu penggunaan metode *machine learning* yang bebas asumsi seringkali menghasilkan nilai prediksi curah hujan yang negatif, sedangkan secara logika hal tersebut sangat tidak mungkin curah hujan bernilai negatif. Oleh karena itu untuk hasil prediksi curah hujan yang bernilai negatif ditransformasi atau diasumsikan bernilai 0.

BAB II

TINJAUAN PUSTAKA

Tinjauan pustaka dalam penelitian ini memuat statistika deskriptif, *pre-processing* data, peramalan data *time series* dengan metode *Holt-Winters*, prediksi dengan *Nu-Support Vector Regression*, prediksi dengan *Artificial Neural Network*, perhitungan matematis banjir serta pengetahuan non statistika mengenai banjir, curah hujan dan faktor-faktor yang mempengaruhinya sebagai sebuah landasan yang akan digunakan dalam melakukan analisis dan pembahasan pada Bab 4.

2.1 Statistika Deskriptif

Statistika deskriptif adalah analisis yang berkaitan dengan cara pengumpulan dan penyajian suatu data sehingga dapat memberikan informasi yang berguna. Analisis statistika deskriptif berguna untuk memberikan informasi awal yang telah dikumpulkan dan disajikan. Statistika deskriptif dapat menggambarkan karakteristik atau menguraikan tentang sifat-sifat suatu keadaan serta dapat memberikan gambaran yang sistematis mengenai fakta-fakta dari fenomena yang sedang diselidiki. Statistika deskriptif hanya untuk memberikan informasi mengenai seputar data tanpa mengambil keputusan atau menarik kesimpulan (inferensia) dari data tersebut (Walpole, 2007). Selain itu statistika deskriptif juga dibagi menjadi ukuran pemusatan data dan penyebaran data. Pada penelitian ini ukuran pemusatan data yang akan digunakan adalah *mean*, maksimum dan minimum. Sedangkan untuk penyebaran data akan digunakan nilai varian, dan untuk grafik yang akan digunakan sebagai visualisasi analisis statistika deskriptif adalah *time series plot*, *boxplot* dan *scatterplot*.

2.2 Pre-Processing Data

Pre-processing data perlu dilakukan terlebih dahulu sebelum menganalisis suatu data. Hal ini dikarenakan data memiliki kualitas yang sesuai dengan persyaratan penggunaan data. Terdapat banyak faktor yang mencakup kualitas data antara lain akurasi, kelengkapan, konsistensi, *timeliness*, *believability*, dan *interpretability*. Proses *pre-processing* data meliputi pembersihan

data, integrasi data, reduksi data, dan transformasi (Han, Kamber, & Pei, 2012). Dalam penelitian ini dilakukan proses imputasi *missing value*, PCA untuk mengatasi kasus multikolinieritas atau dependensi antar data, serta dilakukan proses normalisasi dan denormalisasi data sebelum melakukan prediksi.

2.2.1 Missing Value

Missing value (data yang hilang) merupakan informasi yang tidak tersedia pada sebuah objek atau kasus. Hal tersebut sering terjadi akibat informasi tentang objek tidak diberikan, sulit dicari, atau memang informasi tersebut tidak ada. Jika data yang mengandung *missing value* dipaksa untuk tetap diproses, maka *output* yang akan dihasilkan bisa sangat berbeda dengan data yang tidak mengandung *missing value* (Santoso, 2010). Pada umumnya jika *missing value* kurang dari 10% maka dapat diabaikan, akan tetapi jika terdapat lebih dari 15% data per variabel hilang maka variabel tersebut menjadi kandidat untuk dihapuskan atau dihilangkan. Selain itu jika *missing value* lebih dari 50% maka harus dihapus atau dihilangkan. Penghapusan variabel harus memperhatikan beberapa hal seperti jumlah variabel dan jumlah data. Dikarenakan data dalam penelitian ini berbasis *time series* maka proses imputasi *missing value* dilakukan menggunakan imputasi *mean* berdasarkan waktu. Adapun perhitungan yang digunakan dijabarkan pada persamaan (2.1).

$$Y_{dd/mm/yyyy_t}^* = \frac{\sum_{i=1}^n Y_{dd/mm/yyyy_i}}{n} \quad (2.1)$$

dimana :

t' = indeks waktu yang tidak *missing*

$Y_{dd/mm/yyyy}$ = nilai variabel Y pada hari ke-dd, bulan ke-mm, dan tahun ke-yyyy

$Y_{dd/mm/yyyy}^*$ = nilai variabel Y pada hari ke-dd, bulan ke-mm, dan tahun ke-yyyy hasil imputasi

2.2.2 Principal Component Analysis (PCA)

Principal Component (PCA) adalah suatu metode statistika yang dapat digunakan dalam menjelaskan struktur variansi kovariansi dari sekumpulan variabel melalui beberapa variabel baru dimana variabel baru ini saling bebas. Variabel baru yang terbentuk merupakan kombinasi linier dari variabel asal. Selanjutnya variabel baru ini dinamakan komponen utama (PC) (Johnson & Wichern, 2007). PCA digunakan untuk mereduksi dimensi data dengan cara mentransformasi variabel-variabel asli yang berkorelasi menjadi satu set variabel baru yang tidak berkorelasi, dengan tetap mempertahankan sebanyak mungkin keragaman yang dapat dijelaskan.

Vektor random $x = [X_1, X_2, \dots, X_t]^T$ mempunyai matriks varians – kovarians Σ dengan akar ciri (*eigenvalue*) berturut-turut yaitu $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$. Model *principal component* ke- i secara umum dapat ditulis sebagai berikut :

$$z_i = \gamma'_{ij} x_j \quad (2.2)$$

maka kombinasi linier utama dapat dijabarkan dalam persamaan dibawah ini.

$$z_1 = \gamma'_1 x = \gamma_{11} x_1 + \gamma_{12} x_2 + \dots + \gamma_{1t} x_t \quad (2.3)$$

$$z_2 = \gamma'_2 x = \gamma_{21} x_1 + \gamma_{22} x_2 + \dots + \gamma_{2t} x_t \quad (2.4)$$

⋮

$$z_p = \gamma'_p x = \gamma_{p1} x_1 + \gamma_{p2} x_2 + \dots + \gamma_{pt} x_t \quad (2.5)$$

dimana :

$i = 1, 2, \dots, p$; $j = 1, 2, \dots, t$

z_p = PC ke- p , yang mempunyai varians terbesar ke- p

x_p = variabel asal ke- p

γ_p = vektor karakteristik (*eigenvector*) data ke- p

Bila total varians populasi adalah :

$$\sigma_{11} + \sigma_{22} + \dots + \sigma_{pp} = \lambda_1 + \lambda_2 + \dots + \lambda_p \quad (2.6)$$

maka porsi total varians ke- i adalah :

$$\frac{\lambda_i}{\lambda_1 + \lambda_2 + \dots + \lambda_p} \quad (2.7)$$

Penentuan banyaknya *principal component*, antara lain :

1. Melihat *scatterplot*. Garis yang terbentuk mengalami *range* yang cukup besar maka *principal component* sejumlah garis tersebut.
2. Banyaknya *principal component* dipilih sesuai dengan banyaknya *eigenvalue* yang lebih besar dari satu (bila *principal component* diperoleh dari matriks korelasi).
3. Sebaiknya jumlah *principal component* yang dipilih adalah yang mampu memberikan kumulatif persen varians 80% - 90%.

(Johnson & Wichern, 2007)

Pada penelitian ini PCA digunakan untuk mengatasi dependensi antar 3 stasiun BMKG yaitu Juanda, Perak 2, dan Perak 1 untuk masing masing variabel prediktor. Adapun jumlah PC yang akan dipilih adalah yang mampu memberikan kumulatif persen varians sebesar 80%.

2.2.3 Normalisasi dan Denormalisasi

Data curah hujan yang memiliki banyak nilai 0 dapat mengakibatkan beberapa hal diantaranya nilai kebaikan model yang tidak bisa tinggi ataupun hasil prediksinya yang dapat bernilai negatif. Hal tersebut disebabkan karena curah hujan merupakan suatu fenomena alam yang penuh ketidakpastian dan sulit untuk diprediksi. Oleh karena itu, untuk menangani hal tersebut maka perlu dilakukan beberapa penanganan diantaranya melakukan analisis secara terpisah antara curah hujan pada musim kemarau dan curah hujan pada musim hujan agar dihasilkan nilai prediksi yang positif dengan nilai kebaikan model yang tinggi. Akan tetapi disisi lain cara tersebut juga memiliki kekurangan yaitu dapat menghilangkan unsur waktu (*timeseries*). Sehingga untuk melakukan prediksi curah hujan tanpa menghilangkan unsur waktunya dapat digunakan normalisasi data sebelum dimodelkan kemudian hasil prediksi di denormalisasi kembali dengan harapan hasil nilai prediksi bernilai positif. Secara sederhana, agar dapat membawa *range* nilai *output* ke dalam *range input*, maka data

input harus dilakukan normalisasi data ke dalam *range* 0 sampai 1, sehingga *output*-nya dapat di denormalisasi ke dalam *range* nilai *input*. Selain itu normalisasi data perlu dilakukan untuk mempercepat komputasi suatu metode yang sedang digunakan. Metode normalisasi ada bermacam-macam, salah satunya yaitu *min-max normalization* dengan rumus sebagai berikut (Kusumadewi, 2004) :

$$Y_t^* = \frac{Y_t - Y_{\min}}{Y_{\max} - Y_{\min}}, Y_t^* \in (0,1) \quad (2.8)$$

$$Y_{\max} = \max\left(\{Y_t\}_{t=1}^n\right) \quad (2.9)$$

$$Y_{\min} = \min\left(\{Y_t\}_{t=1}^n\right) \quad (2.10)$$

dimana :

Y_t = nilai variabel Y pada waktu ke- t

Y_t^* = nilai variabel Y hasil normalisasi pada waktu ke- t .

Dari persamaan normalisasi diatas maka Y_t^* akan dijadikan *input* untuk memprediksi variabel Y yang nantinya akan menghasilkan nilai \hat{Y}_t^* . Kemudian nilai \hat{Y}_t^* akan dilakukan denormalisasi atau mengembalikan (transformasi) kembali ke nilai *range* awal dengan rumus sebagai berikut :

$$\hat{Y}_t = \hat{Y}_t^* (Y_{\max} - Y_{\min}) + Y_{\min} \quad (2.11)$$

dimana :

\hat{Y}_t = nilai prediksi variabel Y pada waktu ke- t

\hat{Y}_t^* = nilai prediksi variabel Y hasil normalisasi waktu ke- t .

Dalam penelitian ini, normalisasi hanya akan dilakukan pada penggunaan metode ANN dengan tujuan untuk membawa *range output* ke *range* input sekaligus untuk mempercepat proses komputasi yang ada, dimana normalisasi yang akan digunakan adalah *min-max normalization* dengan rumus yang telah dijabarkan pada persamaan (2.8).

2.3 Definisi Peramalan (*Forecasting*)

Peramalan adalah suatu kegiatan untuk memperkirakan ataupun menggambarkan nilai, kejadian, atau kondisi di masa depan berdasarkan data-data masa lalu. Adapun asumsi yang digunakan dalam peramalan yaitu pola masa lampau akan berlanjut ke masa (Brown, 1963). Hampir dari seluruh peramalan didasarkan pada asumsi bahwa masa lampau akan berulang kembali. Dalam kehidupan sehari-hari banyak contoh peristiwa iklim yang dapat diramalkan seperti curah hujan, banjir, kecepatan angin dan lain sebagainya. Berdasarkan jangka waktunya, peramalan dibagi menjadi 3 yaitu peramalan jangka pendek (kurang dari satu tahun), menengah (1 hingga 3 tahun), dan jangka panjang (lebih dari 3 tahun).

2.4 Analisis *Time Series*

Menurut Wei (2006), analisis *time series* atau metode deret berkala merupakan suatu rangkaian kegiatan dalam melakukan pengamatan terhadap variabel yang berbasis urutan waktu, dimana untuk memprediksi masa depan dibutuhkan data-data masa lalu. Adapun tujuan dari peramalan deret waktu yaitu untuk menemukan pola deret data masa lalu berdasarkan waktu dan mengekstrapolasikan pola dalam deret data masa lalu dan masa depan. Setiap pengamatan dalam analisis *time series* dinyatakan dalam variabel random Z_t berdasarkan indeks waktu tertentu t_i dengan $i=1,2,\dots,n$, sehingga penulisan dari data *time series* yaitu $Z_{t_1}, Z_{t_2}, \dots, Z_{t_n}$.

2.4.1 Stationeritas

Salah satu hal yang harus diperhatikan dalam analisis *time series* yaitu kestasioneran data. Stationeritas data dalam *time series* ditunjukkan apabila ekspektasi dari rata-rata dan variansnya konstan setiap waktu. Ketidakstationeran dalam varians dapat diatasi dengan cara transformasi, sedangkan ketidakstationeran dalam rata-rata dapat diatasi dengan cara melakukan *differencing* (Wei, 2006). Apabila data *time series* tidak stationer dalam rata-rata dan varians, maka disarankan mengatasi stationer dalam varians terlebih dahulu kemudian dilanjutkan stationer dalam rata-

rata. Hal tersebut dikarenakan jika melakukan *differencing* terlebih dahulu maka kemungkinan terdapat hasil yang negatif dan akan menyebabkan kendala teknis saat akan melakukan transformasi.

2.4.2 ACF dan PACF

Autocorrelation Function (ACF) atau fungsi autokorelasi merupakan suatu fungsi yang biasanya digunakan untuk melihat kestasioneran dan menjelaskan hubungan linier atau korelasi antara Z_t dan Z_{t+k} dari suatu proses yang sama dan hanya terpisah oleh lag waktu ke- k . Secara umum rumus ACF dapat ditunjukkan pada persamaan (2.12) (Wei, 2006).

$$\hat{\rho}_k = \frac{\sum_{t=1}^{n-k} (Z_t - \bar{Z})(Z_{t+k} - \bar{Z})}{(Z_t - \bar{Z})^2}, k = 0, 1, 2, \dots \quad (2.12)$$

dengan $\bar{Z} = \frac{\sum_{t=1}^n Z_t}{n}$ adalah rata-rata dari sampel.

Partial Autocorrelation Function (PACF) atau fungsi autokorelasi parsial digunakan untuk mengukur korelasi antara Z_t dan Z_{t+k} setelah pengaruh variabel $Z_{t+1}, Z_{t+2}, \dots, Z_{t+k-1}$, dihilangkan. Perhitungan PACF dapat diperoleh melalui perhitungan sebagai berikut :

$$\hat{\phi}_{k+1, k+1} = \frac{\hat{\rho}_{k+1} - \sum_{j=1}^k \hat{\phi}_{kj} \hat{\rho}_{k+1-j}}{1 - \sum_{j=1}^k \hat{\phi}_{kj} \hat{\rho}_j} \quad (2.13)$$

dan

$$\hat{\phi}_{k+1, j} = \hat{\phi}_{k, j} - \hat{\phi}_{k+1, k+1} \hat{\phi}_{k, k+1-j}, j = 1, 2, \dots, k. \quad (2.14)$$

2.4.3 Metode Peramalan *Time-Series Holt-Winters*

Metode peramalan *Holt-Winters* merupakan gabungan dari metode *Holt* dan *Winters* serta merupakan salah satu metode peramalan yang mampu mengatasi permasalahan adanya indikasi *trend* dan *seasonal* pada data *time-series*. Metode *Holt-Winters*

pada dasarnya terdiri dari tiga persamaan pemulusan yaitu pemulusan keseluruhan, pemulusan *trend*, dan pemulusan *seasonal*. Jika ditinjau dari sifat *seasonal*, metode *Holt-Winters* terdiri dari 2 jenis diantaranya yaitu *Holt-Winters Additive* dan *Holt-Winters Multiplicative*. Perbedaan mendasar antara *Holt-Winters Additive* dan *Holt-Winters Multiplicative* terletak pada fluktuasi *seasonal*. *Holt-Winters Additive* memiliki ukuran fluktuasi *seasonal* yang bersifat tetap (*steady seasonal fluctuations*), sedangkan *Holt-Winters Multiplicative* memiliki ukuran fluktuasi *seasonal* yang bersifat variasi. Pada penelitian ini akan digunakan metode *Holt-Winters Additive*, sebab pola data dari curah hujan sendiri bersifat *steady seasonal fluctuations* artinya berfluktuasi *seasonal* tetap. Adapun persamaan yang digunakan untuk *Holt-Winters Additive* dijabarkan pada persamaan (2.15) (Makridarkis, Wheelwright, & Mc GEE, 1999).

$$L_t = \alpha(Y_t - S_{t-s}) + (1 - \alpha)(L_{t-1} + b_{t-1}) \quad (2.15)$$

$$b_t = \beta(L_t - L_{t-1}) + (1 - \beta)b_{t-1} \quad (2.16)$$

$$S_t = \gamma(Y_t - S_{t-s}) + (1 - \gamma)S_{t-s} \quad (2.17)$$

$$F_{t+m} = L_t + b_t m + S_{t-s+m} \quad (2.18)$$

dimana :

Y_t = data *actual* pada periode ke- t

s = panjang musiman

F_{t+m} = peramalan untuk m periode berikutnya

L_t = nilai pemulusan keseluruhan

b_t = komponen *trend*

S_t = komponen *seasonal*

m = jumlah periode kedepan yang diramalkan

α = konstanta pemulusan

β = konstanta untuk *trend*

γ = konstanta untuk *seasonal*

Berdasarkan persamaan diatas maka diketahui bahwa metode *Holt-Winters* membutuhkan tiga parameter pemulusan diantaranya yaitu *alpha*, *beta*, dan *gamma* yang perlu dioptimalkan dengan mencobakkan banyak kombinasi dari ketiga parameter tersebut. Sehingga dalam penggunaan metode *Holt-Winters*

diperlukan nilai awal dengan cara mengambil secara sembarang beberapa nilai sebagai inisiasi awal sebagai berikut.

$$L_s = \frac{1}{s}(Y_1 + Y_2 + \dots + Y_s) \quad (2.19)$$

$$b_t = \frac{1}{K} \left(\frac{Y_{s+1} - Y_1}{s} + \frac{Y_{s+2} - Y_2}{s} + \dots + \frac{Y_{s+k} - Y_k}{s} \right) \quad (2.20)$$

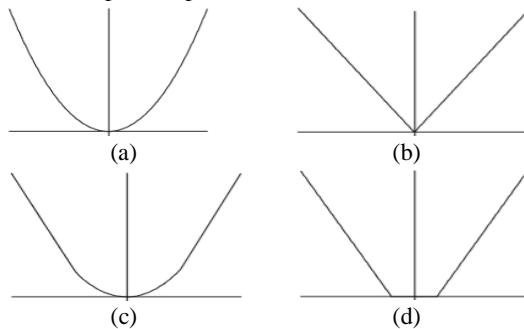
dengan K adalah konstanta pembagi terhadap panjang *seasonal*, sedangkan pemulusan *seasonal* untuk *Holt-Winters Additive* dapat menggunakan nilai awal pada persamaan (2.21) (Montgomery, 2008).

$$S_k = Y_k - L_s \quad (2.21)$$

dengan $k = 1, 2, 3, \dots$

2.5 Support Vector Regression (SVR)

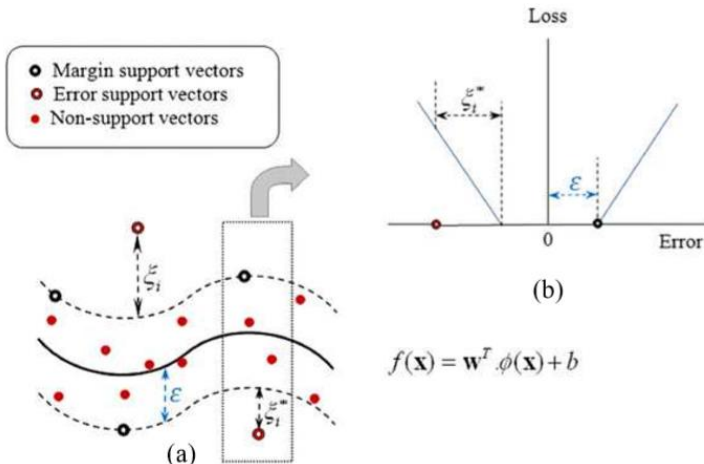
Support Vector Regression merupakan salah satu metode pengembangan dari *Support Vector Machine* (SVM) yang diperkenalkan oleh Vapnik pada tahun 1995. SVM adalah salah satu metode *supervised learning* dengan sistem pembelajaran yang menggunakan ruang hipotesis berupa fungsi-fungsi linier dalam sebuah ruang fitur (*feature space*) berdimensi tinggi dan dilatih dengan algoritma pembelajaran yang didasarkan pada teori optimasi dengan mengimplementasikan *learning bias*. Menurut Gunn (1998) SVM dapat diterapkan pada kasus regresi dengan memperkenalkan *alternative loss function*. *Loss function* dimodifikasi dengan memasukkan ukuran jarak, 4 kemungkinan *loss function* ditampilkan pada Gambar 2.1.



Gambar 2.1 Loss Function (a) Quadratic, (b) Least Modulus, (c) Huber, dan (d) ϵ -Insensitive

Loss function adalah fungsi yang menunjukkan hubungan antara *error* dengan bagaimana *error* dikenai penalti. Perbedaan *loss function* akan menghasilkan formulasi yang berbeda. *Loss function* pada Gambar 2.1 (a) dapat disamakan dengan kriteria *least square error* konvensional, Gambar 2.1 (b) merupakan *loss function Laplacian* yang kurang sensitif terhadap data *outlier*. Huber merupakan *loss function* yang memiliki sifat optimal meskipun distribusi yang mendasari data tidak diketahui. Ketiga *loss function* tersebut menghasilkan *support vector* yang tidak menyebar, sehingga untuk mengatasi masalah tersebut, Vapnik menawarkan ϵ -insensitive *loss function* untuk mendapatkan sekumpulan *support vector* yang tersebar (Gunn, 1998).

Berdasarkan penjelasan Gambar 2.1 maka dapat diketahui bahwa konsep SVM menggunakan konsep ϵ -insensitive *loss function* yang dapat digeneralisasi untuk melakukan pendekatan fungsi yang dikenal dengan SVR (Gunn, 1998). Seperti halnya dengan SVM, SVR juga bertujuan untuk menciptakan *hyperplane* dan *support vector* dalam menentukan nilai prediksi. Konsep SVR didasarkan pada *structural risk minimization*, yaitu untuk mengestimasi suatu fungsi dengan cara meminimalkan batas atas dari *generalization error*, sehingga SVR mampu mengatasi *overfitting*. Berikut merupakan ilustrasi perhitungan *error* pada SVR.



Gambar 2.2 Ilustrasi Perhitungan *Error* pada SVR

Pada Gambar 2.2 dapat diketahui bahwa nilai *error* dianggap sama dengan 0 jika masih berada diantara garis 0 dan garis $f(x)+\varepsilon$, sedangkan apabila berada diluar garis batas toleransi maka nilai *error* akan dihitung sebagai ξ_i dan ξ_i^* . Semakin sempurna suatu fungsi regresi SVR maka batas deviasinya akan bernilai mendekati 0. Jadi intinya dengan menggunakan SVR kita ingin menemukan suatu fungsi $f(x)$ yang mempunyai deviasi paling besar dari target. Jika diketahui pasangan data sebanyak l pengamatan seperti dibawah ini.

$$(y_1, x_1), \dots, (y_l, x_l), \mathbf{x} \in R^n, \mathbf{y} \in R \quad (2.22)$$

Maka fungsi regresi dari metode SVR tersebut dapat dituliskan dalam persamaan (2.23).

$$f(\mathbf{x}) = \mathbf{w}^T \varphi(\mathbf{x}) + b \quad (2.23)$$

dimana :

\mathbf{w} = *vector* pembobot

$\varphi(\mathbf{x})$ = fungsi yang memetakan x dalam satu dimensi

b = bias

Pada persamaan (2.23) simbol $\varphi(\mathbf{x})$ menunjukkan suatu titik dalam ruang fitur F yang merupakan hasil pemetaan \mathbf{x} dalam *input space*. Koefisien \mathbf{w} dan b disini berperan dalam meminimalkan *risk function* sehingga fungsi tersebut dapat terkontrol dan dapat disebut dengan regularisasi. Koefisien \mathbf{w} dan b dapat diestimasi dengan cara meminimalkan *risk function* seperti pada persamaan (2.24) dan (2.25).

$$R(f(\mathbf{x})) = \frac{C}{n} \sum_{t=1}^n L_\varepsilon(\mathbf{y}_t, f(\mathbf{x}_t)) + \frac{1}{2} \|\mathbf{w}\|^2 \quad (2.24)$$

dengan

$$L_\varepsilon(\mathbf{y}_t, f(\mathbf{x}_t)) = \begin{cases} 0, & \text{untuk } |\mathbf{y}_t - f(\mathbf{x}_t)| - \varepsilon \\ |\mathbf{y}_t - f(\mathbf{x}_t)| - \varepsilon, & \text{lainnya} \end{cases} \quad (2.25)$$

dimana :

L_ε = ε -insensitive loss function

\mathbf{y}_t = vektor dari nilai sebenarnya

ε = *hyper*-parameter yang sudah ditentukan nilainya

Fungsi f diasumsikan dapat mendekati semua titik (x_t, y_t) dengan presisi ε . Jika semua titik berada pada rentang $f(x) + \varepsilon$ maka

disebut sebagai kondisi *feasible*, namun jika ada beberapa titik yang berada diluar rentang $f(x) + \varepsilon$ maka kondisi tersebut disebut kondisi *infeasible*. Pada titik-titik yang *infeasible* tersebut dapat ditambahkan variabel *slack* ξ, ξ^* untuk mengatasi *infeasible constrain*.

Jika ditinjau dari jenisnya, SVR terdiri dari dua tipe diantaranya yaitu *Epsilon-Support Vector Regression* (ε -SVR) serta *Nu-Support Vector Regression* (ν -SVR). Perbedaan antara keduanya terletak pada optimasi persamaan (2.24), jika ε -SVR persamaan (2.24) ditransformasi kedalam bentuk yang ditampilkan pada persamaan (2.26), sedangkan ν -SVR persamaan (2.24) ditransformasi kedalam bentuk persamaan (2.30).

$$\frac{1}{2} w^T w + C \sum_{i=1}^N \xi_i + C \sum_{i=1}^N \xi_i^* \quad (2.26)$$

dengan batasan

$$w^T \phi(x_i) + b - y_i \leq \varepsilon + \xi_i^* \quad (2.27)$$

$$y_i - w^T \phi(x_i) - b_i \leq \varepsilon + \xi_i \quad (2.28)$$

$$\xi_i, \xi_i^* \geq 0, i = 1, \dots, N \quad (2.29)$$

$$\frac{1}{2} w^T w + C \left(\nu \varepsilon + \frac{1}{N} \sum_{i=1}^N (\xi_i + \xi_i^*) \right) \quad (2.30)$$

dengan batasan

$$(w^T \phi(x_i) + b) - y_i \leq \varepsilon + \xi_i^* \quad (2.31)$$

$$y_i - (w^T \phi(x_i) - b_i) \leq \varepsilon + \xi_i \quad (2.32)$$

$$\xi_i, \xi_i^* \geq 0, i = 1, \dots, N, \varepsilon \geq 0 \quad (2.33)$$

Dalam ν -SVR parameter ν digunakan untuk menentukan proporsi jumlah *support vector* yang akan digunakan sebagai solusi. Dimana parameter ε dalam ν -SVR dimasukkan kedalam formulasi masalah optimasi dan akan terestimasi secara otomatis. Disisi lain jika dalam ε -SVR, kita dapat mengontrol berapa banyak data *vector* dalam dataset yang akan menjadi *support vector*. Meskipun demikian, kita dapat mengontrol berapa banyak *error* yang dapat terjadi pada model.

Hong (2008) menjelaskan bahwa untuk mengoptimalkan batasan-batasan pada persamaan (2.27), (2.28), (2.31), dan (2.32) tersebut dapat diselesaikan menggunakan *dual lagrangian*. Dimana didalam *dual lagrangian* terdapat fungsi kernel yang berfungsi untuk mengatasi kasus non-linieritas pada data dengan cara mentransformasi data kedalam dimensi ruang yang lebih tinggi, sehingga dapat dipisahkan pada *feature space* yang baru. Pilihan fungsi kernel yang dapat digunakan pada metode SVR dapat dilihat pada Tabel 2.1 (Gunn, 1998).

Tabel 2.1 Macam-macam Fungsi Kernel

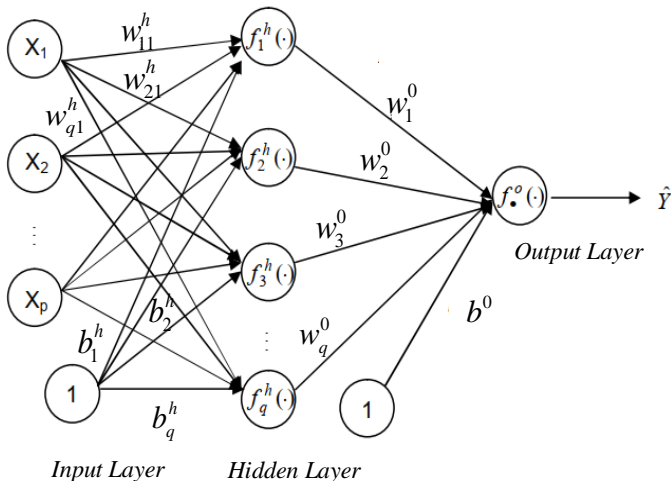
Kernel	Fungsi Kernel
<i>Linear</i>	$K(\mathbf{x}, \mathbf{x}_i) = \mathbf{x}'\mathbf{x}_i$
<i>Polynomial</i>	$K(\mathbf{x}, \mathbf{x}_i) = [(\mathbf{x}'\mathbf{x}_i) + 1]^d$
<i>Radial Basis Function (RBF) atau Gaussian</i>	$K(\mathbf{x}, \mathbf{x}_i) = \exp(-\gamma \ \mathbf{x} - \mathbf{x}_i\ ^2), \gamma > 0$
<i>Sigmoid</i>	$K(\mathbf{x}, \mathbf{x}_i) = \tanh[(\mathbf{x}'\mathbf{x}_i) + b]$
<i>Inverse Multiquadric Function</i>	$K(\mathbf{x}, \mathbf{x}_i) = \frac{1}{\sqrt{\ \mathbf{x} - \mathbf{x}_i\ ^2 + \beta}}$

Pada penelitian fungsi kernel yang akan digunakan adalah fungsi kernel RBF, sebab RBF merupakan fungsi kernel yang sering digunakan dan dapat mengatasi kasus non-linieritas pada data sehingga menghasilkan hasil yang cukup bagus. Lippi, Bertini dan Frasconi (2013) dan Harafani (2015) sebelumnya telah membuktikan bahwa fungsi kernel RBF menghasilkan nilai akurasi peramalan yang lebih baik dibandingkan fungsi kernel lain.

2.6 Artificial Neural Network (ANN)

Artificial neural network atau yang biasa disebut dengan Jaringan Syaraf Tiruan (JST) merupakan salah satu sistem pemroses informasi yang memiliki karakteristik serupa dengan jaringan syaraf pada makhluk hidup atau dengan kata lain merupakan representasi buatan (tiruan) dari otak manusia yang mensimulasikan proses pembelajaran pada otak manusia. Secara sederhana *artificial neural network* merupakan suatu alat pemodelan dan prediksi data statistik non linier yang mampu memperkirakan semua fungsi non-linier dan menyelesaikan

sejumlah proses perhitungan selama proses pembelajaran menggunakan program komputer secara efisien dan stabil. *Artificial neural network* terdiri dari *neuron-neuron* yang saling berhubungan, dimana di dalam *neuron* tersusun atas beberapa lapisan diantaranya yaitu lapisan input (*input layer*), satu atau lebih lapisan tersembunyi (*hiddens layer*) dan lapisan output (*output layer*) (Chong & Zak, 2001). Adapun cara belajar atau proses dari *artificial neural network* yaitu setiap *neuron* akan menerima informasi dari *neuron-neuron* pada *layer* sebelumnya dimana *input neuron* berasal dari bobot *output* pada *layer* sebelumnya. *Input* yang masuk akan diteruskan ke *neuron* dengan bobot tertentu yang biasa disimbolkan dengan w dan akan diproses dengan fungsi perambatan tertentu dengan menjumlahkan semua bobot yang datang. Kemudian hasil penjumlahan akan dibandingkan dengan nilai *threshold* tertentu dengan fungsi aktivasi pada setiap *neuron*. Perubahan bobot yang datang dilakukan secara berulang sampai diperoleh hasil yang diharapkan. Arsitektur dari *artificial neural network* ditunjukkan pada Gambar 2.3.



Gambar 2.3 Arsitektur Neural Network

Fungsi aktivasi pada *neural network* merupakan sebuah fungsi yang digunakan untuk mentransformasikan sebuah *input* menjadi *output* dengan cara mengaktifkan *neuron* yang ada. Beberapa fungsi aktivasi yang sering di gunakan dalam *neural network* diantaranya yaitu fungsi *sigmoid*, fungsi bipolar *sigmoid*, dan fungsi *hyperbolic tangent*. Salah satu *neural network* yang sering digunakan yaitu jaringan dengan banyak lapisan atau biasa disebut dengan *Feed Forward Neural Network* (FFNN). FFNN yang akan digunakan adalah FFNN dengan satu lapis tersembunyi (*hidden layer*) yang terdiri dari q unit *neuron* dan lapis output (*output layer*) terdiri dari satu unit neuron seperti yang di tunjukkan pada Gambar 2.3. Berikut merupakan persamaan untuk menghitung nilai-nilai respon.

$$\hat{Y}_{(t)} = f^o \left[\sum_{j=1}^q \left[w_j^o f_j^h \left(\sum_{i=1}^p w_{ji}^h X_{i(t)} + b_j^h \right) + b^o \right] \right] \quad (2.34)$$

dimana :

$X_{i(t)}$ = variabel *input* sebanyak p , ($i = 1, 2, \dots, p$)

t = indeks pasangan data *input-target* ($X_{i(t)}, \hat{Y}_{(t)}$),
 $t = 1, 2, \dots, n$

w_{ji}^h = bobot dari *input* ke- i yang menuju *neuron* ke- j pada *hidden layer*, dengan $j = 1, 2, \dots, q$

b_j^h = bias pada *neuron* ke- j pada *hidden layer*, $j = 1, 2, \dots, q$

f_j^h = fungsi aktivasi di *neuron* ke- j pada *hidden layer*

w_j^o = bobot dari *neuron* ke- j di *hidden layer* yang menuju *neuron* pada *output layer*

b^o = bias pada *neuron* di *output layer*

f^o = fungsi aktivasi pada *neuron* di *output layer*

$\hat{Y}_{(t)}$ = nilai dugaan dari variabel *output* atau respon

Adapun untuk mengatasi dependensi waktu pada residual maka Y_{t-k} dapat dijadikan sebagai input. Bagian terpenting dari

neural network adalah pada proses *training*. Tujuan dari proses *training* yaitu untuk pengaturan terhadap bobot-bobot yang terbentuk sehingga diperoleh bobot akhir yang tepat sesuai dengan pola data yang dilatih. Pada penelitian ini akan digunakan fungsi aktivasi *logistic* dan *tanh* dengan replikasi sebanyak 5 kali. Adapun fungsi aktivasi *logistic* atau sigmoid merupakan fungsi standar dalam *training Neural Network* dengan nilai fungsi antara 0 sampai dengan 1 yang dituliskan pada persamaan (2.35).

$$f(x) = \frac{1}{1 + e^{-x}} . \quad (2.35)$$

Sedangkan persamaan fungsi aktivasi *tanh* (*Hyperbolic Tangent Function*) dituliskan pada persamaan (2.36).

$$f(x) = \frac{\sinh(x)}{\cosh(x)} = \frac{e^x - e^{-x}}{e^x + e^{-x}} . \quad (2.36)$$

Fungsi aktivasi *tanh* memiliki akurasi yang paling tinggi saat digunakan melakukan *training* pada NN (Karlik & Olgac, 2011).

2.7 Validasi Model dan Pemilihan Metode Terbaik

Validasi model adalah proses untuk mengetahui apakah suatu model yang terbentuk menghasilkan nilai prediksi yang akurat. Dalam melakukan pemodelan *time series* data dibagi menjadi dua bagian, yaitu data *in sample* (data *training*) dan data *out sample* (data *testing*). Data *in sample* digunakan untuk mendapatkan model, sedangkan data *out sample* digunakan untuk melakukan validasi terhadap model tersebut. Begitu pula pemilihan metode terbaik dalam pemodelan *time series* dapat dibedakan menjadi dua kriteria, yaitu kriteria *in sample* dan *out sampel*. Kriteria *in sample* dapat dilihat melalui *Root of Mean Square Error* (RMSE), *Mean Absolute Percentage Error* (MAPE), atau *Symmetric Mean Absolute Percentage Error* (sMAPE). Sedangkan pemilihan metode terbaik untuk kriteria *out sample* dapat menggunakan nilai RMSEP yang terkecil (Wei, 2006). Selain itu pemilihan model terbaik berdasarkan data *out sample* juga dapat dilakukan menggunakan $R^2_{\text{prediction}}$ atau koefisien determinasi. Semakin kecil nilai sMAPE dan RMSEP menunjukkan bahwa model yang didapatkan semakin baik. Sebaliknya, semakin besar nilai $R^2_{\text{prediction}}$

menunjukkan bahwa model yang didapatkan semakin baik. sMAPE dapat dituliskan melalui persamaan (2.37) (Wei, 2006).

$$sMAPE = \left(\frac{1}{n} \sum_{t=1}^n \frac{|Z_t - \hat{Z}_t|}{\left(|Z_t| + |\hat{Z}_t| \right) / 2} \right) \times 100\% \quad (2.37)$$

Sedangkan rumus untuk RMSEP dapat dituliskan dalam persamaan (2.38) (Wei, 2006).

$$RMSEP = \sqrt{\frac{\sum_{t=1}^n (Z_t - \hat{Z}_t)^2}{n}} \quad (2.38)$$

Rumus MAPE ditunjukkan melalui rumus dibawah ini.

$$MAPE = \left(\frac{1}{n} \sum_{t=1}^n \frac{|Z_t - \hat{Z}_t|}{Z_t} \right) \times 100\% \quad (2.39)$$

Adapun rumus untuk nilai $R^2_{\text{prediction}}$ adalah sebagai berikut :

$$R^2_{\text{prediction}} = \left(\frac{\sum_{t=1}^n (\hat{Z}_t - \bar{Z})^2}{\sum_{t=1}^n (Z_t - \bar{Z})^2} \right) \times 100\% \quad (2.40)$$

dimana :

Z_t = nilai data aktual pada waktu ke- t

\hat{Z}_t = data hasil ramalan

n = banyaknya data

Model terbaik adalah model yang memiliki nilai RMSE, MAPE, dan sMAPE terkecil. Semakin kecil nilai RMSE, MAPE, dan sMAPE menunjukkan semakin akurat hasil prediksi. Pada penelitian ini hanya akan digunakan kriteria kebaikan model RMSE, sebab kriteria R^2 tidak valid digunakan pada data non-linier seperti halnya data curah hujan. Sedangkan untuk MAPE dan sMAPE tidak digunakan karena akan menghasilkan nilai *infinite*. Hal tersebut disebabkan data curah hujan ada yang bernilai 0,

sehingga jika dimasukkan kedalam persamaan (2.37) dan (2.39) akan menghasilkan nilai *infinite* atau dengan kata lain jika sesuatu nilai dibagi dengan nilai 0 maka tidak akan dapat didefinisikan.

2.8 Perhitungan Matematis Banjir

Terdapat banyak metode perhitungan dalam menentukan banjir rencana, salah satu diantaranya yaitu hubungan empiris antara curah hujan dan limpasan menggunakan metode rasional. Metode rasional sudah dipakai sejak abad pertengahan 19 dan merupakan metode yang sering dipakai untuk perencanaan banjir daerah perkotaan. Walaupun banyak yang mengkritik, namun metode ini tetap dipakai karena kesederhanaannya. Metode ini juga menunjukkan parameter-parameter yang dipakai metode-metode perkiraan banjir lainnya, yaitu koefisien *run-off*, intensitas hujan dan luas DPS. Metode rasional digunakan untuk daerah perkotaan dengan luas DPS kurang dari 200 *acres* atau lebih kurang 81 ha, sehingga sesuai dengan kondisi Kota Surabaya yang memiliki luas kurang lebih 35,050 ha. Adapun persamaan dari metode rasional sebagai berikut :

$$Q = 0,278 C I A \quad (2.41)$$

$$\frac{Q}{0,278 C A} = I \quad (2.42)$$

sehingga,

$$Q^* = I \quad (2.43)$$

dengan

$$I = \left[\frac{R_{24}}{24} \right] \left[\frac{24}{t_c} \right]^{2/3} \quad (2.44)$$

dimana :

C = koefisien *run-off* (dari tabel atau dengan rumus)

I = intensitas maksimum selama waktu konsentrasi (mm/jam)

A = luas daerah aliran (km²)

Q = debit maksimum(m³/detik)

R = hujan maksimum (mm)

t_c = waktu konsentrasi (jam).

Persamaan (2.41) menunjukkan bahwa dalam perhitungan debit hidrologi, hal yang harus dipertimbangkan adalah nilai I sebab nilai C dan A akan selalu tetap sedangkan nilai I akan selalu

berubah-ubah bergantung pada curah hujan maksimum. Koefisien *run-off* dari masing-masing karakteristik wilayah dijelaskan pada Tabel 2.2.

Tabel 2.2 Tabel Koefisien *Run-off* (C)

No	Tata Guna Lahan	Koefisien Pengaliran
Daerah Perdagangan		
1	Perkotaan (<i>downtown</i>)	0,70-0,90
	Pinggiran	0,50-0,70
Pemukiman		
2	Perumahan satu keluarga	0,30-0,50
	Perumahan berkelompok, terpisah-pisah	0,40-0,60
	Perumahan berkelompok, bersambungan	0,60-0,75
	Suburban	0,25-0,40
	Daerah apartemen industri	0,50-0,70
	Daerah industri ringan	0,50-0,80
	Daerah industri berat	0,60-0,90
3	Taman, perkuburan	0,10-0,25
4	Tempat Bermain	0,20-0,35
5	Daerah stasiun kereta api	0,20-0,35
6	Daerah belum diperbaiki	0,10-0,30
Jalan		
7	Jalan hampanan	0,75-0,85
	Atap	0,75-0,95

Nilai dari t_c dapat dihitung melalui persamaan berikut :

$$t_c = t_0 + t_f \quad (2.45)$$

dimana :

- t_0 = waktu yang dibutuhkan untuk mengalir di permukaan untuk mencapai inlet (*overland flow time, inlet time*)
 t_f = waktu yang diperlukan untuk mengalir di sepanjang saluran.

Dalam menentukan harga t_0 digunakan asumsi bahwa permukaan yang dianalisis memiliki penutup *homogeny* sehingga

untuk perhitungan digunakan rumus Kerby dengan persamaan berikut :

$$t_0 = 1,44 \left(n_d \times \frac{l}{\sqrt{s}} \right)^{0,467} \quad (2.46)$$

dimana :

l = jarak dari titik terjauh ke inlet (m)

n_d = koefisien setara koefisien kekasaran

s = kemiringan saluran.

Asumsi-asumsi metode ini diantaranya (Kodoatie & Sugiyanto, 2002) :

- Curah hujan mempunyai intensitas yang merata di seluruh daerah aliran untuk durasi tertentu.
- Debit yang terjadi (debit puncak) bukan hasil dari intensitas hujan yang lebih tinggi dengan durasi yang lebih pendek dimana hal ini berlangsung hanya pada sebagian DPS yang berkontribusi debit puncak tersebut.
- Lamanya curah hujan = waktu konsentrasi dari daerah aliran. Dengan kata lain waktu konsentrasi merupakan waktu terjadinya *run-off* dan mengalir dari jarak antara titik terjauh dari DPS ke titik *inflow* yang ditinjau.
- Puncak banjir dan intensitas curah hujan mempunyai tahun berulang yang sama.

Air meluap yang disebabkan oleh saluran air atau sungai yang tidak mampu menampung air hujan dengan baik sehingga mengakibatkan banjir terjadi pada saat $Q_{hidrolika} < Q_{hidrologi}$ atau debit hidrolika kurang dari debit hidrologi. Sehingga untuk mengantisipasi terjadinya banjir maka seharusnya desain ideal untuk saluran-saluran yang ada maksimal harus memiliki $Q_{hidrologi} \approx Q_{hidrolika}$. Debit hidrolika dapat dirumuskan melalui persamaan (2.47).

$$Q_{hidrolika} = V_{(m/detik)} \times A_{(m^2)} = \frac{1}{n} \times R^{\frac{2}{3}} \times S^{\frac{1}{2}} \times A \quad (2.47)$$

dengan

$$A_{(m^2)} = b_{(m)} \times h_{(m)} \quad (2.48)$$

dimana :

V = kecepatan saluran

A = luas penampang sungai

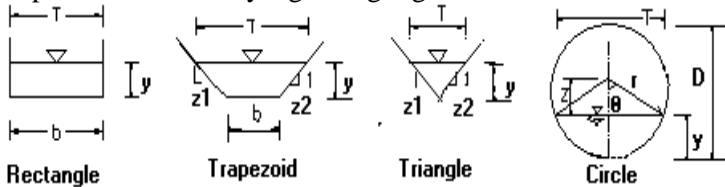
b = lebar sungai

h = tinggi sungai

$b_{(m)}$ diperoleh dari rata-rata lebar sungai minimum dan maksimum yang diperoleh dari data Dinas Pekerjaan Umum Bina Marga dan Pematuan Kota Surabaya. Sedangkan untuk nilai $h_{(m)}$ dari tabel *u-ditch* tergantung nilai $b_{(m)}$ -nya. Dalam memudahkan analisis, maka dapat diketahui $h_{(m)}$ suatu saluran adalah sama dengan $b_{(m)}$ jika $b_{(m)}$ lebih besar sama dengan 12. Jika $b_{(m)} < 12$ maka :

$$h_{(m)} = b_{(m)} \times \frac{10}{6} \quad (2.49)$$

Hal ini dikarenakan, keseluruhan saluran di Kota Surabaya menggunakan *u-ditch* atau saluran *rectangle*. Berikut merupakan beberapa bentuk saluran yang sering digunakan.



Gambar 2.4 Macam-macam Bentuk Saluran

Adapun kecepatan saluran dapat dihitung menggunakan Rumus *Manning* dibawah ini.

$$V_{(m/detik)} = \frac{1}{n} R^{2/3} S^{1/2} \quad (2.50)$$

Keterangan

n = koefisien kekerasan saluran

R = jari-jari hidrolis

S = kemiringan salur

Sedangkan untuk jari-jari hidrolis dapat dirumuskan dalam persamaan (2.51).

$$R_{(m)} = \frac{A_{(m^2)}}{P_{(m)}} \quad (2.51)$$

Dimana $P_{(m)}$ atau penampang yang terkena air dapat dituliskan melalui persamaan (2.52).

$$P_{(m)} = b_{(m)} + 2h_{(m)} \quad (2.52)$$

Dalam menentukan kecepatan aliran di saluran tergantung material saluran yang digunakan. Semakin halus permukaan saluran maka kecepatan semakin besar. Saluran juga harus dapat menahan kecepatan agar tidak tergerus aliran. Kekasaran juga merupakan hambatan pada aliran, berikut tabel koefisien kekasaran saluran yang sering digunakan.

Tabel 2.3 Tabel Kekasaran *Manning* untuk Saluran

Saluran	Keterangan	<i>n</i> Manning
Tanah	Lurus, baru, seragam, landai & bersih	0,016-0,033
	Berkelok, landai & berumput	0,023-0,040
	Tidak terawat & kotor	0,050-0,140
	Tanah berbatu, kasar & tidak teratur	0,035-0,045
Pasangan	Batu kosong	0,023-0,035
	Pasangan batu belah	0,017-0,030
Beton	Halus, sambungan baik dan rata	0,014-0,018
	Kurang halus dan sambungan kurang rata	0,018-0,030

2.9 Curah Hujan

Curah hujan merupakan ketinggian air hujan yang terkumpul dalam tempat yang datar, tidak menguap, tidak meresap, dan tidak mengalir. Curah hujan satu milimeter artinya adalah dalam luasan satu meter persegi tempat yang datar, tertampung air setinggi satu milimeter atau tertampung air sebanyak satu liter. Intensitas hujan merupakan banyaknya curah hujan per-satuan jangka waktu tertentu. Jadi, apabila intensitas hujan di katakan besar, itu tandanya hujan lebat dan dapat menimbulkan banjir. Berdasarkan intensitasnya, curah hujan dibedakan menjadi 3 yaitu hujan sedang berada diantara 20 dan 50 mm perhari, hujan lebat berada diantara 50 dan 100 mm perhari, dan hujan sangat lebat berada diatas 100 mm perhari (Suriadikusumah, 2007). Di beberapa negara banyaknya curah hujan masih dinyatakan dengan inci. Biasanya pembacaan dilakukan satu kali dalam sehari dan dicatat sebagai curah hujan hari terdahulu atau kemarin. Data curah hujan yang

terkumpul nantinya dapat digunakan sebagai acuan untuk prediksi curah hujan pada kemudian hari (Siswanti, 2011).

2.10 Banjir

Berdasarkan jenisnya, banjir yang sering terjadi di Surabaya merupakan banjir air yaitu jenis banjir yang sangat umum terjadi, biasanya banjir ini terjadi akibat meluapnya air sungai, danau atau selokan akibat curah hujan yang berlebihan. Dikarenakan intensitasnya yang banyak menyebabkan air tidak tertampung dan meluap itulah banjir air. Banjir air sangat sering terjadi saat hujan deras dalam kurun waktu yang lama, sehingga air tidak tertampung, meluap dan menyebabkan genangan dimana-mana (Kodoatie & Sugiyanto, 2002). Pengertian genangan sendiri merupakan terhentinya aliran atau air tidak mengalir. Menurut Peraturan Menteri Pekerjaan Umum Nomor 01/PRT/M/2014 tentang Petunjuk Teknis Standar Pelayanan Minimal Bidang Pekerjaan Umum dan Penataan Ruang, terkait penjabaran drainase, definisi genangan adalah terendamnya suatu kawasan perkotaan lebih dari 30 cm selama lebih dari 2 jam. Sedangkan genangan yang dikategorikan banjir yaitu terendamnya suatu kawasan perkotaan lebih dari 100 cm atau 1 meter dengan durasi lebih dari 24 jam dan menyebabkan perekonomian suatu daerah terhenti. Sedangkan jika di lihat dari sisi *engineering*, banjir dan genangan merupakan suatu kawasan/daerah/dataran yang terendam air, yang memiliki kesamaan pada volume air yang melimpah namun memiliki skala yang berbeda, dimana banjir memiliki skala yang lebih besar daripada genangan.

2.11 Dinas Pekerjaan Umum Bina Marga dan Pematusan Kota Surabaya

Dinas pemerintah daerah yang berkaitan dengan banjir adalah Dinas Pekerjaan Umum Bina Marga dan Pematusan. Adapun tugas Dinas PU&Pematusan yaitu melaksanakan sebagian urusan Pemerintahan dibidang pekerjaan umum, perhubungan, pemberdayaan masyarakat, otonomi daerah, pemerintahan umum, dan administrasi keuangan daerah. Bidang Pematusan mempunyai tugas melaksanakan sebagian tugas Dinas Pekerjaan Umum Bina

Marga dan Pematuan di bidang pematuan. Rincian tugas Bidang Pematuan sebagai berikut.

- a. Pelaksanaan operasi, pemeliharaan dan rehabilitasi pada sungai, danau, waduk dan pantai pada wilayah sungai dalam satu kota.
- b. Penyusunan peraturan daerah mengenai kebijakan dan strategi kota berdasarkan kebijakan nasional dan provinsi.
- c. Penyelesaian masalah dan permasalahan operasionalisasi sistem drainase dan penanggulangan banjir di wilayah kota serta koordinasi dengan daerah sekitarnya.
- d. Penyelenggaraan pembangunan dan pemeliharaan Prasarana dan Sarana drainase di wilayah kota.
- e. Pembangunan, pemeliharaan, pengerukan alur pelayaran sungai dan danau kota.
- f. Pelaksanaan koordinasi dan fasilitasi pemeliharaan prasarana dan sarana serta penyehatan lingkungan skala kota.

2.12 Balai PSAWS Buntung Peketingan Surabaya

Balai PSAWS Buntung Peketingan Surabaya merupakan salah satu subfond atau badan dari beberapa subfond yang berada di bawah naungan Dinas Pekerjaan Umum Pengairan Provinsi Jawa Timur. Adapun salah satu tugas dari Balai PSAWS yang ada di Surabaya ini yaitu melakukan pengukuran mengenai curah hujan yang ada di Kota Surabaya. Balai PSAWS membagi daerah Surabaya menjadi 7 wilayah pengamatan hujan.

2.13 Badan Meteorologi, Klimatologi, dan Geofisika Kota Surabaya

Badan Meteorologi Klimatologi dan Geofisika sebagai Lembaga Pemerintah Non Departemen (LPND) adalah instansi pelayanan dan penyediaan informasi di bidang Meteorologi, Klimatologi dan Geofisika. Adapun beberapa variabel klimatologi yang diukur dan dipantau oleh BMKG diantaranya seperti curah hujan, kelembapan udara, tekana udara, suhu udara, persentase penyinaran matahari dan kecepatan angin.

BAB III

METODOLOGI PENELITIAN

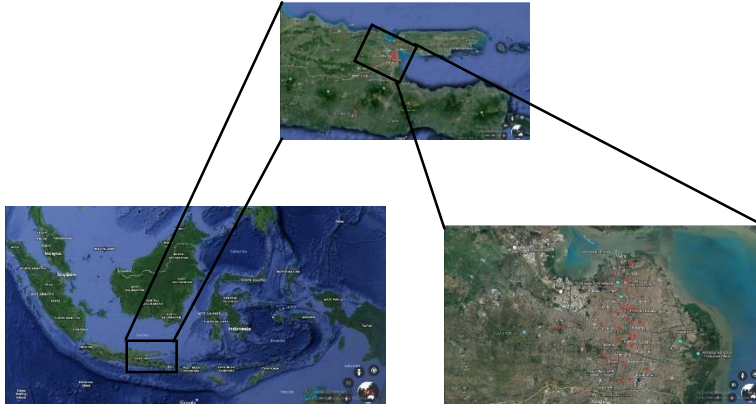
Metodologi penelitian ini akan menjelaskan tentang sumber data, daerah penelitian, variabel penelitian, definisi operasional, struktur data, langkah penelitian, beserta diagram alirnya.

3.1 Sumber Data

Data yang digunakan dalam penelitian tugas akhir ini adalah data sekunder yang diperoleh dari Dinas Pekerjaan Umum Bina Marga dan Pematuan (Dinas PU dan Pematuan) Kota Surabaya, Balai PSAWS Buntung Peketingan Surabaya, dan Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) Kota Surabaya. Pada Dinas PU dan Pematuan Kota Surabaya didapatkan data mengenai informasi saluran-saluran atau drainase yang ada di 5 rayon pematuan. Sedangkan pada Balai PSAWS Buntung Peketingan Surabaya didapatkan data mengenai curah hujan harian dari 7 balai dan pada BMKG Kota Surabaya didapatkan data harian mengenai temperatur, kelembapan udara, dan kecepatan angin pada 3 stasiun BMKG. Data yang digunakan dari Balai PSAWS Buntung Peketingan Surabaya maupun BMKG Kota Surabaya merupakan data harian mulai bulan Januari 2009 sampai bulan Desember 2018, dimana data tersebut akan dibagi menjadi data *in sample (training)* dan *out sample (testing)* terlebih dahulu. Data *in sample* yang digunakan untuk membangun model merupakan data dari bulan Januari 2009 hingga Nopember 2018 untuk semua variabel, sedangkan data *out sample* yang digunakan untuk validasi model adalah data bulan Desember 2018.

3.2 Daerah Penelitian

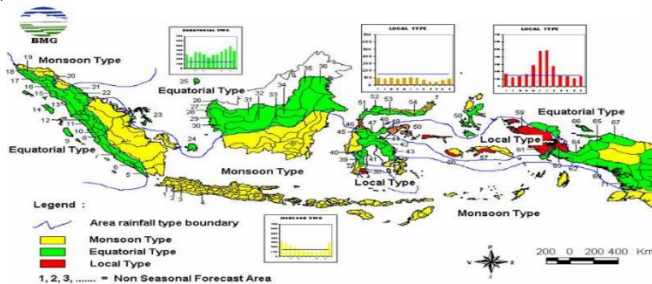
Kota Surabaya dipilih sebagai daerah penelitian untuk memprediksi curah hujan guna mengantisipasi resiko banjir, berikut merupakan letak administratif Kota Surabaya yang ditunjukkan pada Gambar 3.1.



Gambar 3.1 Letak Administratif Kota Surabaya

(Sumber : google earth.com)

Kota Surabaya merupakan Ibukota Provinsi Jawa Timur, dimana yang secara geografis berada pada $7^{\circ} 9' - 7^{\circ} 21'$ Lintang Selatan dan $112^{\circ} 36' - 112^{\circ} 57'$ Bujur Timur. Sebagian besar wilayah Kota Surabaya merupakan dataran rendah dengan ketinggian 3 - 6 meter di atas permukaan laut, sedangkan sebagian lainnya pada sebelah Selatan merupakan kondisi berbukit-bukit dengan ketinggian 25 - 50 meter di atas permukaan laut (BAPEDA, 2013).



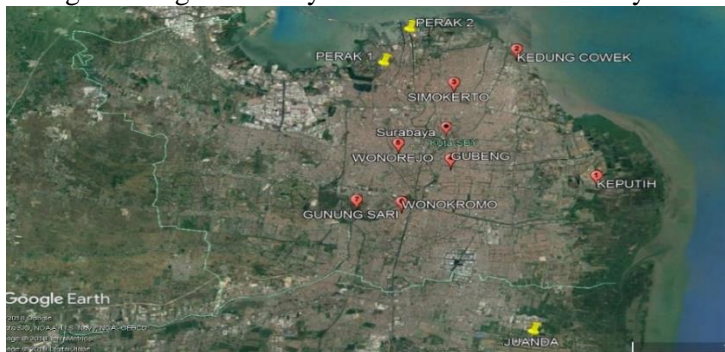
Gambar 3.2 Pola Hujan Surabaya Berdasarkan Jenisnya

(Sumber : Caraka & Tahmid, 2019)

Gambar 3.2 menunjukkan bahwa BMKG membagi wilayah Indonesia menjadi tiga pola hujan, pertama pola hujan monsun, kedua pola hujan equatorial dan terakhir pola hujan lokal. Dimana wilayah Surabaya sendiri masuk kedalam pola hujan monsun yaitu suatu daerah yang memiliki perbedaan sangat jelas antara periode angin monsun barat yang membawa musim hujan dengan periode

musim angin monsun timur yang membawa musim kemarau (Caraka & Tahmid, 2019). Oleh sebab itulah, di Surabaya hanya mengenal 2 musim yaitu musim hujan dan kemarau. Musim kemarau terjadi antara bulan Mei hingga Oktober dan musim hujan terjadi antara bulan November sampai April. Temperatur Kota Surabaya cukup panas, yaitu rata-rata antara $22,60^{\circ}\text{C}$ – $34,10^{\circ}\text{C}$, dengan tekanan udara rata-rata antara 1005,2 – 1013,9 milibar dan kelembaban antara 42% - 97%. Kecepatan angin rata-rata perjam mencapai 12 – 23 km, sedangkan curah hujan rata-rata antara 120 – 190 mm (BPN, 2001).

Berikut juga ditampilkan area pengamatan Balai PSAWS Buntung Peketingan Surabaya dan BMKG Kota Surabaya.



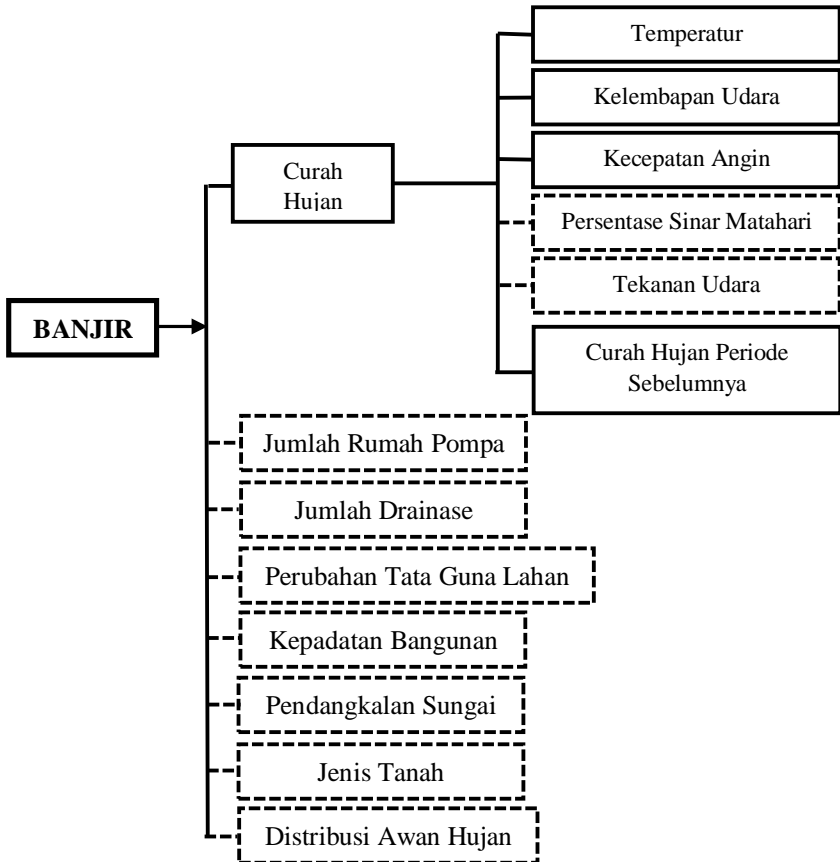
Gambar 3.3 Wilayah Pengamatan Balai PSAWS Buntung Peketingan dan BMKG Kota Surabaya
(Sumber : google earth.com)

Dari Gambar diatas diketahui bahwa daerah penelitian antara data yang diperoleh dari Balai PSAWS Buntung Peketingan Surabaya memiliki skala yang berbeda dengan data yang diperoleh dari BMKG. Balai PSAWS Buntung Peketingan Surabaya membagi wilayah Surabaya menjadi 7 wilayah pengamatan yang selanjutnya disebut dengan balai diantaranya yaitu Keputih, Kedung Cowek, Simo, Gubeng, Wonorejo, Wonokromo dan Gunung Sari. Sedangkan untuk wilayah pengamatan BMKG hanya terdiri dari 3 stasiun di Surabaya yaitu stasiun Juanda, stasiun Perak 1, dan stasiun Perak 2. Sehingga, berdasarkan penjelasan mengenai wilayah pengamatan Balai PSAWS Buntung Peketingan Surabaya dan BMKG Kota Surabaya, maka dapat diketahui bahwa data curah hujan akan memiliki 7 nilai yang berbeda untuk masing-

masing balai sedangkan faktor-faktor yang mempengaruhi curah hujan akan memiliki 3 nilai yang berbeda pada masing-masing stasiun BMKG.

3.3 Variabel Penelitian

Berdasarkan hasil studi literatur mengenai faktor-faktor yang berpengaruh terhadap banjir serta faktor-faktor yang mempengaruhi curah hujan. Maka penentuan variabel pada penelitian ini dapat dilihat pada Gambar 3.4.



Gambar 3.4 Kerangka Berfikir atau Kerangka Penelitian

Gambar 3.4 menunjukkan hubungan antara variabel yang digunakan untuk memprediksi banjir dan curah hujan, dimana variabel yang berada dalam kotak penuh dan kotak putus-putus adalah variabel yang diduga mempengaruhi banjir. Namun karena kesulitan mendapatkan data atau dengan kata lain ketersediaan data tidak mendukung maka variabel-variabel yang berada dalam kotak putus-putus tidak digunakan dalam penelitian ini. Sehingga dalam penelitian ini hanya akan menggunakan variabel yang berada dalam kotak penuh. Untuk lebih jelasnya, variabel yang akan digunakan dalam penelitian ini ditampilkan pada Tabel 3.1.

Tabel 3.1 Kebutuhan Data Penelitian

Jenis / Notasi Variabel	Nama Variabel	Satuan	Jenis Data	Keterangan	
<i>Response</i>	Z	Banjir Genangan	Sentimeter (cm)	Rasio	
<i>Predictant</i>	Y	Curah Hujan	Milimeter (mm)	Rasio	data harian di 7 balai
<i>Predictor</i>	X_1	Temperatur Rata-rata	Derajat Celcius (°C)	Rasio	data harian di 3 stasiun
	X_2	Kelembapan Udara Rata-rata	Persen (%)	Rasio	data harian di 3 stasiun
	X_3	Kecepatan Angin Rata-rata	Meter per detik (m/s)	Rasio	data harian di 3 stasiun
	Y_{t-k}	Curah Hujan Periode Sebelumnya	Milimeter (mm)	Rasio	data harian di 7 balai

Berdasarkan Tabel 3.1 maka dapat disimpulkan bahwa variabel X berpengaruh terhadap variabel Y sedangkan variabel Y berpengaruh terhadap variabel Z atau dengan kata lain dapat dikatakan bahwa variabel Z diprediksi menggunakan variabel Y sedangkan variabel Y sendiri diprediksi menggunakan variabel X. Selain itu pada variabel *predictor* juga akan ditambahkan variabel *dummy* guna untuk menangkap 2 pola musim hujan di Indonesia yaitu kemarau dan hujan. Pada musim kemarau data curah hujan cenderung bernilai 0, sedangkan adanya nilai 0 yang lumayan banyak ketika musim kemarau ini dapat menjadi hambatan dalam

proses analisis sebab akan lebih sulit untuk diprediksi dan akan memperbesar nilai *error* yang terbentuk. Oleh karena itu, penambahan variabel *dummy* disini sangat penting dan bermanfaat untuk membedakan antara musim kemarau dan musim hujan. Diharapkan dengan adanya penambahan variabel *dummy* pada variabel *predictor* dapat membantu menaikkan nilai RMSE dari model yang terbentuk. Berikut penjelasan mengenai variabel *dummy* yang akan digunakan dalam penelitian.

$$M_t = \begin{cases} 0, & \text{untuk bulan kemarau tanggal 1 Mei – 31 Okt} \\ 1, & \text{untuk bulan hujan tanggal 1 Nov – 31 April} \end{cases}$$

dimana :

M_t = variabel *dummy*

3.4 Definisi Operasional

Definisi operasional dari variabel-variabel yang akan digunakan dalam penelitian ini adalah sebagai berikut :

1. Genangan yaitu suatu kawasan/daerah/dataran yang terendam air lebih dari 30 cm selama lebih dari 2 jam dengan cakupan wilayah yang cukup kecil, sedangkan banjir merupakan suatu peristiwa terendamnya kawasan/daerah dataran selama lebih dari 100 cm dengan waktu lebih dari 24 jam dan terjadi pada cakupan wilayah yang lebih besar daripada genangan (Maulina, 2016). Menurut Dinas PU&Pematuan Kota Surabaya, parameter yang merepresentasikan banjir yaitu dalam, luas, dan lama genangan.
2. Curah hujan merupakan tinggi air hujan yang sampai ke permukaan tanah dan diukur berdasarkan volume air hujan persatuan luas. Curah hujan dibatasi sebagai tinggi air hujan yang diterima dipermukaan sebelum mengalami aliran permukaan, evaporasi dan peresapan ke dalam tanah (Manullang, 2014).
3. Temperatur adalah ukuran energi kinetik rata – rata dari pergerakan molekul – molekul (Noer, 2016).
4. Kelembapan udara adalah perbandingan jumlah uap air dalam udara dengan jumlah air maksimum yang dapat dikandung oleh udara tersebut dalam suhu yang sama dan dinyatakan dalam persen (%). Kelembapan udara yang relatif rendah (< 40%) dapat mengakibatkan kekeringan selaput lendir

membran, sedangkan kelembapan yang tinggi ($> 70\%$) dapat meningkatkan pertumbuhan mikroorganisme (Candrasari & Mukono, 2013).

5. Kecepatan angin atau velositas gelombang angin adalah sebuah kuantitas atmosferik fundamental. Kecepatan angin disebabkan oleh pergerakan angin dari tekanan tinggi ke tekanan rendah, biasanya karena perubahan suhu. Adapun satuan kecepatan angin yaitu Knots (Skala Beaufort) dan dapat juga dalam m/s (Derek, Allo, & Tulung, 2016).

3.5 Struktur Data

Struktur data secara umum yang akan digunakan dalam penelitian ini ditunjukkan secara rinci pada Tabel 3.2 dan Tabel 3.3.

Tabel 3.2 Struktur Data Penelitian Keseluruhan

Tanggal	Data PSAWS	Data BMKG		
	Y_b	X_{1s}	X_{2s}	X_{3s}
<i>Data In Sample</i>				
2009/01/01	$Y_{b,1}$	$X_{1s,1}$	$X_{2s,1}$	$X_{3s,1}$
2009/01/02	$Y_{b,2}$	$X_{1s,2}$	$X_{2s,2}$	$X_{3s,2}$
2009/01/03	$Y_{b,3}$	$X_{1s,3}$	$X_{2s,3}$	$X_{3s,3}$
2009/01/04	$Y_{b,4}$	$X_{1s,4}$	$X_{2s,4}$	$X_{3s,4}$
2009/01/05	$Y_{b,5}$	$X_{1s,5}$	$X_{2s,5}$	$X_{3s,5}$
\vdots	\vdots	\vdots	\vdots	\vdots
2018/11/27	$Y_{b,3618}$	$X_{1s,3618}$	$X_{2s,3618}$	$X_{3s,3618}$
2018/11/28	$Y_{b,3619}$	$X_{1s,3619}$	$X_{2s,3619}$	$X_{3s,3619}$
2018/11/29	$Y_{b,3620}$	$X_{1s,3620}$	$X_{2s,3620}$	$X_{3s,3620}$
2018/11/30	$Y_{b,3621}$	$X_{1s,3621}$	$X_{2s,3621}$	$X_{3s,3621}$
<i>Data Out Sample</i>				
2018/12/01	$Y_{b,3622}$	$X_{1s,3622}$	$X_{2s,3622}$	$X_{3s,3622}$
2018/12/02	$Y_{b,3623}$	$X_{1s,3623}$	$X_{2s,3623}$	$X_{3s,3623}$
2018/12/03	$Y_{b,3624}$	$X_{1s,3624}$	$X_{2s,3624}$	$X_{3s,3624}$
2018/12/04	$Y_{b,3625}$	$X_{1s,3625}$	$X_{2s,3625}$	$X_{3s,3625}$
2018/12/05	$Y_{b,3626}$	$X_{1s,3626}$	$X_{2s,3626}$	$X_{3s,3626}$
\vdots	\vdots	\vdots	\vdots	\vdots
2018/12/28	$Y_{b,3649}$	$X_{1s,3649}$	$X_{2s,3649}$	$X_{3s,3649}$
2018/12/29	$Y_{b,3650}$	$X_{1s,3650}$	$X_{2s,3650}$	$X_{3s,3650}$
2018/12/30	$Y_{b,3651}$	$X_{1s,3651}$	$X_{2s,3651}$	$X_{3s,3651}$
2018/12/31	$Y_{b,3652}$	$X_{1s,3652}$	$X_{2s,3652}$	$X_{3s,3652}$

dimana :

s = 1 (Juanda), 2 (Perak 2), 3 (Perak 1) ; stasiun BMKG
 b = 1 (Keputih), 2 (Kedung Cowek), 3 (Gubeng), 4 (Wonorejo), 5 (Wonokromo), 6 (Gunung Sari), 7 (Simo) ; balai PSAWS

Tabel 3.3 Struktur Data Penelitian untuk PCA

t	Y _b	Data BMKG								
		Temperatur			Kelembapan Udara			Kecepatan Angin		
		X ₁₁	X ₁₂	X ₁₃	X ₂₁	X ₂₂	X ₂₃	X ₃₁	X ₃₂	X ₃₃
<i>Data In Sample</i>										
1	Y _{b,1}	X _{11,1}	X _{12,1}	X _{13,1}	X _{21,1}	X _{22,1}	X _{23,1}	X _{31,1}	X _{32,1}	X _{33,1}
2	Y _{b,2}	X _{11,2}	X _{12,2}	X _{13,2}	X _{21,2}	X _{22,2}	X _{23,2}	X _{31,2}	X _{32,2}	X _{33,2}
3	Y _{b,3}	X _{11,3}	X _{12,3}	X _{13,3}	X _{21,3}	X _{22,3}	X _{23,3}	X _{31,3}	X _{32,3}	X _{33,3}
4	Y _{b,4}	X _{11,4}	X _{12,4}	X _{13,4}	X _{21,4}	X _{22,4}	X _{23,4}	X _{31,4}	X _{32,4}	X _{33,4}
5	Y _{b,5}	X _{11,5}	X _{12,5}	X _{13,5}	X _{21,5}	X _{22,5}	X _{23,5}	X _{31,5}	X _{32,5}	X _{33,5}
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
3617	Y _{b,3617}	X _{11,3617}	X _{12,3617}	X _{13,3617}	X _{21,3617}	X _{22,3617}	X _{23,3617}	X _{31,3617}	X _{32,3617}	X _{33,3617}
3618	Y _{b,3618}	X _{11,3618}	X _{12,3618}	X _{13,3618}	X _{21,3618}	X _{22,3618}	X _{23,3618}	X _{31,3618}	X _{32,3618}	X _{33,3618}
3619	Y _{b,3619}	X _{11,3619}	X _{12,3619}	X _{13,3619}	X _{21,3619}	X _{22,3619}	X _{23,3619}	X _{31,3619}	X _{32,3619}	X _{33,3619}
3620	Y _{b,3620}	X _{11,3620}	X _{12,3620}	X _{13,3620}	X _{21,3620}	X _{22,3620}	X _{23,3620}	X _{31,3620}	X _{32,3620}	X _{33,3620}
3621	Y _{b,3621}	X _{11,3621}	X _{12,3621}	X _{13,3621}	X _{21,3621}	X _{22,3621}	X _{23,3621}	X _{31,3621}	X _{32,3621}	X _{33,3621}
<i>Data Out Sample</i>										
3622	Y _{b,3622}	X _{11,3622}	X _{12,3622}	X _{13,3622}	X _{21,3622}	X _{22,3622}	X _{23,3622}	X _{31,3622}	X _{32,3622}	X _{33,3622}
3623	Y _{b,3623}	X _{11,3623}	X _{12,3623}	X _{13,3623}	X _{21,3623}	X _{22,3623}	X _{23,3623}	X _{31,3623}	X _{32,3623}	X _{33,3623}
3624	Y _{b,3624}	X _{11,3624}	X _{12,3624}	X _{13,3624}	X _{21,3624}	X _{22,3624}	X _{23,3624}	X _{31,3624}	X _{32,3624}	X _{33,3624}
3625	Y _{b,3625}	X _{11,3625}	X _{12,3625}	X _{13,3625}	X _{21,3625}	X _{22,3625}	X _{23,3625}	X _{31,3625}	X _{32,3625}	X _{33,3625}
3626	Y _{b,3626}	X _{11,3626}	X _{12,3626}	X _{13,3626}	X _{21,3626}	X _{22,3626}	X _{23,3626}	X _{31,3626}	X _{32,3626}	X _{33,3626}
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
3648	Y _{b,3648}	X _{11,3648}	X _{12,3648}	X _{13,3648}	X _{21,3648}	X _{22,3648}	X _{23,3648}	X _{31,3648}	X _{32,3648}	X _{33,3648}
3649	Y _{b,3649}	X _{11,3649}	X _{12,3649}	X _{13,3649}	X _{21,3649}	X _{22,3649}	X _{23,3649}	X _{31,3649}	X _{32,3649}	X _{33,3649}
3650	Y _{b,3650}	X _{11,3650}	X _{12,3650}	X _{13,3650}	X _{21,3650}	X _{22,3650}	X _{23,3650}	X _{31,3650}	X _{32,3650}	X _{33,3650}
3651	Y _{b,3651}	X _{11,3651}	X _{12,3651}	X _{13,3651}	X _{21,3651}	X _{22,3651}	X _{23,3651}	X _{31,3651}	X _{32,3651}	X _{33,3651}
3652	Y _{b,3652}	X _{11,3652}	X _{12,3652}	X _{13,3652}	X _{21,3652}	X _{22,3652}	X _{23,3652}	X _{31,3652}	X _{32,3652}	X _{33,3652}

Observasi 1 sampai dengan observasi 3621 akan digunakan sebagai data *in sample* guna membangun model dalam memprediksi curah hujan. Kemudian model prediksi yang diperoleh dari hasil pengolahan data *in sample* akan divalidasi menggunakan data *out sample* dari observasi 3622 sampai dengan observasi 3652 (periode Desember 2018).

3.6 Langkah Penelitian

Langkah analisis yang akan dilakukan dalam penelitian ini antara lain sebagai berikut :

1. Melakukan pengumpulan data sekunder.
2. Melakukan *pre-processing* data diantaranya sebagai berikut
 - a. Menggabungkan seluruh excel menjadi satu file dengan bantuan *software* R (R Core Team, 2018).
 - b. Imputasi *time series* untuk data yang *missing*.
3. Mendeskripsikan gambaran umum dari variabel penelitian menggunakan statistika deskriptif dengan bantuan *software R*.
 - a. Membuat *timeseries plot* untuk mengidentifikasi pola variabel *predictant* (curah hujan) dan variabel *predictor* (variabel-variabel klimatologi)
 - b. Membuat *boxplot* untuk mengetahui apakah terdapat data *outlier* serta mengetahui rata-rata terendah dan tertinggi untuk masing-masing variabel dimasing-masing lokasi penelitian.
 - c. Membuat *scatterplot* untuk mengetahui hubungan antar variabel.
 - d. Membuat plot ACF dan PACF variabel curah hujan.
 - e. Membuat *scatterplot* untuk mengetahui hubungan antara curah hujan dengan curah hujan sebelumnya berdasarkan lag yang signifikan pada ACF dan PACF.
4. Melakukan prediksi variabel-variabel *predictor* secara *univariate* menggunakan metode peramalan *Holt-Winters*.
5. Melakukan analisis PCA untuk setiap variabel *predictor* di 3 stasiun BMKG. *Principal Component* (PC) yang terpilih adalah PC dengan nilai kumulatif proporsi melebihi 80%.
6. Melakukan analisis prediksi curah hujan berbasis data BMKG menggunakan ν -SVR. Adapun tahapan prediksi curah hujan yaitu memodelkan variabel klimatologi (variabel *predictor*)

di tiga stasiun BMKG terhadap 6 balai curah hujan (variabel *predictant*), sehingga akan ada 6 analisis secara terpisah dan didapatkan 6 nilai prediksi hujan di Surabaya dari 6 tempat balai hujan yang ada di Surabaya. Setiap pasangan antara 1 balai hujan dengan 3 stasiun BMKG dilakukan berbagai kombinasi model dengan skenario sebagai berikut.

- Curah hujan → Temperatur
- Curah hujan → Temperatur + dummy
- Curah hujan → Kelembapan udara
- Curah hujan → Kelembapan udara+dummy
- Curah hujan → Kecepatan angin
- Curah hujan → Kecepatan angin +dummy
- Curah hujan → Temperatur+Kelembapan udara +Kecepatan Angin
- Curah hujan → Temperatur+Kelembapan udara +Kecepatan Angin+dummy

Dari model diatas kemudian dipilih model dengan RMSE terkecil yang selanjutnya akan dianalisis kembali dengan menambahkan variabel prediktor lag curah hujan yang signifikan dari PACF dan lag yang signifikan dari ARIMA untuk membantu menurunkan nilai RMSE. Adapun tahapan metode ν -SVR yang dilakukan pada setiap model yang akan dianalisis, diantaranya yaitu sebagai berikut.

- a. Membagi data menjadi *in-sample* dan *out-sample*.
- b. Menentukan kernel yang akan digunakan yaitu kernel RBF atau *Gaussian*.
- c. Menentukan nilai parameter *cost*, *gamma*, dan *nu* yang paling optimum melalui optimasi dengan mengkombinasikan berbagai kemungkinan dari *cost*, *gamma*, dan *nu*. Nilai untuk masing-masing parameter yang akan dicobakkan adalah sebagai berikut.

$Cost = 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100$

$Gamma = 0.5, 1, 1.5, 2, 2.5$

$Nu = 0.2, 0.3, 0.4, 0.5$

Sehingga dari nilai-nilai parameter tersebut akan terbentuk 400 kombinasi parameter, kemudian dari 400

kombinasi parameter tersebut dipilih satu kombinasi parameter antara *cost*, *gamma*, dan *nu* yang optimum dengan nilai RMSE terkecil.

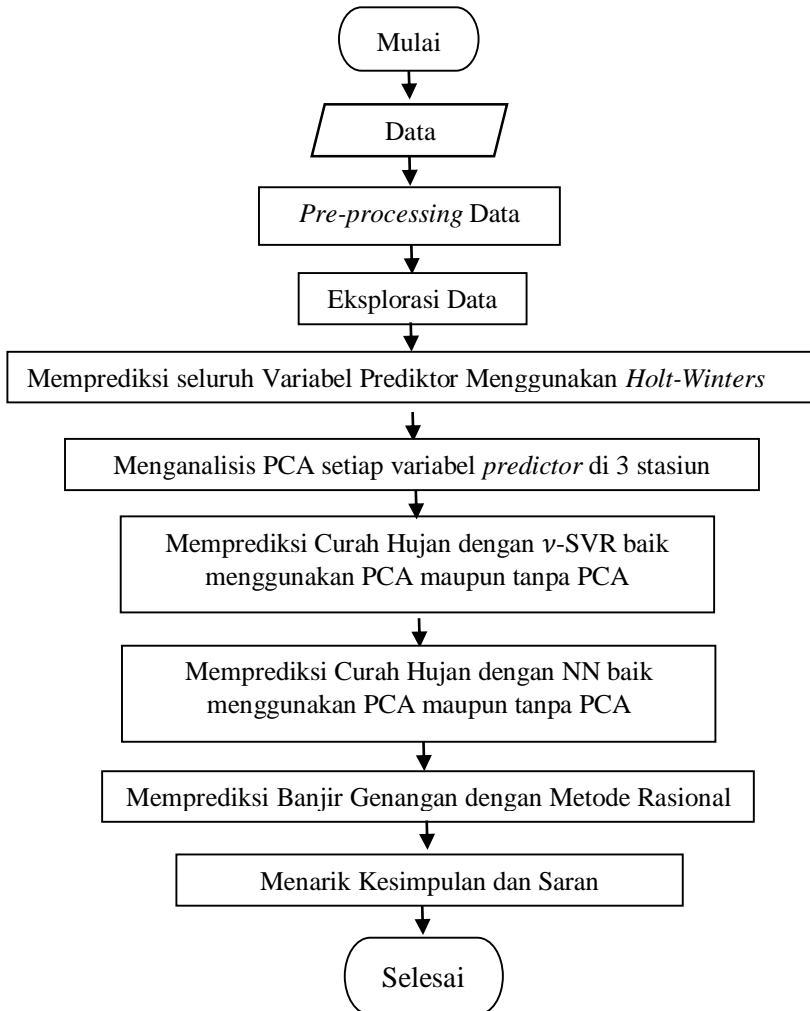
- d. Melakukan pemodelan ν -SVR menggunakan parameter yang optimum dari hasil optimasi parameter sebelumnya.
 - e. Data hasil prediksi yang bernilai negatif ditransformasi menjadi nilai 0.
 - f. Membuat plot antara data *actual* dengan data hasil prediksi.
 - g. Menghitung nilai RMSE *in-sample*.
- Dari setiap model yang terbentuk kemudian dibandingkan nilai RMSE *in-sample*, model terbaik dipilih dari nilai RMSE yang paling kecil. Kemudian dilakukan pemodelan curah hujan untuk memprediksi data *out-sampel* dan dihitung nilai RMSE *out-sample*-nya. Jadi setiap 1 balai hujan akan memiliki 3 model terbaik dari 3 stasiun BMKG.
7. Melakukan analisis ν -SVR menggunakan PC yang telah terbentuk pada langkah 5. Adapun tahapan dan skenarionya sama seperti analisis ν -SVR pada langkah 6.
 8. Melakukan prediksi curah hujan berbasis data BMKG menggunakan *Neural Network*. Adapun skenario yang akan dilakukan sama seperti analisis ν -SVR sebelumnya, sedangkan untuk tahapan NN untuk setiap modelnya dijelaskan sebagai berikut.
 - a. Melakukan normalisasi data dengan *min-max normalization*.
 - b. Membagi data menjadi *in-sample* dan *out-sample*.
 - c. Melakukan pemodelan NN dengan neuron yang telah telah ditentukan. Adapun fungsi aktivasi yang akan digunakan adalah *tanh* dengan replikasi sebanyak 5 kali.
 - d. Membuat plot NN terbaik dari 5 kali replikasi.
 - e. Men-denormalisasi data hasil prediksi.
 - f. Data hasil prediksi yang bernilai negatif ditransformasi menjadi nilai 0.
 - g. Membuat plot antara data *actual* dengan data hasil prediksi.

- h. Menghitung nilai RMSE *in-sample*.
- i. Melakukan langkah 8b kembali namun dengan jumlah neuron yang berbeda.
- j. Membandingkan nilai RMSE *in-sample* dari berbagai kombinasi neuron yang telah dilakukan untuk kemudian dipilih neuron dengan nilai RMSE *in-sample* terkecil.

Dari setiap model yang terbentuk dipilih model terbaik dengan nilai RMSE *in-sample* terkecil. Kemudian dari model terbaik tersebut dilakukan prediksi data *out-sample* dan dihitung nilai RMSE nya. Sama halnya seperti analisis ν -SVR, setiap 1 balai hujan akan memiliki 3 model terbaik dari 3 stasiun BMKG.

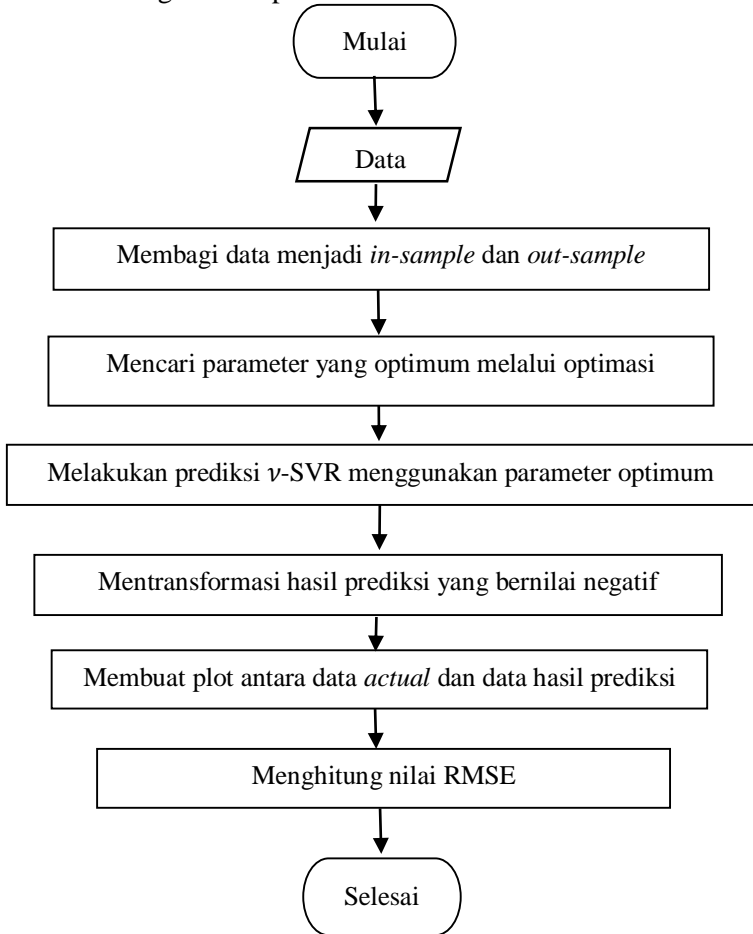
- 9. Melakukan analisis NN menggunakan PC yang telah terbentuk pada langkah 5. Adapun tahapan dan skenarionya sama seperti analisis NN pada langkah 8.
- 10. Membandingkan hasil prediksi menggunakan ν -SVR dan NN.
- 11. Melakukan analisis prediksi banjir genangan berdasarkan hasil prediksi curah hujan pada masing-masing balai menggunakan perhitungan matematis dengan metode rasional.
- 12. Menarik kesimpulan dan memberikan saran.

Berdasarkan langkah-langkah penelitian yang akan dilakukan dan penjelasan sebelumnya, adapun diagram alir untuk analisis ini dapat dilihat pada Gambar 3.5 sebagai berikut.



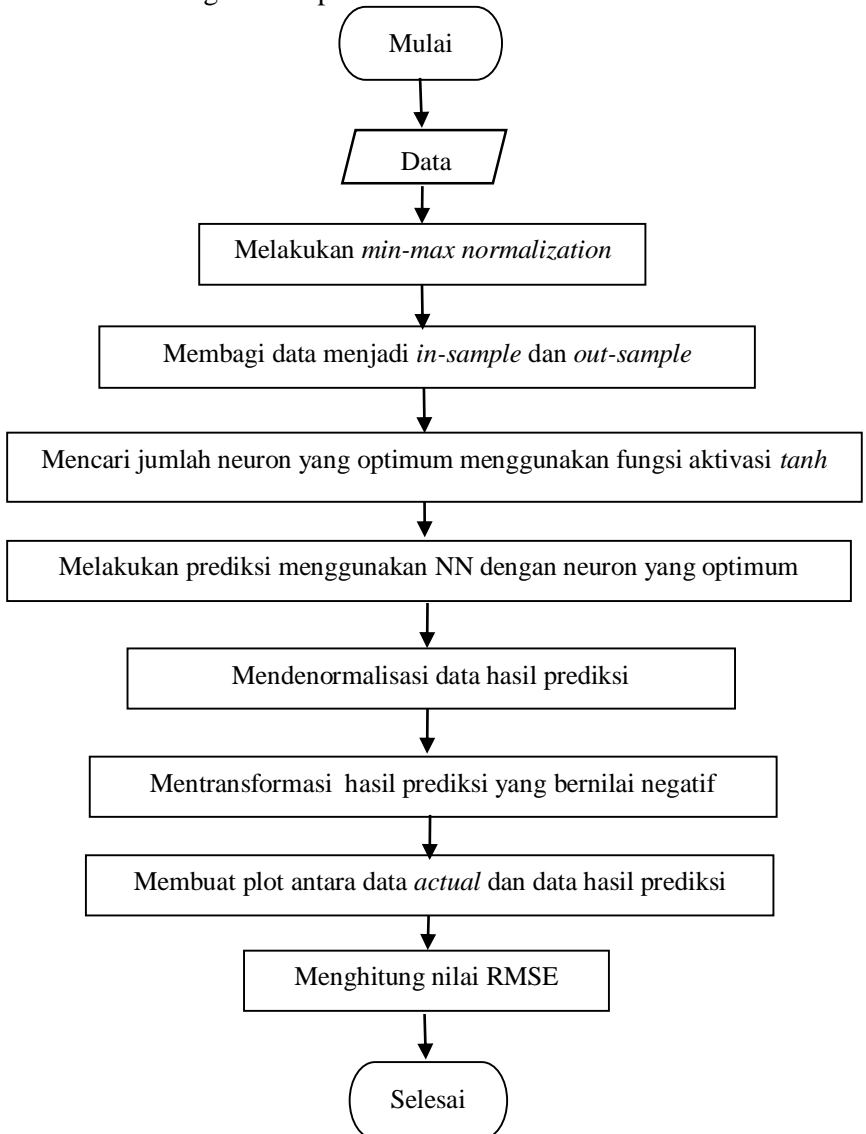
Gambar 3.5 Diagram Alir Penelitian Keseluruhan

Adapun langkah-langkah dalam pemodelan ν -SVR dapat dilihat dalam diagram alir pada Gambar 3.6.



Gambar 3.6 Diagram Alir *Nu-Support Vector Regression*

Sedangkan langkah-langkah dalam pemodelan NN dapat dilihat dalam diagram alir pada Gambar 3.7.



Gambar 3.7 Diagram Alir *Neural Network*

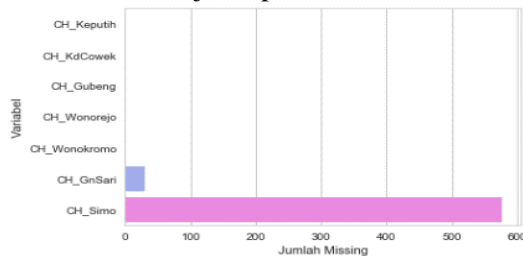
(Halaman ini sengaja dikosongkan)

BAB IV ANALISIS DAN PEMBAHASAN

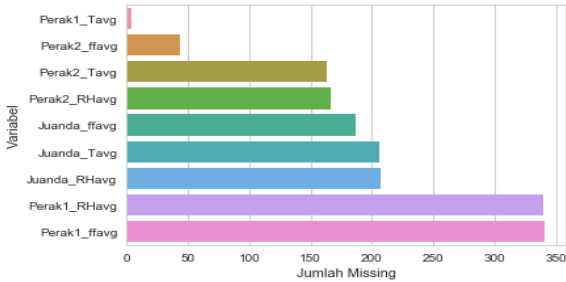
Pada penelitian ini, analisis dan pembahasan berisi mengenai serangkaian proses yang dilakukan untuk menjawab tujuan penelitian diantaranya yaitu *pre-processing* data, eksplorasi data, peramalan *Holt-Winters*, analisis PCA, analisis *v*-SVR, analisis NN, serta dilanjutkan analisis perhitungan banjir dari hasil prediksi curah hujan terbaik yang telah didapatkan.

4.1 Pre-Processing Data Penelitian

Suatu data yang akan digunakan sebagai bahan penelitian harus terlebih dahulu dilakukan *pre-processing* data. Salah satu tujuan dilakukan *pre-processing* data adalah untuk melihat apakah data penelitian yang akan digunakan mengandung *missing value* (data kosong, hilang, tidak dilakukan pengukuran) atau tidak. Hal ini dilakukan agar *missing value* dapat diatasi terlebih dahulu untuk menghindari hasil penelitian yang bias dan terganggu oleh adanya *missing value*. *Pre-processing* pada penelitian ini dilakukan melalui beberapa langkah diantaranya yaitu pertama menggabungkan file-file yang berisi data-data klimatologi BMKG keseluruhan dengan bantuan *software* R. Kemudian kedua dilanjutkan menggabungkan file-file yang berisi data curah hujan di 7 balai PSAWS Buntung Peketingan. Ketiga menggabungkan data BMKG dan data balai PSAWS menjadi satu file excel. Langkah terakhir yaitu melakukan deteksi *missing value* pada data penelitian yang telah tergabung dalam satu file excel. Adapun jumlah *missing value* pada data penelitian untuk masing-masing variabel secara visual di sajikan pada Gambar 4.1 berikut.



(a)



(b)

Gambar 4.1 *Missing Value* (a) Data Balai PSAWS dan (b) Data BMKG

Gambar 4.1 (a) menampilkan jumlah *missing value* data curah hujan di 7 Balai PSAWS dari yang terendah hingga tertinggi, sedangkan Gambar 4.1 (b) menunjukkan jumlah *missing value* pada variabel-variabel klimatologi di 3 stasiun pengamatan BMKG dari yang paling sedikit hingga paling terbanyak *missing value*. Untuk mendukung analisis *missing value* secara visual diatas, berikut disajikan Tabel yang menunjukkan jumlah dan persentase *missing value* untuk masing-masing variabel penelitian.

Tabel 4.1 Jumlah Data *Missing Value* pada Data Penelitian

Variabel	Keterangan	<i>Missing Value</i>		Jumlah Data
		Jumlah	Persentase	
Balai PSAWS				
Y_1	CH Keputih	0	0 %	3652
Y_2	CH Kedung Cowek	0	0 %	
Y_3	CH Gubeng	0	0 %	
Y_4	CH Wonorejo	0	0 %	
Y_5	CH Wonokromo	0	0 %	
Y_6	CH Gunung Sari	30	0,82 %	
Y_7	CH Simo	577	15,8 %	
Stasiun BMKG				
X_{11}	Tavg Juanda	206	5,64 %	3652
X_{12}	Tavg Perak 2	163	4,46 %	
X_{13}	Tavg Perak 1	4	0,11 %	
X_{21}	RHavg Juanda	207	5,67 %	
X_{22}	RHavg Perak 2	167	4,57 %	
X_{23}	RHavg Perak 1	340	9,31 %	
X_{31}	ffavg Juanda	187	5,12 %	
X_{32}	ffavg Perak 2	43	1,18 %	
X_{33}	ffavg Perak 1	341	9,33 %	

Berdasarkan Tabel 4.1 dapat diketahui bahwa variabel Y_7 yaitu curah hujan di daerah pengamatan Simo Balai PSAWS memiliki jumlah *missing value* terbesar dibandingkan keenam balai lainnya yaitu sebesar 15,8% atau 577 observasi. Hal tersebut dikarenakan pada tanggal 1 Januari 2013 sampai dengan 31 Juli 2014 atau setara dengan 19 bulan tidak dilakukan pengukuran pada balai tersebut. Pengukuran tidak bisa dilakukan karena alat pengukur curah hujan yang ada sedang rusak, tidak dapat digunakan sebagaimana mestinya, serta masih dalam proses perbaikan dan baru dapat digunakan kembali tanggal 1 Agustus 2014. Sehingga data curah hujan yang *missing* di balai Simo tersebut tidak mungkin untuk dilakukan proses imputasi, sebab jumlah *missing* yang terlalu banyak dan menghilangkan 1,5 tahun data Simo dari 10 tahun pengamatan yang ada. Jika dipaksakan dilakukan imputasi maka hasil yang diperoleh kurang akurat dan kurang representatif. Oleh karena itu, akibat dari ketidaktersediaan data tersebut maka variabel curah hujan di balai Simo (Y_7) tidak akan diikuti atau dihapus dalam analisis dan hanya akan menggunakan 6 daerah pengamatan curah hujan balai PSAWS.

Adapun balai curah hujan yang memiliki *missing value* terbesar kedua setelah Simo yaitu balai Gunung Sari, dimana terdapat 30 (1 bulan) data yang *missing value* atau sebesar 0,82% dari keseluruhan data yang ada. Sehingga khusus untuk variabel Y_6 yaitu curah hujan di balai Gunung Sari perlu dilakukan penanganan dengan cara imputasi *time series*. Imputasi *time series* pada penelitian ini merupakan imputasi dengan cara mencari rata-rata nilai variabel pada tanggal dan bulan yang sama tapi dengan tahun yang berbeda kemudian memasukkannya pada data yang *missing* tersebut. Sedangkan untuk 5 balai lainnya yaitu Keputih, Kedung Cowek, Gubeng, Wonorejo, dan Wonokromo tidak terdapat data yang *missing*, sehingga tidak perlu dilakukan imputasi dan dapat dilanjutkan atau digunakan dalam analisis.

Variabel-variabel klimatologi BMKG pada 3 stasiun yang ada semuanya mengandung *missing value* dengan persentase yang berbeda-beda. Jika dilihat dari jenis variabelnya, variabel kelembapan udara (X_2) merupakan variabel yang paling banyak terdapat *missing value* dibandingkan variabel temperatur dan

kecepatan angin. Sedangkan jika dilihat dari jenis stasiun BMKG, stasiun Perak 1 memiliki jumlah data yang *missing* paling banyak dibandingkan Perak 2 dan Juanda. Adapun jumlah *missing value* terbesar terdapat pada variabel kecepatan angin di stasiun Perak 1 yaitu 341 observasi atau sebesar 9,33 % dari seluruh observasi. Oleh karena itu, data-data yang *missing* pada variabel-variabel klimatologi juga perlu dilakukan imputasi *time series* dengan merata-rata data pada tanggal dan bulan yang sama dari tahun yang berbeda-beda. Hasil analisis statistika deskriptif dari data balai PSAWS maupun data BMKG yang telah diimputasi dapat dilihat pada Tabel berikut.

Tabel 4.2 Statistika Deskriptif Data Penelitian Hasil Imputasi

Variabel	Mean	Stdev	Minimum	Maximum
Balai PSAWS				
Y_1	5,513	13,195	0,000	164,000
Y_2	5,132	11,401	0,000	123,000
Y_3	6,487	14,009	0,000	116,000
Y_4	5,860	13,782	0,000	122,000
Y_5	6,419	14,104	0,000	114,000
Y_6	6,361	14,052	0,000	120,000
Stasiun BMKG				
X_{11}	27,968	1,177	24,400	33,200
X_{12}	28,999	1,111	25,400	32,700
X_{13}	28,585	1,135	25,000	32,800
X_{21}	76,344	7,521	44,000	95,000
X_{22}	74,452	6,801	52,000	95,000
X_{23}	75,165	7,089	51,000	94,000
X_{31}	3,3343	1,0707	0,0000	24,0000
X_{32}	2,7219	1,3033	0,0000	9,0000
X_{33}	2,5990	1,2092	0,0000	7,0000

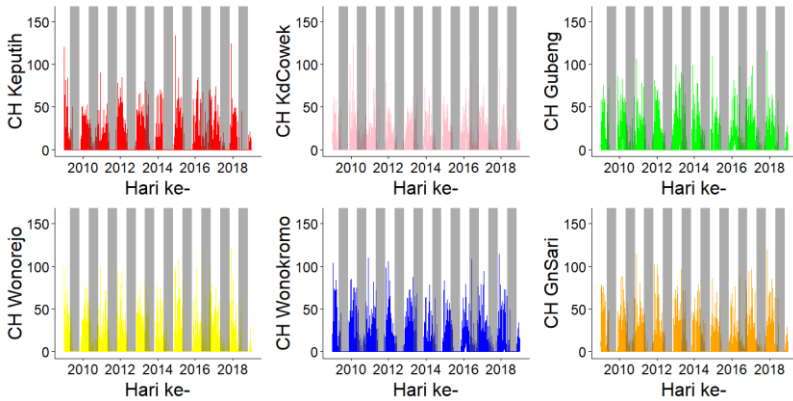
Pada Tabel 4.2 terlihat bahwa nilai rata-rata curah hujan harian tertinggi terjadi di Gubeng yaitu sebesar 6,487 mm dengan nilai curah hujan maksimum sebesar 116 mm. Sedangkan untuk nilai rata-rata curah hujan terendah terjadi di Kedung Cowek yaitu sebesar 5,132 mm dan curah hujan maksimum sebesar 123 mm. Adapun jika nilai curah hujan dilihat dari keseluruhan Balai sebagai representasi Kota Surabaya diketahui bahwa curah hujan maksimum yaitu sebesar 164 mm yang terjadi di sekitar daerah

Keputih tepatnya pada tanggal 30 Mei 2016. Ukuran keragaman atau persebaran data dapat dilihat dari nilai varians atau standar deviasi, dimana dapat diketahui bahwa balai Wonokromo memiliki nilai standar deviasi paling besar dibandingkan ke-lima balai lainnya. Hal tersebut menunjukkan bahwa curah hujan yang terjadi di daerah Wonokromo dan sekitarnya memiliki perbedaan nilai yang cukup beragam setiap harinya selama 10 tahun terakhir.

Sama halnya seperti curah hujan di balai PSWAS, ukuran persebaran data pada variabel variabel klimatologi di BMKG juga dapat dilihat dari nilai varians atau nilai standar deviasinya. Dimana berdasarkan Tabel 4.2 diatas diketahui bahwa variabel kelembapan udara di 3 stasiun memiliki persebaran data yang paling tinggi dibandingkan dengan variabel temperatur dan kecepatan angin. Sedangkan jika dilihat dari nilai rata-ratanya, stasiun Juanda memiliki nilai rata-rata tertinggi untuk kelembapan udara dan kecepatan angin, sedangkan rata-rata tertinggi untuk variabel temperatur terjadi di stasiun Perak 2. Informasi penting lainnya yang dapat diambil dari Tabel 4.2 diatas yaitu nilai-nilai variabel klimatologi yang berada di Perak 2 memiliki selisih yang sangat kecil dengan nilai-nilai variabel klimatologi yang berada di Perak 1. Hal ini dikarenakan letak stasiun pengamatan Perak 2 cukup dekat dengan Perak 1, sehingga berdasarkan kedekatan letaknya tersebut dapat diindikasikan variabel-variabel klimatologi yang ada di Perak 2 memiliki hubungan atau ada keterkaitan dengan variabel-variabel klimatologi yang ada di Perak 1. Oleh karena itu perlu dilakukan analisis lebih lanjut untuk mengatasi dependensi antar wilayah tersebut.

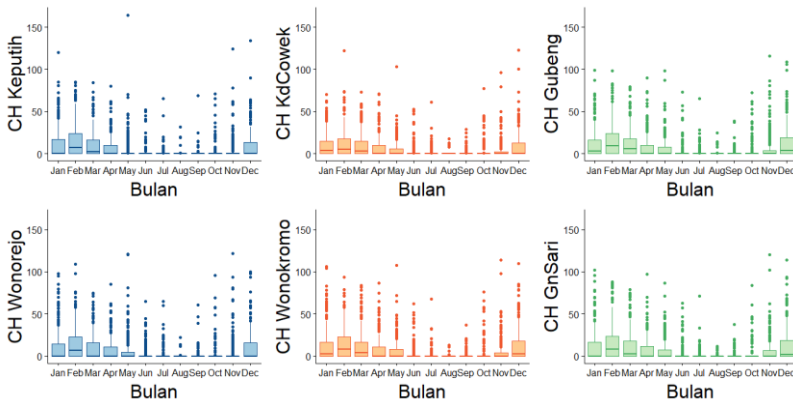
4.2 Gambaran Umum Curah Hujan dan Variabel Klimatologi di Kota Surabaya

Dalam menganalisis suatu permasalahan, hal yang tidak kalah penting harus dilakukan yaitu melakukan eksplorasi data yang bertujuan untuk melihat karakteristik dari data tersebut atau mendapatkan gambaran umum sebagai informasi awal dari sebuah data sebelum menentukan atau menerapkan metode analisis yang tepat. Oleh karena itu, berikut di sajikan visualisasi melalui gambar untuk mengetahui gambaran umum curah hujan di Subaya selama 10 tahun terakhir dari 6 balai pengamatan PSAWS.



Gambar 4.2 Time Series Plot Curah Hujan di 6 Balai PSAWS Surabaya

Time series plot diatas menunjukkan bahwa curah hujan di kota Surabaya sangat bervariasi dan berfluktuatif naik turun setiap harinya. Terdapat indikasi adanya pola atau trend curah hujan di setiap tahunnya di 6 balai PSAWS yang ada di Surabaya. Dimana dalam 10 tahun terakhir dari tahun 2008-2010, rata-rata curah hujan terbesar terjadi pada tahun 2017 dan menyebabkan air menggenang di berbagai daerah di Surabaya. Untuk lebih jelasnya, berikut juga disajikan *boxplot* curah hujan berdasarkan bulan selama 10 tahun terakhir.



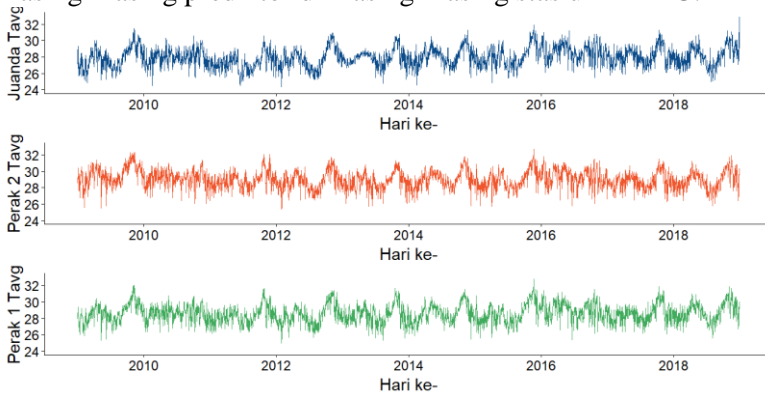
Gambar 4.3 Boxplot Curah Hujan di 6 Balai PSAWS Surabaya

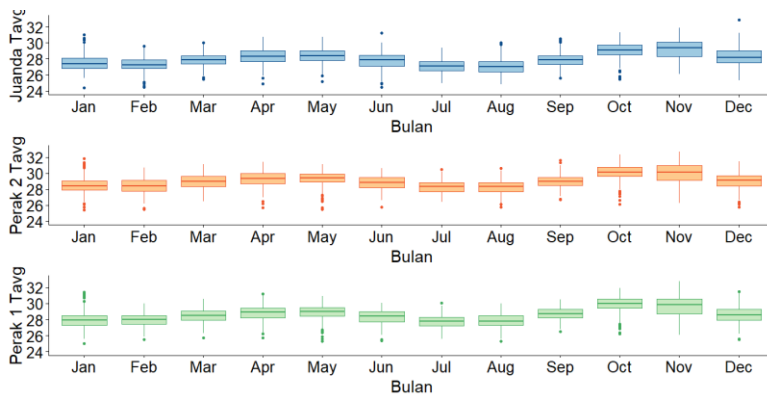
Berdasarkan *boxplot* diatas diketahui terdapat pola atau *trend seasonal* untuk curah hujan yang terjadi di Kota Surabaya di setiap bulannya. Dapat dilihat dari Gambar 4.3 bahwa sekitar bulan

Nopember sampai dengan April memiliki rata-rata curah hujan harian yang cukup tinggi di semua balai curah hujan. Dimana jika dikaitkan dengan pola hujan monsun yang ada di Indonesia, hal tersebut sesuai dengan teori klimatologi yang ada, sebab pada bulan-bulan sekitar Nopember sampai dengan bulan April ini merupakan bulan-bulan hujan atau bulan-bulan terjadinya angin monsun barat yang membawa uap air dan menyebabkan hujan turun dan disebut sebagai bulan-bulan basah. Disisi lain, berkebalikan dengan angin monsun timur yang terjadi sekitar bulan Mei sampai dengan Oktober dan menyebabkan terjadinya musim kemarau di Indonesia. Sehingga curah hujan pada musim kemarau sekitar bulan Mei sampai dengan Oktober tersebut memiliki nilai yang kecil dan berkisar di nilai nol. Informasi lainnya yaitu rata-rata hujan paling banyak terjadi pada bulan Februari sedangkan rata-rata hujan paling sedikit terjadi di bulan April. Hal tersebut dikarenakan bulan April merupakan bulan terakhir pada musim hujan sebelum memasuki musim kemarau pada bulan Mei. Meskipun demikian, masih terlihat terjadi hujan pada bulan Mei di seluruh balai curah hujan. Selain itu pula dapat dilihat bahwa pada bulan Desember curah hujan yang terjadi cenderung tinggi dengan variabilitas yang rendah, hal ini berarti pada bulan Desember terjadi hujan hampir setiap hari. Berbeda halnya dengan bulan Februari yang memiliki curah hujan tinggi namun variabilitasnya tinggi, hal tersebut menandakan bahwa pada bulan Februari terkadang hujan dan terkadang tidak hujan. Sedangkan untuk bulan Juni, Juli, Agustus, dan September memiliki curah hujan yang kecil dengan variabilitas yang sangat rendah pula. Hal tersebut menandakan bahwa pada bulan Juni, Juli, Agustus, dan September hampir selalu tidak terjadi hujan. Informasi lainnya yang dapat diperoleh dari Gambar 4.3 di atas yaitu curah hujan yang terjadi di Surabaya setiap bulannya selama 10 tahun terakhir memiliki distribusi yang tidak normal atau dengan kata lain berdistribusi *skewness* ke-kanan dan *skewness* ke-kiri.

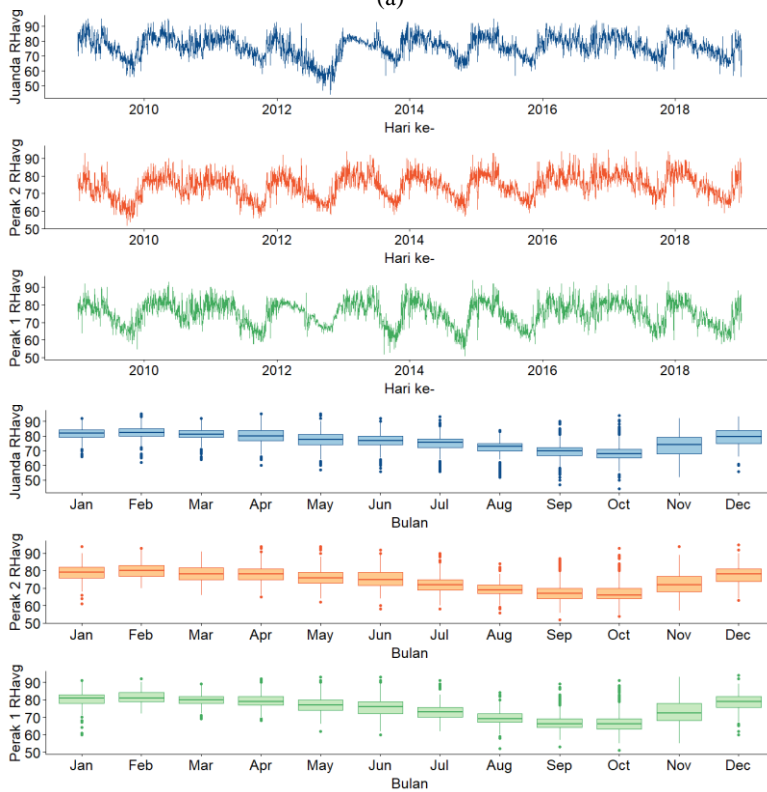
Jadi dari Gambar 4.2 dan Gambar 4.3 dapat disimpulkan bahwasannya antara bulan-bulan basah (musim hujan) dan bulan-bulan kering (musim kemarau) di Kota Surabaya sangat dapat dibedakan dengan jelas baik jika dilihat melalui *time series plot*

maupun *boxplot*. Perbedaan yang sangat jelas tersebut juga didukung oleh penjelasan pada Gambar 3.2, dimana berdasarkan pembagian pola hujan dari BMKG, wilayah Surabaya masuk kedalam pola hujan moonsun yang dapat membedakan secara jelas antara angin monsun barat yang membawa musim hujan dan angin monsun timur yang membawa musim kemarau. Sehingga, jika ingin dilakukan analisis lebih dalam mengenai prediksi curah hujan maka sangat bermanfaat jika ditambahkan variabel *dummy* terlebih dahulu untuk membedakan antara bulan-bulan kering dan bulan-bulan basah. Hal tersebut bertujuan untuk mengakomodasi curah hujan yang bernilai 0 ketika bulan-bulan kering yang mengakibatkan sulitnya analisis dan mengganggu proses prediksi yang sedang dilakukan serta bertujuan untuk mengurangi *error* yang terbentuk dari suatu model. Adapun penambahan variabel *dummy* untuk membedakan bulan-bulan kering dan bulan-bulan basah dapat dilihat pada subbab 3.3 mengenai variabel penelitian. Berikut juga ditampilkan *time series plot* dan *boxplot* untuk masing-masing prediktor di masing-masing stasiun BMKG.

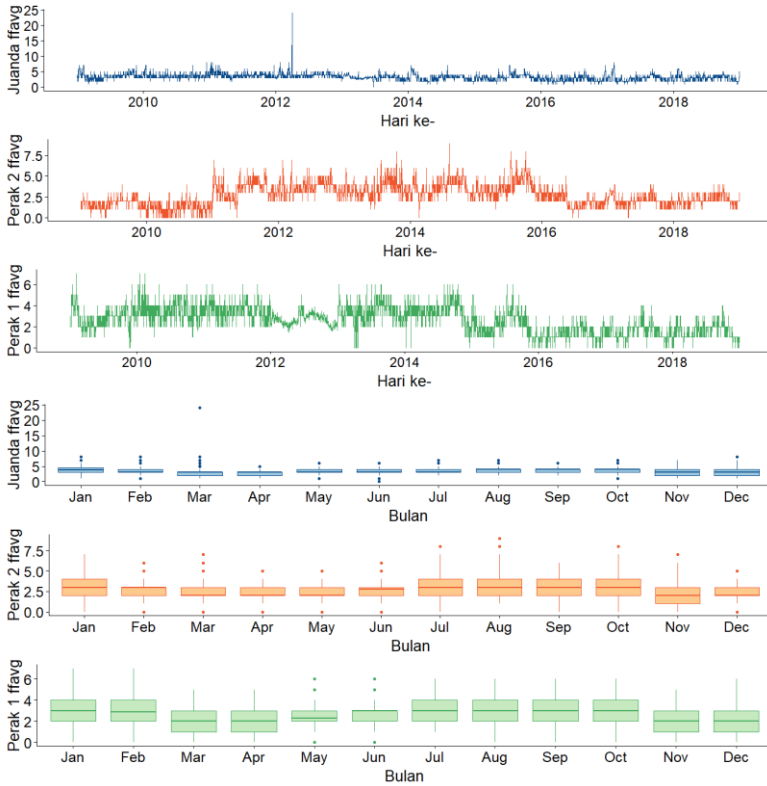




(a)



(b)



(c)

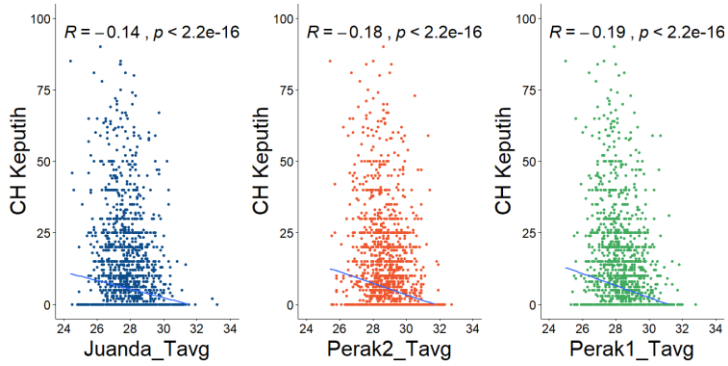
Gambar 4.4 Time Series Plot (Atas) dan Boxplot (Bawah) Variabel (a) Temperatur, (b) Kelembapan Udara, dan (c) Kecepatan Angin

Gambar 4.4 (a) memberikan informasi bahwa temperatur di Surabaya memiliki pola atau *trend seasonal* setiap bulan dan setiap tahunnya, dimana pada bulan-bulan hujan yaitu sekitar bulan Desember, Januari, Februari dan Maret, temperatur di Surabaya cenderung memiliki rata-rata yang rendah. Hal tersebut dikarenakan pada bulan-bulan itu merupakan bulan-bulan hujan dan menyebabkan temperatur menjadi dingin. Sedangkan pada sekitar bulan Juli dan Juni temperatur cenderung tinggi atau panas, sebab pada bulan-bulan tersebut sedang terjadi musim kemarau. Selain itu pula dapat diketahui bahwa Juanda memiliki temperatur

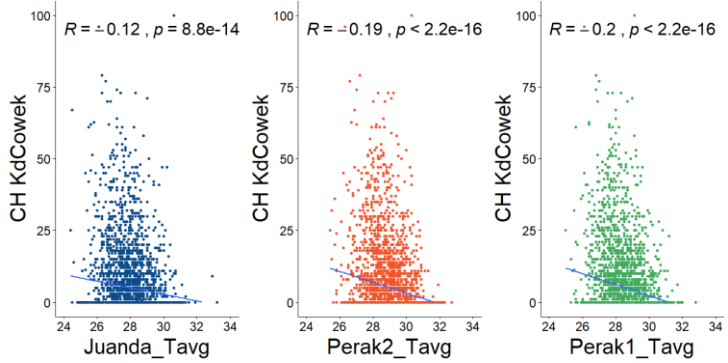
yang cenderung lebih rendah dibandingkan dengan Perak 1 dan Perak 2.

Adapun Gambar 4.4 (b) memberikan informasi bahwa kelembapan udara yang ada di Surabaya memiliki pola atau *trend seasonal* yang sama dengan curah hujan. Sekitar bulan Nopember, Desember, Januari, Februari, Maret dan April, kelembapan udara cenderung tinggi dibandingkan bulan lainnya. Hal tersebut dikarenakan pada bulan Nopember, Desember, Januari, Februari, Maret dan April sedang terjadi curah hujan yang cukup tinggi atau dengan kata lain merupakan bulan-bulan musim penghujan sehingga kelembapan udaranya akan semakin tinggi atau semakin lembap. Selain itu pula dapat diketahui bahwa stasiun Juanda memiliki nilai rata-rata kelembapan udara paling tinggi dibandingkan Perak 1 dan Perak 2.

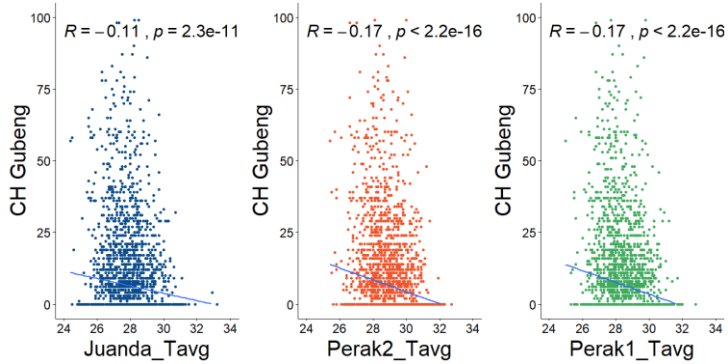
Variabel klimatologi ketiga yaitu kecepatan angin, dimana dapat dilihat pada Gambar 4.4 (c) bahwasannya kecepatan angin di Surabaya memiliki pola atau *trend seasonal* yang berkebalikan dengan curah hujan. Rata-rata kecepatan angin tertinggi terjadi pada bulan-bulan musim kemarau di sekitar Juni, Juli, Agustus, dan September. Sedangkan rata-rata kecepatan angin terendah terjadi pada bulan-bulan musim penghujan yaitu di sekitar bulan Desember, Januari, dan Februari. Adapun rata-rata kecepatan angin tertinggi berada di stasiun Juanda, kemudian disusul Perak 2 dan Perak 1. Hal yang tidak kalah penting sebelum melakukan analisis yaitu perlu diketahui terlebih dahulu hubungan antar variabel *predictor* dan variabel *predictant*. Hubungan antar variabel dapat dilihat melalui dua cara, pertama yaitu dapat dilihat secara visual melalui *scatterplot* dan secara obyektif menggunakan uji korelasi *pearson*. Berikut merupakan hubungan antara masing-masing variabel *predictor* dengan variabel *predictant* dimasing-masing balai dan stasiun.



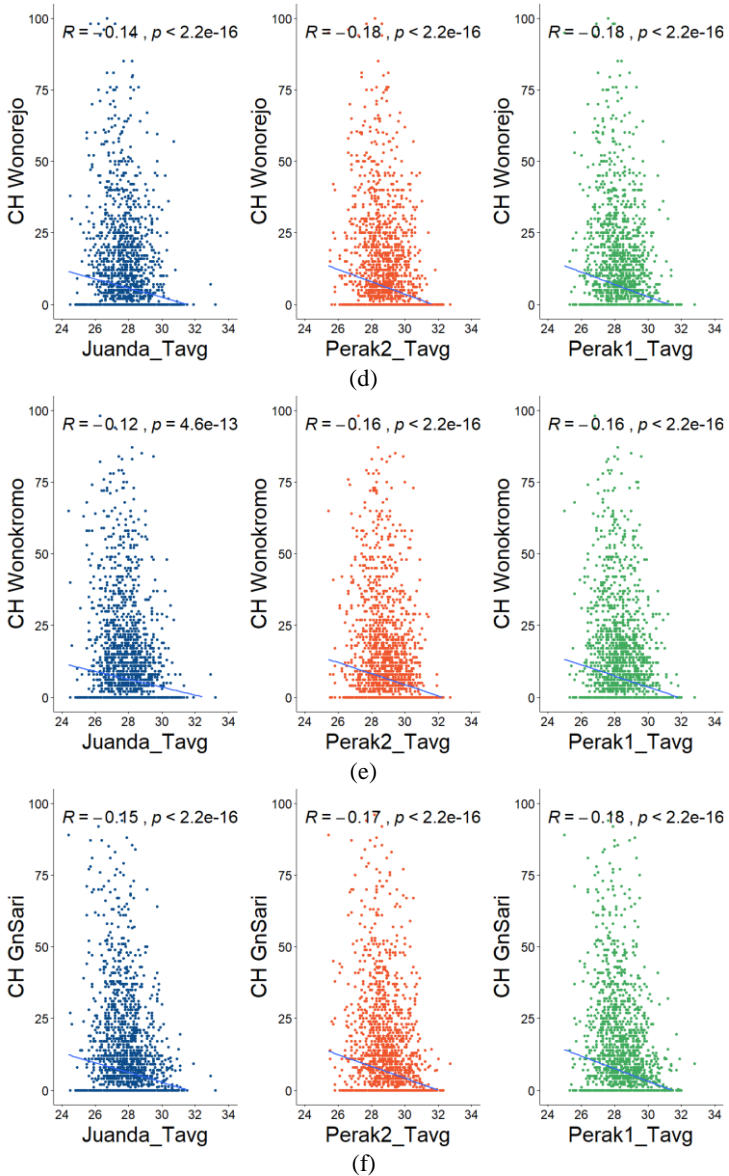
(a)



(b)

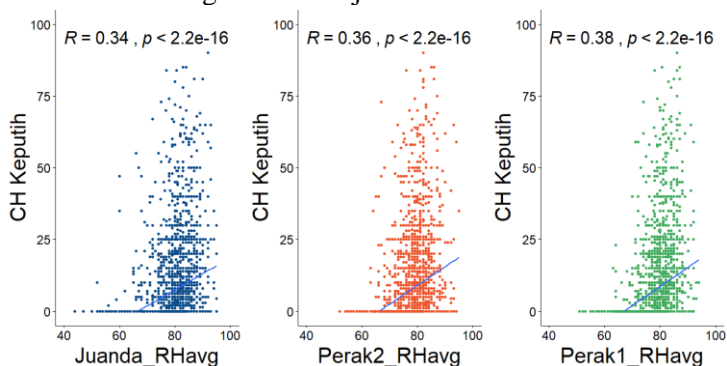


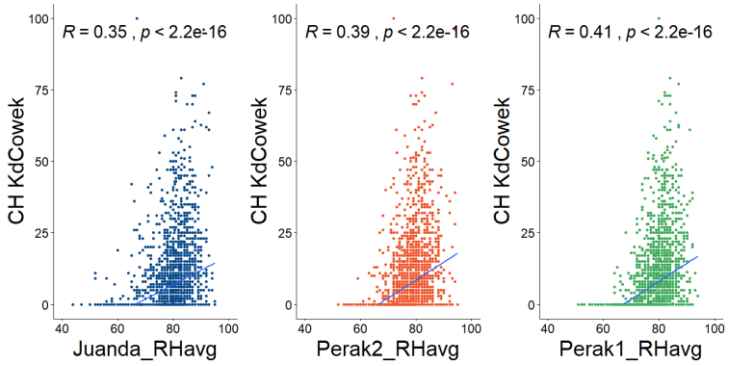
(c)



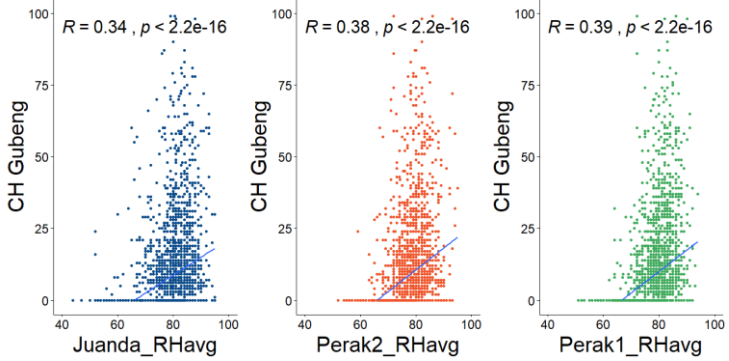
Gambar 4.5 Scatterplot Temperatur di 3 Stasiun BMKG dan Curah Hujan (a) Keputih, (b) Kedung Cowek, (c) Gubeng, (d) Wonorejo, (e) Wonokromo, dan (f) Gunung Sari

Berdasarkan Gambar 4.5 dapat diketahui bahwa hubungan antara temperatur dan curah hujan pada masing masing kombinasi balai dan stasiun memiliki hasil yang sama yaitu memiliki hubungan yang berbanding terbalik, dimana jika curah hujan tinggi di Surabaya maka temperatur akan rendah atau suhu akan semakin dingin dan begitu pula sebaliknya. Adapun korelasi antara temperatur di Juanda dengan curah hujan di semua balai memiliki nilai korelasi yang paling rendah dibandingkan dengan hubungan antara curah hujan di semua balai dan temperatur di Perak 2 serta Perak 1. Nilai signifikansi atau *p-value* dari semua korelasi pada Gambar 4.5 bernilai sangat kecil dan mendekati 0, sehingga dapat diputuskan tolak H_0 yang berarti terdapat hubungan yang signifikan antara temperatur dengan curah hujan di semua balai yang ada di Surabaya. Oleh karena itu, berdasarkan analisis tersebut dapat diindikasikan bahwasannya bisa saja terdapat hubungan yang signifikan antara curah hujan saat ini dengan temperatur satu hari sebelumnya dan untuk membuktikan hal tersebut maka juga dilakukan uji korelasi *pearson* antara curah hujan hari ini dengan temperatur 1 hari sebelumnya (lag1), temperatur 2 hari sebelumnya (lag2), temperatur 3 hari sebelumnya (lag3), temperatur 4 hari sebelumnya (lag4), serta temperatur 5 hari sebelumnya (lag5) yang dapat di lihat pada Lampiran 1. Dari Lampiran 1 tersebut dapat diketahui bahwa terdapat beberapa lag temperatur yang signifikan terhadap curah hujan dimasing-masing balai. Selanjutnya akan dilakukan analisis mengenai hubungan kelembapan udara di 3 stasiun BMKG dengan curah hujan di 6 balai PSAWS.

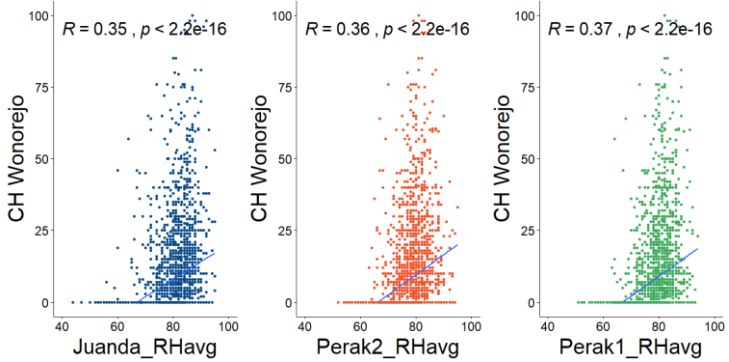




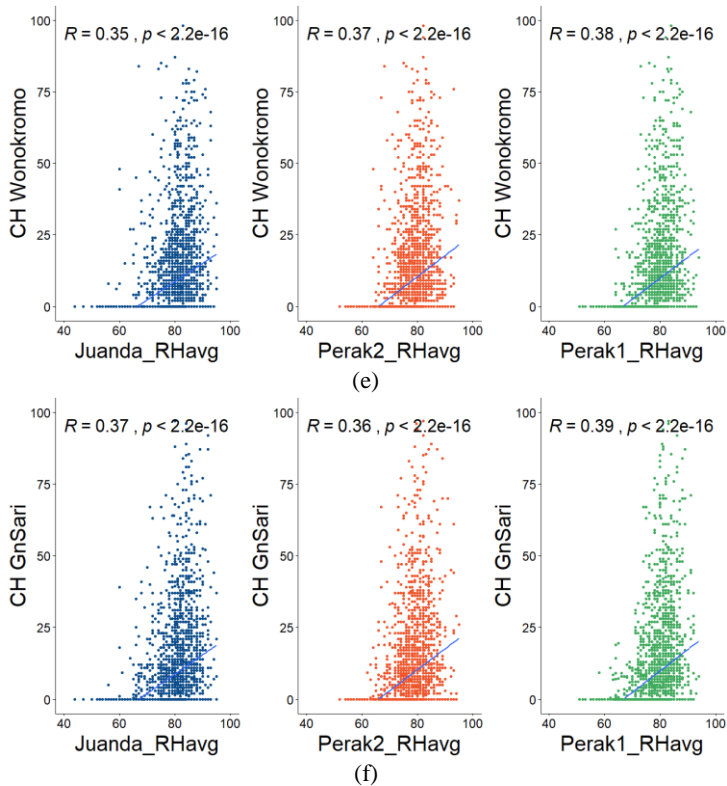
(b)



(c)



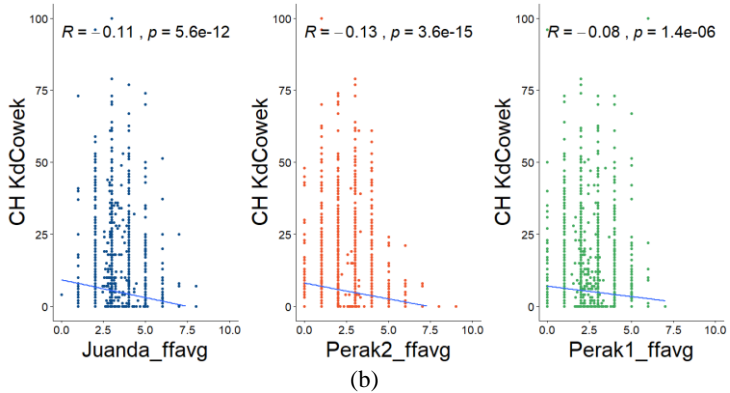
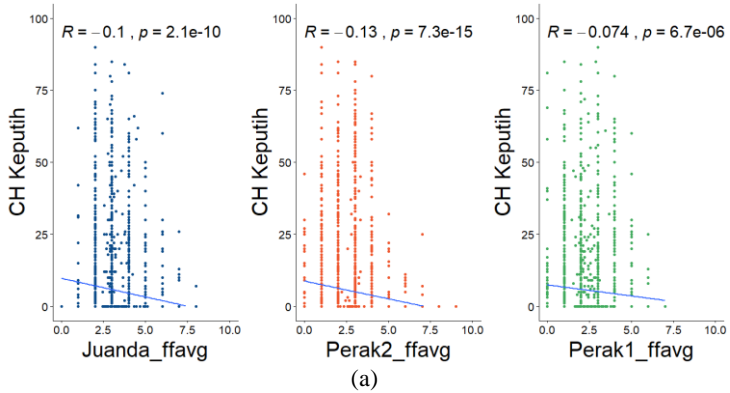
(d)

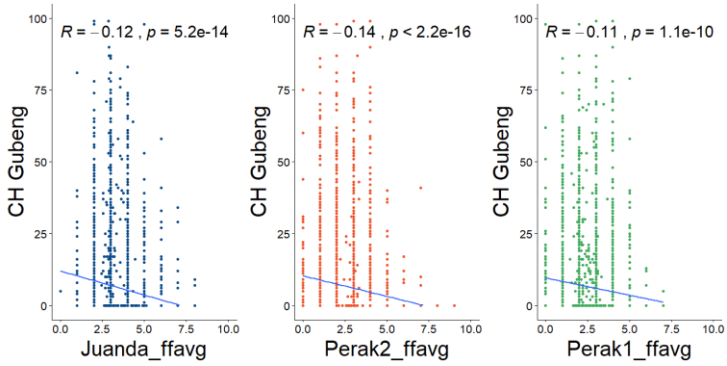


Gambar 4.6 Scatterplot Kelembapan Udara di 3 Stasiun BMKG dan Curah Hujan (a) Keputih, (b) Kedung Cowek, (c) Gubeng, (d) Wonorejo, (e) Wonokromo, dan (f) Gunung Sari

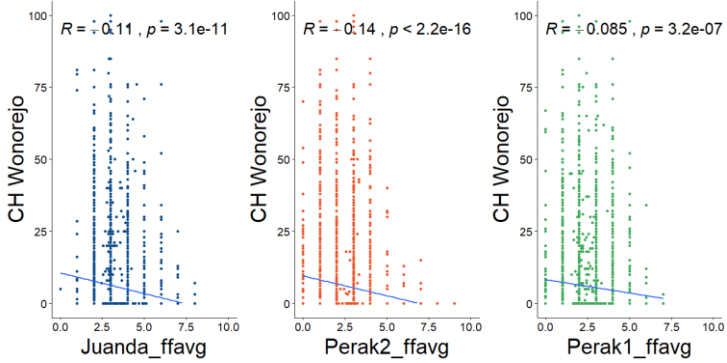
Gambar 4.6 menginformasikan bahwa hubungan antara kelembapan udara di masing-masing stasiun BMKG dengan curah hujan dimasing-masing balai yaitu berbanding lurus. Jika curah hujan hari ini sedang tinggi maka kelembapan udara hari ini juga akan tinggi, begitu pula sebaliknya. Nilai signifikansi uji korelasi *pearson* semuanya menunjukkan nilai kurang dari α 0,05 sehingga dapat dikatakan bahwa terdapat hubungan yang signifikan antara kelembapan udara di 3 stasiun BMKG dengan curah hujan di 6 balai PSAWS. Selain itu pula dilakukan analisis mengenai hubungan antara curah hujan hari ini dengan kelembapan udara 1 hari sebelumnya, 2 hari sebelumnya, 3 hari sebelumnya, 4 hari sebelumnya, dan 5 hari sebelumnya yang dapat

dilihat pada Lampiran 2. Pada Lampiran 2 didapatkan informasi bahwa nilai korelasi *pearson* antara kelembapan udara di Juanda dengan curah hujan memiliki nilai yang lebih kecil daripada korelasi antara kelembapan udara di Perak 1 dan Perak 2 dengan curah hujan di seluruh balai. Berikut juga disajikan hubungan antara kecepatan angin dengan curah hujan.

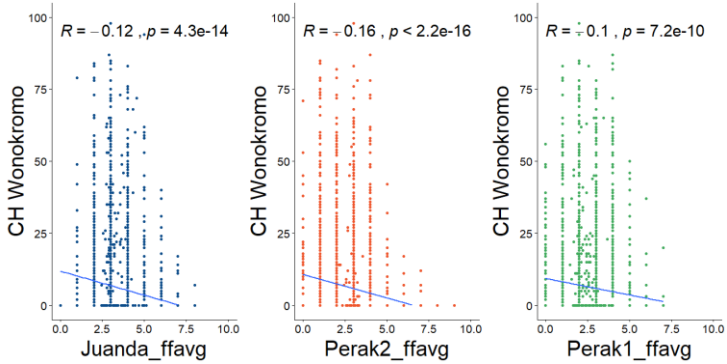




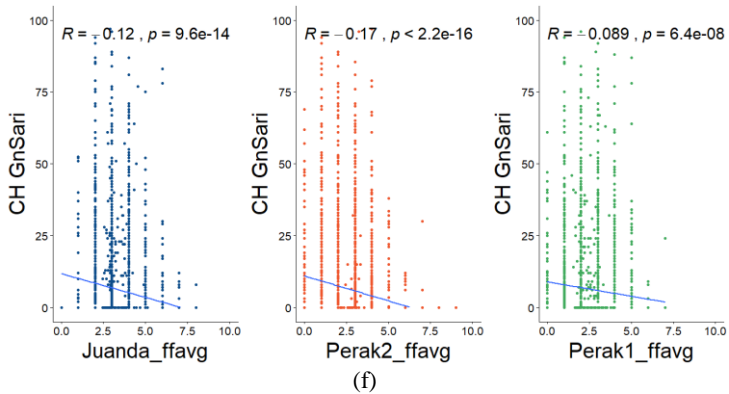
(c)



(d)



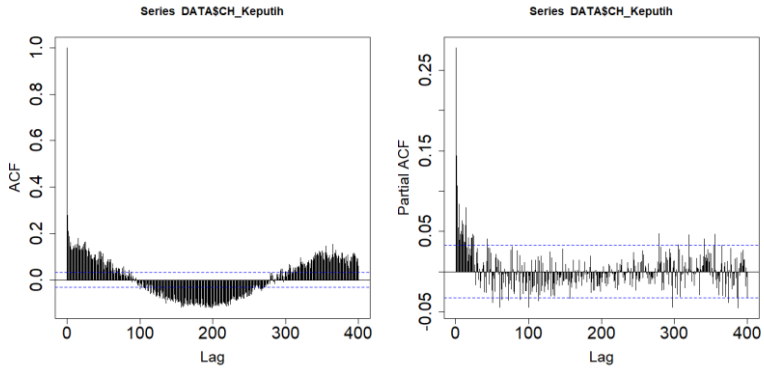
(e)



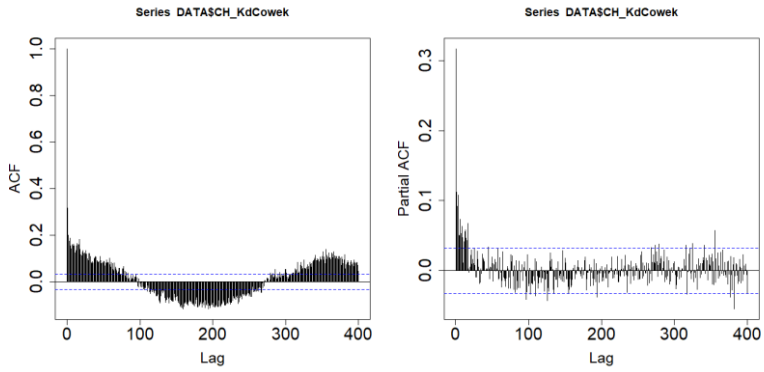
Gambar 4.7 Scatterplot Kecepatan Angin di 3 Stasiun BMKG dan Curah Hujan (a) Keputih, (b) Kedung Cowek, (c) Gubeng, (d) Wonorejo, (e) Wonokromo, dan (f) Gunung Sari

Hubungan antara curah hujan dengan kecepatan angin adalah berbanding terbalik, hal tersebut dapat dilihat dari nilai korelasi yang bernilai negatif. Nilai korelasi yang negatif menandakan bahwa jika curah hujan tinggi maka kecepatan angin cenderung akan rendah, begitu pula sebaliknya. Sedangkan untuk hubungan antara curah hujan hari ini dengan kecepatan angin hari-hari sebelumnya dapat dilihat pada Lampiran 3. Lampiran 3 menunjukkan bahwa terdapat beberapa lag dari kecepatan angin yang berpengaruh signifikan terhadap curah hujan.

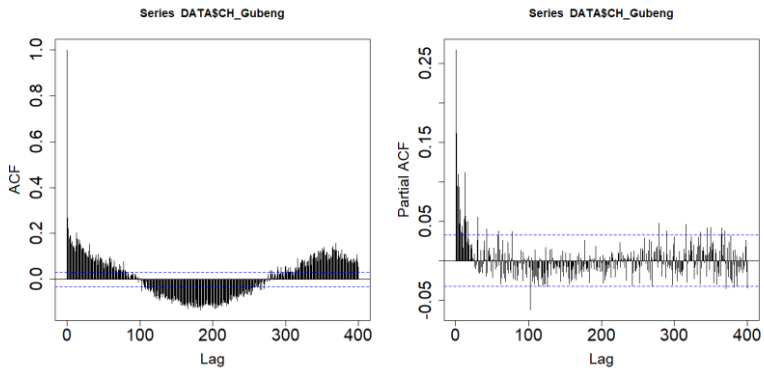
Terjadinya curah hujan selain disebabkan oleh variabel klimatologi, juga disebabkan atau berkaitan erat dengan curah hujan periode sebelumnya atau satu hari sebelumnya. Oleh karena itu perlu dilakukan eksplorasi dengan melihat plot ACF dan PACF curah hujan pada masing-masing balai yang akan ditampilkan pada Gambar 4.8 berikut.



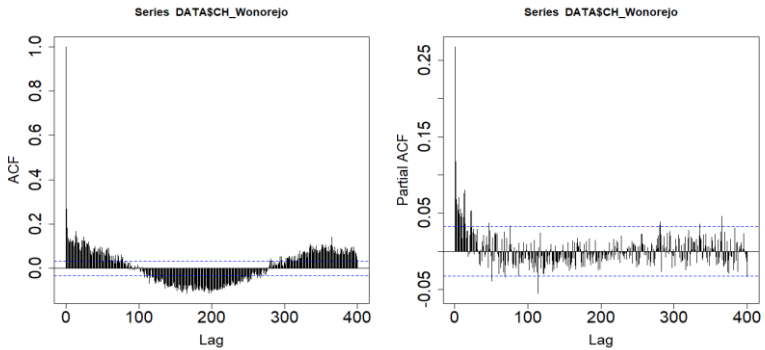
(a)



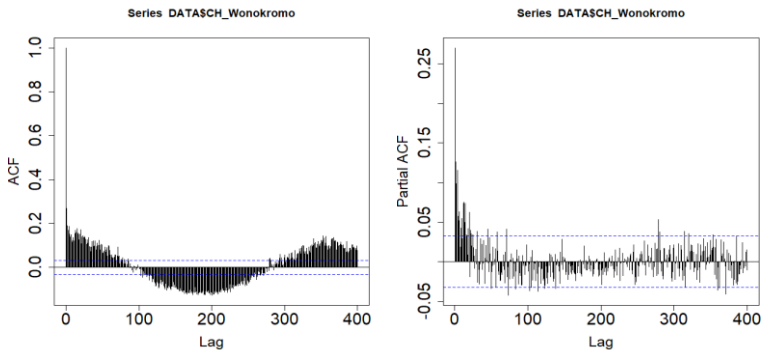
(b)



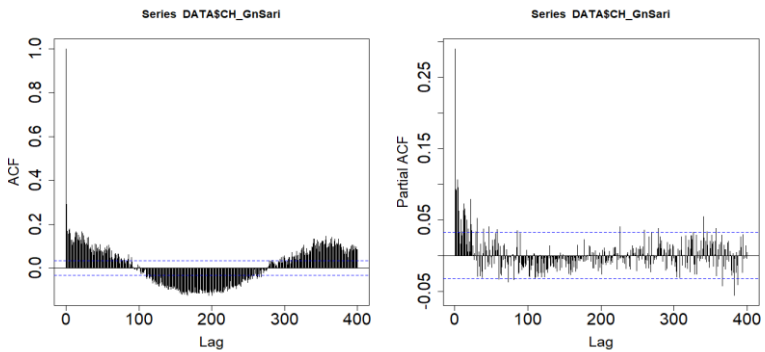
(c)



(d)



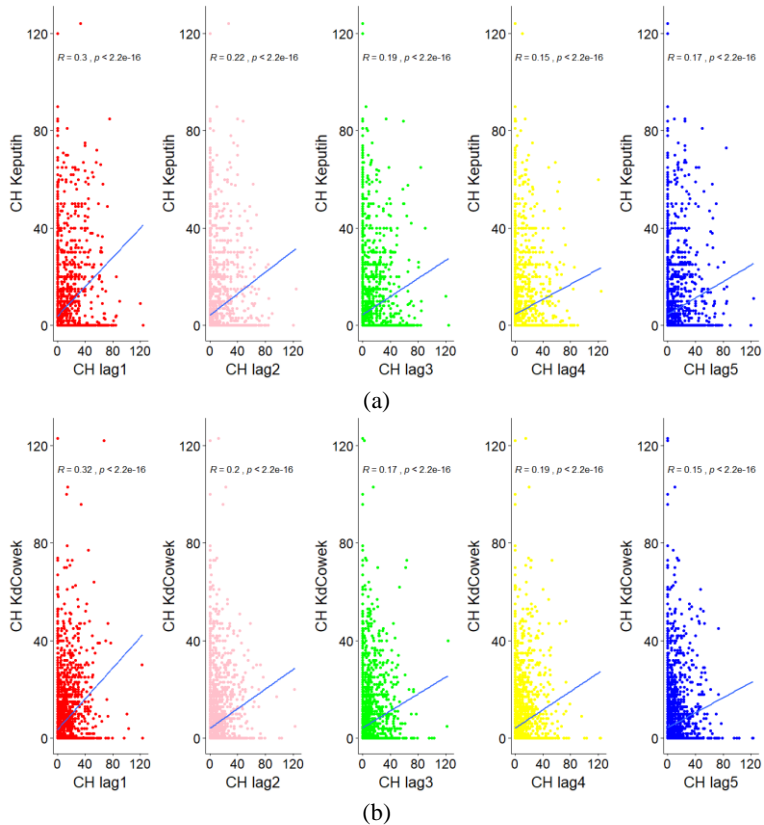
(e)

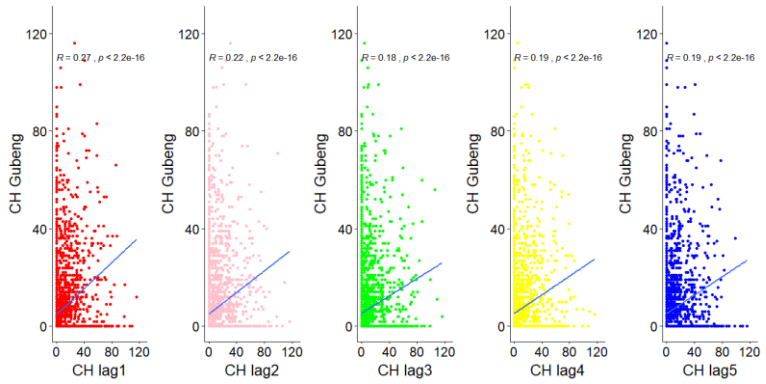


(f)

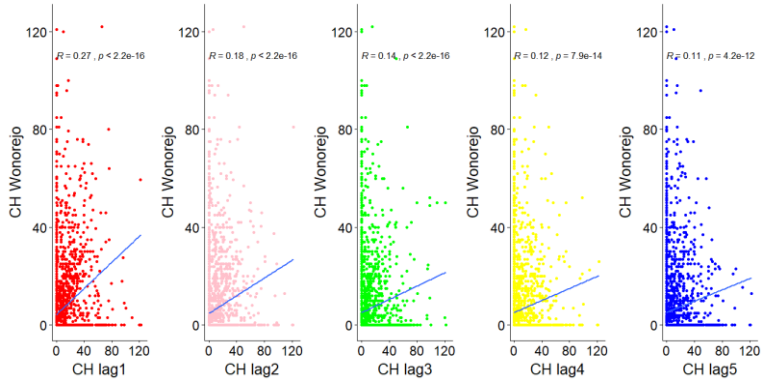
Gambar 4.8 ACF (Kiri) dan PACF (Kanan) Curah Hujan Balai (a) Keputih, (b) Kedung Cowek, (c) Gubeng, (d) Wonorejo, (e) Wonokromo, dan (f) Gunung Sari

Dari Gambar 4.8 dapat dilihat bahwa secara keseluruhan plot PACF pada masing-masing balai menunjukkan lag yang paling signifikan adalah lag 1. Sehingga dapat diindikasikan bahwa curah hujan periode sebelumnya atau sehari sebelumnya berpengaruh signifikan terhadap curah hujan saat ini. Selain lag 1, juga terdapat banyak lag-lag lainnya yang signifikan pada masing-masing balai hujan. Oleh karena itu perlu di cari tahu hubungan antara curah hujan hari ini dengan curah hujan beberapa hari sebelumnya melalui *scatterplot*. Namun pada penelitian ini hanya akan dilakukan uji korelasi antara curah hujan hari ini dengan curah hujan lag 1, lag 2, lag 3, lag 4, dan lag 5 yang akan ditampilkan pada Gambar 4.9 berikut.

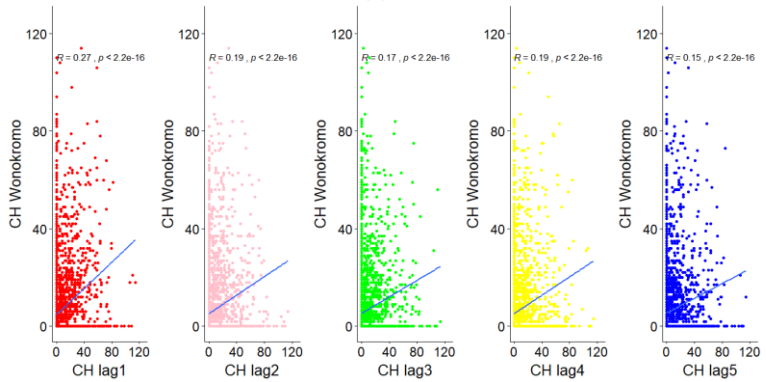




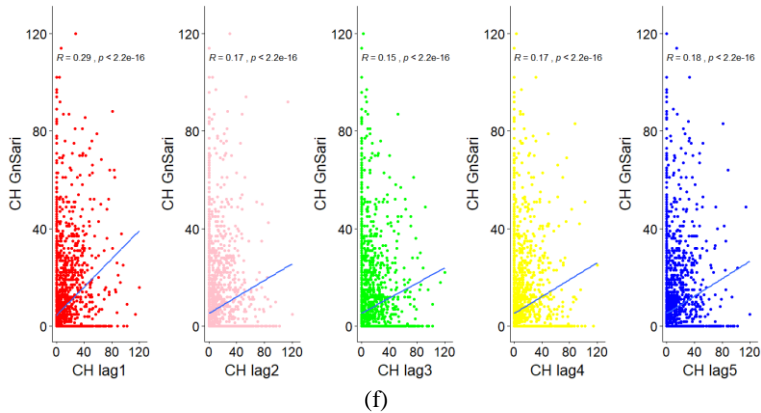
(c)



(d)



(e)

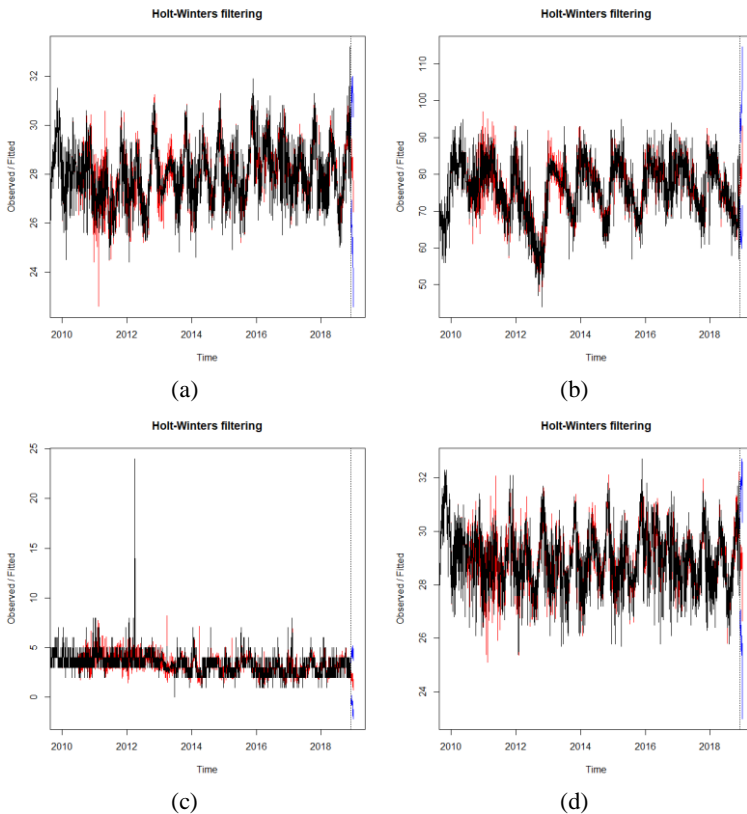


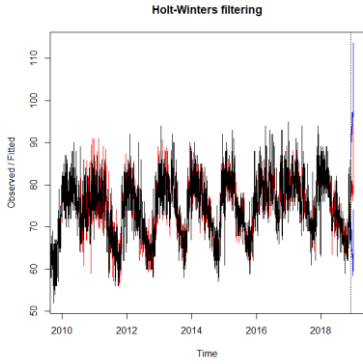
Gambar 4.9 Scatterplot antara CH dengan CH lag 1, lag 2, lag 3, lag 4, lag 5, dan lag 6 di Balai (a) Keputih, (b) Kedung Cowek, (c) Gubeng, (d) Wonorejo, (e) Wonokromo, dan (f) Gunung Sari

Gambar 4.9 memberikan informasi bahwasannya curah hujan hari ini sangat berkaitan erat atau memiliki hubungan yang berbanding lurus dengan curah hujan periode sebelumnya. Hal ini berarti bahwa jika curah hujan kemaren (sehari sebelumnya) tinggi, maka curah hujan hari ini cenderung akan tinggi, begitu pula sebaliknya. Hubungan curah hujan dengan curah hujan lag 1 memiliki nilai yang lebih tinggi daripada hubungan antara curah hujan dengan variabel temperatur maupun kecepatan angin. Sedangkan untuk hubungan antara kelembapan udara dengan curah hujan masih memiliki nilai yang lebih tinggi daripada hubungan antara curah hujan hari ini dengan curah hujan sebelum-sebelumnya. Selain itu pula, dari Gambar 4.9 didapatkan informasi lain bahwa semakin naik lag maka akan semakin menurunkan nilai korelasi atau dengan kata lain semakin jauh curah hujan sebelumnya maka akan semakin tidak berpengaruh terhadap curah hujan yang terjadi hari ini. Dari uraian analisis yang telah dijelaskan diatas maka dapat disimpulkan bahwa variabel-variabel yang berpengaruh signifikan dari mulai yang tertinggi sampai yang terendah berdasarkan nilai korelasi diantaranya yaitu kelembapan udara, lag dari curah hujan, kemudian disusul dengan temperatur, dan yang terakhir adalah kecepatan angin. Kecepatan angin disini memiliki nilai korelasi yang paling kecil dengan curah hujan dibandingkan dengan variabel-variabel *predictor* lainnya.

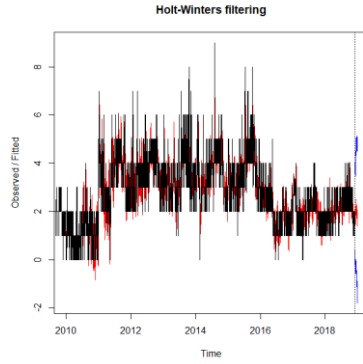
4.3 Peramalan *Univariate Variabel Predictor* dengan *Holt-Winters*

Metode *Holt-Winters* merupakan metode peramalan sederhana yang dapat menangkap adanya kasus *trend* dan *seasonal*. Dikarenakan variabel-variabel penelitian berbasis *time-series*, maka terlebih dahulu perlu dilakukan peramalan terhadap variabel *predictor* yang nantinya akan digunakan untuk meramalkan curah hujan pada data *out-sample*. Berikut merupakan hasil peramalan variabel *predictor* untuk 31 hari kedepan (periode Desember 2018) menggunakan metode *Holt-Winters*.

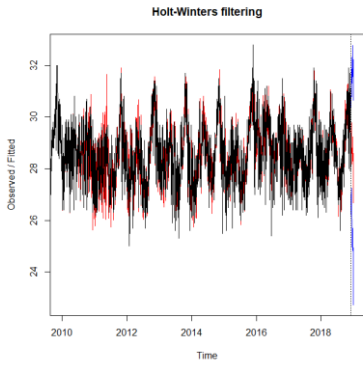




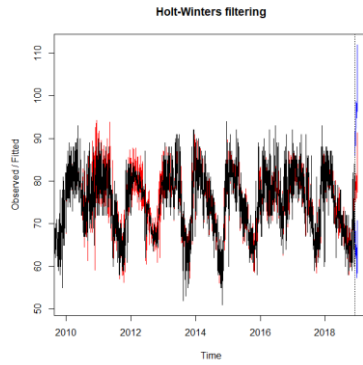
(e)



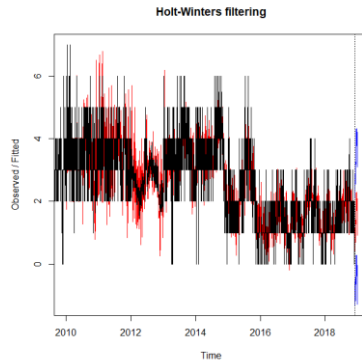
(f)



(g)



(h)



(i)

Gambar 4.10 Peramalan *Univariate* Menggunakan *Holt-Winters* Variabel *Predictor* (a) Temperatur Juanda (RMSE = 2,926509), (b) Kelembapan Udara Juanda (RMSE = 16,87781), (c) Kecepatan Angin Juanda (RMSE = 2,431392), (d) Temperatur Perak 2 (RMSE = 2,760564), (e) Kelembapan Udara Perak 2 (RMSE = 15,01662), (f) Kecepatan Angin Perak 2 (2,174094), (g) Temperatur Perak 1 (RMSE = 2,829435), (h) Kelembapan Udara Perak 1 (RMSE = 15,04157), (i) Kecepatan Angin Perak 1 (RMSE = 1,84213)

Gambar 4.10 menunjukkan bahwa hasil peramalan secara *univariate* variabel prediktor menggunakan metode *Holt-Winters* menghasilkan nilai prediksi 31 hari kedepan dengan nilai RMSE yang sudah cukup kecil, meskipun nilai RMSE untuk variabel kelembapan udara berbeda jauh dan justru lebih besar daripada temperatur dan kecepatan angin. Adapun hasil prediksi lengkapnya dapat dilihat pada Lampiran 4.

4.4 *Principal Component Analysis*

Dalam rangka mengatasi dependensi antar stasiun BMKG untuk masing-masing varabel *predictor* maka perlu dilakukan *principal component analysis* (PCA). Struktur data yang akan digunakan dalam proses PCA ini dapat dilihat pada Tabel 3.3. Tahapan awal yang dilakukan dalam PCA yaitu menentukan berapa banyak jumlah PC yang akan terbentuk, di mana penentuan jumlah PC tersebut dapat dilakukan dengan dua cara yaitu secara visual dengan melihat *scree plot* dan bisa juga dengan melihat nilai kumulatif proporsi variabilitas proses. Pada penelitian ini penentuan jumlah PC ditentukan hanya dengan melihat nilai kumulatif proporsi variabilitas proses yang melebihi angka 80%.

PCA dilakukan untuk masing-masing variabel *predictor*, untuk variabel temperatur akan dilakukan PCA antara temperatur di Juanda, temperatur di Perak 2, dan temperatur di Perak 1, begitupula seterusnya untuk variabel kelembapan udara dan kecepatan angin. Berikut merupakan hasil dari kumulatif proporsi yang diperoleh dengan bantuan *software R*.

Tabel 4.3 Hasil Kumulatif Proporsi

Variabel	PC1	PC2	PC3
Temperatur	0,8953	0,97173	1,00000
Kelembapan Udara	0,8693	0,96166	1,00000
Kecepatan Angin	0,6041	0.8389	1,00000

Pada Tabel 4.3 terlihat bahwa untuk variabel temperatur dan kelembapan udara, nilai kumulatif proporsi variabilitasnya telah mencapai angka 80% pada komponen pertama. Maka dapat diputuskan bahwa jumlah dari komponen utama yang akan digunakan adalah sebesar satu komponen (1 PC) saja. Dimana dari satu PC yang terbentuk tersebut dapat menjelaskan variabel temperatur sebesar 89,53% dari keragaman total dan dapat menjelaskan variabel kelembapan udara sebesar 86,93% dari keragaman total. Berbeda dengan temperatur dan kelembapan udara, variabel kecepatan angin untuk nilai kumulatif proporsinya mencapai angka 80% pada komponen kedua. Hal tersebut menunjukkan bahwasannya akan terdapat 2 PC untuk variabel kecepatan angin dan dapat menjelaskan 83,89 dari keragaman total.

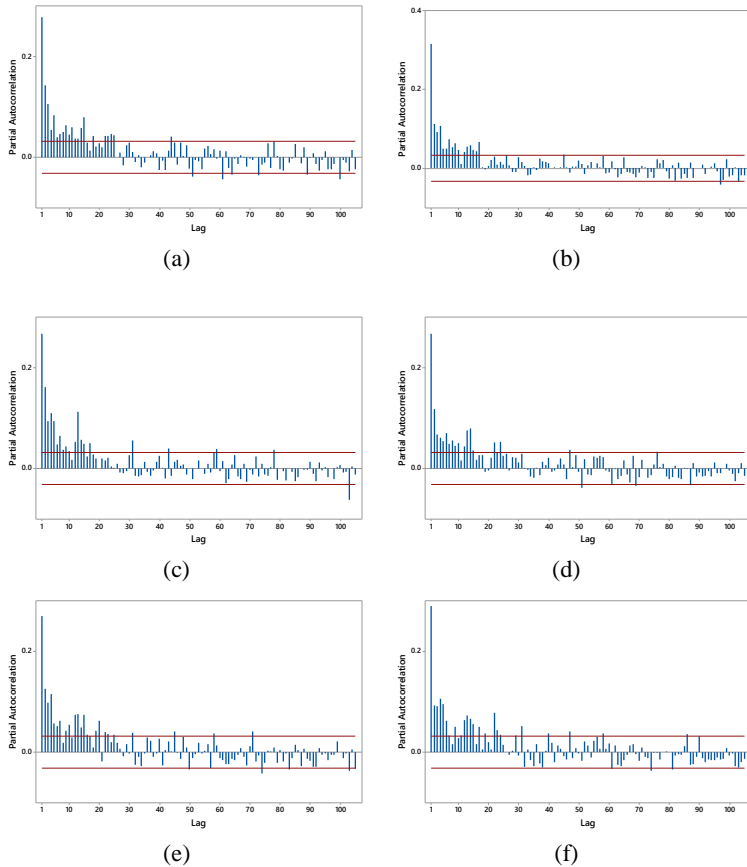
4.5 Penentuan *Input Lag* yang Signifikan

Dalam melakukan analisis menggunakan *Nu-Support Vector Regression* dan *Neural Network* yang tergolong sebagai metode non parametrik yang bebas asumsi, khususnya dalam penggunaan *v-SVR* dan NN untuk kasus prediksi maka proses pemilihan *input* merupakan salah satu hal yang paling *crusial* sebelum memulai analisis. *Input* yang digunakan dalam pemodelan *time-series v-SVR* dan NN adalah lag-lag yang signifikan. Dimana penentuan lag-lag yang signifikan untuk dijadikan sebagai *input* dapat diperoleh melalui dua cara, pertama yaitu dapat dilihat dari lag-lag yang signifikan pada plot PACF sebagai representasi model AR

dan yang kedua input lag-lag yang signifikan dari model ARIMA. Berikut akan dijabarkan lebih lengkap mengenai pemilihan *input* dari lag-lag yang signifikan pada plot PACF dan dalam model ARIMA untuk semua balai PSAWS.

4.5.1 Penentuan *Input* dari PACF

Penentuan *input* *v*-SVR dan NN dari PACF yaitu berdasarkan lag-lag yang signifikan pada PACF. Berikut disajikan plot PACF untuk curah hujan di semua balai PSAWS.



Gambar 4.11 Plot PACF Curah Hujan di Balai (a) Keputih, (b) Kedung Cowek, (c) Gubeng, (d) Wonorejo, (e) Wonokromo, dan (f) Gunung Sari

Gambar 4.11 (a) adalah plot PACF curah hujan di balai Keputih yang menunjukkan plot antara lag dan nilai PACF. Dapat dilihat bahwa lag-lag yang signifikan atau lag-lag yang nilainya berada diluar batas kepercayaan diantaranya yaitu lag 1, lag 2, lag 3, lag 4, lag 5, lag 6, lag 7, lag 8, lag 9, lag 10, ... , lag 61, lag 64, lag 73, lag 89, lag 100 dan seterusnya. Namun pada penelitian kali ini lag-lag yang akan digunakan sebagai *input* dibatasi maksimal sampai lag 100. Sehingga lag-lag yang akan digunakan sebagai *input* dalam *v*-SVR dan NN dirangkum dalam Tabel 4.4 berikut.

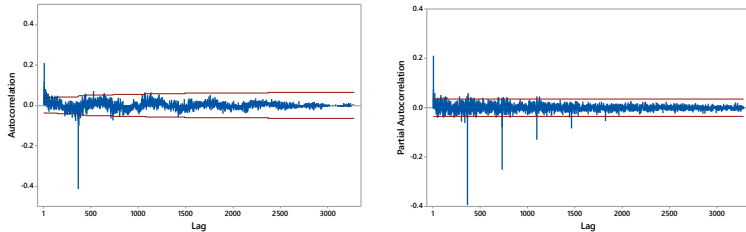
Tabel 4.4 *Input* Lag yang Signifikan dari PACF

<i>Predictant</i>	<i>Input Lag dari PACF</i>
CH Keputih	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 18, 22, 23, 24, 25, 44, 51, 61, 64, 73, 89, 100
CH KdCowek	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 13, 14, 15, 17, 45, 97
CH Gubeng	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 13, 14, 15, 17, 43, 59, 78
CH Wonorejo	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 13, 14, 15, 47, 51, 76
CH Wonokromo	1, 2, 3, 4, 5, 6, 7, 9, 10, 12, 13, 14, 15, 16, 45, 50, 58, 71, 74, 83
CH GnSari	1, 2, 3, 4, 5, 6, 9, 11, 12, 13, 14, 15, 17, 40, 47, 58, 74, 86

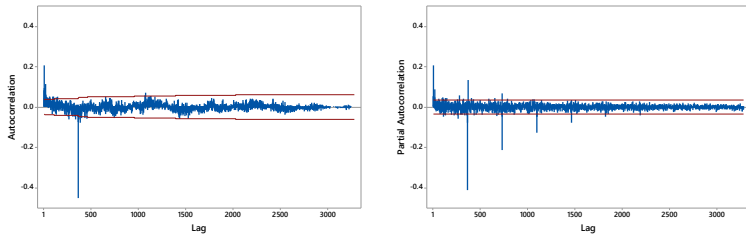
4.5.2 Penentuan *Input* dari model ARIMA

Tahapan awal yang harus dilakukan adalah identifikasi *time series plot* untuk mengetahui pola data. Sebelum identifikasi model, asumsi yang harus dipenuhi dalam pembentukan model ARIMA yaitu data stasioner dalam varians dan rata-rata. Dimana jika data tidak stationer dalam varians maka dilakukan transformasi sedangkan jika data tidak stationer dalam *mean* maka harus dilakukan *differencing*. Kestasioneran dalam varians ditunjukkan dengan fluktuasi data yang tidak terlalu jauh. Kestasioneran dalam rata-rata dapat dilakukan dengan melihat plot *time series* yang menunjukkan bahwa data curah hujan berada di sekitar rata-rata dan tidak mempunyai pola musiman. Sehingga dari Gambar 4.2 sebelumnya maka dapat disimpulkan bahwa data curah hujan di 6 balai PSAWS tidak stasioner dalam varians dan rata-rata, sebab masih terjadi fluktuasi data yang cukup jauh dan data curah hujan tidak berada di sekitar rata-rata serta terdapat pola

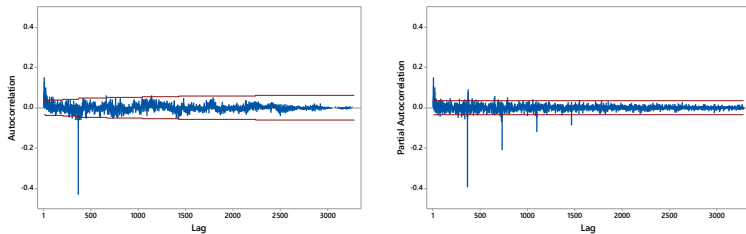
musiman pada Gambar 4.2. Stasioneritas dalam rata-rata juga dapat diselidiki lebih lanjut dengan melihat plot ACF dan PACF dari data curah hujan di 6 balai PSAWS yang pernah disajikan pada Gambar 4.8 sebelumnya. Pada Gambar 4.8 terlihat bahwa pada plot ACF masih terdapat banyak lag yang keluar dari batas signifikansi dan membentuk pola musiman, sedangkan plot PACF juga masih terdapat banyak lag yang keluar. Hal ini menunjukkan bahwa data curah hujan di semua balai PSAWS tidak memenuhi asumsi stasioner. Berdasarkan hasil analisis dari *time series plot*, ACF, dan PACF curah hujan di semua balai PSAWS, maka untuk mengatasi stasioneritas data perlu dilakukan *differencing* musiman yaitu pada lag 365 dan didapatkan hasil seperti pada Gambar 4.12.



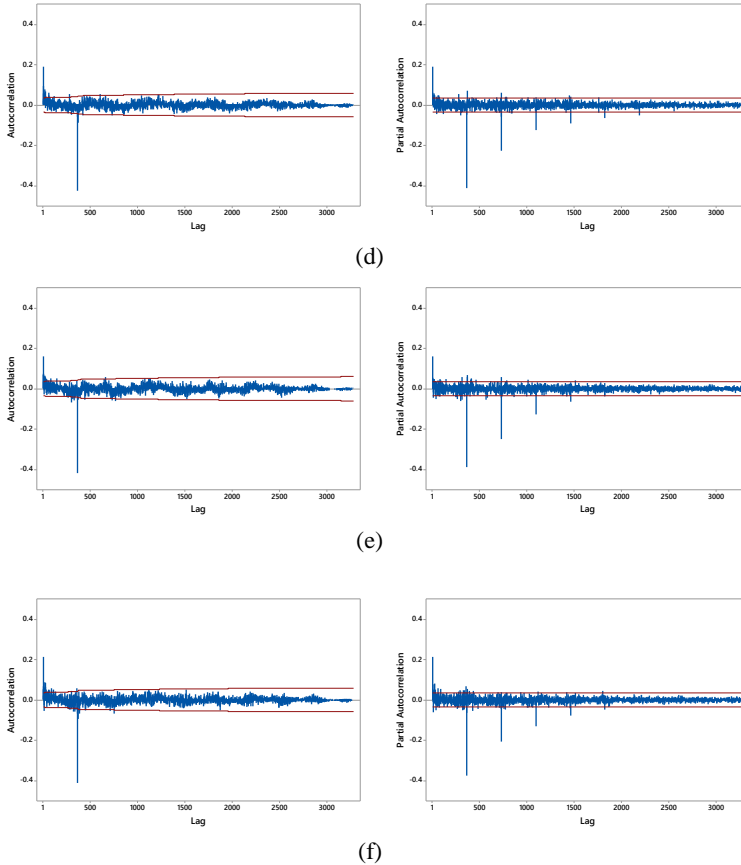
(a)



(b)



(c)



Gambar 4.12 Plot ACF (Kiri) dan PACF (Kanan) Curah Hujan di Balai (a) Keputih, (b) Kedung Cowek, (c) Gubeng, (d) Wonorejo, (e) Wonokromo, dan (f) Gunung Sari setelah *Differencing* Musiman Lag 365

Tahap selanjutnya adalah menentukan orde $(p,d,q)(P,D,Q)^s$ yang digunakan untuk membentuk model dugaan awal ARIMA. Sehingga berdasarkan plot ACF dan PACF pada Gambar 4.8 (a) dan 4.12 (a), maka untuk curah hujan Keputih dapat diduga bahwa model ARIMA yang terbentuk adalah $ARIMA(1,0,1)(5,1,1)^{365}$, $ARIMA(2,0,2)(5,1,1)^{365}$, $ARIMA(3,0,3)(5,1,1)^{365}$ dan lain-lain. Namun pada penelitian ini dikarenakan model ARIMA hanya untuk mendapatkan lag-lag yang signifikan bukan untuk estimasi parameter, maka untuk memudahkan analisis model ARIMA yang

akan digunakan adalah model ARIMA dengan menghiraukan model MA sebab jika MA dimasukkan dalam analisis akan sulit dilakukan dan harus memodelkan eror terlebih dahulu. Sehingga model ARIMA yang akan digunakan untuk menentukan lag-lag yang signifikan pada curah hujan di Keputih yaitu ARIMA $(1,0,0)(5,1,0)^{365}$. Cara yang sama dilakukan untuk curah hujan di 5 balai PSAWS lainnya dan didapatkan model ARIMA sebagai berikut.

Tabel 4.5 Model ARIMA untuk Curah Hujan di 6 Balai PSAWS

<i>Predictant</i>	Model ARIMA
CH Keputih	$(1,0,0)(5,1,0)^{365}$
CH KdCowek	$(1,0,0)(6,1,0)^{365}$
CH Gubeng	$(1,0,0)(4,1,0)^{365}$
CH Wonorejo	$(1,0,0)(6,1,0)^{365}$
CH Wonokromo	$(1,0,0)(5,1,0)^{365}$
CH GnSari	$(1,0,0)(5,1,0)^{365}$

Tabel 4.5 memberikan informasi bahwa curah hujan di balai Keputih, Wonokromo dan Gunung Sari memiliki model ARIMA yang sama yaitu ARIMA $(1,0,0)(5,1,0)^{365}$. Sedangkan curah hujan di balai Kedung Cowek memiliki model ARIMA yang sama dengan curah hujan di balai Wonorejo yaitu ARIMA $(1,0,0)(6,1,0)^{365}$ sedangkan untuk curah hujan di balai Gubeng memiliki model ARIMA $(1,0,0)(4,1,0)^{365}$. Selanjutnya untuk mendapatkan lag-lag yang signifikan dari model ARIMA diatas maka perlu dilakukan perhitungan matematis sebagai berikut :

Penjabaran Model $(1,0,0)(4,1,0)^{365}$:

$$Y_t = ARIMA(1,0,0)(4,1,0)^{365}$$

$$\phi_p(B)\Phi_p(B^s)(1-B)^d(1-B^s)^D Y_t = \theta_0 + \theta_q(B)\Theta_Q(B^s)a_t$$

$$(1-\phi_1 B)(1-\Phi_1 B^{365} - \Phi_2 B^{2(365)} - \Phi_3 B^{3(365)} - \Phi_4 B^{4(365)})(1-B^{365})^1 Y_t = a_t$$

$$(1-\phi_1 B)(1-\Phi_1 B^{365} - \Phi_2 B^{2(365)} - \Phi_3 B^{3(365)} - \Phi_4 B^{4(365)})(Y_t - Y_{t-365}) = a_t$$

$$((1-\Phi_1 B^{365} - \Phi_2 B^{730} - \Phi_3 B^{1095} - \Phi_4 B^{1460}) - (\phi_1 B - \phi_1 \Phi_1 B B^{365} - \phi_1 \Phi_2 B B^{730} - \phi_1 \Phi_3 B B^{1095} - \phi_1 \Phi_4 B B^{1460}))(Y_t - Y_{t-365}) = a_t$$

$$(1-\phi_1 B - \Phi_1 B^{365} + \Psi_1 B^{366} - \Phi_2 B^{730} + \Psi_2 B^{731} - \Phi_3 B^{1095} + \Psi_3 B^{1096}$$

$$- \Phi_4 B^{1460} + \Psi_4 B^{1461})(Y_t - Y_{t-365}) = a_t$$

$$W_t - \phi_1 W_{t-1} - \Phi_1 W_{t-365} + \Psi_1 W_{t-366} - \Phi_2 W_{t-730} + \Psi_2 W_{t-731} - \Phi_3 W_{t-1095} + \Psi_3 W_{t-1096}$$

$$- \Phi_4 W_{t-1460} + \Psi_4 W_{t-1461} = a_t$$

$$W_t = \phi_1 (Y_{t-1} - Y_{t-1-365}) + \Phi_1 (Y_{t-365} - Y_{t-365-365}) - \Psi_1 (Y_{t-366} - Y_{t-366-365})$$

$$+ \Phi_2 (Y_{t-730} - Y_{t-730-365}) - \Psi_2 (Y_{t-731} - Y_{t-731-365}) + \Phi_3 (Y_{t-1095} - Y_{t-1095-365})$$

$$- \Psi_3 (Y_{t-1096} - Y_{t-1096-365}) + \Phi_4 (Y_{t-1460} - Y_{t-1460-365}) - \Psi_4 (Y_{t-1461} - Y_{t-1461-365}) + a_t$$

$$W_t = \phi_1 (Y_{t-1} - Y_{t-366}) + \Phi_1 (Y_{t-365} - Y_{t-730}) - \Psi_1 (Y_{t-366} - Y_{t-731}) + \Phi_2 (Y_{t-730} - Y_{t-1095})$$

$$- \Psi_2 (Y_{t-731} - Y_{t-1096}) + \Phi_3 (Y_{t-1095} - Y_{t-1460}) - \Psi_3 (Y_{t-1096} - Y_{t-1461}) + \Phi_4 (Y_{t-1460} - Y_{t-1825})$$

$$- \Psi_4 (Y_{t-1461} - Y_{t-1826}) + a_t$$

$$Y_t = f(Y_{t-1}, Y_{t-365}, Y_{t-366}, Y_{t-730}, Y_{t-731}, Y_{t-1095}, Y_{t-1096}, Y_{t-1460}, Y_{t-1461}, Y_{t-1825}, Y_{t-1826})$$

Penjabaran Model $(1,0,0)(5,1,0)^{365}$:

$$Y_t = ARIMA(1,0,0)(5,1,0)^{365}$$

$$\phi_p(B)\Phi_p(B^s)(1-B)^d(1-B^s)^D Y_t = \theta_0 + \theta_q(B)\Theta_Q(B^s)a_t$$

$$(1-\phi_1 B)(1-\Phi_1 B^{365} - \Phi_2 B^{2(365)} - \Phi_3 B^{3(365)} - \Phi_4 B^{4(365)}$$

$$-\Phi_5 B^{5(365)})(1-B^{365})^1 Y_t = a_t$$

$$(1-\phi_1 B)(1-\Phi_1 B^{365} - \Phi_2 B^{2(365)} - \Phi_3 B^{3(365)} - \Phi_4 B^{4(365)}$$

$$-\Phi_5 B^{5(365)})(Y_t - Y_{t-365}) = a_t$$

$$((1-\Phi_1 B^{365} - \Phi_2 B^{730} - \Phi_3 B^{1095} - \Phi_4 B^{1460} - \Phi_5 B^{1825})$$

$$-(\phi_1 B - \phi_1 \Phi_1 B B^{365} - \phi_1 \Phi_2 B B^{730} - \phi_1 \Phi_3 B B^{1095} - \phi_1 \Phi_4 B B^{1460}$$

$$-\phi_1 \Phi_5 B B^{1825}))(Y_t - Y_{t-365}) = a_t$$

$$(1-\phi_1 B - \Phi_1 B^{365} + \Psi_1 B^{366} - \Phi_2 B^{730} + \Psi_2 B^{731} - \Phi_3 B^{1095} + \Psi_3 B^{1096}$$

$$-\Phi_4 B^{1460} + \Psi_4 B^{1461} - \Phi_5 B^{1825} + \Psi_5 B^{1826})(Y_t - Y_{t-365}) = a_t$$

$$W_t - \phi_1 W_{t-1} - \Phi_1 W_{t-365} + \Psi_1 W_{t-366} - \Phi_2 W_{t-730} + \Psi_2 W_{t-731} - \Phi_3 W_{t-1095}$$

$$+ \Psi_3 W_{t-1096} - \Phi_4 W_{t-1460} + \Psi_4 W_{t-1461} - \Phi_5 W_{t-1825} + \Psi_5 W_{t-1826} = a_t$$

$$W_t = \phi_1 (Y_{t-1} - Y_{t-1-365}) + \Phi_1 (Y_{t-365} - Y_{t-365-365}) - \Psi_1 (Y_{t-366} - Y_{t-366-365})$$

$$+ \Phi_2 (Y_{t-730} - Y_{t-730-365}) - \Psi_2 (Y_{t-731} - Y_{t-731-365}) + \Phi_3 (Y_{t-1095} - Y_{t-1095-365})$$

$$- \Psi_3 (Y_{t-1096} - Y_{t-1096-365}) + \Phi_4 (Y_{t-1460} - Y_{t-1460-365}) - \Psi_4 (Y_{t-1461} - Y_{t-1461-365})$$

$$+ \Phi_5 (Y_{t-1825} - Y_{t-1825-365}) - \Psi_5 (Y_{t-1826} - Y_{t-1826-365}) + a_t$$

$$W_t = \phi_1 (Y_{t-1} - Y_{t-366}) + \Phi_1 (Y_{t-365} - Y_{t-730}) - \Psi_1 (Y_{t-366} - Y_{t-731})$$

$$+ \Phi_2 (Y_{t-730} - Y_{t-1095}) - \Psi_2 (Y_{t-731} - Y_{t-1096}) + \Phi_3 (Y_{t-1095} - Y_{t-1460})$$

$$- \Psi_3 (Y_{t-1096} - Y_{t-1461}) + \Phi_4 (Y_{t-1460} - Y_{t-1825}) - \Psi_4 (Y_{t-1461} - Y_{t-1826})$$

$$+ \Phi_5 (Y_{t-1825} - Y_{t-2190}) - \Psi_5 (Y_{t-1826} - Y_{t-2191}) + a_t$$

$$Y_t = f(Y_{t-1}, Y_{t-365}, Y_{t-366}, Y_{t-730}, Y_{t-731}, Y_{t-1095}, Y_{t-1096}, Y_{t-1460}, Y_{t-1461}, Y_{t-1825},$$

$$Y_{t-1826}, Y_{t-2190}, Y_{t-2191})$$

Penjabaran Model $(1,0,0)(6,1,0)^{365}$:

$$Y_t = ARIMA(p, d, q)(P, D, Q)^s$$

$$Y_t = ARIMA(1, 0, 0)(6, 1, 0)^{365}$$

$$\phi_p(B)\Phi_p(B^s)(1-B)^d(1-B^s)^DY_t = \theta_0 + \theta_q(B)\Theta_Q(B^s)a_t$$

$$(1 - \phi_1 B)(1 - \Phi_1 B^{365} - \Phi_2 B^{2(365)} - \Phi_3 B^{3(365)} - \Phi_4 B^{4(365)}$$

$$- \Phi_5 B^{5(365)} - \Phi_6 B^{6(365)})(1 - B^{365})^1 Y_t = a_t$$

$$(1 - \phi_1 B)(1 - \Phi_1 B^{365} - \Phi_2 B^{2(365)} - \Phi_3 B^{3(365)} - \Phi_4 B^{4(365)}$$

$$- \Phi_5 B^{5(365)} - \Phi_6 B^{6(365)})(Y_t - Y_{t-365}) = a_t$$

$$((1 - \Phi_1 B^{365} - \Phi_2 B^{730} - \Phi_3 B^{1095} - \Phi_4 B^{1460} - \Phi_5 B^{1825} - \Phi_6 B^{2190})$$

$$- (\phi_1 B - \phi_1 \Phi_1 B B^{365} - \phi_1 \Phi_2 B B^{730} - \phi_1 \Phi_3 B B^{1095} - \phi_1 \Phi_4 B B^{1460}$$

$$- \phi_1 \Phi_5 B B^{1825} - \phi_1 \Phi_6 B B^{2190}))(Y_t - Y_{t-365}) = a_t$$

$$(1 - \phi_1 B - \Phi_1 B^{366} + \Psi_1 B^{366} - \Phi_2 B^{730} + \Psi_2 B^{731} - \Phi_3 B^{1095} + \Psi_3 B^{1096}$$

$$- \Phi_4 B^{1460} + \Psi_4 B^{1461} - \Phi_5 B^{1825} + \Psi_5 B^{1826} - \Phi_6 B^{2190} + \Psi_6 B^{2191})(Y_t - Y_{t-365}) = a_t$$

$$W_t - \phi_1 W_{t-1} - \Phi_1 W_{t-365} + \Psi_1 W_{t-366} - \Phi_2 W_{t-730} + \Psi_2 W_{t-731} - \Phi_3 W_{t-1095} + \Psi_3 W_{t-1096}$$

$$- \Phi_4 W_{t-1460} + \Psi_4 W_{t-1461} - \Phi_5 W_{t-1825} + \Psi_5 W_{t-1826} - \Phi_6 W_{t-2190} + \Psi_6 W_{t-2191} = a_t$$

$$W_t = \phi_1 (Y_{t-1} - Y_{t-1-365}) + \Phi_1 (Y_{t-365} - Y_{t-365-365}) - \Psi_1 (Y_{t-366} - Y_{t-366-365})$$

$$+ \Phi_2 (Y_{t-730} - Y_{t-730-365}) - \Psi_2 (Y_{t-731} - Y_{t-731-365}) + \Phi_3 (Y_{t-1095} - Y_{t-1095-365})$$

$$- \Psi_3 (Y_{t-1096} - Y_{t-1096-365}) + \Phi_4 (Y_{t-1460} - Y_{t-1460-365}) - \Psi_4 (Y_{t-1461} - Y_{t-1461-365})$$

$$+ \Phi_5 (Y_{t-1825} - Y_{t-1825-365}) - \Psi_5 (Y_{t-1826} - Y_{t-1826-365}) + \Phi_6 (Y_{t-2190} - Y_{t-2190-365})$$

$$- \Psi_6 (Y_{t-2191} - Y_{t-2191-365}) + a_t$$

$$W_t = \phi_1 (Y_{t-1} - Y_{t-366}) + \Phi_1 (Y_{t-365} - Y_{t-730}) - \Psi_1 (Y_{t-366} - Y_{t-731}) + \Phi_2 (Y_{t-730} - Y_{t-1095})$$

$$- \Psi_2 (Y_{t-731} - Y_{t-1096}) + \Phi_3 (Y_{t-1095} - Y_{t-1460}) - \Psi_3 (Y_{t-1096} - Y_{t-1461}) + \Phi_4 (Y_{t-1460} - Y_{t-1825})$$

$$- \Psi_4 (Y_{t-1461} - Y_{t-1826}) + \Phi_5 (Y_{t-1825} - Y_{t-2190}) - \Psi_5 (Y_{t-1826} - Y_{t-2191})$$

$$+ \Phi_6 (Y_{t-2190} - Y_{t-2555}) - \Psi_6 (Y_{t-2191} - Y_{t-2556}) + a_t$$

$$Y_t = f(Y_{t-1}, Y_{t-365}, Y_{t-366}, Y_{t-730}, Y_{t-731}, Y_{t-1095}, Y_{t-1096}, Y_{t-1460}, Y_{t-1461}, Y_{t-1825}, Y_{t-1826},$$

$$Y_{t-2190}, Y_{t-2191}, Y_{t-2555}, Y_{t-2556})$$

Sehingga berdasarkan penjabaran diatas, dapat diambil kesimpulan yang dirangkum dalam Tabel 4.6.

Tabel 4.6 *Input Lag yang Signifikan dari ARIMA*

<i>Predictant</i>	<i>Input Lag dari ARIMA</i>
CH Keputih	1, 365, 366, 730, 731, 1095, 1096, 1460, 1461, 1825, 1826, 2190, 2191
CH KdCowek	1, 365, 366, 730, 731, 1095, 1096, 1460, 1461, 1825, 1826, 2190, 2191, 2555, 2556
CH Gubeng	1, 365, 366, 730, 731, 1095, 1096, 1460, 1461, 1825, 1826
CH Wonorejo	1, 365, 366, 730, 731, 1095, 1096, 1460, 1461, 1825, 1826, 2190, 2191, 2555, 2556
CH Wonokromo	1, 365, 366, 730, 731, 1095, 1096, 1460, 1461, 1825, 1826, 2190, 2191
CH GnSari	1, 365, 366, 730, 731, 1095, 1096, 1460, 1461, 1825, 1826, 2190, 2191

4.6 Analisis Curah Hujan di Balai Keputih

Balai Keputih merupakan bagian dari balai PSAWS yang mempunyai salah satu pos hujan di daerah Keputih Surabaya. Pada bagian ini akan dilakukan prediksi curah hujan di balai keputih berdasarkan variabel *predictor* dari 3 stasiun BMKG. Adapun metode yang akan digunakan dan kemudian dibandingkan diantaranya yaitu ν -SVR, ν -SVR dengan PCA, NN, dan NN dengan PCA menggunakan *input lag* dari PACF dan *input lag* dari ARIMA. Berikut penjabaran lebih lengkapnya.

4.6.1 Analisis Nu-Support Vector Regression

Seperti yang telah dijelaskan pada bagian langkah penelitian di bab 3, analisis ν -SVR dilakukan dengan mencobakkan berbagai kemungkinan model yang terbentuk seperti skenario pada bagian langkah penelitian. Namun sebelum memodelkan menggunakan ν -SVR terlebih dahulu harus dilakukan optimasi parameter *gamma*, *cost*, dan *nu*. Dimana range untuk nilai *gamma* yang akan dioptimasi yaitu antara 1-2,5 dengan *seq* 0,5. Nilai *cost* dioptimasi pada range 1-100 dengan *seq* 5, dan yang terakhir nilai *nu* pada range 0.2-0.5 dengan *seq* 0,1. Dikarenakan pada penelitian ini digunakan SVR jenis *Nu* bukan *Eps*, maka *epsilon* tidak perlu dioptimasi, sebab dalam ν -SVR nilai *epsilon* sudah dioptimasi

secara otomatis. Sehingga kombinasi optimasi antara parameter γ , $cost$, dan ν terbentuk sebanyak 400 kombinasi. Adapun fungsi kernel yang akan digunakan adalah RBF atau *Gaussian* sebab sudah banyak penelitian yang membuktikan bahwa kernel RBF merupakan kernel terbaik diantara kernel-kernel yang lainnya. Selain itu dasar digunakannya kernel RBF daripada polynomial dan linier yaitu karena setelah dilakukan beberapa percobaan sampling (studi pendahuluan di curah hujan Keputih) untuk membandingkan ketiganya diperoleh hasil lebih baik menggunakan kernel RBF, dimana kernel RBF mampu memberikan hasil prediksi yang lebih mendekati data aktual. Oleh karena itu analisis pada proses optimasi parameter digunakan dan dipilihlah kernel RBF. Langkah-langkah prediksi menggunakan ν -SVR untuk setiap model dapat dilihat pada langkah penelitian di subbab 3.6.

a. Pemodelan ν -SVR dengan *Input* Variabel Prediktor

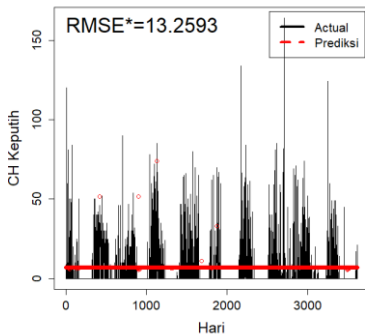
Pemodelan curah hujan di Keputih berdasarkan variabel variabel klimatologi di 3 stasiun Juanda menggunakan ν -SVR dilakukan dengan skenario sebagai berikut.

Tabel 4.7 Skenario Pemodelan ν -SVR CH Keputih dengan *Input* Prediktor

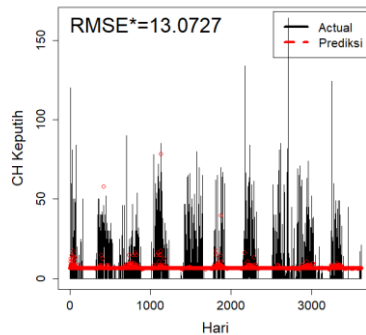
Model	Persamaan
1	$Y_{1,t} = f(X_{11,t})$
2	$Y_{1,t} = f(X_{11,t}, M_t)$
3	$Y_{1,t} = f(X_{21,t})$
4	$Y_{1,t} = f(X_{21,t}, M_t)$
5	$Y_{1,t} = f(X_{31,t})$
6	$Y_{1,t} = f(X_{31,t}, M_t)$
7	$Y_{1,t} = f(X_{11,t}, X_{21,t}, X_{31,t})$
8	$Y_{1,t} = f(X_{11,t}, X_{21,t}, X_{31,t}, M_t)$
9	$Y_{1,t} = f(X_{12,t})$
10	$Y_{1,t} = f(X_{12,t}, M_t)$
11	$Y_{1,t} = f(X_{22,t})$
12	$Y_{1,t} = f(X_{22,t}, M_t)$
13	$Y_{1,t} = f(X_{32,t})$
14	$Y_{1,t} = f(X_{32,t}, M_t)$
15	$Y_{1,t} = f(X_{12,t}, X_{22,t}, X_{32,t})$
16	$Y_{1,t} = f(X_{12,t}, X_{22,t}, X_{32,t}, M_t)$

- 17 $Y_{1,t} = f(X_{13,t})$
 18 $Y_{1,t} = f(X_{13,t}, M_t)$
 19 $Y_{1,t} = f(X_{23,t})$
 20 $Y_{1,t} = f(X_{23,t}, M_t)$
 21 $Y_{1,t} = f(X_{33,t})$
 22 $Y_{1,t} = f(X_{33,t}, M_t)$
 23 $Y_{1,t} = f(X_{13,t}, X_{23,t}, X_{33,t})$
 24 $Y_{1,t} = f(X_{13,t}, X_{23,t}, X_{33,t}, M_t)$

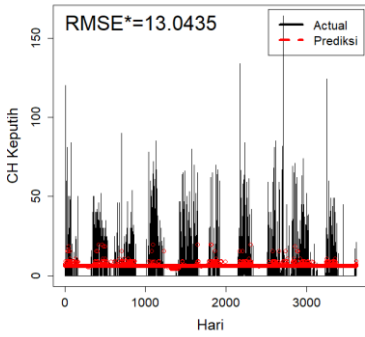
Berdasarkan skenario pada Tabel 4.7, maka 24 model di atas dilakukan optimasi parameter γ , $cost$, dan nu sesuai dengan rentang yang telah ditentukan dalam langkah penelitian. Adapun nilai parameter optimum yang digunakan hasil dari 400 kombinasi parameter untuk ke-24 model di atas serta nilai RMSE dari data training dapat dilihat pada Lampiran 5. Berikut merupakan plot antara data aktual dan data prediksi pada data *training* dari model 1 - model 8 yang telah melalui proses optimasi parameter.



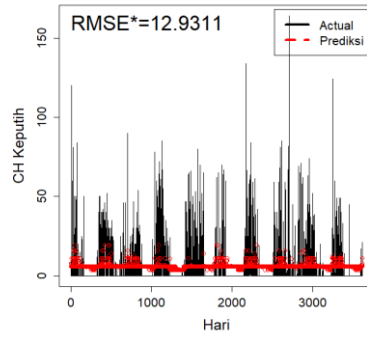
(a)



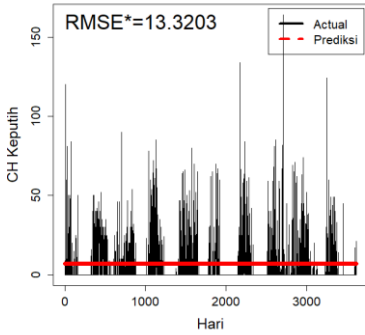
(b)



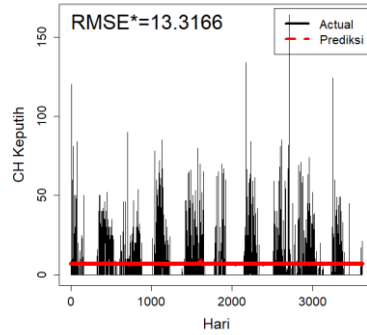
(c)



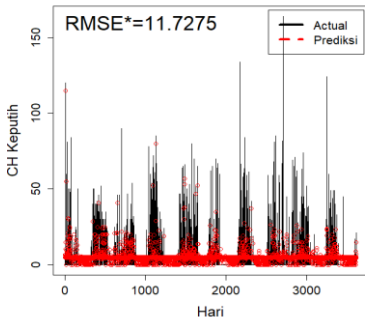
(d)



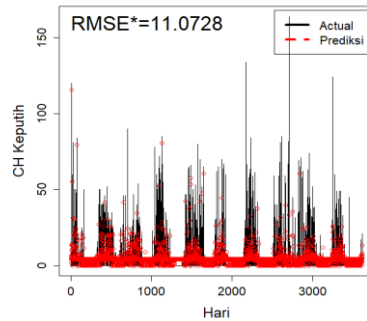
(e)



(f)



(g)



(h)

Gambar 4.13 Plot (a) Model 1, (b) Model 2, (c) Model 3, (d) Model 4, (e) Model 5, (f) Model 6, (g) Model 7, (h) Model 8

Nilai RMSE* pada Gambar 4.13 dan Lampiran 5 merupakan nilai RMSE yang dihitung dari data aktual yang ditandai oleh garis hitam dengan data hasil prediksi berwarna merah yang telah melalui proses transformasi 0. Berdasarkan optimasi parameter γ , $cost$, dan ν dapat disimpulkan bahwa modus γ optimum dari seluruh kombinasi model adalah γ dengan nilai 2,5, sedangkan modus dari parameter $cost$ dari seluruh kombinasi model adalah 100, dan yang terakhir modus dari parameter ν adalah 0,3. Hasil analisis ν -SVR pada Gambar 4.13 menunjukkan bahwa jika curah hujan Keputih dimodelkan secara *univariate* menggunakan variabel temperatur, kelembapan udara, dan kecepatan angin di Juanda, maka model antara curah hujan Keputih dan kelembapan udara Juanda menghasilkan nilai RMSE paling kecil. Hal tersebut juga berlaku pada pemodelan curah hujan Keputih dengan stasiun Perak 2 maupun Perak 1 yang dapat dilihat pada Lampiran 6 dan Lampiran 7. Kemudian disusul dengan model antara curah hujan Keputih dengan temperatur Juanda dan yang terakhir model antara curah hujan Keputih dengan kecepatan angin Juanda. Hasil pemodelan tersebut didukung oleh hasil eksplorasi data yang telah dilakukan sebelumnya, dimana pada bagian eksplorasi data sebelumnya diketahui bahwa kelembapan udara memiliki nilai korelasi yang paling tinggi diantara temperatur dan kecepatan angin, sehingga model antara curah hujan dengan kelembapan udara baik di Juanda, Perak 2, maupun Perak 1 dapat menghasilkan nilai RMSE dan RMSE* paling kecil diantara model-model lainnya.

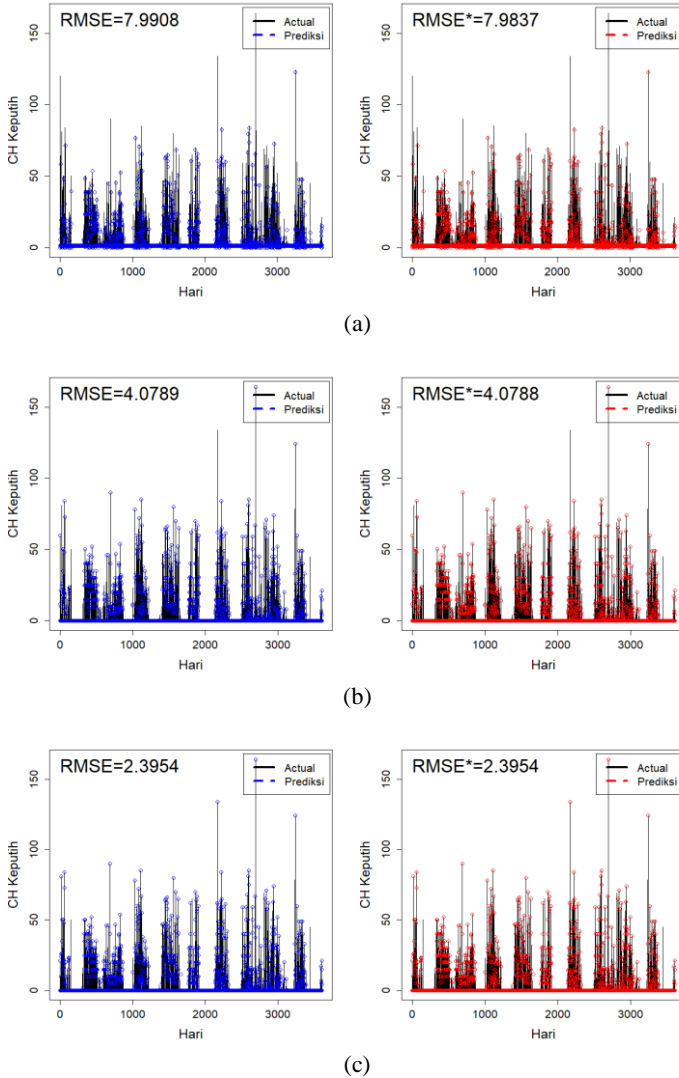
Namun, meskipun demikian nilai RMSE yang dihasilkan dari pemodelan *univariate* di masing-masing stasiun BMKG tersebut masih tergolong kurang baik. Oleh sebab itu, analisis dilanjutkan dengan menambahkan variabel *dummy* pada masing masing model yang telah terbentuk. Dari hasil penambahan *dummy* pada masing-masing model ternyata sangat membantu menurunkan nilai RMSE dan RMSE*. Akan tetapi dengan penambahan variabel *dummy* pun belum dapat menurunkan nilai RMSE secara signifikan. Hal tersebut dapat dilihat dari plot hasil prediksi data *training* untuk masing-masing model diatas yang terdapat pada Gambar 4.13.

Model yang ditawarkan berikutnya adalah dengan menggabungkan seluruh variabel prediktor yaitu temperatur, kelembapan udara, dan kecepatan angin untuk masing-masing stasiun BMKG kemudian dimodelkan dengan curah hujan di Keputih. RMSE yang didapatkan masih berkisar diangka 11. Sehingga model berikutnya adalah dengan mencobakkan variabel *dummy* kedalamnya, hasil yang diperoleh yaitu variabel *dummy* mampu menurunkan RMSE meskipun tidak terlalu signifikan. Sehingga dari pemodelan ν -SVR didapatkan prediksi curah hujan keputih terbaik dengan *input* yaitu gabungan seluruh variabel prediktor klimatologi (temperatur, kelembapan udara, dan kecepatan angin) ditambah *dummy* untuk stasiun Juanda, Perak 2, maupun Perak 1.

b. Pemodelan ν -SVR dengan *Input Lag PACF*

Pada pemodelan dengan *input* variabel prediktor dan penambahan variabel *dummy* pada analisis sebelumnya masih belum mampu menghasilkan model ν -SVR dengan nilai RMSE yang kecil, maka muncul indikasi bahwa curah hujan dipengaruhi oleh curah hujan pada periode atau hari-hari sebelumnya. Hal tersebut didukung oleh hasil eksplorasi data yang menyatakan hubungan antara curah hujan dengan curah hujan sebelumnya lebih tinggi dari pada hubungan antara curah hujan dengan variabel kelembapan udara dan kecepatan angin. Sehingga model selanjutnya yang akan di analisis yaitu memodelkan curah hujan dengan lag-lag yang signifikan pada plot PACF yang terdapat pada subbab 4.5.1 dan penjelasan dalam eksplorasi data sebelumnya. Hasil analisis menyebutkan bahwa penambahan lag yang signifikan secara terus menerus sampai lag tertentu mampu menurunkan nilai RMSE secara signifikan. Mengingat kebutuhan praktis akan hasil analisis prediksi curah hujan ini, maka kombinasi model selanjutnya yang akan dicobakan yaitu dengan menggunakan *input* dari lag-lag PACF yang signifikan pada Tabel 4.4. Namun sebelum melakukan analisis menggunakan lag-lag yang signifikan, terlebih dahulu dilakukan analisis untuk mengetahui pengaruh pertambahan lag yang digunakan didalam model. Berikut merupakan pemodelan ν -SVR untuk memprediksi

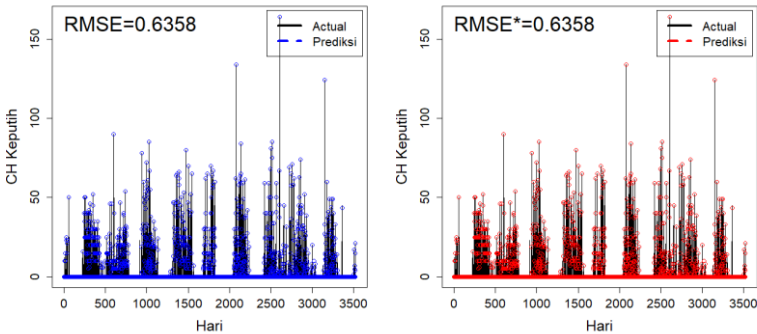
curah hujan di Keputih menggunakan lag-lag yang signifikan pada PACF.



Gambar 4.14 Pengaruh Penambahan Lag pada RMSE (a) $Y_{1,t} = f(Y_{1,t-1} + Y_{1,t-2} + Y_{1,t-3} + Y_{1,t-4} + Y_{1,t-5})$, (b) $Y_{1,t} = f(Y_{1,t-1} + Y_{1,t-2} + Y_{1,t-3} + \dots + Y_{1,t-10})$, (c) $Y_{1,t} = f(Y_{1,t-1} + Y_{1,t-2} + Y_{1,t-3} + \dots + Y_{1,t-15})$

Gambar 4.14 menunjukkan prediksi curah hujan Keputih dari lag-lag yang signifikan, dimana plot prediksi yang berwarna biru merupakan hasil prediksi yang masih mengandung nilai negatif, sedangkan plot prediksi yang berwarna merah merupakan hasil prediksi negatif yang telah ditransformasi menjadi 0 dan dihitung kebaikan modelnya menggunakan $RMSE^*$. Dari Gambar 4.14 maka dapat ditarik kesimpulan bahwa penambahan lag sangat berkontribusi dalam menurunkan nilai $RMSE$. Selain itu dilihat dari plot antara data aktual dan prediksi terlihat bahwa model dengan jumlah lag yang semakin banyak lebih mampu mendekati nilai aktual dan bahkan sangat *fitting* terhadap data aktual serta sangat mampu memprediksi data curah hujan yang sangat tinggi.

Analisis selanjutnya yaitu memprediksi curah hujan Keputih menggunakan ν -SVR berdasarkan *input* lag-lag PACF yang signifikan pada Tabel 4.4. dan didapatkan plot antara data aktual dengan data hasil prediksi pada data *training* sebagai berikut.

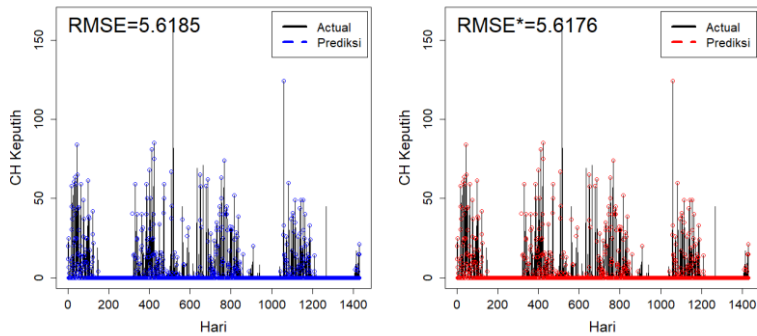


Gambar 4.15 Plot *Training* Pemodelan CH Keputih dengan *Input* lag PACF yang Signifikan

Penggunaan lag-lag yang signifikan pada PACF sebagai *input* dalam memprediksi curah hujan di balai Keputih sangat berkontribusi besar dalam menghasilkan prediksi yang mendekati data aktual pada data *training*, hal tersebut dapat dilihat dari titik-titik berwarna merah yang menunjukkan hasil prediksi sangat mampu mencapai puncak-puncak pada data aktual. Selain itu dari model tersebut dihasilkan nilai $RMSE$ dan $RMSE^*$ yang sangat kecil sekali.

c. Pemodelan ν -SVR dengan *Input Lag ARIMA*

Analisis selanjutnya yaitu memprediksi curah hujan di Keputih menggunakan variabel *input* dari lag-lag signifikan pada model ARIMA yang telah dijabarkan matematis pada subbab 4.5.2 dan didapatkan plot antara data aktual dan data hasil prediksi sebagai berikut.



Gambar 4.16 Plot *Training* Pemodelan CH Keputih dengan *Input lag ARIMA* yang Signifikan

Gambar 4.16 menunjukkan bahwa data hasil prediksi sangat *fitting* mencapai puncak data aktual, hal ini berarti bahwa model yang dibangun dari data *training* sangat baik digunakan untuk memprediksi curah hujan di Keputih. Selain itu jika dibandingkan dengan *input lag PACF* dapat disimpulkan bahwa *input lag PACF* mampu menurunkan nilai RMSE dan menghasilkan nilai prediksi yang lebih mendekati data aktual.

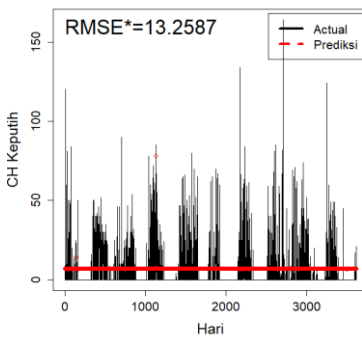
d. Pemodelan ν -SVR dengan *Input PCA*

Kombinasi model yang akan dicoba selanjutnya adalah model antara curah hujan Keputih dengan *predictor* hasil analisis PCA pada subbab 4.4 yaitu PC1X1, PC1X2, PC1X3, dan PC2X3. Pemodelan dilakukan secara *univariate* dan juga bersama sama sekaligus maupun dengan menambahkan variabel dummy. Adapun skenario yang digunakan adalah sebagai berikut.

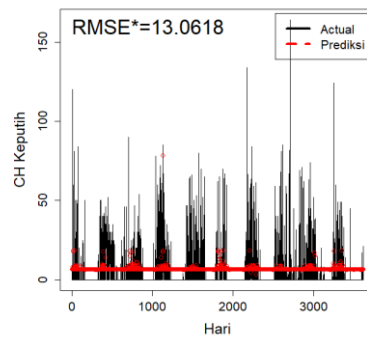
Tabel 4.8 Skenario Pemodelan ν -SVR CH Keputih dengan *Input* PCA

Model	Persamaan
1	$Y_{1,t} = f(PC1X_{1,t})$
2	$Y_{1,t} = f(PC1X_{1,t}, M_t)$
3	$Y_{1,t} = f(PC1X_{2,t})$
4	$Y_{1,t} = f(PC1X_{2,t}, M_t)$
5	$Y_{1,t} = f(PC1X_{3,t}, PC2X_{3,t})$
6	$Y_{1,t} = f(PC1X_{3,t}, PC2X_{3,t}, M_t)$
7	$Y_{1,t} = f(PC1X_{1,t}, PC1X_{2,t}, PC1X_{3,t}, PC2X_{3,t})$
8	$Y_{1,t} = f(PC1X_{1,t}, PC1X_{2,t}, PC1X_{3,t}, PC2X_{3,t}, M_t)$

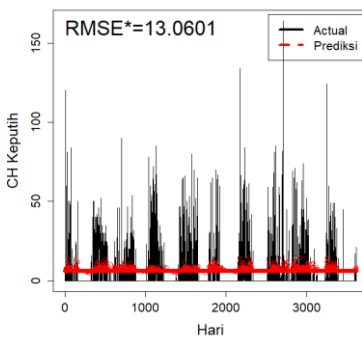
Dari skenario yang terdapat dalam Tabel 4.8 dapat diketahui plot antara data aktual dan prediksi yang ditampilkan pada Gambar 4.17 berikut.



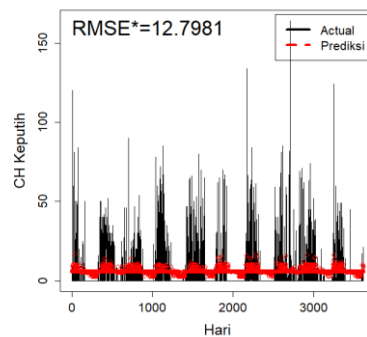
(a)



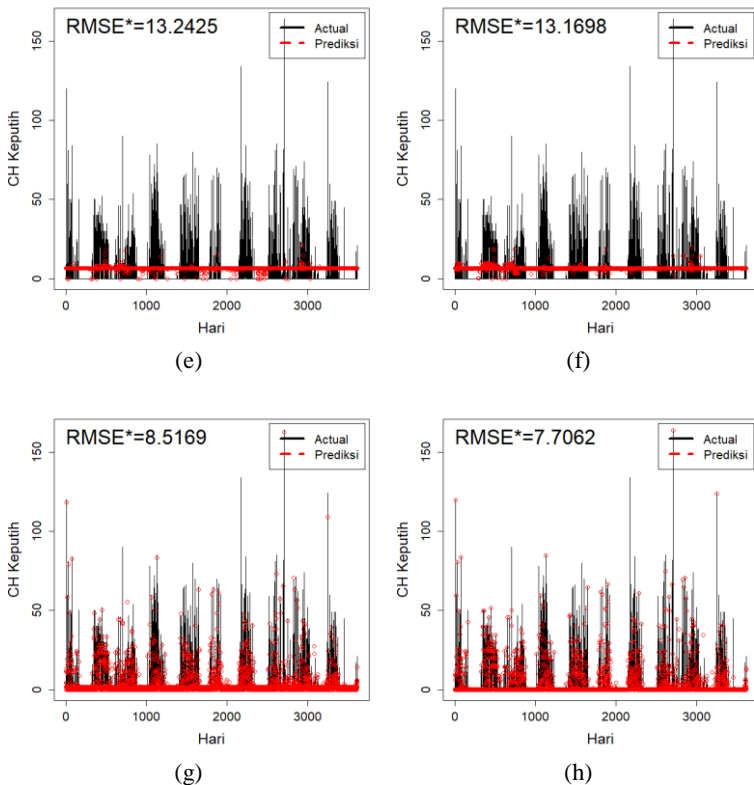
(b)



(c)



(d)



Gambar 4. 17 Plot (a) Model 1, (b) Model 2, (c) Model 3, (d) Model 4, (e) Model 5, (f) Model 6, (g) Model 7, (h) Model 8

Pemodelan secara *univariate* dengan *input* PCA baik dengan menambahkan variabel *dummy* atau tidak tetap belum mampu menghasilkan nilai RMSE yang baik meskipun penambahan variabel *dummy* berkontribusi dalam menurunkan RMSE. Disisi lain pemodelan bersama antara curah hujan Keputih dengan PC1X1, PC1X2, PC1X3, dan PC2X3 sekaligus *dummy* mampu menurunkan RMSE secara signifikan, meskipun tidak sebaik RMSE sebelumnya yang menggunakan *input* dari lag PACF, tapi setidaknya lebih baik daripada pemodelan *univariate* dengan penambahan *dummy*. Untuk lebih jelasnya dapat dilihat penurunan atau pergerakan dari nilai RMSE pada Lampiran 8.

e. Pemodelan ν -SVR dengan *Input* Kombinasi antara Variabel Prediktor, PCA, Lag PACF, dan Lag ARIMA

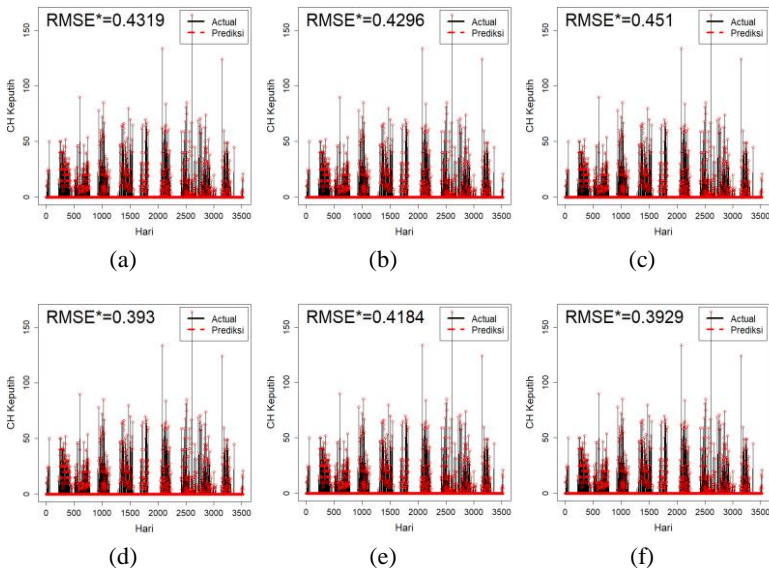
Dari analisis sebelumnya mengenai prediksi curah hujan Keputih pada data *training* menggunakan *input* variabel prediktor, PCA, lag signifikan dari PACF, dan lag signifikan dari ARIMA masih belum mampu menghasilkan nilai RMSE yang cukup baik, meskipun *input* dari lag signifikan PACF sudah mampu mencapai RMSE dibawah nilai 1. Oleh karena itu, model selanjutnya yang akan dianalisis adalah dengan mengkombinasikan ke-empat macam variabel *input* diantaranya mengkombinasikan *input* prediktor dengan lag PACF, *input* prediktor dengan lag ARIMA, *input* PCA dengan lag PACF, dan *input* PCA dengan lag ARIMA. Adapun skenario atau skema pemodelan lengkap yang akan digunakan dapat dilihat pada Tabel 4.9.

Tabel 4.9 Skenario Pemodelan ν -SVR CH Keputih dengan Kombinasi Keempat Macam *Input*

Model	Persamaan
1	$Y_{1,t} = f(X_{11,t}, X_{21,t}, X_{31,t}, Y_{1,t-1}, Y_{1,t-2}, \dots, Y_{1,t-89}, Y_{1,t-100})$
2	$Y_{1,t} = f(X_{11,t}, X_{21,t}, X_{31,t}, M_t, Y_{1,t-1}, Y_{1,t-2}, \dots, Y_{1,t-89}, Y_{1,t-100})$
3	$Y_{1,t} = f(X_{12,t}, X_{22,t}, X_{32,t}, Y_{1,t-1}, Y_{1,t-2}, \dots, Y_{1,t-89}, Y_{1,t-100})$
4	$Y_{1,t} = f(X_{12,t}, X_{22,t}, X_{32,t}, M_t, Y_{1,t-1}, Y_{1,t-2}, \dots, Y_{1,t-89}, Y_{1,t-100})$
5	$Y_{1,t} = f(X_{13,t}, X_{23,t}, X_{33,t}, Y_{1,t-1}, Y_{1,t-2}, \dots, Y_{1,t-89}, Y_{1,t-100})$
6	$Y_{1,t} = f(X_{13,t}, X_{23,t}, X_{33,t}, M_t, Y_{1,t-1}, Y_{1,t-2}, \dots, Y_{1,t-89}, Y_{1,t-100})$
7	$Y_{1,t} = f(X_{11,t}, X_{21,t}, X_{31,t}, Y_{1,t-1}, Y_{1,t-365}, \dots, Y_{1,t-2190}, Y_{1,t-2191})$
8	$Y_{1,t} = f(X_{11,t}, X_{21,t}, X_{31,t}, M_t, Y_{1,t-1}, Y_{1,t-365}, \dots, Y_{1,t-2190}, Y_{1,t-2191})$
9	$Y_{1,t} = f(X_{12,t}, X_{22,t}, X_{32,t}, Y_{1,t-1}, Y_{1,t-365}, \dots, Y_{1,t-2190}, Y_{1,t-2191})$
10	$Y_{1,t} = f(X_{12,t}, X_{22,t}, X_{32,t}, M_t, Y_{1,t-1}, Y_{1,t-365}, \dots, Y_{1,t-2190}, Y_{1,t-2191})$
11	$Y_{1,t} = f(X_{13,t}, X_{23,t}, X_{33,t}, Y_{1,t-1}, Y_{1,t-365}, \dots, Y_{1,t-2190}, Y_{1,t-2191})$
12	$Y_{1,t} = f(X_{13,t}, X_{23,t}, X_{33,t}, M_t, Y_{1,t-1}, Y_{1,t-365}, \dots, Y_{1,t-2190}, Y_{1,t-2191})$

$$\begin{aligned}
 13 \quad & Y_{1,t} = f(PC1X_{1,t}, PC1X_{2,t}, PC1X_{3,t}, PC2X_{3,t}, Y_{1,t-1}, \\
 & \quad Y_{1,t-2}, \dots, Y_{1,t-89}, Y_{1,t-100}) \\
 14 \quad & Y_{1,t} = f(PC1X_{1,t}, PC1X_{2,t}, PC1X_{3,t}, PC2X_{3,t}, M_t, Y_{1,t-1}, \\
 & \quad Y_{1,t-2}, \dots, Y_{1,t-89}, Y_{1,t-100}) \\
 15 \quad & Y_{1,t} = f(PC1X_{1,t}, PC1X_{2,t}, PC1X_{3,t}, PC2X_{3,t}, Y_{1,t-1}, \\
 & \quad Y_{1,t-365}, \dots, Y_{1,t-2190}, Y_{1,t-2191}) \\
 16 \quad & Y_{1,t} = f(PC1X_{1,t}, PC1X_{2,t}, PC1X_{3,t}, PC2X_{3,t}, M_t, Y_{1,t-1}, \\
 & \quad Y_{1,t-365}, \dots, Y_{1,t-2190}, Y_{1,t-2191})
 \end{aligned}$$

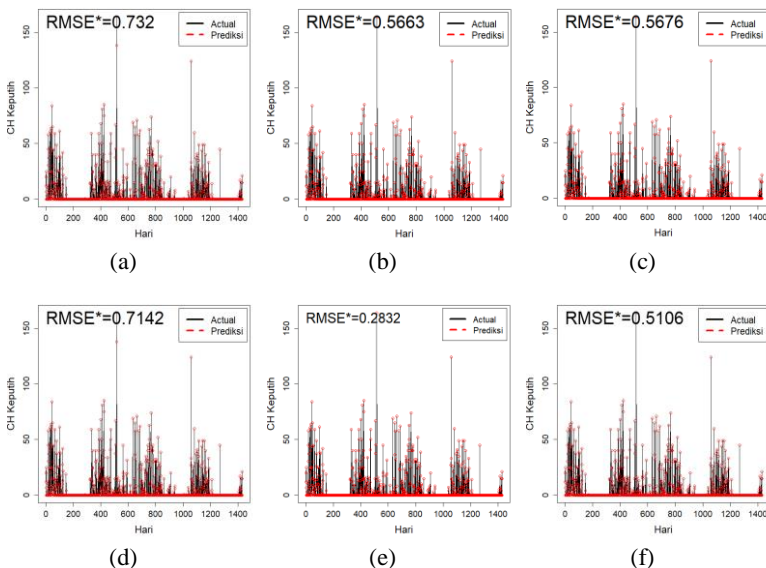
Skenario yang ditampilkan pada Tabel 4.9 perlu dilakukan optimasi parameter terlebih dahulu dengan 400 kombinasi parameter untuk mendapatkan parameter-parameter yang optimum dan kemudian dilakukan pemodelan serta prediksi pada data *training* menggunakan ν -SVR dan didapatkan hasil prediksi untuk model 1 sampai model 6 sebagai berikut.



Gambar 4.18 Plot *Training* Pemodelan CH Keputih dengan Kombinasi *Input* Prediktor dan lag PACF (a) Model 1, (b) Model 3, (c) Model 5, (d) Model 2, (e) Model 4, dan (f) Model 6

Prediksi curah hujan di balai Keputih menggunakan kombinasi *input* antara variabel prediktor dengan lag-lag yang signifikan pada PACF menghasilkan kesimpulan bahwa

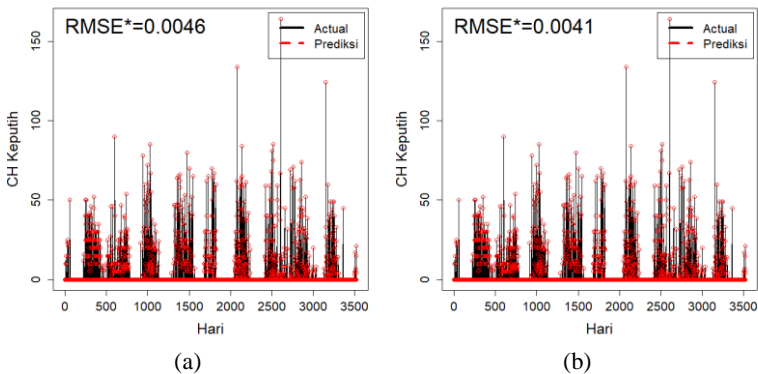
penambahan variabel lag PACF yang signifikan terhadap *input* prediktor temperatur, kelembapan udara, dan kecepatan angin mampu menghasilkan nilai RMSE yang sangat kecil dengan plot data prediksi dapat mendekati data aktual, hal tersebut dapat dilihat dari titik-titik merah pada Gambar 4.18 yang sangat mengikuti pola garis hitam dan mampu mencapai puncak-puncak garis hitam. Selain itu dari Gambar 4.18 dapat diketahui model yang terbaik untuk memprediksi curah hujan di keputih adalah model 6 yaitu model yang menggunakan semua variabel-variabel klimatologi (temperatur, kelembapan udara, dan kecepatan angin) di stasiun Perak 1, variabel *dummy* dan lag yang signifikan pada PACF sebagai *input*. Nilai RMSE yang dihasilkan dari model tersebut yaitu sebesar 0,3929. Sedangkan hasil prediksi untuk kombinasi *input* prediktor dan lag dari ARIMA ditunjukkan pada Gambar 4.19 berikut.

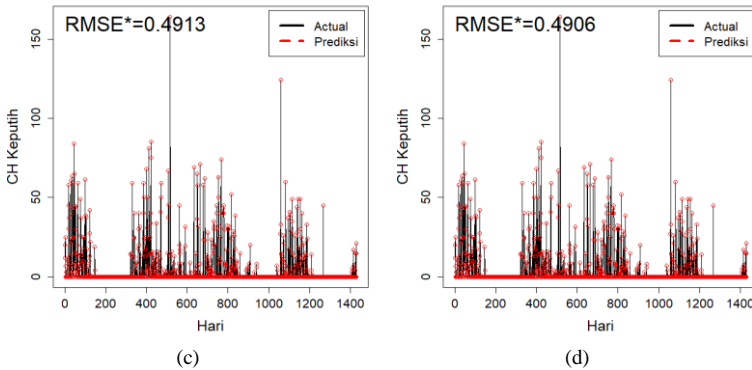


Gambar 4.19 Plot *Training* Pemodelan CH Keputih dengan Kombinasi *Input* Prediktor dan lag ARIMA (a) Model 7, (b) Model 9, (c) Model 11, (d) Model 8, (e) Model 10, dan (f) Model 12

Hasil-hasil prediksi dari kombinasi model pada Gambar 4.19 memberikan informasi bahwa penambahan lag-lag ARIMA yang signifikan pada model dan digabungkan dengan variabel prediktor klimatologi dapat menghasilkan prediksi yang lebih akurat pada data *training*. Hal tersebut dapat dilihat dari nilai RMSE yang mampu mencapai nilai dibawah 1. Model terbaik untuk memprediksi curah hujan di Keputih yang dapat diperoleh dari kombinasi model ini adalah model dengan variabel *input* seluruh variabel klimatologi di stasiun Perak 2 dengan penambahan lag-lag yang signifikan pada ARIMA dan juga penambahan variabel *dummy*. Adapun nilai RMSE yang didapatkan dari model terbaik tersebut yaitu sebesar 0,2832. Kesimpulan yang dapat diambil dari Gambar 4.18 dan 4.19 yaitu prediksi curah hujan di balai Keputih dapat dilakukan menggunakan variabel-variabel klimatologi di stasiun Perak 2, sebab dari hasil analisis yang telah dilakukan dihasilkan bahwa pemodelan curah hujan keputih dengan stasiun BMKG Perak 2 mampu memberikan nilai RMSE yang paling kecil dibandingkan stasiun BMKG Perak 1 dan Juanda.

Berikut merupakan hasil prediksi dari kombinasi *input* PCA, lag signifikan PACF, dan lag signifikan ARIMA.



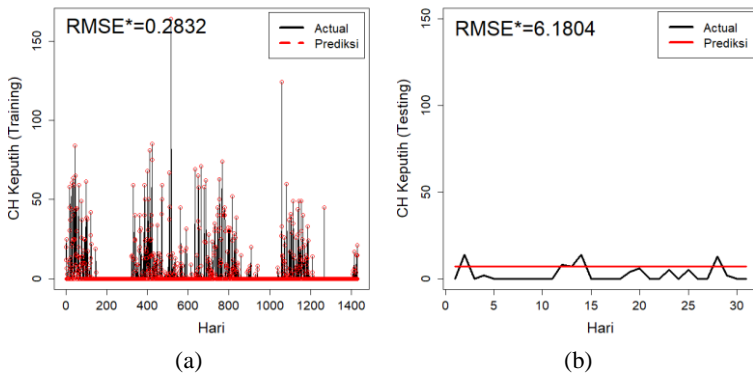


Gambar 4.20 Plot *Training* Pemodelan CH Keputih dengan Kombinasi *Input* Prediktor dan lag ARIMA (a) Model 13, (b) Model 14, (c) Model 15, dan (d) Model 16

Prediksi curah hujan Keputih menggunakan kombinasi *input* PCA, lag PACF dan lag ARIMA dapat menghasilkan nilai RMSE yang sangat kecil seperti halnya kombinasi model-model sebelumnya. Namun, untuk model terbaik yang akan digunakan untuk memprediksi curah hujan di balai Keputih adalah kombinasi model dari prediktor dengan lag ARIMA, sebab kombinasi model dengan *input* PCA menghasilkan prediksi yang *overfitting*, dimana sangat tepat memprediksi di data *training* tetapi tidak cukup baik untuk memprediksi data *testing*.

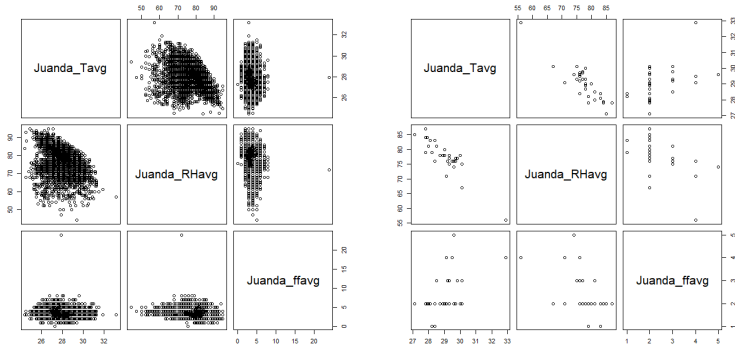
f. Model Terbaik Prediksi Curah Hujan di Keputih Menggunakan ν -SVR

Berdasarkan analisis-*analisis* sebelumnya diketahui model terbaik untuk memprediksi curah hujan di balai Keputih adalah dengan menggunakan variabel *input* temperatur Perak 2, kelembapan udara Perak 2, kecepatan angin Perak 2, *dummy*, dan lag-lag yang signifikan pada ARIMA. Adapun hasil prediksi curah hujan di Keputih menggunakan model ν -SVR terbaik pada data *training* dan *testing* ditampilkan pada Gambar 4.21 berikut.

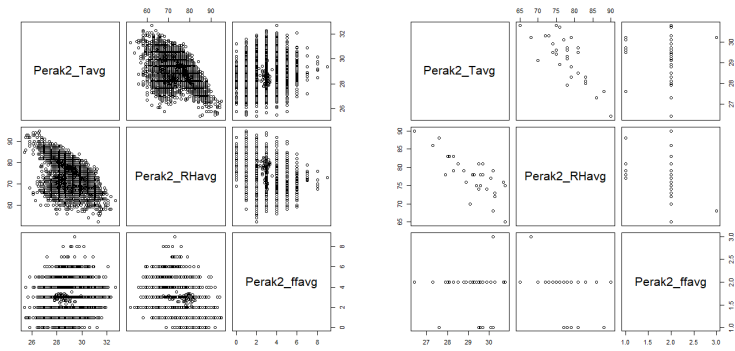


Gambar 4.21 Model Terbaik untuk Prediksi Curah Hujan Keputih (a) *Training* dan (b) *Testing*

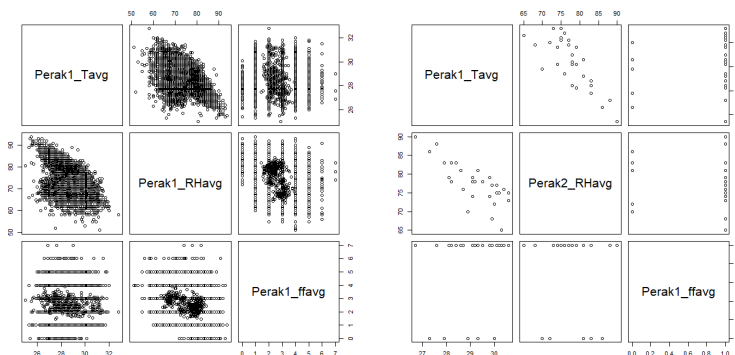
Pada Gambar 4.21 dapat dilihat bahwa hasil prediksi dari model terbaik pada data *training* sangat mampu mengikuti pola data aktual dan banyak titik-titik prediksi berwarna merah yang *fitting* sampai titik-titik puncak data aktual. Sedangkan jika dilihat hasil prediksi pada data *testing* menunjukkan nilai prediksi yang tidak cukup baik yaitu bernilai hampir konstan. Sehingga dapat diindikasikan terjadi kasus *overfitting* pada model terbaik menggunakan metode ν -SVR. Kasus *overfitting* yang terjadi dapat disebabkan oleh berbagai macam hal, salah satunya yaitu kurang tepatnya pembagian data *training* dan data *testing* dan mengakibatkan karakteristik variabel-variabel prediktor pada data *training* tidak sama dengan *testing*. Oleh karena itu perlu dilakukan peninjauan kembali mengenai data *training* dan *testing* yang digunakan untuk mengetahui karakteristik variabel-variabel prediktor pada keduanya. Berikut merupakan *matrix plot* variabel-variabel prediktor yang digunakan dalam analisis pada data *training* dan *testing*.



(a)



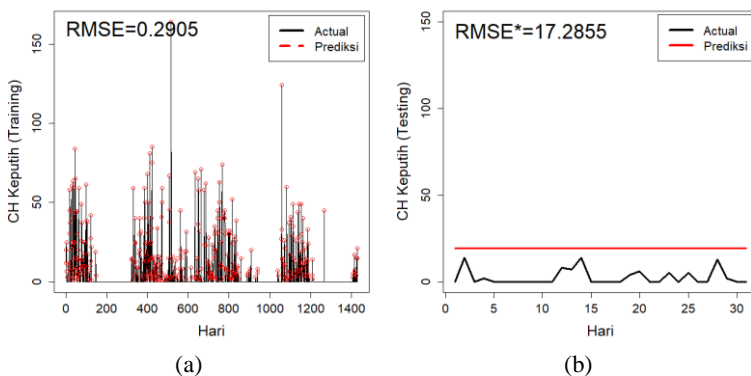
(b)



(c)

Gambar 4.22 Karakteristik Variabel Prediktor pada *Training* (Kiri) dan *Testing* (Kanan) di Stasiun (a) Juanda, (b) Perak 2, dan (c) Perak 1

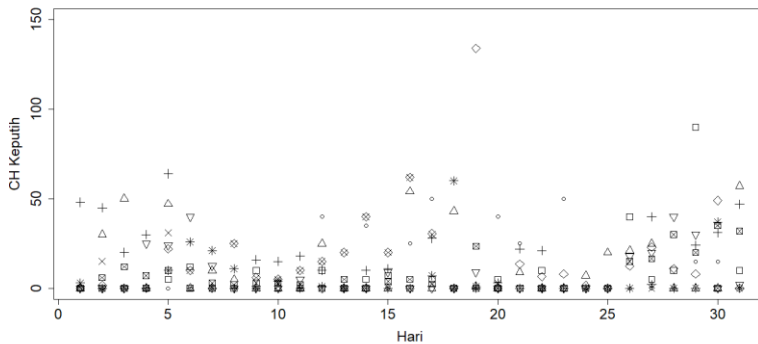
Berdasarkan Gambar 4.22 terlihat jelas bahwa karakteristik variabel-variabel prediktor pada data *testing* hampir sama dengan karakteristik variabel-variabel prediktor pada data *training*. Dari *matrix plot* diatas dapat dikatakan bahwa sebenarnya data *testing* yang digunakan secara segi karakteristik sudah baik digunakan. Sehingga muncul indikasi lain dari penyebab konstannya hasil prediksi pada data *testing* yaitu karena data curah hujan mengandung banyak nilai 0 dan bersifat *chaotic* atau sangat sulit diprediksi. Kemudian solusi lain yang ditawarkan berdasarkan saran dari Rob J. Hyndman sebagai guru besar *forecasting* yaitu mengganti nilai 0 yang banyak tersebut dengan NA dan kemudian dimodelkan kembali. Berikut hasil analisis ulang dengan mengganti nilai 0 menjadi NA.



Gambar 4.23 Prediksi Model Terbaik dengan NA untuk Curah Hujan Keputih
(a) *Training* dan (b) *Testing*

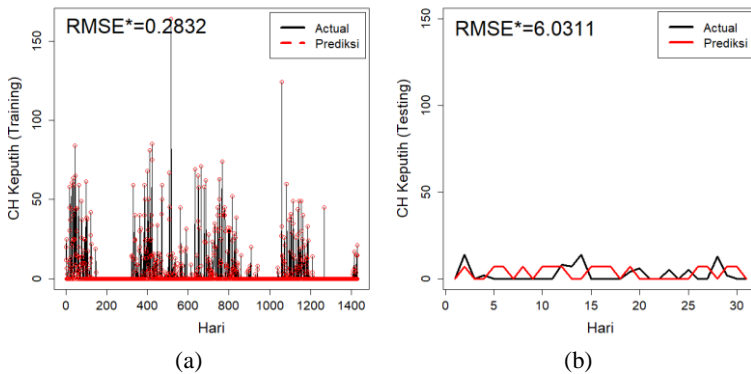
Solusi yang ditawarkan dengan mengganti nilai 0 menjadi NA dalam analisis menghasilkan prediksi pada data *testing* yang tidak jauh berbeda ketika menggunakan nilai 0 yaitu bernilai konstan. Dapat dilihat pada Gambar 4.23 prediksi pada data *training* sudah sangat bagus dengan nilai RMSE yang sangat kecil, namun pada data *testing* bernilai konstan untuk 31 periode kedepan. Selain itu, terjadi peningkatan RMSE pada data *training* dari semula yang bernilai 0,2832 menjadi 0,2905. Jika ditinjau ulang kembali data *testing* (Desember 2018) yang digunakan, diketahui bahwa curah hujan pada bulan Desember 2018 memiliki banyak nilai 0, hal

tersebut juga dapat dilihat dari histori curah hujan pada bulan Desember ditahun-tahun sebelumnya. Berikut merupakan *boxplot* curah hujan bulan Desember 2009-2017 jika dilihat per hari-nya.



Gambar 4.24 *Boxplot* Curah Hujan Bulan Desember 2009-2017

Boxplot diatas menunjukkan bahwa setiap hari di bulan Desember selama 9 tahun terakhir selalu terjadi hujan setidaknya sekali. Sehingga solusi atau pendekatan terakhir yang diusulkan untuk mengatasi prediksi pada data *testing* yang konstan yaitu dengan cara menggunakan informasi atau histori dari rata-rata curah hujan bulan Desember tahun sebelum-sebelumnya. Jadi data *testing* yang digunakan adalah data *testing* yang telah mendapatkan *prior knowledge*. Dimana *prior knowledge* ini berguna sebagai *prior* atau acuan pada hari tersebut akan terjadi hujan atau tidak yang dasarnya dari data histori bulan Desember tahun-tahun sebelumnya. Adapun *prior knowledge* yang akan digunakan yaitu apabila terjadinya hujan kurang dari 5 pada tanggal yang sama di bulan desember selama 9 tahun sebelumnya maka otomatis akan bernilai 0. Tanpa informasi tambahan dari *prior knowledge* maka akan susah atau sulit melakukan prediksi pada data *testing* secanggih apapun hasil prediksi pada data *training*. Berikut merupakan hasil prediksi dengan menggunakan informasi tambahan dari *prior knowledge*.



Gambar 4.25 Model Terbaik untuk Prediksi Curah Hujan Keputih menggunakan Informasi *Prior Knowledge* (a) *Training* dan (b) *Testing*

Prediksi curah hujan di balai Keputih menggunakan informasi tambahan dari *prior knowledge* menghasilkan nilai prediksi pada data *testing* yang lebih baik daripada model sebelumnya yang tanpa menggunakan *prior knowledge*. Penambahan *prior knowledge* mampu menurunkan nilai RMSE pada data *testing* yang semula bernilai 6,1804 menjadi 6,0311. Jadi dari keseluruhan analisis yang telah dilakukan mengenai prediksi curah hujan di balai Keputih menggunakan metode *v*-SVR dapat ditarik kesimpulan bahwa model terbaik yang dapat digunakan untuk memprediksi curah hujan di balai Keputih adalah model yang menggunakan *input* temperatur di Perak 2, kelembapan udara di Perak 2, kecepatan angin di Perak 2, *dummy*, dan lag yang signifikan pada model ARIMA. Dimana parameter-parameter yang digunakan dalam model terbaik ini antara lain yaitu nilai $\gamma = 2,5$, $cost = 100$, $\nu = 0,3$, jenis kernel = radial dan jumlah dari *support vectors* yang digunakan yaitu 1391 dari 1430 observasi pengamatan. Selain itu, model terbaik yang dihasilkan juga sesuai dengan kondisi wilayah Surabaya, dimana letak balai Keputih memiliki jarak yang lebih dekat dengan stasiun BMKG Perak 2 (17 km) dari pada stasiun Perak 1 (18 km) dan Juanda (18 km). Analisis tambahan mengenai alternatif lain yang dapat dilakukan untuk mengatasi hasil prediksi yang konstan yaitu dengan cara menghilangkan observasi tanggal 29 Februari dan didapatkan hasil bahwa penghapusan observasi

setiap tanggal 29 Februari tersebut tidak berpengaruh signifikan terhadap hasil prediksi. Analisis lebih lengkapnya dapat dilihat pada Lampiran 21.

4.6.2 Analisis *Neural Network*

Pada analisis prediksi menggunakan *Neural Network* kombinasi model yang digunakan atau skenario yang digunakan tidak jauh berbeda dengan analisis prediksi menggunakan ν -SVR pada subbab sebelumnya. Jika pada SVR yang dioptimasi adalah parameter γ , $cost$, dan ν maka pada NN yang harus dioptimasi adalah jumlah *hidden layer* dan jumlah *neuron* tiap *hidden layer* menggunakan konsep *trial and error*. Jumlah *hidden layer* yang akan digunakan yaitu 1, 2, dan 3. Dimana dengan 1 *hidden layer* jumlah *neuron* yang akan digunakan diantaranya 1, 2, 3, 4, 5, dan 10. Kombinasi *neuron* pada 2 *hidden layer* diantaranya yaitu (1,1), (1,2), (2,1), (2,2), (3,2), (4,3), (5,4), dan (10,5). Sedangkan kombinasi *neuron* pada 3 *hidden layer* diantaranya yaitu (2,1,1), (2,1,2), (2,2,1), dan (2,2,2). Adapun fungsi aktivasi yang akan digunakan adalah fungsi aktivasi *tanh*, hal tersebut didasarkan pada studi pendahuluan yang telah dilakukan oleh peneliti. Kesimpulan yang didapatkan dari studi pendahuluan yaitu bahwa dengan menggunakan kombinasi *hidden layer* dan *neuron* yang sama, fungsi aktivasi *tanh* memberikan nilai RMSE dan RMSE* lebih kecil dibandingkan fungsi aktivasi *logistic*. Oleh karena itu pada penelitian ini, fungsi aktivasi yang akan digunakan adalah fungsi aktivasi *tanh* dengan 5 replikasi, untuk kemudian dicari model terbaik dari 5 replikasi tersebut.

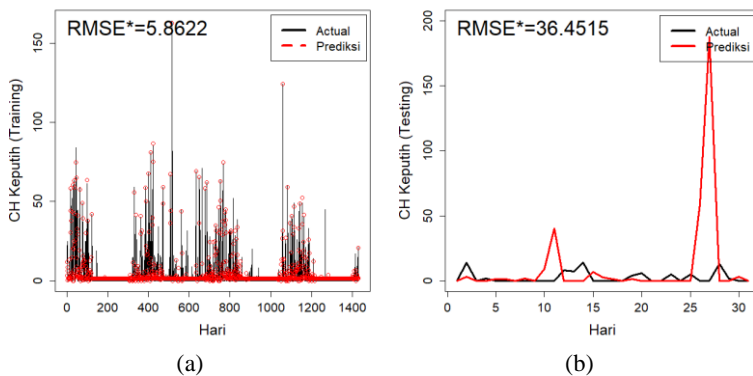
Berikut merupakan hasil analisis menggunakan NN dari berbagai skenario model yang digunakan untuk memprediksi curah hujan di Keputih.

Tabel 4.10 Hasil Analisis *Neural Network*

No	Model	Hidden/ Neuron opt	Best Rep ke-	RMSE*
1	$Y_{1,t} = f(X_{11,t})$	2 10 5	2	13.028
2	$Y_{1,t} = f(X_{11,t}, M_t)$	1 10	2	12.198
3	$Y_{1,t} = f(X_{21,t})$	2 3 2	2	12.252
4	$Y_{1,t} = f(X_{21,t}, M_t)$	3 2 2 1	2	12.080
5	$Y_{1,t} = f(X_{31,t})$	3 2 1 1	3	13.101
6	$Y_{1,t} = f(X_{31,t}, M_t)$	3 2 1 1	5	12.584
7	$Y_{1,t} = f(X_{11,t}, X_{21,t}, X_{31,t})$	2 3 2	1	11.965
8	$Y_{1,t} = f(X_{11,t}, X_{21,t}, X_{31,t}, M_t)$	2 10 5	4	11.663
9	$Y_{1,t} = f(X_{11,t}, X_{21,t}, X_{31,t}, M_t, Y_{1,t-1}, \dots, Y_{1,t-100})$	2 10 5	3	7.885
10	$Y_{1,t} = f(X_{12,t}, X_{22,t}, X_{32,t}, M_t, Y_{1,t-1}, \dots, Y_{1,t-100})$	2 10 5	4	8.084
11	$Y_{1,t} = f(X_{13,t}, X_{23,t}, X_{33,t}, M_t, Y_{1,t-1}, \dots, Y_{1,t-100})$	2 10 5	1	8.128
12	$Y_{1,t} = f(PC1X_{1,t}, PC1X_{2,t}, PC1X_{3,t}, PC2X_{3,t}, M_t)$	2 10 5	5	7.515
13	$Y_{1,t} = f(X_{11,t}, X_{21,t}, X_{31,t}, M_t, Y_{1,t-1}, \dots, Y_{1,t-2191})$	2 10 5	1	5.862
14	$Y_{1,t} = f(X_{12,t}, X_{22,t}, X_{32,t}, M_t, Y_{1,t-1}, \dots, Y_{1,t-2191})$	2 10 5	5	6.28
15	$Y_{1,t} = f(X_{13,t}, X_{23,t}, X_{33,t}, M_t, Y_{1,t-1}, \dots, Y_{1,t-2191})$	2 10 5	2	5.886

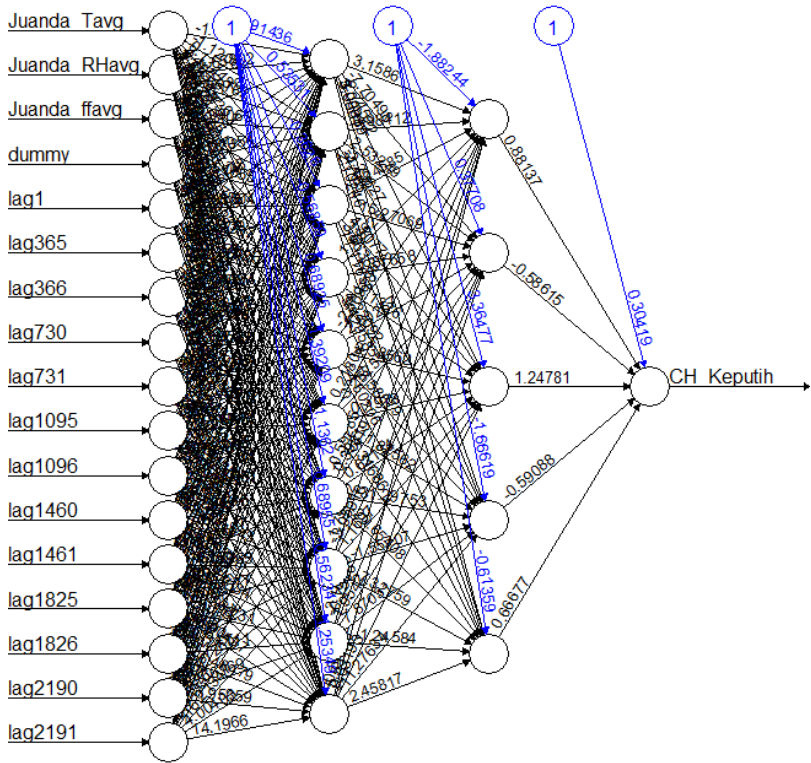
Berdasarkan hasil analisis pada Tabel 4.10 menggunakan skenario yang sama dengan analisis ν -SVR sebelumnya dihasilkan kesimpulan bahwa penambahan lag-lag curah hujan yang berpengaruh signifikan pada PACF dan ARIMA, penambahan variabel *dummy* untuk membedakan bulan-bulan basah dan bulan-bulan kering, serta variabel prediktor hasil analisis PCA mampu menghasilkan nilai RMSE dan RMSE* paling kecil dibandingkan model-model lainnya. Adapun plot antara data aktual dan data hasil prediksi untuk seluruh model yang dicobakkan dapat dilihat pada Lampiran 9. Informasi yang dapat diperoleh dari Lampiran 9 yaitu

bahwa prediksi curah hujan di balai keputih menggunakan *input* variabel prediktor belum mampu memberikan nilai RMSE yang baik dan belum dapat menghasilkan nilai prediksi yang mendekati nilai aktual, sedangkan prediksi dengan *input* kombinasi dari lag yang signifikan ARIMA dan variabel prediktor mampu menghasilkan nilai prediksi pada *training* yang sangat mendekati aktual, model mampu menangkap data-data yang *ekstrem* sangat tepat. Sehingga model yang terbaik menggunakan NN untuk prediksi curah hujan di wilayah Keputih adalah model dengan *input* variabel prediktor di Juanda dan penambahan lag-lag signifikan ARIMA. Dimana spesifikasi dari model terbaik yang digunakan yaitu menggunakan fungsi aktivasi *tanh* dengan 2 *hidden layer* yang masing masing *hidden layer* terdiri dari 10 neuron dan 5 neuron serta direplikasi sebanyak 5 kali. Berikut merupakan plot data *training* dan *testing* untuk model terbaik NN.



Gambar 4.26 Model Terbaik CH Keputih dengan NN (a) *Training* dan (b) *Testing*

Hasil prediksi menggunakan model terbaik menunjukkan bahwa pada data *testing* nilai prediksi sangat bagus dan mampu mendekati nilai data aktual. Hal tersebut dapat terlihat jelas dari plot diatas terdapat titik titik merah yang mampu mencapai puncak. Sedangkan pada data *testing* terdapat beberapa prediksi yang sangat *outlier* atau jauh dari data aktual. Berikut merupakan arsitektur jaringan NN untuk model terbaik dengan *best* replikasi ke-1.



Gambar 4.27 Arsitektur Jaringan NN dari Model Terbaik Prediksi Curah Hujan Keputih

Berdasarkan nilai estimasi parameter model NN diatas maka dapat dituliskan persamaan dari NN yang terbentuk adalah sebagai berikut :

$$Y_{1,t}^* = 0,304 + 0,881f_1^{h_2} (.) - 0,586f_2^{h_2} (.) + 1,248f_3^{h_2} (.) - 0,591f_4^{h_2} (.) + 0,667f_5^{h_2} (.)$$

dengan

$$Y_{1,t}^* = \frac{Y_t - Y_{\min}}{Y_{\max} - Y_{\min}}$$

$$f_1^{h_2}(\cdot) = \tanh(-1,882 + 3,158 f_1^{h_1}(\cdot) + 0,984 f_2^{h_1}(\cdot) + \dots - 0,852 f_{10}^{h_1}(\cdot))$$

$$f_2^{h_2}(\cdot) = \tanh(0,377 - 7,705 f_1^{h_1}(\cdot) + 2,532 f_2^{h_1}(\cdot) + \dots + 5,531 f_{10}^{h_1}(\cdot))$$

$$f_3^{h_2}(\cdot) = \tanh(3,365 - 2,745 f_1^{h_1}(\cdot) - 3,296 f_2^{h_1}(\cdot) + \dots + 0,632 f_{10}^{h_1}(\cdot))$$

$$f_4^{h_2}(\cdot) = \tanh(-1,666 + 3,003 f_1^{h_1}(\cdot) - 4,165 f_2^{h_1}(\cdot) + \dots - 11,276 f_{10}^{h_1}(\cdot))$$

$$f_5^{h_2}(\cdot) = \tanh(-0,613 + 5,086 f_1^{h_1}(\cdot) + 2,102 f_2^{h_1}(\cdot) + \dots + 2,458 f_{10}^{h_1}(\cdot))$$

dengan

$$f_1^{h_1}(\cdot) = \tanh(-1,914 X_{11} - 1,869 X_{21} + \dots + 0,376 Y_{t-2191})$$

$$f_2^{h_1}(\cdot) = \tanh(0,535 X_{11} - 1,121 X_{21} + \dots - 1,007 Y_{t-2191})$$

.....

$$f_{10}^{h_1}(\cdot) = \tanh(1,253 X_{11} + 4,278 X_{21} + \dots + 14,197 Y_{t-2191})$$

4.6.3 Membandingkan Metode ν -SVR dan NN untuk Prediksi Curah Hujan di Keputih

Berdasarkan hasil uraian diatas, masing-masing metode ν -SVR dan NN memiliki model terbaik dengan RMSE yang berbeda. Perbandingan antara kedua metode tersebut dapat dilihat pada Tabel 4.11.

Tabel 4.11 Perbandingan Metode ν -SVR dan NN pada data *Testing* Curah Hujan Keputih

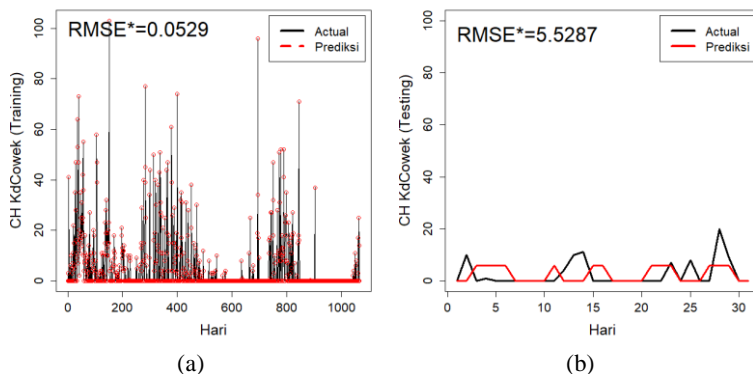
<i>Input</i>	<i>Metode</i>	RMSE <i>Training</i>	RMSE* <i>Testing</i>
$Y_{1,t} = f(X_{12,t}, X_{22,t}, X_{32,t}, M_t, Y_{1,t-1}, Y_{1,t-365}, Y_{1,t-366}, \dots, Y_{1,t-2190}, Y_{1,t-2191})$	ν -SVR	0,2832	6,0311
$Y_{1,t} = f(X_{11,t}, X_{21,t}, X_{31,t}, M_t, Y_{1,t-1}, Y_{1,t-365}, Y_{1,t-366}, \dots, Y_{1,t-2190}, Y_{1,t-2191})$	NN	5,8622	36,4515

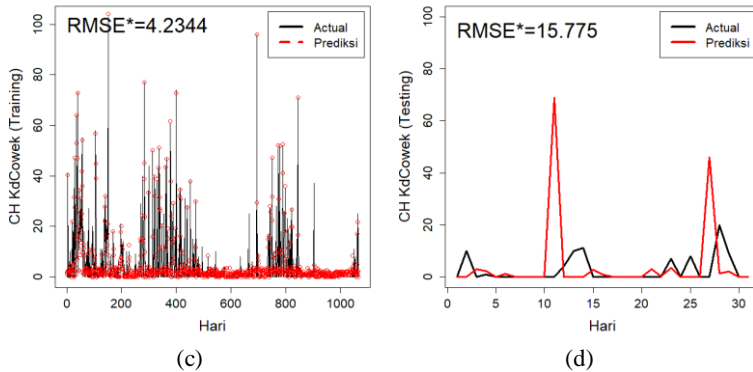
Dari Tabel 4.11 maka dapat diambil kesimpulan bahwa metode ν -SVR lebih baik digunakan untuk memprediksi curah hujan di balai Keputih daripada metode NN, hal tersebut dikarenakan metode ν -SVR memiliki nilai RMSE dan RMSE* lebih kecil daripada NN pada saat di data *training* maupun *testing*. Selain itu keunggulan dari metode ν -SVR juga didukung oleh Gambar 4.25 dan Gambar 4.26 mengenai plot antara data aktual dan data hasil prediksi, dimana Gambar 4.25 yang merujuk pada

metode ν -SVR memiliki hasil prediksi yang sangat mendekati data aktual dan mampu mencapai data-data aktual yang ekstrem atau tinggi. Sehingga metode ν -SVR dipilih sebagai metode terbaik untuk memprediksi curah hujan di balai Keputih. Hasil prediksi curah hujan Keputih untuk periode Desember 2018 dapat dilihat pada Gambar 4.25 (b), dimana dari hasil prediksi tersebut diketahui bahwa curah hujan tertinggi yang dapat terjadi di bulan Desember 2018 adalah sebesar 7,004366 mm.

4.7 Analisis Curah Hujan di Balai Kedung Cowek

Analisis curah hujan di balai Kedung Cowek dilakukan dengan pola pikir dan cara yang sama seperti analisis curah hujan di balai Keputih. Dimana model terbaik untuk prediksi curah hujan di Kedung Cowek menggunakan metode ν -SVR adalah model dengan menggunakan *input* variabel klimatologi temperatur, kelembapan udara, kecepatan angin di stasiun Perak 2. Sedangkan model terbaik untuk prediksi curah hujan di Kedung Cowek menggunakan metode NN yaitu dengan *input* seluruh variabel klimatologi di stasiun Perak 2. Berikut merupakan hasil prediksi menggunakan kedua metode *machine learning*.



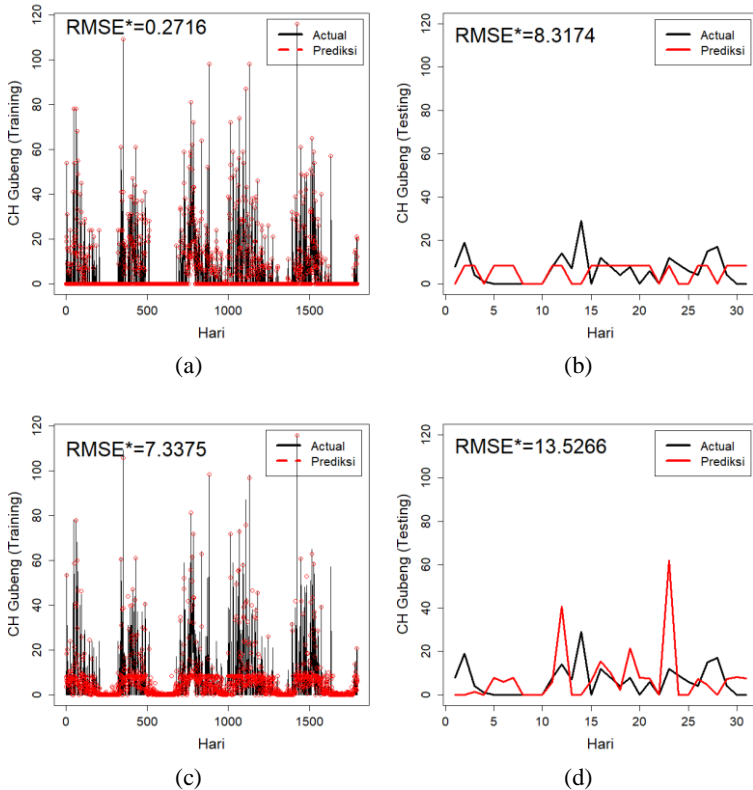


Gambar 4.28 Model Terbaik Prediksi Curah Hujan Kedung Cowek (a) *Training* SVR, (b) *Testing* SVR, (c) *Training* NN, dan (d) *Testing* NN

Prediksi menggunakan metode ν -SVR sangat baik digunakan daripada metode NN untuk memprediksi curah hujan di balai Kedung Cowek, sebab metode ν -SVR mampu menghasilkan nilai RMSE terkecil baik pada data *training* maupun pada data *testing*. Dari hasil prediksi untuk bulan Desember 2019 (31 hari) menggunakan ν -SVR didapatkan prediksi curah hujan tertinggi sebesar 5,895104 mm.

4.8 Analisis Curah Hujan di Balai Gubeng

Prediksi Curah Hujan untuk balai Gubeng melalui proses dan langkah-langkah yang sama seperti balai Keputih dan Kedung Cowek. Hasil terbaik yang diperoleh dari kedua metode tersebut yaitu model dengan *input* kombinasi seluruh variabel klimatologi, *dummy*, dan lag-lag yang signifikan pada ARIMA. Perbedaannya adalah metode ν -SVR menghasilkan model terbaik ketika menggunakan *input* variabel klimatologi dari stasiun Perak 2, sedangkan metode NN dari Perak 1. Berikut hasil prediksi yang didapat dari metode ν -SVR dan NN

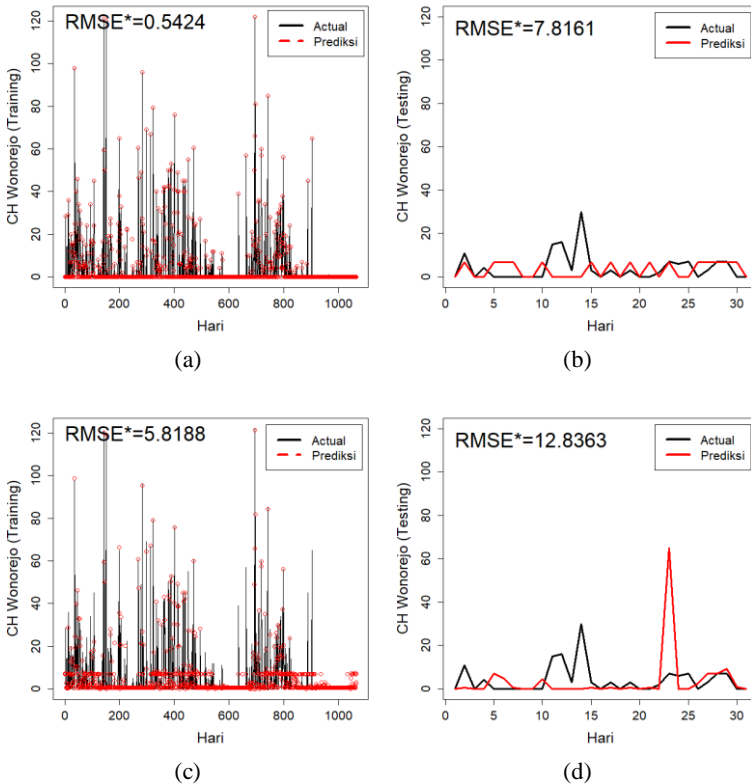


Gambar 4.29 Model Terbaik Prediksi Curah Hujan Gubeng (a) *Training SVR*, (b) *Testing SVR*, (c) *Training NN*, dan (d) *Testing NN*

Pada data *training* metode ν -SVR memiliki hasil prediksi dengan nilai RMSE lebih kecil dibandingkan metode NN, begitu pula untuk hasil prediksi pada data *testing*. Akan tetapi meskipun hasil dari metode ν -SVR dipilih sebagai model terbaik, prediksi pada data *testing* masih cenderung bernilai konstan dibandingkan dengan metode NN yang cenderung *fluktuatif* dengan nilai RMSE yang cukup besar. Dikarenakan pemilihan metode terbaik dilihat dari nilai RMSE yang kecil maka metode ν -SVR dipilih sebagai metode yang layak untuk memprediksi curah hujan di balai Gubeng. Adapun hasil prediksi curah hujan tertinggi yang akan terjadi pada bulan Desember 2018 adalah sebesar 8,327121 mm.

4.9 Analisis Curah Hujan di Balai Wonorejo

Penggunaan metode ν -SVR untuk memprediksi curah hujan di balai Wonorejo memberikan hasil prediksi terbaik jika menggunakan *input* variabel –variabel klimatologi dari stasiun Perak 1. Namun berbeda halnya dengan metode NN yang akan memberikan medel terbaik jika menggunakan input dari variabel-variabel klimatologi di stasiun Juanda. Berikut model terbaik dari kedua metode ν -SVR dan NN untuk memprediksi curah hujan di balai Wonorejo menggunakan proses dan langkah-langkah yang sama seperti balai Keputih.

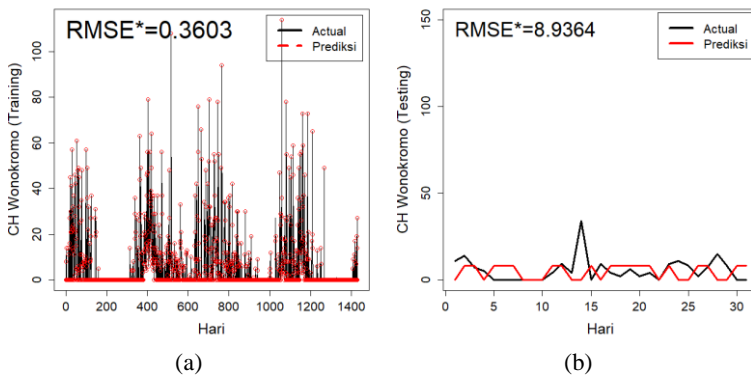


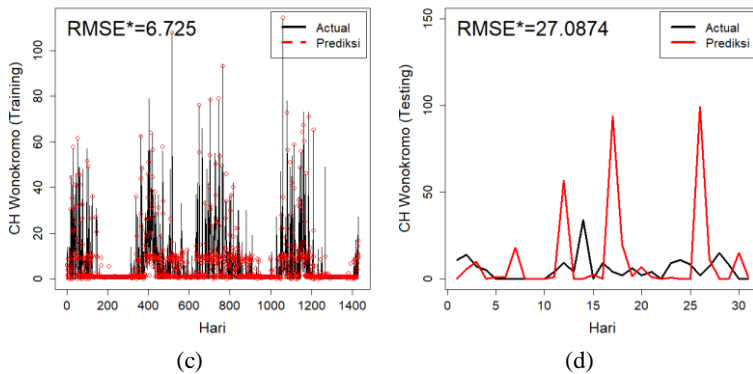
Gambar 4.30 Model Terbaik Prediksi Curah Hujan Wonorejo (a) *Training* SVR, (b) *Testing* SVR, (c) *Training* NN, dan (d) *Testing* NN

Berdasarkan Gambar 4.30 diketahui bahwa pada data *training* dan *testing*, metode NN menghasilkan nilai RMSE lebih besar dibandingkan dengan metode ν -SVR. Meskipun demikian nilai prediksi pada data *testing* menggunakan ν -SVR terdapat beberapa prediksi yang bernilai konstan. Kesimpulan yang dapat diambil adalah model terbaik yang dapat digunakan untuk memprediksi curah hujan di balai Wonorejo adalah model dengan metode ν -SVR. Dimana hasil prediksi curah hujan tertinggi pada bulan Desember 2018 sebesar 6,796896 mm.

4.10 Analisis Curah Hujan di Balai Wonokromo

Model terbaik dari metode ν -SVR maupun NN untuk memprediksi curah hujan di Wonokromo adalah model dengan *input* prediktor dari stasiun Perak 2 dan lag-lag yang signifikan pada ARIMA. Berikut merupakan perbandingan dari model terbaik metode ν -SVR dan model terbaik dengan metode NN dalam memprediksi curah hujan di balai Wonokromo.



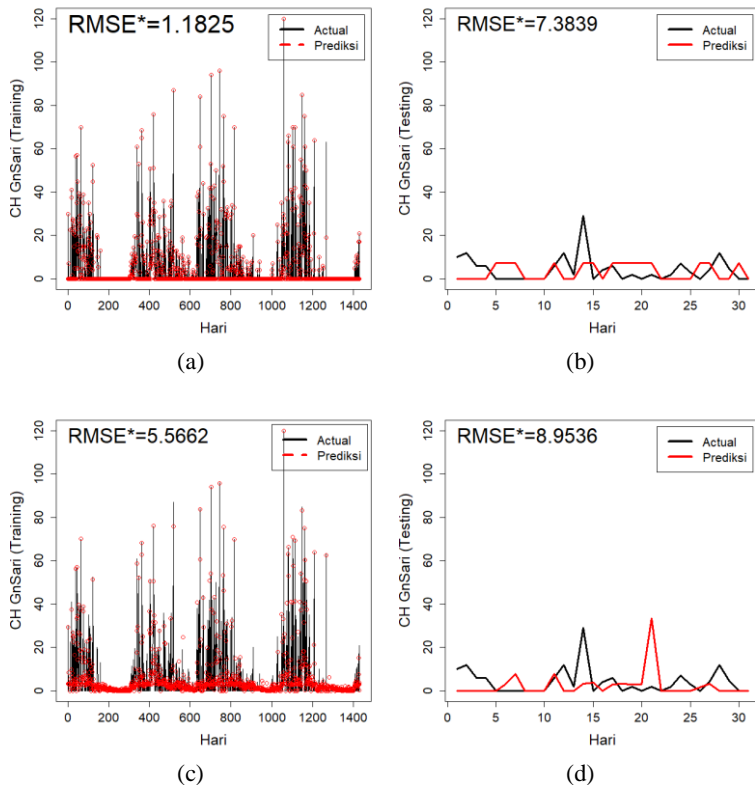


Gambar 4.31 Model Terbaik Prediksi Curah Hujan Wonokromo (a) *Training* SVR, (b) *Testing* SVR, (c) *Training* NN, dan (d) *Testing* NN

Perbandingan metode ν -SVR dan NN jika dilihat dari hasil prediksi pada data *training* dan *testing* menghasilkan kesimpulan bahwa metode ν -SVR menghasilkan nilai RMSE paling kecil baik pada data *training* maupun pada data *testing*. Namun jika dilihat dari hasil prediksi data *testing*, metode ν -SVR memberikan nilai prediksi yang cenderung konstan (mirip) dengan nilai RMSE yang kecil, sedangkan metode NN memberikan hasil prediksi yang berfluktuatif dengan nilai RMSE yang lebih besar dari ν -SVR. Sehingga diputuskan, metode terbaik yang akan digunakan untuk memprediksi curah hujan di balai Wonokromo adalah metode ν -SVR dan menghasilkan nilai prediksi tertinggi pada bulan Desember 2018 sebesar 8,125746 mm.

4.11 Analisis Curah Hujan di Balai Gunung Sari

Analisis curah hujan di balai Gunung Sari menggunakan metode ν -SVR dan NN dengan skenario dan cara berfikir sama seperti analisis-analisis sebelumnya menghasilkan nilai prediksi seperti yang ditampilkan pada Gambar 4.32.



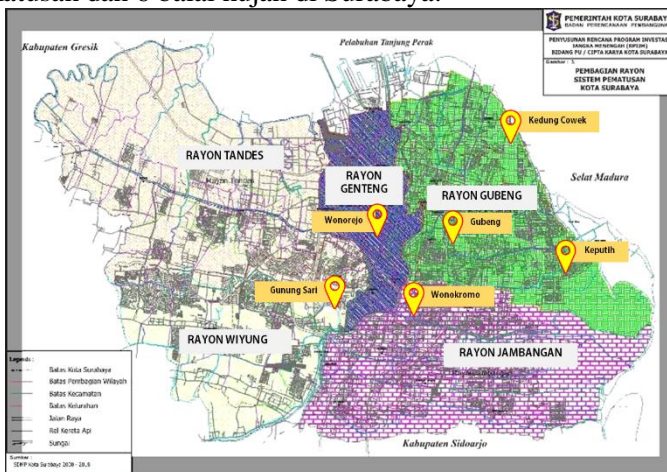
Gambar 4.32 Model Terbaik Prediksi Curah Hujan Gunung Sari (a) *Training SVR*, (b) *Testing SVR*, (c) *Training NN*, dan (d) *Testing NN*

Model terbaik untuk memprediksi curah hujan di balai Gunung Sari adalah model yang diperoleh dari metode ν -SVR yaitu dengan memasukkan *input* variabel-variabel klimatologi pada stasiun Perak 2, variabel *dummy*, dan lag signifikan pada ARIMA. Adapun hasil prediksi curah hujan tertinggi pada bulan Desember 2018 sebesar 7,406693 mm.

4.12 Prediksi Banjir di Kota Surabaya

Banjir merupakan bencana yang selalu terjadi setiap tahunnya di Kota Surabaya dan belum dapat diatasi secara tuntas sampai saat ini. Permasalahan banjir yang selalu terjadi setiap tahunnya diakibatkan oleh curah hujan yang tinggi. Salah satu pencegahan yang dapat dilakukan untuk mengantisipasi atau mitigasi banjir di Kota Surabaya adalah dengan memprediksi kapan banjir tersebut akan terjadi dengan mempertimbangkan faktor klimatologi yaitu curah hujan ekstrem. Oleh karena itu, dalam penelitian ini akan dilakukan perhitungan banjir berdasarkan prediksi curah hujan pada masing - masing balai hujan PSAWS yang berada di kota Surabaya.

Namun perlu diketahui terlebih dahulu bahwa berdasarkan perhitungan matematis rasional, banjir terjadi jika Q Hidrologi lebih tinggi daripada Q Hidrolika. Hidrolika dan hidrologi akan dilihat berdasarkan saluran-saluran di setiap rayon. Dimana diketahui bahwa berdasarkan PU Binamarga & Pematusan Kota Surabaya membagi wilayah Surabaya ke dalam lima rayon untuk pengevaluasian saluran di Kota Surabaya. Sedangkan untuk pos curah hujan di Kota Surabaya, balai PSAWS membagi wilayah Surabaya ke dalam 6 balai hujan. Berikut merupakan 5 rayon Pematusan dan 6 balai hujan di Surabaya.



Gambar 4.33 Rayon PU Binamarga & Pematusan Kota Surabaya dan Balai PSAWS Buntung Paketingan

Setiap rayon memiliki saluran sungai (primer maupun sekunder) yang memiliki nilai hidrolika masing-masing. Curah hujan di keenam balai hujan diduga akan berpengaruh terhadap Q hidrologi setiap saluran di Kota Surabaya. Pada Gambar 4.33 diketahui letak balai hujan tidak berada di setiap rayon, maka perlu dilakukan sinkronisasi terlebih dahulu antara saluran di setiap rayon berdasarkan letak balai hujan. Sehingga dapat diketahui bahwa saluran di setiap rayon akan menggunakan curah hujan sesuai lokasi balai (dapat dilihat berdasarkan kecamatan yang sesuai). Sinkronisasi antara lokasi balai terhadap lima rayon dapat dilihat pada Tabel 4.12 berikut.

Tabel 4.12 Sinkronisasi Enam Balai Hujan terhadap 5 Rayon Pematusan

No	Rayon	Balai	Jumlah Saluran
1	Gubeng	Keputih, Kedung Cowek, Gubeng	47
2	Genteng	Wonorejo	37
3	Jambangan	Wonokromo	31
4	Wiyung	Gunung Sari	18
5	Tandes	-	37

Tabel 4.12 menunjukkan bahwa pada penelitian ini hanya akan mengamati 4 dari 5 rayon yang ada. Hal tersebut dikarenakan keenam pos hujan tidak dapat mewakili rayon tandes. Sehingga pada penelitian ini, prediksi banjir hanya akan dilihat dari 133 saluran yang ada, dimana dalam 133 saluran tersebut terdapat saluran primer dan sekunder.

Setiap saluran nantinya akan memiliki Q hidrolika, dimana Q hidrolika atau debit saluran sangat dipengaruhi oleh luas penampang basah dan kecepatan aliran. Pada penelitian kali ini diasumsikan keseluruhan koefisien kekerasan *Manning* di setiap saluran adalah 0,03. Sedangkan untuk kemiringan saluran di asumsikan konstan (0,0001). Dalam praktik lapangan, nilai S pada umumnya diperoleh dari hasil pengukuran *long-section* dan *cross-section*. Jika Q hidrolika $<$ Q hidrologi, maka dapat dikatakan saluran tersebut akan diprediksi menimbulkan banjir dalam 31 hari ke depan. Perhitungan rasional yang digunakan untuk memprediksi banjir dapat dilihat pada subbab 2.8.

Berikut merupakan hasil prediksi curah hujan maksimum yang telah dianalisis berdasarkan metode terbaik sebelumnya maka saluran yang berada pada pos hujan ini yang berpotensi untuk terjadi banjir di kota Surabaya.

Tabel 4.13 Hasil Prediksi Curah Hujan Maksimum pada 6 Balai PSAWS

Balai Hujan	Curah Hujan Maksimum (mm)
Keputih	7,004366
Kedungcowek	5,895104
Gubeng	8,327121
Wonorejo	6,796896
Wonokromo	8,125746
Gunungsari	7,406693

Curah Hujan maksimum dari masing-masing balai PSAWS pada Tabel 4.13 akan digunakan untuk menghitung Q hidrologi.

4.12.1 Prediksi Banjir di Keputih

Berikut merupakan perhitungan prediksi banjir di Keputih berdasarkan hasil prediksi maksimum curah hujan Keputih.

Tabel 4.14 Perhitungan Banjir di Keputih

Saluran	Q Hidrolika	Q Hidrologi
Saluran HKSN / Kalijudan	12.2285	6.1787
Saluran Kalijudan	116.6803	4.9027
Saluran Kalisari Timur	116.6803	6.3567
Saluran laban Sari	303.0052	6.5166
Saluran Larangan	898.7190	6.7548
Saluran Manyar Kertoadi	116.6803	6.0282
Saluran Manyar Sabrangan	116.6803	6.2704
Saluran Mulyorejo	1685.2932	7.6898
Saluran Ring Road ITS	495.3067	6.4333
Saluran Sutorejo	48.2730	5.9154

Banjir akan terjadi jika Q Hidrologi lebih besar dari Q Hidrolika. Pada saluran yang terletak di wilayah Keputih diketahui bahwa perhitungan Q Hidrologi semuanya memiliki nilai yang lebih kecil dari Q Hidrolika sehingga Keputih tidak berpotensi banjir.

4.12.2 Prediksi Banjir di Kedung Cowek

Berikut merupakan perhitungan prediksi banjir di Kedung Cowek berdasarkan hasil prediksi curah hujan di balai hujan Kedung Cowek.

Tabel 4.15 Perhitungan Banjir di Kedung Cowek

Saluran	Q Hidrolika	Q Hidrologi
Saluran Gading	1258.7434	5.3408
Saluran Jeblokan	1462.3046	5.0446
Saluran Kenjeran AL	1462.3046	5.7400
Saluran Kenjeran Lama	2493.5391	5.7172
Saluran Kenjeran Pantai Ria	749.9850	5.6646
Saluran Pegirian	4355.3306	4.5933
Saluran Kapas Madya	1462.3046	6.0123
Saluran Kedung Sroko	303.0052	6.0828
Saluran Putra Agung	898.7190	5.9011
Saluran Tambaksari	77.8259	5.2565
Saluran Tambak Segaran	611.1439	5.9120
Saluran Sidotopo Wetan	109.4458	4.4733
Saluran Tambak Wedi	12196.1379	5.3014

Kawasan Kedung Cowek diprediksi tidak mengalami banjir sebab berdasarkan perhitungan banjir yang didapat dari prediksi curah hujan didapatkan bahwa perhitungan Q Hidrologi lebih kecil dari Q Hidrolika sehingga Kedung Cowek tidak berpotensi banjir.

4.12.3 Prediksi Banjir di Gubeng

Berikut merupakan perhitungan perhitungan banjir di balai Gubeng.

Tabel 4.16 Perhitungan Banjir di Gubeng

Saluran	Q Hidrolika	Q Hidrologi	Status
Saluran Jeblokan	1258.7434	6.3839	OK
Saluran Juwangan	229.7671	7.5760	OK
Saluran Kalibokor	4021.6534	5.8489	OK
Saluran Kalidami	18126.9675	7.6218	OK
Saluran Kalisari THR	77.8259	7.9033	OK
Saluran Kamboja	77.8259	7.7766	OK
Saluran Kampung Seng	156.8025	7.5488	OK
Saluran Kedung Tarukan	2493.5391	8.0191	OK
Saluran Kertajaya Viaduck	303.0052	8.5923	OK
Saluran Menur	229.7671	7.7586	OK
Saluran Mojo	4.2377	7.1538	BANJIR
Saluran Pacar	77.8259	8.4668	OK
Saluran Pucang Adi	48.2730	7.9163	OK
Saluran Pucang Jajar	168.7372	7.3854	OK
Saluran Pucang Rinenggo	26.0566	7.5942	OK
Saluran Simokerto / Donorejo	229.7671	8.1064	OK
Saluran Simolawang	116.6803	8.2561	OK
Saluran Slamet	48.2730	7.5052	OK
Saluran Srikana	116.6803	7.4546	OK
Saluran Sawah Pulo	77.8259	8.0624	OK
Saluran Sidoluhur dsk	5.2302	7.1972	BANJIR
Saluran Sidotopo Lor	749.9850	7.9714	OK
Saluran Tenggumung karya	12.2285	5.8341	OK
Saluran Wonosari Lor	61.9510	7.5601	OK

Berdasarkan perhitungan banjir di Gubeng diketahui bahwa terdapat saluran yang berpotensi terjadi banjir yaitu pada saluran Mojo dan Sidoluhur. Hal tersebut dikarenakan pada kedua saluran itu diketahui nilai Q Hidrologi lebih besar dari nilai Q Hidrolika dan menyebabkan air meluap atau banjir.

4.12.4 Prediksi Banjir di Wonorejo

Berikut merupakan perhitungan perhitungan banjir di Wonorejo berdasarkan hasil prediksi curah hujan di pos hujan Wonorejo.

Tabel 4.17 Perhitungan Banjir di Wonorejo

Saluran	Q Hidrolika	Q Hidrologi
Saluran Wonorejo III - IV	25.0757	4.9673
Saluran Jatisari	48.2730	6.5221
Saluran Kedung Turi	5.2302	4.9175
Saluran Kedungsari	77.8259	5.7244
Saluran Indrapura	244.8774	6.5038
Saluran Rembang	303.0052	6.5238
Saluran Pelemahan	80131.7048	7.0852
Saluran Jl.Semarang	14.7379	4.8953
Saluran Keputran	61.9510	5.9057
Saluran Wonorejo I - II	65.2665	6.1541
Saluran Sumber Mulyo	182.8993	6.6189
Saluran Tembok Dukuh	4305760.0685	7.3913
Saluran Jalan Opak	527661.2756	7.3383
Saluran Jepara	689.4963	6.9414
Saluran Dupak Ps.Turi	106.0394	6.1435
Saluran Kutilang PA.Pesapen	48029.8409	7.1249
Saluran Ronggolawe	96.0270	5.7202
Saluran Mojopahit	22440.0937	6.7956
Saluran Bubutan	9.4670	5.5210
Saluran Embong Malang	18.3336	5.9784
Saluran Komerling	61.9510	7.0159
Saluran Petemon Kuburan	48.2730	6.5840
Saluran Pregolan Bunder	26.0566	5.5440
Saluran Krembangan Jaya Sel	61.9510	6.5133
Saluran Ikan Mungsing	61.9510	6.0921
Saluran Juwono	116.6803	6.6347
Saluran Embong Kemiri	26.0566	5.5440
Saluran Keputran	61.9510	7.0159
Saluran Cempaka	61.9510	6.8813
Saluran Ciliwung	77.8259	6.4510

Saluran Ikan Cucut	61.9510	6.0441
Saluran Anwari	116.6803	6.3448
Saluran Sriwijaya	1718.4481	7.3653
Saluran Brawijaya	52.6400	6.0497
Saluran Serayu	2206.9159	7.2415
Saluran Greges	4355.3306	5.6313
Saluran Kali Mas	24572.5980	5.2087

Daerah Wonorejo diprediksi tidak mengalami banjir sebab berdasarkan perhitungan banjir yang didapat dari prediksi curah hujan, perhitungan Q Hidrologi lebih kecil dari Q Hidrolika sehingga Wonorejo tidak berpotensi terjadi banjir.

4.12.5 Prediksi Banjir di Wonokromo

Berikut merupakan perhitungan prediksi banjir di Wonokromo berdasarkan hasil prediksi maksimum curah hujan Wonokromo.

Tabel 4.18 Perhitungan Banjir di Wonokromo

Saluran	Q Hidrolika	Q Hidrologi
Saluran Kebon Agung	2493.5391	4.8818
Saluran Avoor Wonorejo	2206.9159	4.2110
Saluran Medokan Ayu	495.3067	5.2001
Saluran Dukuh Menanggal	229.7671	6.0492
Saluran Gayungsari	229.7671	7.0658
Saluran Gayung Kebonsari	264.8045	6.9675
Saluran Wonokromo Tangkis	349.0536	6.1818
Saluran Jetis Kulon	26.0566	6.6279
Saluran Margorejo	26.0566	6.5168
Saluran Bendul Merisi Timur	229.7671	7.5254
Saluran Bendul Merisi Besar	229.7671	7.8096
Saluran Tepi Jl. A. Yani	36.6585	6.5460
Saluran Jemursari Prapen	349.0536	6.0366
Saluran Jemur Wonosari	61.9510	7.9991
Saluran Kutisari	61.9510	7.4013
Saluran Saroni Jiwo	264.8045	7.1153
Saluran Kali Rejo	116.6803	7.2743
Saluran Kali Rungkut	495.3067	7.3206
Saluran Tenggilis Mejoyo	61.9510	7.2831
Saluran Rungkut Kidul	303.0052	7.9215

Saluran Rungkut Asri Lor	495.3067	7.7705
Saluran Kedung Asem	26.0566	6.7439
Saluran Penjarangan Sari	96.0270	6.3488
Saluran Medokan Semampir	495.3067	5.8120
Saluran Semolowaru	495.3067	6.4839
Saluran Manyar	394.2743	7.6717
Saluran Manyar Rejo	197.7822	7.6242
Saluran Ngagel Wasono	36.6585	7.8866
Saluran Kalisumo	1941.8341	7.8227
Saluran Ngagel Jaya Selatan	229.7671	7.3928
Saluran Krukah	18.3336	6.8625

Daerah Wonokromo diprediksi tidak mengalami banjir sepanjang periode Desember 2018, sebab nilai Q Hidrologi disemua saluran bernilai kurang dari Q Hidrolika.

4.12.6 Prediksi Banjir di Gunung Sari

Berikut merupakan hasil perhitungan prediksi banjir di balai Gunung Sari yang ditampilkan pada Tabel 4.19.

Tabel 4.19 Perhitungan Banjir di Gunung Sari

Saluran	Q Hidrolika	Q Hidrologi
Lidah wetan - Kulon	1258.7434	7.2423
Jeruk	229.7671	6.8513
Babadan Indah	26.0566	6.2129
Babadan Karang	168.7372	6.6013
Wiyung	394.2743	6.6058
Menganti	48.2730	5.9120
Gemol Kedurus	2206.9159	7.4876
Gogor	611.1439	7.4995
jajar Tunggal	442.9908	7.0577
Gunung Sari	749.9850	6.6425
Bogangin I	77.8259	6.4146
Bogangin Baru	36.6585	5.7048
Kebraon	96.0270	6.1485
Kebraon Manis	96.0270	6.4818
Prima Kebraon	168.7372	6.7341
Pondok Maritim	96.0270	6.7634
Karang Klumprik Barat	168.7372	6.6503
Balas Klumprik	5751773.0885	8.0079

Daerah disekitar wilayah balai Gunungsari diprediksi tidak mengalami banjir atau tidak berpotensi banjir, sebab semua saluran yang ada mampu menampung debit hidrologi. Jadi, dari keseluruhan analisis yang telah dilakukan di 6 balai hujan PSAWS maka dapat disimpulkan bahwa tidak terdapat satupun saluran yang diduga akan menimbulkan banjir untuk beberapa hari ke depan di-kelima (Keputih, Kedung Cowek, Wonorejo, Wonokromo, dan Gunung Sari) balai hujan, hal tersebut dikarenakan seluruh saluran yang ada di kelima balai dianggap sudah cukup baik dalam menampung curah hujan pada 31 hari tersebut. Sedangkan untuk saluran Gubeng terdapat 2 saluran yaitu Mojo dan Sidoluhur yang tidak mampu menampung curah hujan dengan baik dikarenakan Q Hidrologi lebih besar dari nilai Q Hidrolika yang ada sehingga menyebabkan air meluap atau banjir.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan analisis yang telah dilakukan, diperoleh kesimpulan sebagai berikut :

1. Curah hujan di 6 balai PSAWS yang ada di Surabaya memiliki pola musiman yaitu musim kemarau dan musim penghujan, hal tersebut dikarenakan letak Surabaya yang berada pada wilayah dengan tipe hujan monsun, sehingga antara musim hujan dan kemarau dapat dibedakan dengan sangat jelas. Musim hujan terjadi dari bulan Nopember hingga April, sedangkan musim kemarau terjadi di pertengahan tahun antara bulan Mei sampai Oktober. Selain itu jika dilihat dari nilai korelasi dapat diketahui bahwa hubungan antara lag curah hujan dengan curah hujan itu sendiri memiliki nilai korelasi yang lebih besar daripada hubungan curah hujan dengan variabel prediktor klimatologi lainnya.
2. Metode ν -SVR merupakan metode sangat direkomendasikan untuk memprediksi curah hujan di 6 balai PSAWS Buntung Paketingan Surabaya daripada metode NN, sebab ν -SVR menghasilkan RMSE lebih kecil dibandingkan dengan metode NN. Meskipun demikian hasil prediksi pada data *out-sample* menggunakan metode ν -SVR menunjukkan adanya kasus *overfitting*, karena hasil prediksi di data *in-sample* sangat bagus, sedangkan hasil prediksi pada data *out-sample* terdapat beberapa prediksi yang bernilai hampir sama. Sedangkan metode NN tidak sebaik metode ν -SVR yang mampu menangkap data aktual ekstrem pada data *training*.
3. Prediksi banjir di kota Surabaya yang diperoleh melalui prediksi nilai curah hujan maksimum pada bulan Desember 2018 menghasilkan kesimpulan bahwa terdapat dua saluran yaitu saluran Mojo dan saluran Sidoluhur di balai Gubeng yang diduga akan menimbulkan banjir untuk beberapa hari ke depan dalam periode Desember 2018. Hal tersebut dikarenakan saluran Mojo dan Sidoluhur dianggap tidak cukup baik dalam menampung curah hujan pada 31 periode

kedepan atau dengan kata lain disebabkan oleh nilai Q Hidrologi yang lebih besar dari nilai Q Hidrolika yang ada.

5.2 Saran

Saran untuk penelitian selanjutnya yaitu agar menggunakan *input lag-lag* yang signifikan dari hasil analisis fungsi transfer sehingga diharapkan mampu didapatkan *input* yang sesuai sebelum menggunakan metode *Nu-Support Vector Regression* dan *Neural Network*. Selain itu dalam proses optimasi parameter dalam *Nu-Support Vector Regression* lebih memperbesar *range* yang digunakan sehingga *tuning* yang dilakukan dapat lebih maksimal. Dapat dilakukan juga analisis curah hujan dalam dasarian ataupun bulanan untuk mendapatkan hasil prediksi yang lebih baik tanpa melibatkan banyak data 0.

DAFTAR PUSTAKA

- Baboo, & Shereef. (2010). An Efficient Weather Forecasting System using Artificial Neural Network. *International Journal of Neural Networks and Advanced Applications*, 1, 321-326.
- BAPEDA. (2013). *Potensi Kota Surabaya*. Surabaya: Pemerintah Kota Surabaya.
- Benestad, R. E. (2004). Empirical-Statistical Downscaling in Climate Modeling. *Eos Trans. Am. Geophys*, 85, 42-45.
- BMKG. (2010). *Buku Pedoman Pelayanan Peringatan Dini Tsunami INATEWS*. Jakarta: Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) dan GTZ-IS GITEWS.
- BNPB. (2018). *Pengetahuan Bencana Meliputi Definisi dan Jenis Bencana*. Jakarta: Graha BNPB.
- BPN. (2001). *Profil Kota Surabaya, Jawa Timur*. Surabaya: Pemerintah Kota Surabaya.
- Brown, R. G. (1963). *Smoothing, Forecasting, and Prediction of Discrete Time Series*. Englewood Cliffs: Prentice-Hall.
- Cahyadi, A. (2015). Pemodelan Spasial Banjir Berdasarkan Skenario Kenaikan Muka Air Laut Akibat Perubahan Iklim Global. *Klimatologi*, 1, 8-15.
- Candrasari, C. R., & Mukono, J. (2013). Correlation of Indoor Air Quality with Prisoners Health Complaints in the County Jail Sidoarjo. *Jurnal Kesehatan Lingkungan*, 7(1), 21-25.
- Caraka, R. E., & Tahmid, M. (2019). *Statistika Klimatologi*. Yogyakarta: Mobius.
- Chong, E., & Zak, Z. (2001). *An Introduction To Optimization*. Canada: John Wiley & Sons.
- Cryer, J. D., & Chan, K. S. (2008). *Time Series Analysis With Applications in R, Second Edition*. New York: Springer.
- Damanik, C. (2019). *Surabaya Terima Penghargaan Sekaligus*. Diakses pada tanggal 12 Februari 2019 dari Kompas: <https://regional.kompas.com/read/2019/01/14/13203031/surabaya-terima-tiga-penghargaan-sekaligus>.

- Derek, O., Allo, E. K., & Tulung, N. M. (2016). Rancang Bangun Alat Monitoring Kecepatan Angin Dengan Koneksi Wireless Menggunakan Arduino Uno. *Journal Teknik Elektro dan Komputer*, 5(4), 1-2.
- Gunn, S. (1998). *Support Vector Machines for Classification. Technical Report*. Southampton: University of Southampton.
- Gupta, A. (2003). Flood and Floodplain Management in North East India. *Ecological*, 1, 10-14.
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining Concepts and Techniques (3rd ed.)*. USA : Elsevier.
- Harafani, H. (2015). Optimasi Parameter pada Support Vector Machine Berbasis Algoritma Genetika untuk Estimasi Kebakaran Hutan. *Journal of Intelligent System*, 1(2), 82-90.
- Hong, W. C. (2008). Electric Load Forecasting by Support Vector Model. *Applied Mathematical Modelling*, 33, 2444-2454.
- Johnson, R., & Wichern, D. (2007). *Applied Multivariate Statistical Analysis Sixth Edition*. New York: Prentice Hall Inc.
- Karomah, Y., & Hendikawati, P. (2014). Estimasi Parameter Bootstrap pada Proses ARMA dan Aplikasinya pada Harga Saham. *UNNES Journal of Mathematics*, 3(2), 126-135.
- Khosravi, K., Nohani, E., Maroufinia, E., & Pourghasemi, H. R. (2016). A GIS-Based Flood Susceptibility Assessment and its Mapping in Iran: a Comparison between Frequency Ratio and Weights-of-Evidence Bivariate Statistical Models with Multi-Criteria Decision-making technique. *Journal of Hydrology*, 83(2), 947-987.
- Kodoatie, R., & Sugiyanto. (2002). *Banjir Beberapa Penyebab dan Metoda Pengendaliannya Dalam Perspektif Lingkungan*. Yogyakarta: Pustaka Pelajar.
- Kusumadewi, S. (2004). *Membangun Jaringan Syaraf Tiruan Menggunakan Matlab & Exel Link*. Yogyakarta: Graha Ilmu.

- Lippi, M., Bertini, M., & Frasconi, B. (2013). Short-Term Traffic Flow Forecasting: An Experimental Comparison of Time-Series Analysis and Supervised Learning. *IEEE Trans. Intell. Transp. Syst.*, 14(2), 871-882.
- Maharesi, R. (2013). Penggunaan Support Vector Regression pada Return Saham Syariah BEI. *Proceeding PESAT*, 5, 1-9.
- Makridarkis, S., Wheelwright, S. C., & Mc GEE. (1999). *Forecasting Methods and Application 2nd Edition*. New York : John Wiley and Sons, Inc.
- Manullang, V. S. (2014). *Modifikasi Penakar Hujan Otomatis Tipe Tipping Bucket Dengan Hall Effect Sensor AT276*. Medan: Universitas Sumatera Utara.
- Maulina, C. (2016). *Bencana Banjir*. Surabaya : Jatiarjo.
- McCulloch, M. (2010). *Chapter 1 Econometric Computing with R. Dalam Vinod (Editor) Advance in Social Science Research Using R*. Canada: John Wiley & Sons.
- Montgomery. (2008). *Introduction to Time Series Analysis and Forecasting*. Canada: Wiley Interscience.
- Navianti, D. R., Usadha, I. N., & Widjajati, F. A. (2012). Penerapan Fuzzy Inference System pada Prediksi Curah Hujan di Surabaya Utara. *Sains dan Seni*, 1(1), 23-28.
- Noer, R. (2016). *Tekanan dan Suhu Udara*. Yogyakarta : Graha Ilmu.
- Patmawati, E. (2013). *Prediksi Time Series untuk Curah Hujan Menggunakan Artificial Neural Network*. Bandung: Universitas Widyatama.
- Pirasteh, S., Pradhan, B., Mahmud, A. R., Sulaiman, W. N., & Kia, M. B. (2011). An Artificial Neural Network Model for Flood Simulation using GIS: Johor River Basin, Malaysia. *Environmental Earth Sciences*, 67, 251-264.
- Purwidhiati, A. (2015). *Prediksi Pergerakan Harga Saham Menggunakan Epsilon SVR dan Nu SVR*. Yogyakarta: Universitas Gajah Mada.
- R Core Team. (2018). *R: A language and environment for statistical computing. R Foundation for Statistical Computing*. Vienna, Austria. URL <https://www.R-project.org/>.

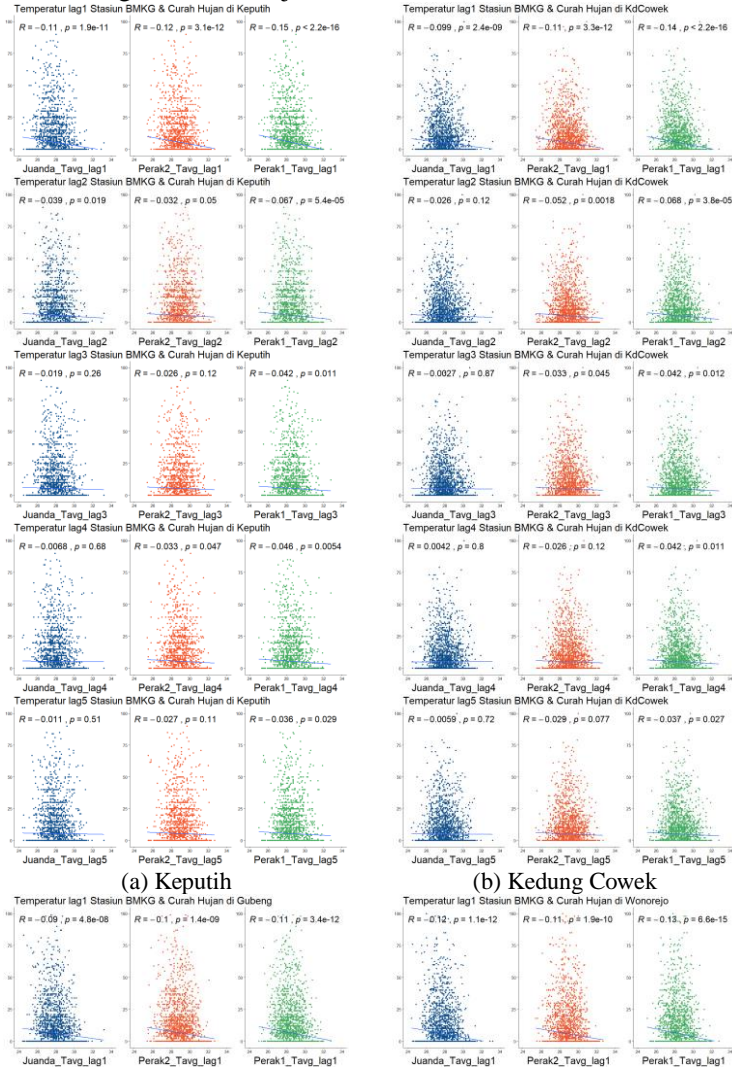
- Salowong, D. (2008). *Keuntungan dan Kerugian Letak Geografis Indonesia*. Yogyakarta : Graha Ilmu.
- Santoso, S. (2010). *Statistik Multivariate Konsep dan Aplikasi dengan SPSS*. Jakarta: Elex Media Komputindo.
- Siswanti, Y. (2011). *Model Fungsi Transfer Multivariant dan Aplikasinya untuk Meramalkan Curah Hujan di Kota Yogyakarta*. Fakultas Matematika Dan Ilmu Pengetahuan Alam . Yogyakarta: Universitas Negeri Yogyakarta.
- Suci, K. W. (2017). Peramalan Curah Hujan sebagai Pendukung Kalender Tanam Padi di Kabupaten Bojonegoro menggunakan metode ARIMA dan Support Vector Regression. *Jurnal Sains dan Seni ITS*, 6(1), 9-15.
- Suhartono, D. (2017). *Tumpukan Sampah Penyebab Banjir di Surabaya*. Diakses pada tanggal 30 Januari 2019 dari Republika: <https://nasional.republika.co.id/berita/nasional/daerah/17/11/25/ozyy71428-risma-tumpukan-sampah-penyebab-banjir-di-surabaya>.
- Suriadikusumah, A. (2007). *Analisis Curah Hujan Perhitungan & Penggunaannya*. Bandung: Rekayasa Sains.
- Tehrany, M. S., Pradhan, B., Mansor, S., & Ahmad, N. (2015). Flood Susceptibility Assessment using GIS-based Support Vector Machine Model with Different Kernel Types. *CATENA. Journal of Hidrology*, 125, 91-101.
- Walpole, R. (2007). *Pengantar Statistika Edisi ke-3*. Terjemahan oleh Ir. Bambang Sumantri. Jakarta: PT Gramedia Pustaka Utama.
- Wei, W. S. (2006). *Time Series Analysis: Univariate and Multivariate Methods, Second Edition*. United States: Pearson Education, Inc.
- Wijaya, A. S. (2017). *Optimasi Parameter Support Vector Regression untuk Pemodelan Beban Listrik*. Surabaya: Program Studi Sarjana Departemen Statistika Fakultas Matematika, Komputasi, dan Sains Data ITS.
- Zhai, Y. (2005). *Time Series Forecasting Competition Among Three Sophisticated Paradigms*. Wilmington: University of North Carolina Wilmington.

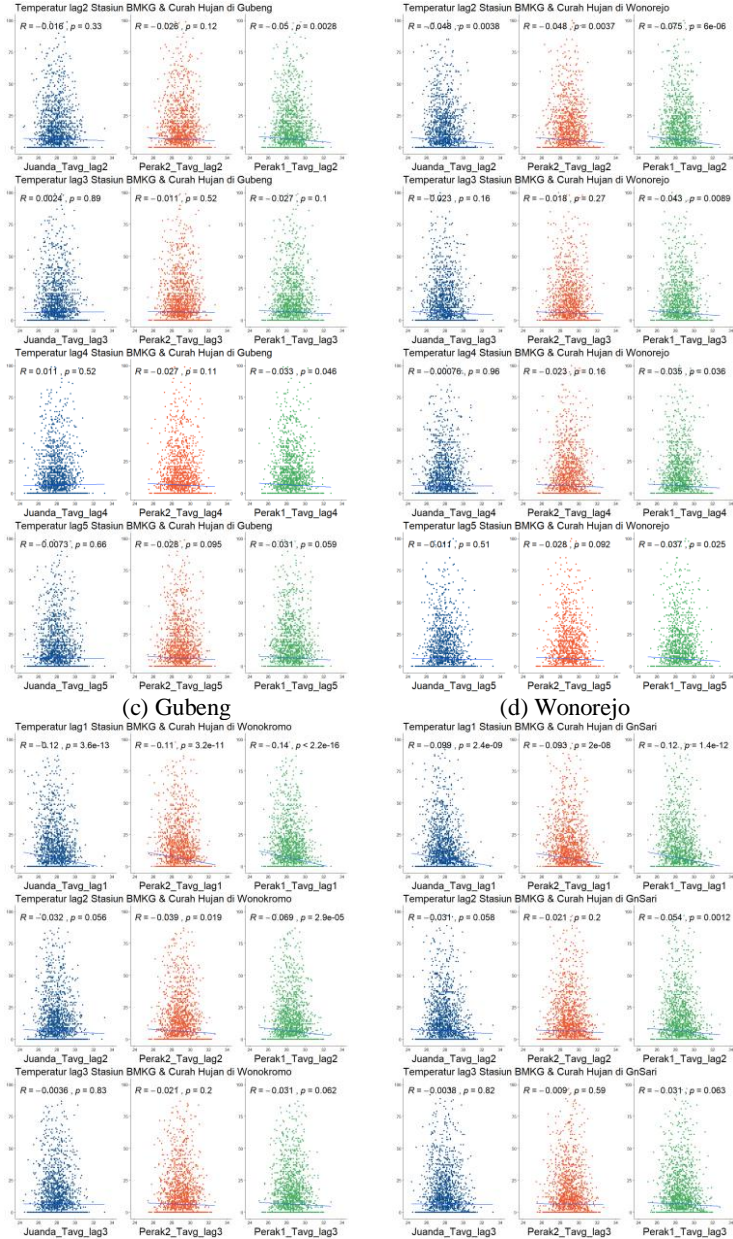
- Zhang, G. (2003). Time Series Forecasting Using a Hybrid ARIMA and Neural Network Model. *Journal Neurocomputing*, 50, 159-175.
- Zhao, W., Tao, T., & Zio, E. (2015). System Reliability Prediction by Support Vector Regression with Analytic Selection and Genetic Algorithm Parameters Selection. *Applied Soft Computing*, 30, 792-802.

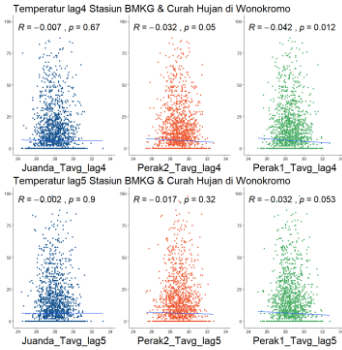
(Halaman ini sengaja dikosongkan)

LAMPIRAN

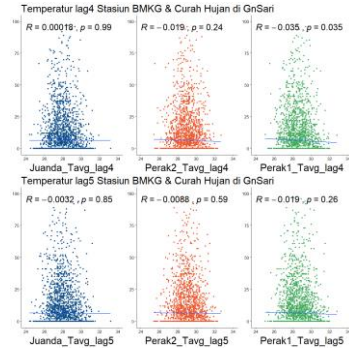
Lampiran 1. Scatterplot antara Temperatur Lag 1-5 di 3 Stasiun BMKG dengan Curah Hujan di 6 Balai PSAWS





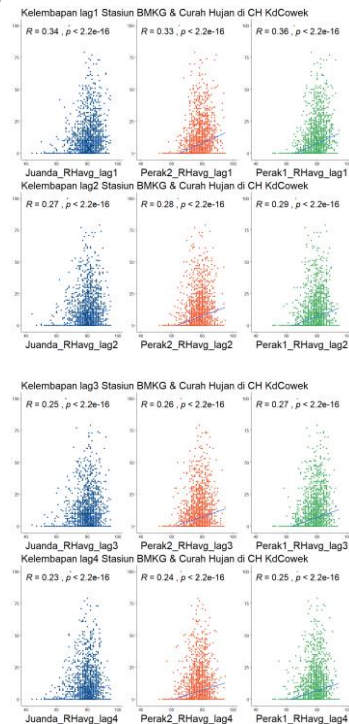
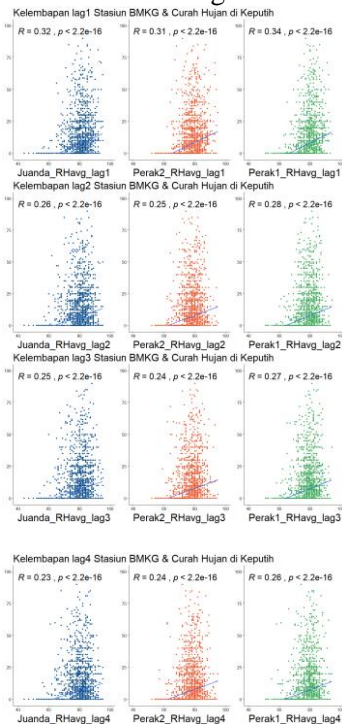


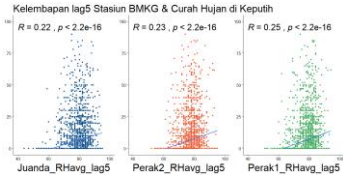
(e) Wonokromo



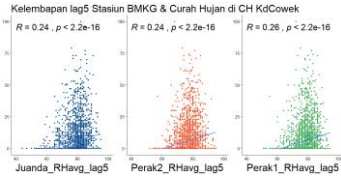
(f) Gunung Sari

Lampiran 2. Scatterplot antara Kelembapan Udara Lag 1-5 di 3 Stasiun BMKG dengan Curah Hujan di 6 Balai PSAWS

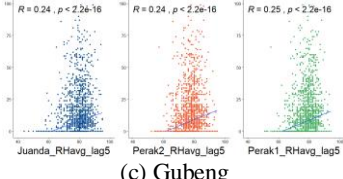
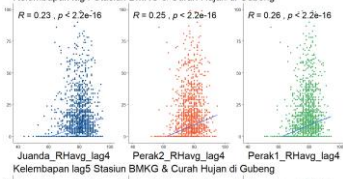
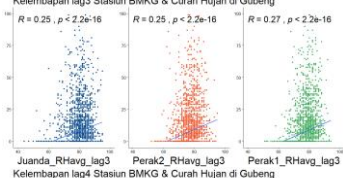
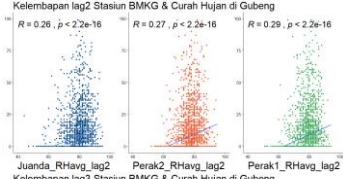
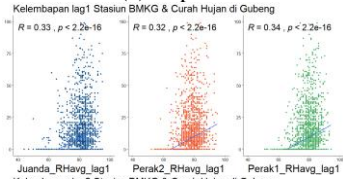




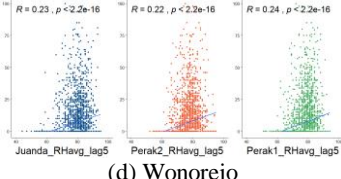
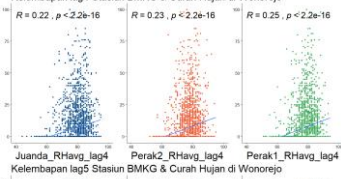
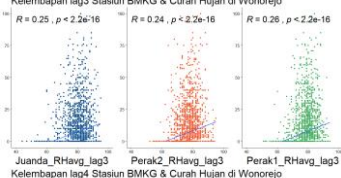
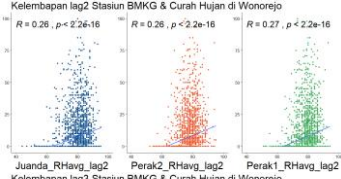
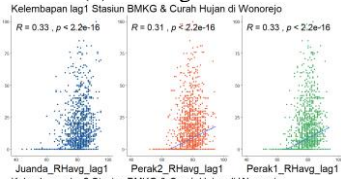
(a) Keputih



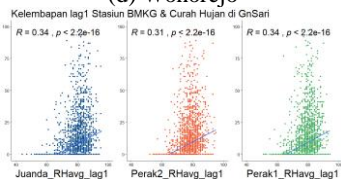
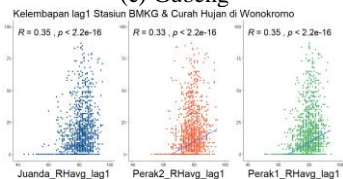
(b) Kedung Cowek

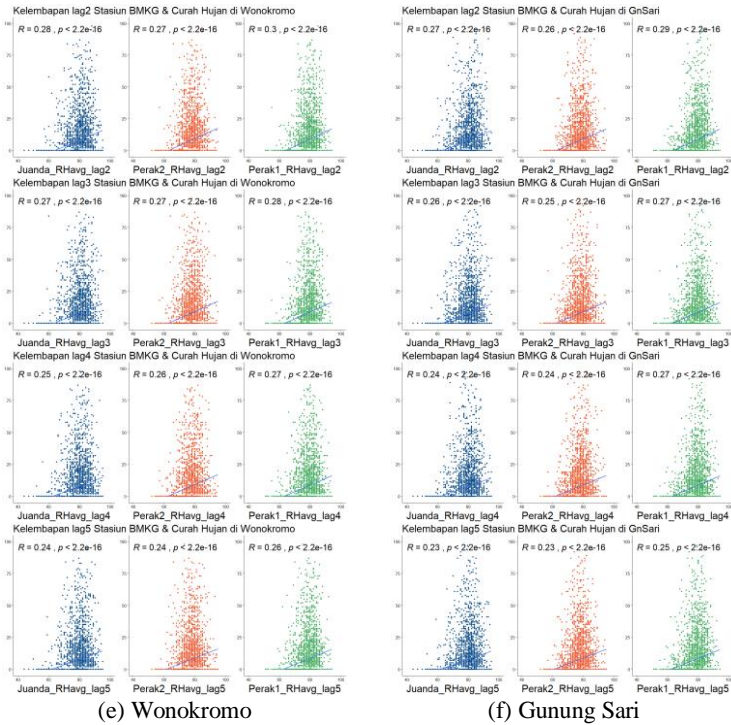


(c) Gubeng

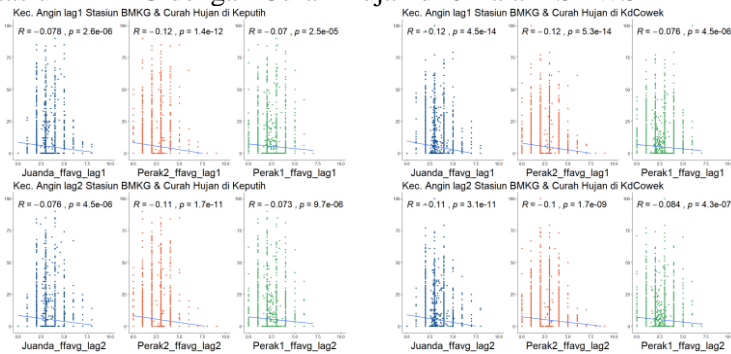


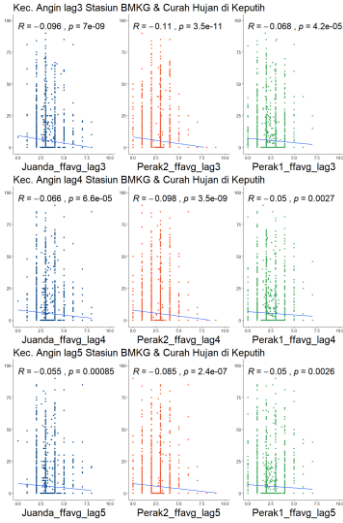
(d) Wonorejo



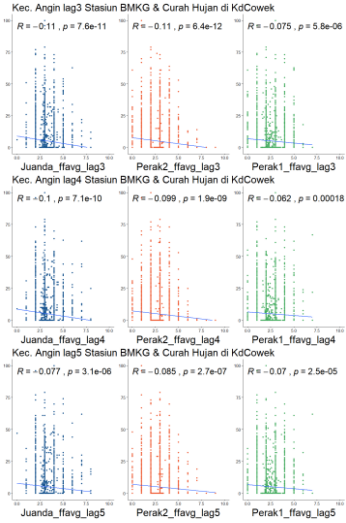


Lampiran 3. Scatterplot antara Kecepatan Angin Lag 1-5 di 3 Stasiun BMKG dengan Curah Hujan di 6 Balai PSAWS

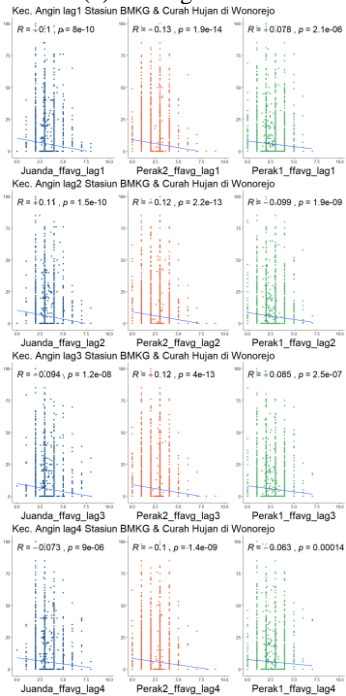
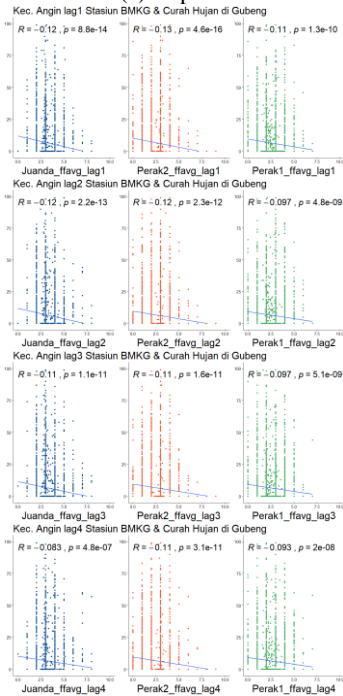


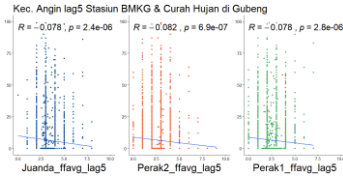


(a) Keputih

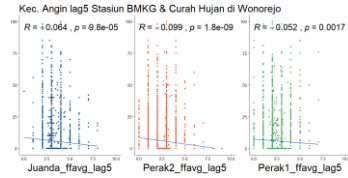


(b) Kedung Cowek

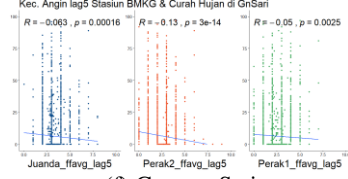
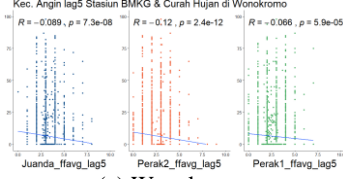
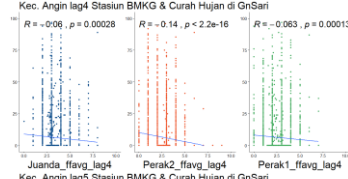
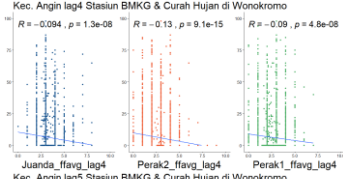
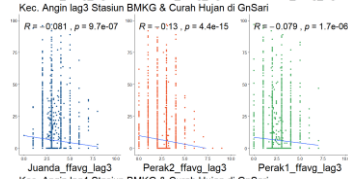
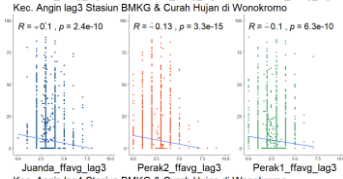
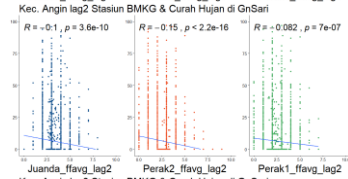
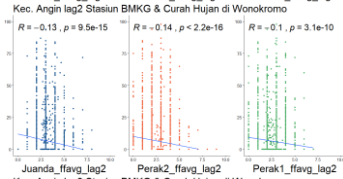
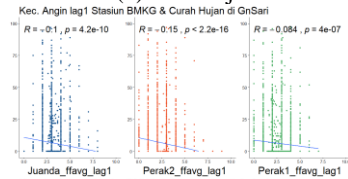
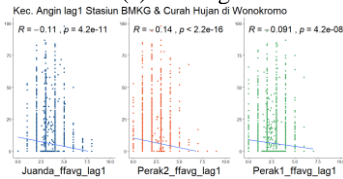




(c) Gubeng



(d) Wonorejo



(e) Wonokromo

(f) Gunung Sari

Lampiran 4. Hasil Prediksi *Variabel Predictor* Menggunakan *Holt-Winters*

X11	X21	X31	X12	X22	X32	X13	X23	X33
27.88	82.37	1.83	28.41	80.01	2.04	28.66	79.08	0.40
28.41	81.37	1.86	28.74	78.69	1.85	28.85	78.06	1.17
28.33	81.13	1.40	28.94	76.48	1.78	28.86	75.90	1.10
28.05	83.11	1.83	29.05	81.08	2.08	28.75	81.05	0.81
27.77	84.91	1.83	28.46	80.89	2.07	28.37	80.24	1.09
28.46	81.14	1.71	28.71	80.82	1.46	28.77	79.47	0.97
29.08	77.05	2.45	28.55	79.97	1.90	28.86	79.87	1.18
28.72	77.95	1.78	28.79	79.80	1.80	28.70	79.26	1.41
29.44	74.81	1.61	29.50	77.26	2.21	29.78	75.15	1.31
28.75	77.48	2.00	28.59	80.52	2.07	28.85	78.40	0.69
28.49	77.73	1.65	28.94	79.14	2.20	28.80	78.32	1.24
28.66	76.74	1.67	29.13	77.94	2.10	28.80	78.01	1.41
28.37	79.79	1.83	29.19	78.69	1.90	28.60	80.03	1.27
28.46	79.80	1.90	29.00	78.32	1.86	28.38	80.05	1.33
28.24	80.74	2.17	28.98	78.25	2.30	28.78	78.35	1.77
28.49	79.87	2.33	29.14	78.35	2.50	28.77	78.64	1.52
27.84	81.79	2.18	28.79	78.48	2.32	28.42	79.21	1.58
28.14	81.15	2.27	28.55	80.95	2.22	28.19	81.87	2.28
28.63	78.79	2.34	28.76	79.16	2.26	28.38	80.92	2.30
28.70	79.29	2.33	29.09	79.37	1.73	28.84	78.70	1.77
28.35	79.74	2.39	29.01	79.16	2.02	28.81	78.27	1.96
28.40	80.09	1.90	28.62	79.73	1.70	28.31	79.24	1.84
28.25	80.99	1.67	29.41	76.18	2.23	29.24	76.33	1.70
28.35	79.54	1.24	29.33	76.95	2.16	28.65	76.05	2.08
27.13	86.20	0.76	29.12	78.33	2.03	28.57	78.46	1.93
27.33	89.47	1.47	28.51	84.91	1.88	27.87	85.88	2.08
26.98	91.26	1.42	28.39	84.74	1.90	28.06	84.34	1.97
27.38	88.79	1.06	28.41	82.93	1.41	27.83	81.83	1.70
27.24	88.21	1.82	28.29	85.81	1.65	27.90	84.60	1.98

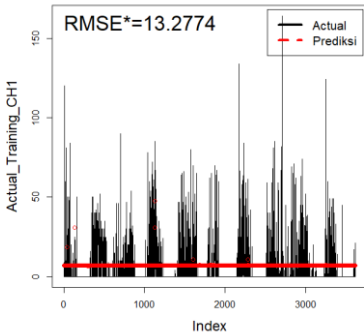
26.45	93.12	1.42	26.66	93.92	1.40	26.68	91.38	0.91
26.65	90.59	1.65	27.32	88.96	1.85	27.28	87.27	1.43

Lampiran 5. Pemodelan ν -SVR CH Keputih dengan *Input* Prediktor BMKG yang telah di Optimasi Parameter dengan 400 Kombinasi

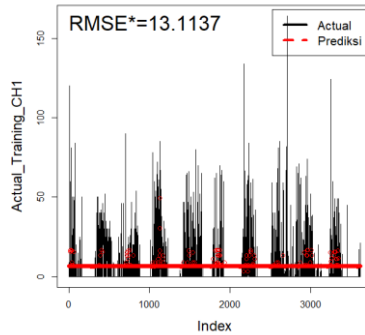
No	Model	Γ _{opt}	C _{opt}	ν _{opt}	RMSE	RMSE*
1	Y1 - X11	2	90	0.3	13.259	13.259
2	Y1 - X11 + Mt	0.5	10	0.4	13.073	13.073
3	Y1 - X21	1.5	65	0.4	13.044	13.044
4	Y1 - X21 +Mt	0.5	90	0.4	12.931	12.931
5	Y1 - X31	0.5	90	0.3	13.320	13.320
6	Y1 - X31 +Mt	2.5	95	0.3	13.317	13.317
7	Y1 - X11+X21+X31	2.5	100	0.3	11.752	11.728
8	Y1 - X11+X21+X31 +Mt	2.5	100	0.3	11.118	11.073
9	Y1 - X12	2.5	95	0.3	13.277	13.277
10	Y1 - X12 + Mt	2	100	0.4	13.114	13.114
11	Y1 - X22	1.5	5	0.4	13.231	13.050
12	Y1 - X22 +Mt	1	95	0.4	13.132	13.132
13	Y1 - X32	2	85	0.3	13.317	13.317
14	Y1 - X32 +Mt	0.5	10	0.3	13.320	13.320
15	Y1 - X12+X22+X32	2.5	100	0.3	11.859	11.836
16	Y1 - X12+X22+X32 +Mt	2.5	100	0.3	11.229	11.186
17	Y1 - X13	2.5	90	0.3	13.264	13.264
18	Y1 - X13 + Mt	1	50	0.4	13.120	13.120
19	Y1 - X23	1	35	0.4	13.054	13.054
20	Y1 - X23 +Mt	2.5	100	0.4	12.711	12.711
21	Y1 - X33	2.5	100	0.3	13.318	13.318

22	Y1 - X33 +Mt	2.5	100	0.4	13.305	13.305
23	Y1 - X13+X23+X33	2.5	100	0.3	11.538	11.512
24	Y1 - X13+X23+X33 +Mt	2.5	100	0.3	10.802	10.772

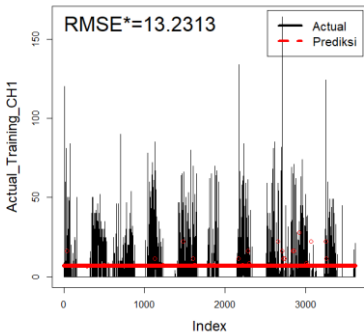
Lampiran 6. Plot *Training* Pemodelan *v*-SVR CH Keputih dengan Prediktor Klimatologi Perak 2



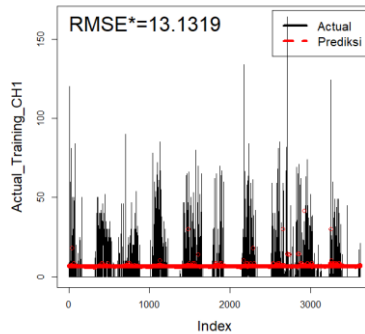
Model 9



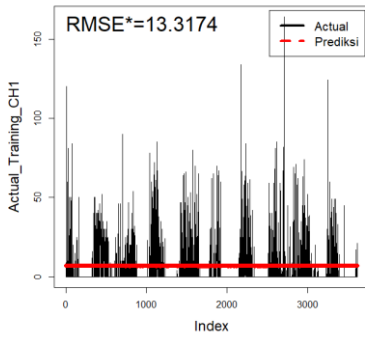
Model 10



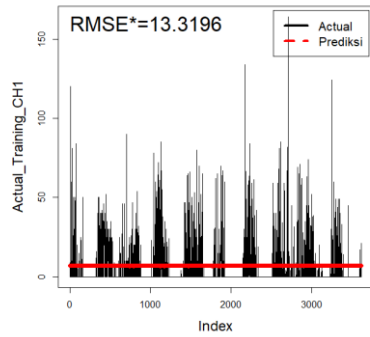
Model 11



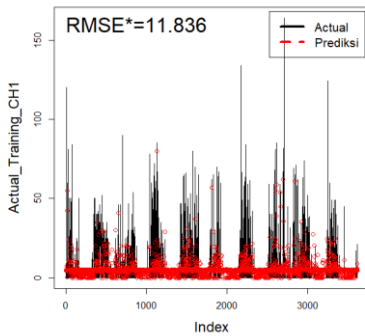
Model 12



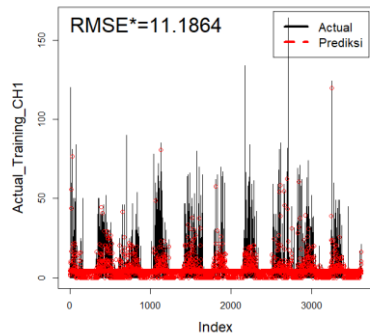
Model 13



Model 14

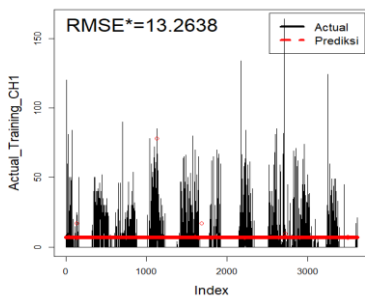


Model 15

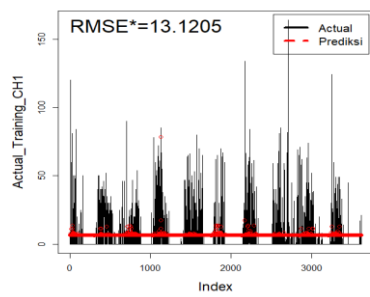


Model 16

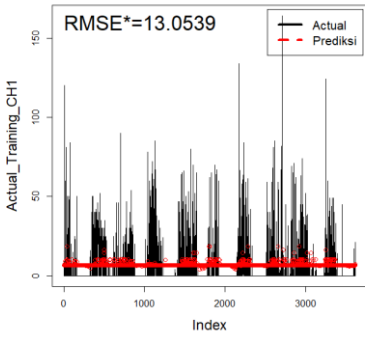
Lampiran 7. Plot *Training* Pemodelan ν -SVR CH Keputih dengan Prediktor Klimatologi Perak 1



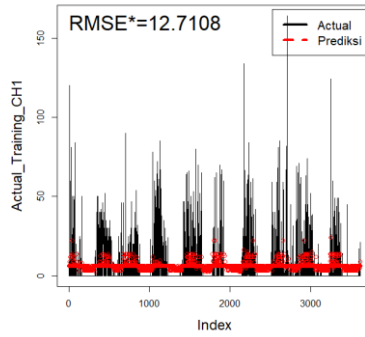
Model 17



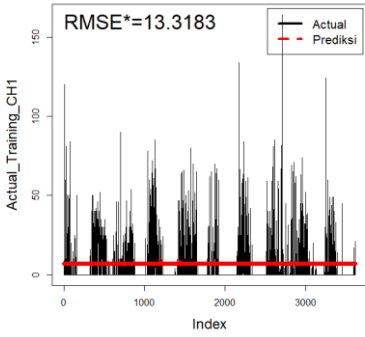
Model 18



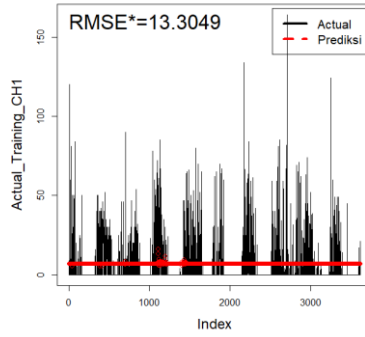
Model 19



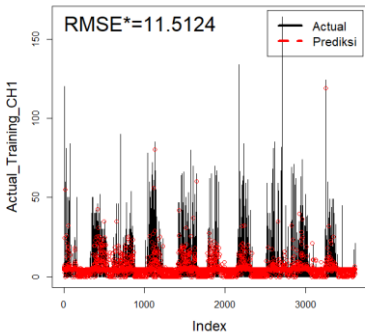
Model 20



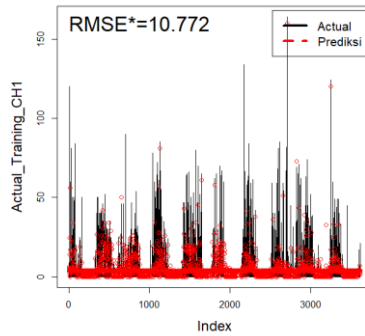
Model 21



Model 22



Model 23

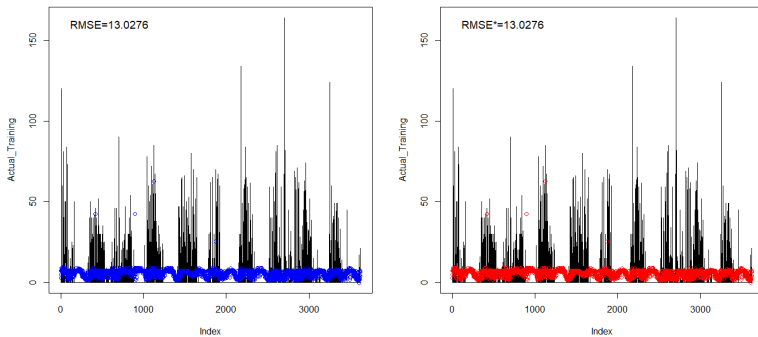


Model 24

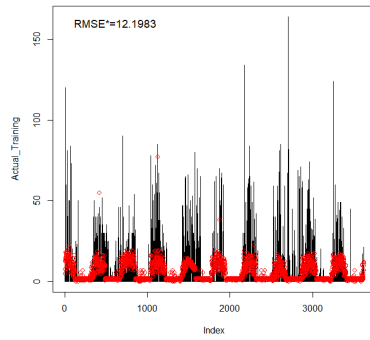
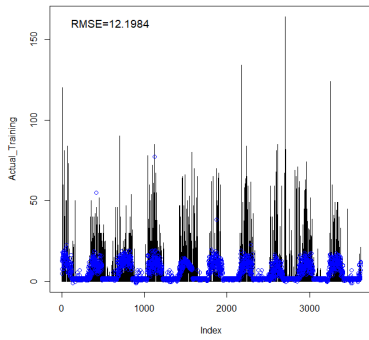
Lampiran 8. Pemodelan ν -SVR CH Keputih dengan Prediktor PCA

No	Model	Gamma opt	Cost opt	Nu opt	RMSE	RMSE*
1	Y1~PC1X1	2.5	95	0.3	13.259	13.259
2	Y1~PC1X1 + Mt	2.5	50	0.4	13.062	13.062
3	Y1~PC1X2	2.5	10	0.4	13.060	13.060
4	Y1~PC1X2 + Mt	2	20	0.4	12.798	12.798
5	Y1~PC1X3 + PC2X3	2.5	85	0.3	13.248	13.242
6	Y1~PC1X3 + PC2X3 + Mt	2.5	100	0.3	13.170	13.17
7	Y1~PC1X1 + PC1X2 + PC1X3 + PC2X3	2.5	100	0.3	8.558	8.517
8	Y1~PC1X1 + PC1X2 + PC1X3 + PC2X3 + Mt	2.5	100	0.3	7.727	7.706

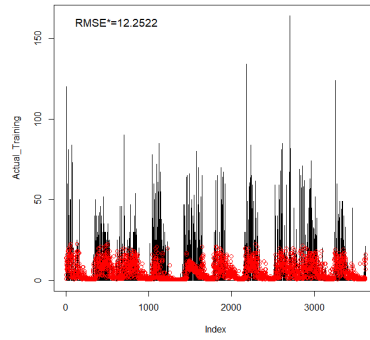
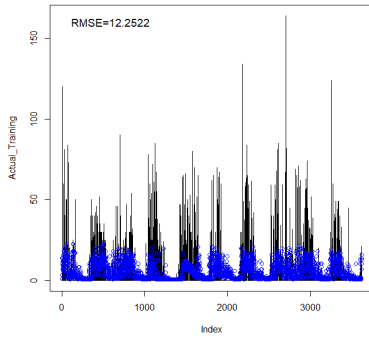
Lampiran 9. Plot Hasil Prediksi dan Aktual Menggunakan *Neural Network* untuk Prediksi CH Keputih



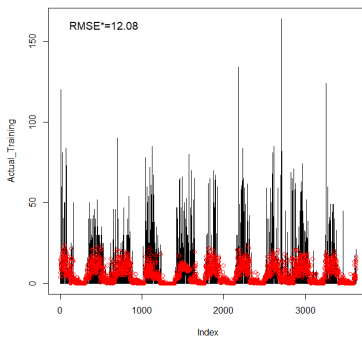
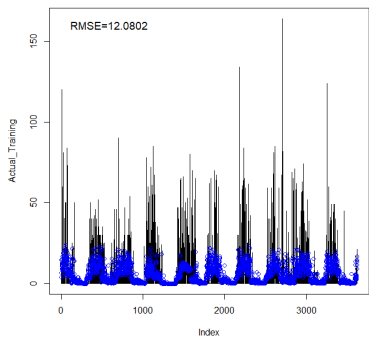
Model 1



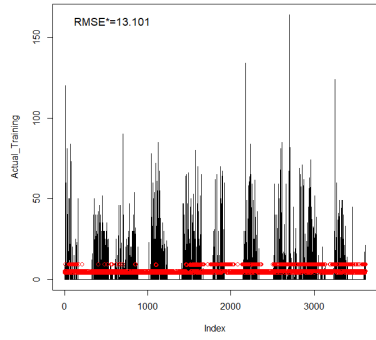
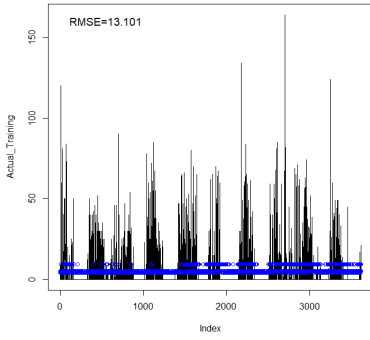
Model 2



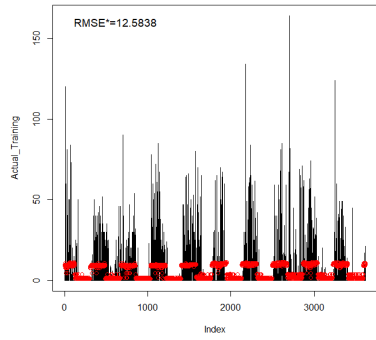
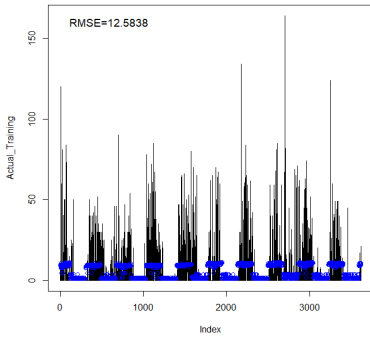
Model 3



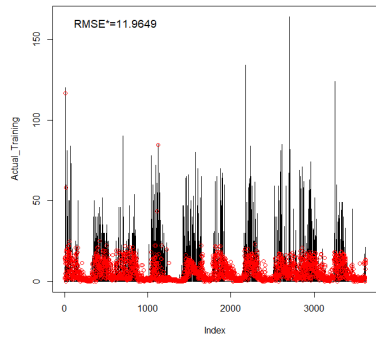
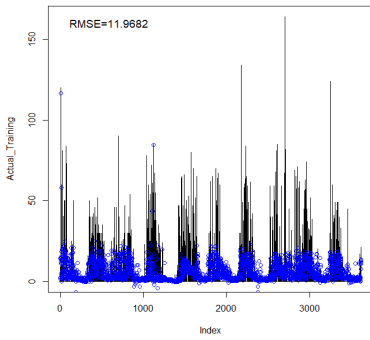
Model 4



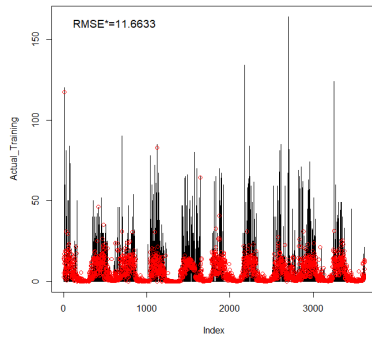
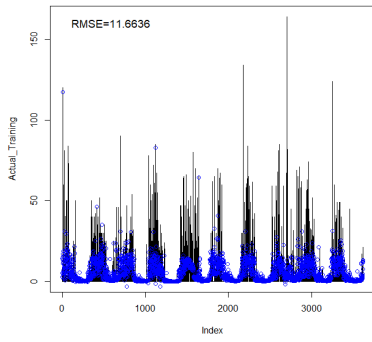
Model 5



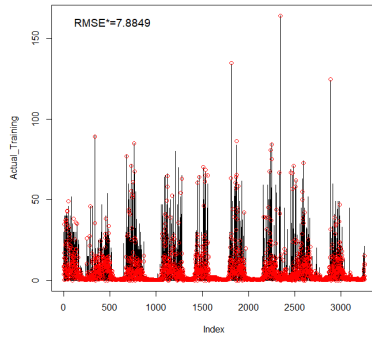
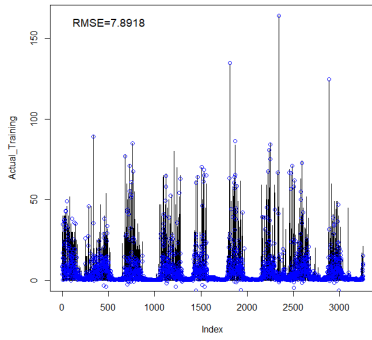
Model 6



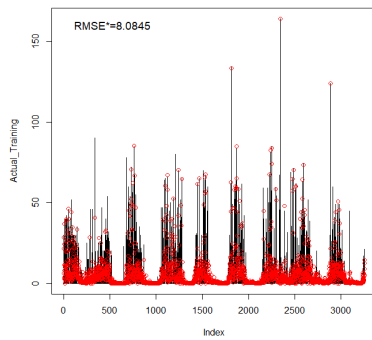
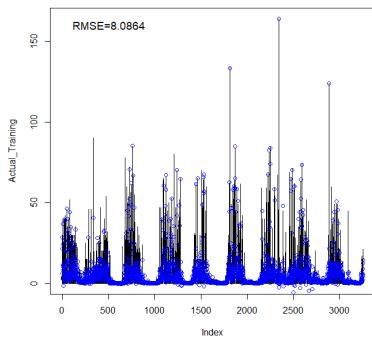
Model 7



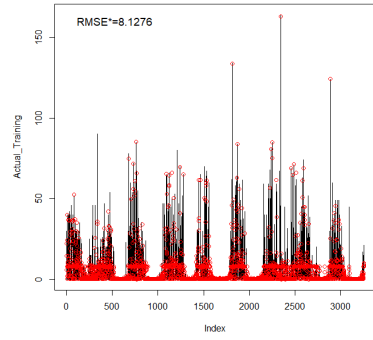
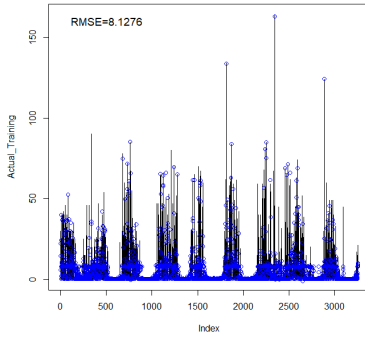
Model 8



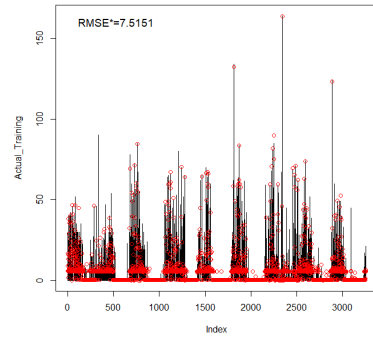
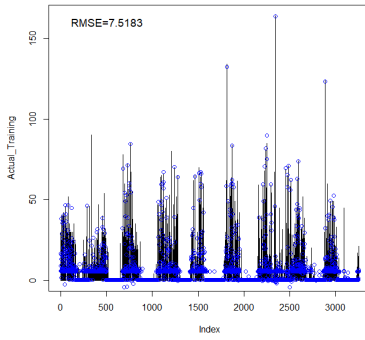
Model 9



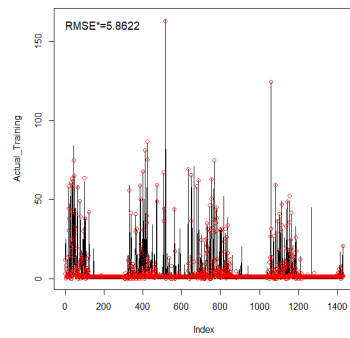
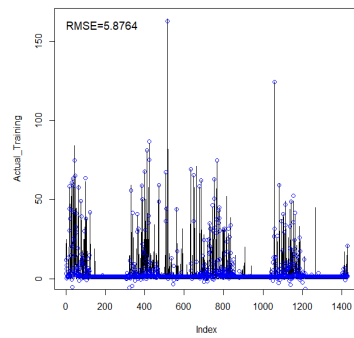
Model 10



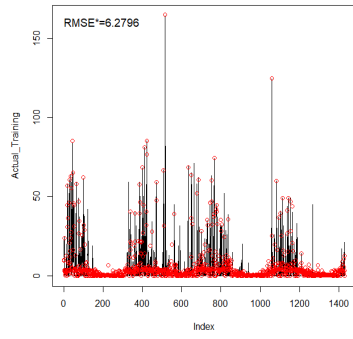
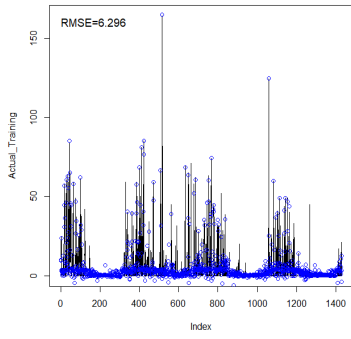
Model 11



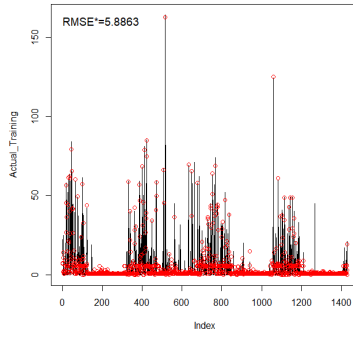
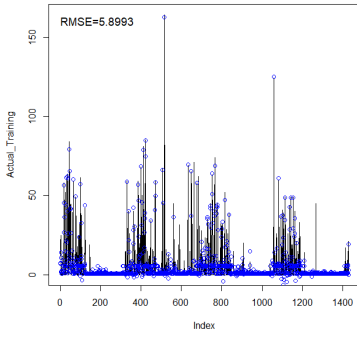
Model 12



Model 13

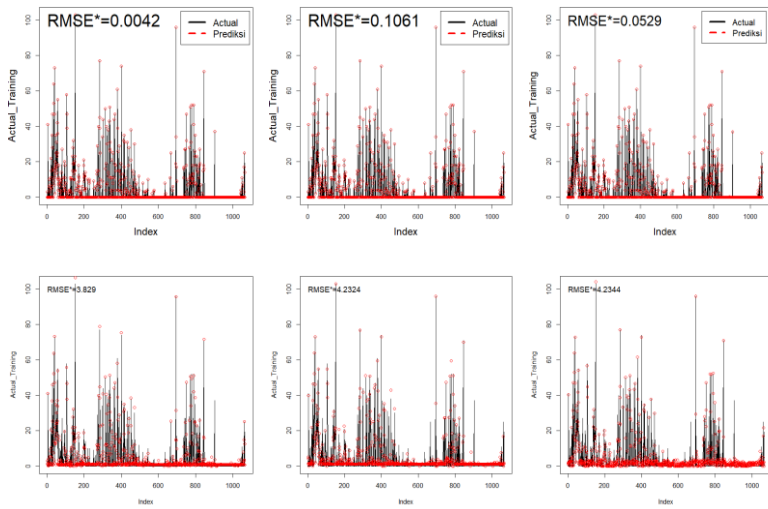


Model 14

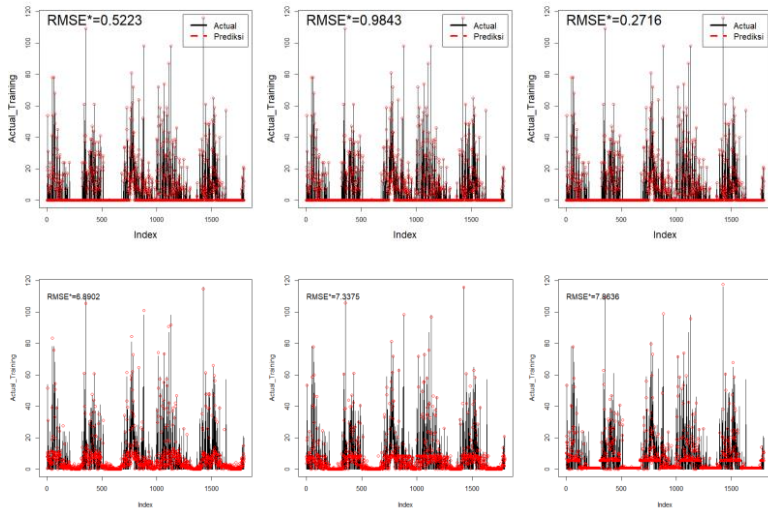


Model 15

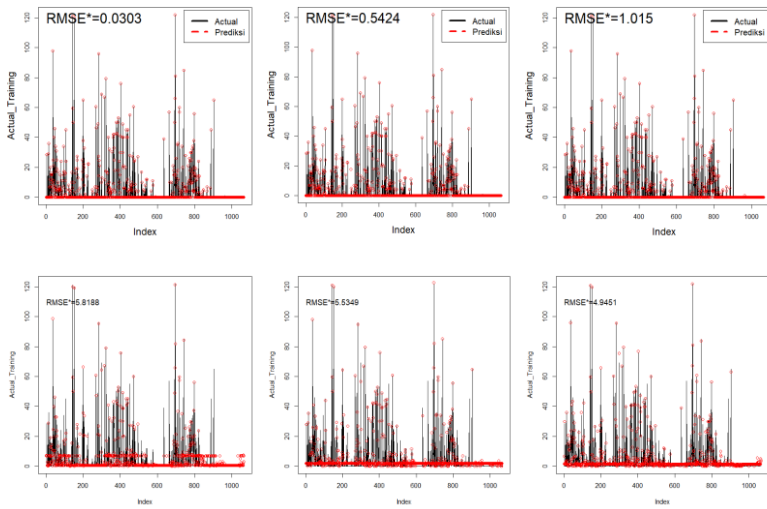
Lampiran 10. Plot 6 Percobaan Terbaik Prediksi Curah Hujan di Kedung Cowek



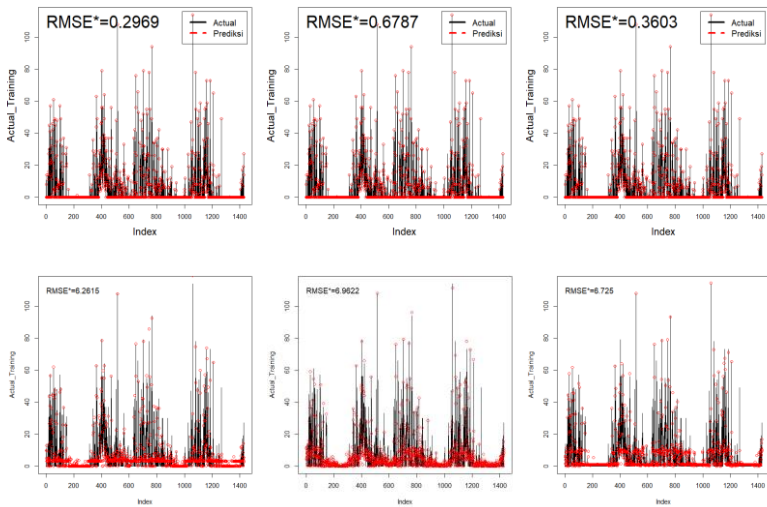
Lampiran 11. Plot 6 Percobaan Terbaik Prediksi Curah Hujan di Gubeng



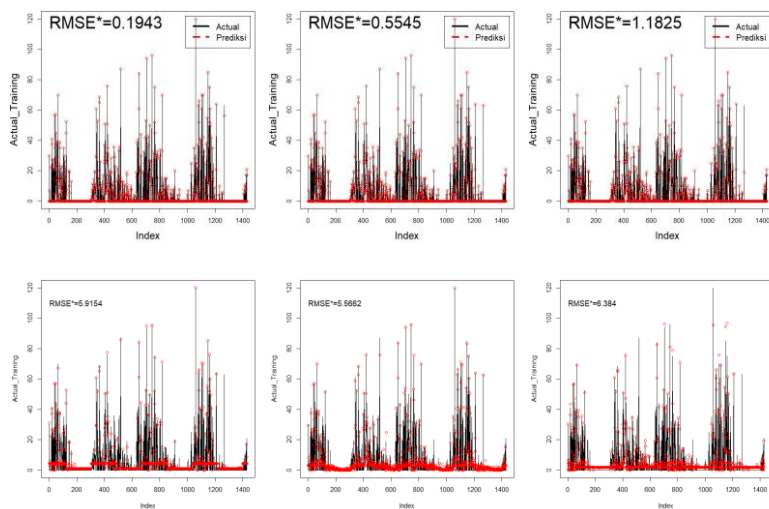
Lampiran 12. Plot 6 Percobaan Terbaik Prediksi Curah Hujan di Wonorejo



Lampiran 13. Plot 6 Percobaan Terbaik Prediksi Curah Hujan di Wonokromo



Lampiran 14. Plot 6 Percobaan Terbaik Prediksi Curah Hujan di Gunung Sari



Lampiran 15. Syntax Pre-Processing Data

```
# Membuka library yang dibutuhkan
library(openxlsx)
library(readxl)

### PRE-PROCESSING SELURUH DATA BMKG DAN DATA BALAI PSAWS

## PENGGABUNGAN SELURUH DATA BMKG STASIUN JUANDA

# Mengambil file excel keseluruhan dari 1 folder (Juanda) menjadi 1 file
setwd('E:/Juanda All')
files1 = list.files()

# Membaca seluruh data dari file yang berbeda
Juanda_R = NA
for(x in files1){
  BMKG1 = read_excel(x)
  Juanda_R = rbind(Juanda_R, BMKG1[9:39,])
}

# Menghapus missing satu baris
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")
```

```

# Mentransformasikan ke numeric
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)
)

# Mendeskripsikan variabel
sapply(Juanda_R, mode)
summary(Juanda_R)

# Meng-export data hasil gabungan ke dalam csv
?write.table
write.table(Juanda_R, file="E:/Juanda_R.csv", sep=",")

# Menyortir Data
Juanda = Juanda_R[c(1,4,5,6,10)]
Juanda = Juanda[-c(3653:3742),]

# Menghapus Data 8888 dan 9999 menjadi NA
Juanda[Juanda == 8888] = 0
Juanda[Juanda == 9999] = NA
Juanda[Juanda == 8888.0] = 0
Juanda[Juanda == 9999.0] = NA

# Meng-export data hasil sortir ke dalam csv
?write.table
write.table(Juanda, file="E:/Juanda.csv", sep=",")

## IMPUTASI MISSING VALUE DATA BMKG

# Membuka library yang dibutuhkan
library(dplyr)
library(lubridate)

# Proses Imputasi Data
Im_Juanda = Juanda %>%
  group_by(Bulan = month(Tanggal), Hari = day(Tanggal)) %>%
  mutate(Tavg = replace(Tavg, is.na(Tavg), mean(Tavg, na.rm = TRUE))) %>%
  mutate(RH_avg = replace(RH_avg, is.na(RH_avg), mean(RH_avg, na.rm = TRUE)))
  %>%
  mutate(RR = replace(RR, is.na(RR), mean(RR, na.rm = TRUE))) %>%
  mutate(ff_avg = replace(ff_avg, is.na(ff_avg), mean(ff_avg, na.rm = TRUE)))

```

```

Im_Perak2 = Perak2 %>%
  group_by(Bulan = month(Tanggal), Hari = day(Tanggal)) %>%
  mutate(Tavg = replace(Tavg, is.na(Tavg), mean(Tavg, na.rm = TRUE))) %>%
  mutate(RH_avg = replace(RH_avg, is.na(RH_avg), mean(RH_avg, na.rm = TRUE)))
%>%
  mutate(RR = replace(RR, is.na(RR), mean(RR, na.rm = TRUE))) %>%
  mutate(ff_avg = replace(ff_avg, is.na(ff_avg), mean(ff_avg, na.rm = TRUE)))
Im_Perak1 = Perak1 %>%
  group_by(Bulan = month(Tanggal), Hari = day(Tanggal)) %>%
  mutate(Tavg = replace(Tavg, is.na(Tavg), mean(Tavg, na.rm = TRUE))) %>%
  mutate(RH_avg = replace(RH_avg, is.na(RH_avg), mean(RH_avg, na.rm = TRUE)))
%>%
  mutate(RR = replace(RR, is.na(RR), mean(RR, na.rm = TRUE))) %>%
  mutate(ff_avg = replace(ff_avg, is.na(ff_avg), mean(ff_avg, na.rm = TRUE)))

## PENGGABUNGAN DATA CURAH HUJAN (CH) BALAI PSAWS DI 7 STASIUN

# Membuka library yang dibutuhkan
library(openxlsx)
library(readxl)

# Mengambil file excel keseluruhan dari 1 folder 1 file
setwd('E:/Curah Hujan All')
CH <- list.files()

# CH_Keputih
# Membaca seluruh data CH_Keputih dari file yang berbeda
CH_Keputih = NA
for(x in CH){
  PSAWS1 = read_excel(x, sheet = 1)
  CH_Keputih = rbind(CH_Keputih, stack(PSAWS1[2:33,2:13]))
}
CH_Keputih = CH_Keputih[complete.cases(CH_Keputih), ]
# Menghapus missing satu baris
CH_Keputih = cbind((seq(as.Date("2009/1/1"), as.Date("2018/12/31"), "day")),
CH_Keputih)
colnames(CH_Keputih) = c("Tanggal", "CH")
CH_Keputih = CH_Keputih[,1:2]
CH_Keputih = transform(CH_Keputih, CH=as.numeric(CH))
# Meng-export data hasil gabungan ke dalam csv
?write.table
write.table(CH_Keputih, file="E:/1. CH_Keputih.csv", sep=",")

# CH_GnSari
# Membaca seluruh data CH_GnSari dari file yang berbeda
CH_GnSari = NA
for(x in CH){
  PSAWS7 = read_excel(x, sheet = 7)
  CH_GnSari = rbind(CH_GnSari, stack(PSAWS7[2:33,2:13]))
}

```

```

CH_GnSari = CH_GnSari[complete.cases(CH_GnSari), ]
# Menghapus missing satu baris
CH_GnSari = cbind((seq(as.Date("2009/1/1"), as.Date("2018/12/31"), "day")),
CH_GnSari)
colnames(CH_GnSari) = c("Tanggal", "CH")
CH_GnSari = CH_GnSari[,1:2]
CH_GnSari = transform(CH_GnSari, CH=as.numeric(CH))
# Meng-export data hasil gabungan ke dalam csv
?write.table
write.table(CH_GnSari, file="E:/7. CH_GnSari.csv", sep=",")
# Imputasi Missing Value pada Data Curah Hujan di Gunung Sari
Im_GnSari = CH_GnSari %>%
  group_by(Bulan = month(Tanggal), Hari = day(Tanggal)) %>%
  mutate(CH = replace(CH, is.na(CH), mean(CH, na.rm = TRUE)))

## PENGGABUNGAN DATA BMKG DAN DATA BALAI PSAWS SEBELUM
IMPUTASI
DATA_MENTAH = cbind((seq(as.Date("2009/1/1"), as.Date("2018/12/31"), "day")),
  Juanda[2:5], Perak2[2:5], Perak1[2:5],
  CH_Keputih[2], CH_KdCowek[2], CH_Simo[2], CH_Gubeng[2],
CH_Wonorejo[2], CH_Wonokromo[2], CH_GnSari[2])
# Penamaan Variabel pada BIG DATA
colnames(DATA_MENTAH) = c("Tanggal",
  "Juanda_Tavg", "Juanda_RHavg", "Juanda_RR", "Juanda_ffavg",
  "Perak2_Tavg", "Perak2_RHavg", "Perak2_RR", "Perak2_ffavg",
  "Perak1_Tavg", "Perak1_RHavg", "Perak1_RR", "Perak1_ffavg",
  "CH_Keputih", "CH_KdCowek", "CH_Simo", "CH_Gubeng",
"CH_Wonorejo", "CH_Wonokromo", "CH_GnSari"
)
# Meng-export Data Mentah ke dalam csv
?write.table
write.table(DATA_MENTAH, file="E:/DATA_MENTAH.csv", sep=",")
summary(DATA_MENTAH)

## PENGGABUNGAN DATA BMKG DAN DATA BALAI PSAWS SETELAH
IMPUTASI
DATA = cbind((seq(as.Date("2009/1/1"), as.Date("2018/12/31"), "day")),
  Im_Juanda[2:5], Im_Perak2[2:5], Im_Perak1[2:5],
  CH_Keputih[2], CH_KdCowek[2], CH_Simo[2], CH_Gubeng[2],
CH_Wonorejo[2], CH_Wonokromo[2], Im_GnSari[2])
# Penamaan Variabel pada BIG DATA
colnames(DATA) = c("Tanggal",
  "Juanda_Tavg", "Juanda_RHavg", "Juanda_RR", "Juanda_ffavg",
  "Perak2_Tavg", "Perak2_RHavg", "Perak2_RR", "Perak2_ffavg",
  "Perak1_Tavg", "Perak1_RHavg", "Perak1_RR", "Perak1_ffavg",
  "CH_Keputih", "CH_KdCowek", "CH_Simo", "CH_Gubeng",
"CH_Wonorejo", "CH_Wonokromo", "CH_GnSari"
)
# Meng-export Big Data ke dalam csv
write.table(DATA, file="E:/DATA.csv", sep=",")

```


Lampiran 16. *Syntax* Eksplorasi Data, *Missing Value*, *Boxplot* Desember dan *Matrix Plot* Prediktor

```
## TIME SERIES PLOT

# Curah Hujan Balai PSAWS
Kemarau = data.frame(
  from = as.Date(c("2009-05-01", "2010-05-01", "2011-05-01", "2012-05-01", "2013-
05-01",
    "2014-05-01", "2015-05-01", "2016-05-01", "2017-05-01", "2018-05-01")),
  to = as.Date(c("2009-10-31", "2010-10-31", "2011-10-31", "2012-10-31", "2013-10-
31",
    "2014-10-31", "2015-10-31", "2016-10-31", "2017-10-31", "2018-10-31"))
)
ggplot() + geom_line(data=DATA, aes(x = Tanggal, y = CH_Keputih)) +
  geom_rect(data = Kemarau, aes(xmin = from - 1, xmax = to, ymin = -Inf, ymax = Inf),
alpha = 0.4)
win.graph()
plot_grid(
  ggplot() +
    geom_line(data=DATA, aes(x = Tanggal, y = CH_Keputih), color = "red") +
    labs(x = "Hari ke-") +
    geom_rect(data=Kemarau, aes(xmin = from, xmax = to, ymin = 0, ymax = Inf), alpha
= 0.3) +
    theme_minimal(),
  ggplot() +
    geom_line(data=DATA, aes(x = Tanggal, y = CH_KdCowek), color = "pink") +
    labs(x = "Hari ke-") +
    geom_rect(data=Kemarau, aes(xmin = from, xmax = to, ymin = 0, ymax = Inf), alpha
= 0.3) +
    theme_minimal(),
  ggplot() +
    geom_line(data=DATA, aes(x = Tanggal, y = CH_Gubeng), color = "green") +
    labs(x = "Hari ke-") +
    geom_rect(data=Kemarau, aes(xmin = from, xmax = to, ymin = 0, ymax = Inf), alpha
= 0.3) +
    theme_minimal(),
  ggplot() +
    geom_line(data=DATA, aes(x = Tanggal, y = CH_Wonorejo), color = "yellow") +
    labs(x = "Hari ke-") +
    geom_rect(data=Kemarau, aes(xmin = from, xmax = to, ymin = 0, ymax = Inf), alpha
= 0.3) +
    theme_minimal(),
  ggplot() +
    geom_line(data=DATA, aes(x = Tanggal, y = CH_Wonokromo), color = "blue") +
    labs(x = "Hari ke-") +
    geom_rect(data=Kemarau, aes(xmin = from, xmax = to, ymin = 0, ymax = Inf), alpha
= 0.3) +
    theme_minimal(),
  ggplot() +
    geom_line(data=DATA, aes(x = Tanggal, y = CH_GnSari), color = "orange") +
```

```

labs(x = "Hari ke-" ) +
  geom_rect(data=Kemarau, aes(xmin = from, xmax = to, ymin = 0, ymax = Inf), alpha
= 0.3) +
  theme_minimal(),
labels = NULL, nrow = 2, ncol = 3)

## Box PLOT

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_GnSari") +
    ylim(0.00000,165)+

```

```

geom_boxplot(color = "#41ab5d", fill = "#c7e9c0") +
  theme_minimal(),
labels = NULL, nrow = 2, ncol = 3)

#SCATTERPLOT
# Temperatur vs CH Keputih
win.graph()
plot_grid(
  ggplot(data = DATA) +
    aes(x = Juanda_Tavg, y = CH_Keputih) +
    labs(x = "Juanda_Tavg", y = NULL, subtitle = "Temperatur Stasiun BMKG v.s Curah
Hujan di Keputih") +
    xlim(24, 34) +
    ylim(0, 100) +
    geom_point(color = "#0c4c8a") +
    stat_cor(method = "pearson", label.x = 30, label.y = 95, size=4) +
    theme_minimal() + geom_smooth(method="lm",se=F),
  ggplot(data = DATA) +
    aes(x = Perak2_Tavg, y = CH_Keputih) +
    labs(x = "Perak2_Tavg", y = NULL, subtitle = " ") +
    xlim(24, 34) +
    ylim(0, 100) +
    geom_point(color = "#ef562d") +
    stat_cor(method = "pearson", label.x = 30, label.y = 95, size=4) +
    theme_minimal() + geom_smooth(method="lm",se=F),
  ggplot(data = DATA) +
    aes(x = Perak1_Tavg, y = CH_Keputih) +
    labs(x = "Perak1_Tavg", y = NULL, subtitle = " ") +
    xlim(24, 34) +
    ylim(0, 100) +
    geom_point(color = "#41ab5d") +
    stat_cor(method = "pearson", label.x = 30, label.y = 95, size=4) +
    theme_minimal() + geom_smooth(method="lm",se=F),
labels = NULL, nrow = 1, ncol = 3)

#BOXPLOT DESEMBER
#Meng-import data
DATA = read.csv ("E:/DATA.csv",sep=";",header=T)
DATA = DATA[1:3287,]
win.graph()
plot(DATA$CH_Keputih[335:365], type = "p", ylim = c(0,170))
points(DATA$CH_Keputih[730:700], pch = 0)
points(DATA$CH_Keputih[1065:1095], pch = 2)
points(DATA$CH_Keputih[1461:1431], pch = 3)
points(DATA$CH_Keputih[1796:1826], pch = 4)
points(DATA$CH_Keputih[2161:2191], pch = 5)
points(DATA$CH_Keputih[2526:2556], pch = 6)
points(DATA$CH_Keputih[2892:2922], pch = 7)
points(DATA$CH_Keputih[3257:3287], pch = 8)

```

```

#MATRIX PLOT
#Meng-import data
DATA = read.csv ("E:/DATA.csv",sep="," ,header=T)

#Mengambil data yang dibutuhkan
data = DATA[,-1]
training=data[1:3621,]
testing=data[3622:3652,]

win.graph()
pairs(~Juanda_Tavg+Juanda_RHavg+Juanda_ffavg, data=training)
win.graph()
pairs(~Juanda_Tavg+Juanda_RHavg+Juanda_ffavg, data=testing)

win.graph()
pairs(~Perak2_Tavg+Perak2_RHavg+Perak2_ffavg, data=training)
win.graph()
pairs(~Perak2_Tavg+Perak2_RHavg+Perak2_ffavg, data=testing)

win.graph()
pairs(~Perak1_Tavg+Perak1_RHavg+Perak1_ffavg, data=training)
win.graph()
pairs(~Perak1_Tavg+Perak2_RHavg+Perak1_ffavg, data=testing)

#Meng-import data
DATA = read.csv ("E:/DATA.csv",sep="," ,header=T)

#Mengambil data yang dibutuhkan
data = DATA[,-1]
max=apply(data,2,max)
min=apply(data,2,min)
scaled=as.data.frame(scale(data,center=min,scale=max-min))
head(scaled)
training=scaled[1:3621,]
testing=scaled[3622:3652,]

win.graph()
pairs(~Juanda_Tavg+Juanda_RHavg+Juanda_ffavg, data=training)
win.graph()
pairs(~Juanda_Tavg+Juanda_RHavg+Juanda_ffavg, data=testing)

win.graph()
pairs(~Perak2_Tavg+Perak2_RHavg+Perak2_ffavg, data=training)
win.graph()
pairs(~Perak2_Tavg+Perak2_RHavg+Perak2_ffavg, data=testing)

win.graph()
pairs(~Perak1_Tavg+Perak1_RHavg+Perak1_ffavg, data=training)
win.graph()
pairs(~Perak1_Tavg+Perak2_RHavg+Perak1_ffavg, data=testing)

```

Lampiran 17. Syntax Holt-Winters

```

library(Metrics)

#Meng-import data
data = read.csv ("E:/DATA.csv",sep=" ",header=T)
#Membagi data menjadi training dan testing
nrow(data)
training=data[1:3621,]
testing=data[3622:3652,]

#SEASONAL HOLTWINTERS

#Prediksi Juanda_Tavg
JTavg = ts(training$Juanda_Tavg,start = c(2009,1),frequency = 365)
(TrainJTavg = HoltWinters(JTavg))
win.graph()
plot(TrainJTavg)
lines(fitted(TrainJTavg)[,1], col="red")
predtestJTavg = predict(TrainJTavg, n.ahead=31, prediction.interval = TRUE)
win.graph()
plot(TrainJTavg, predtestJTavg)

dan seterusnya

```

Lampiran 18. Syntax PCA

```

#Meng-import data
data = read.csv ("E:/DATA.csv",sep=" ",header=T)
#Membagi data menjadi training dan testing
nrow(data)
training=data[1:3621,]
testing=data[3622:3652,]

#Temperatur
Temp = as.data.frame(cbind(training$Juanda_Tavg, training$Perak2_Tavg,
training$Perak1_Tavg))
data.pca_Tavg = prcomp(Temp)
summary(data.pca_Tavg)
data.pca_Tavg
result_X1=as.data.frame(data.pca_Tavg$x)
PC1_X1= result_X1[,1]
PC2_X1= result_X1[,2]

#Kelembapan Udara
Lembap = as.data.frame(cbind(training$Juanda_RHavg, training$Perak2_RHavg,
training$Perak1_RHavg))
data.pca_RHavg = prcomp(Lembap)
summary(data.pca_RHavg)
data.pca_RHavg
result_X2=as.data.frame(data.pca_RHavg$x)

```

```

PC1_X2= result_X2[,1]
PC2_X2= result_X2[,2]

#Kecepatan angin
KecAngin = as.data.frame(cbind(training$Juanda_ffavg, training$Perak2_ffavg,
training$Perak1_ffavg))
data.pca_ffavg = prcomp(KecAngin)
summary(data.pca_ffavg)
data.pca_ffavg
result_X3=as.data.frame(data.pca_ffavg$x)
PC1_X3= result_X3[,1]
PC2_X3= result_X3[,2]
PC3_X3= result_X3[,3]

```

Lampiran 19. *Syntax* v-SVR

```

#Membuka package yang dibutuhkan
library(e1071)
library(DataCombine)
library(Metrics)
library(zoo)
library(dplyr)
library(lubridate)
library(openxlsx)
library(readxl)

#Meng-import data
DATA = read.csv ("E:/DATA.csv",sep="," ,header=T)

#Penambahan variabel dummy
Data_d = data.frame(DATA)
Data_d = mutate(Data_d, dummy = ifelse(month(Tanggal) %in% 05:10, 0, 1))

#Penambahan Lag
Data=slide(Data_d, Var = "CH_Keputih", slideBy = -
c(1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,365))
colnames(Data)=c("Tanggal",
"Juanda_Tavg", "Juanda_RHavg", "Juanda_RR", "Juanda_ffavg",
"Perak2_Tavg", "Perak2_RHavg", "Perak2_RR", "Perak2_ffavg",
"Perak1_Tavg", "Perak1_RHavg", "Perak1_RR", "Perak1_ffavg",
"CH_Keputih", "CH_KdCowek", "CH_Simo", "CH_Gubeng",
"CH_Wonorejo", "CH_Wonokromo", "CH_GnSari",

"dummy", "lag1", "lag2", "lag3", "lag4", "lag5", "lag6", "lag7", "lag8", "lag9", "lag10", "lag1
1", "lag12",
"lag13", "lag14", "lag15", "lag365")
write.csv(Data, file="E:/Model_CH_Keputih.csv")

#Mengambil data yang dibutuhkan
data = Data[,-1]
Actual_Training=data[366:3621,13]

```

```

Actual_Testing=data[3622:3652,13]

#Membagi data menjadi training dan testing
nrow(data)
training=data[366:3621,]
testing=data[3622:3652,]

#Membuat matrix sebagai tempat menyimpan hasil iterasi
Gamma=matrix(ncol=1,nrow=400)
Cost=matrix(ncol=1,nrow=400)
Nu=matrix(ncol=1,nrow=400)
mse_Training=matrix(ncol=1,nrow=400)
rmse_Training=matrix(ncol=1,nrow=400)
mape_Training=matrix(ncol=1,nrow=400)
smape_Training=matrix(ncol=1,nrow=400)
mse_Trainingfin=matrix(ncol=1,nrow=400)
rmse_Trainingfin=matrix(ncol=1,nrow=400)
mape_Trainingfin=matrix(ncol=1,nrow=400)
smape_Trainingfin=matrix(ncol=1,nrow=400)
#Mencari parameter yang optimal dari berbagai kombinasi parameter
for (i in 1:5){
  for (j in 1:20){
    for (k in 1:4){
      cat("Gamma", i, "Cost", j, "Nu", k, "\n")
      Model2 = svm(CH_Keputih ~ Juanda_Tavg + Juanda_RHavg + Juanda_ffavg +
dummy +
      lag1 + lag2 + lag3 + lag4 + lag5 + lag6 + lag7 + lag8 + lag9 +
      lag10 + lag11 + lag12 + lag13 + lag14 + lag15 + lag365,
      data = training, type = "nu-regression",
      kernel = "radial", gamma = i*0.5, cost = j*5, nu = k*0.1+0.1, epsilon = 0.1)
      Gamma[80*(i-1)+4*(j-1)+k,]=j*0.5
      Cost[80*(i-1)+4*(j-1)+k,]=j*5
      Nu[80*(i-1)+4*(j-1)+k,]=k*0.1+0.1
      Predictawal_Training=Model2[[28]]
      mse_Training[80*(i-1)+4*(j-1)+k,]=mse(Actual_Training,Predictawal_Training)
      rmse_Training[80*(i-1)+4*(j-1)+k,]=rmse(Actual_Training,Predictawal_Training)
      mape_Training[80*(i-1)+4*(j-
1)+k,]=mape(Actual_Training,Predictawal_Training)
      smape_Training[80*(i-1)+4*(j-
1)+k,]=smape(Actual_Training,Predictawal_Training)
      Predictfinal_Training=ifelse(Predictawal_Training<0, 0, Predictawal_Training)
      mse_Trainingfin[80*(i-1)+4*(j-1)+k,]=mse(Actual_Training,Predictfinal_Training)
      rmse_Trainingfin[80*(i-1)+4*(j-
1)+k,]=rmse(Actual_Training,Predictfinal_Training)
      mape_Trainingfin[80*(i-1)+4*(j-
1)+k,]=mape(Actual_Training,Predictfinal_Training)
      smape_Trainingfin[80*(i-1)+4*(j-
1)+k,]=smape(Actual_Training,Predictfinal_Training)
    }
  }
}

```

```

Hasil=cbind.data.frame(Gamma, Cost, Nu, mse_Training, rmse_Training,
mape_Training, smape_Training,
                        mse_Trainingfin, rmse_Trainingfin, mape_Trainingfin,
smape_Trainingfin)
?write.table
write.table(Hasil, file="E:/LengkapJuanda.csv", sep=",")

#Model Fix (berdasarkan hasil optimasi)
ModelFix = svm(CH_Keputih ~ Juanda_Tavg + Juanda_RHavg + Juanda_ffavg +
dummy +
lag1 + lag2 + lag3 + lag4 + lag5 + lag6 + lag7 + lag8 + lag9 +
lag10 + lag11 + lag12 + lag13 + lag14 + lag15 + lag365,
data = training, type = "nu-regression",
kernel = "radial", gamma = 2.5, cost = 100, nu = 0.3, epsilon = 0.1)

#Hasil prediksi Model nu-SVR Terbaik (Fitting Model)
Predictawal_Training=ModelFix[[28]]
#Evaluasi kebaikan Prediksi Awal
mse_Training=mse(Actual_Training,Predictawal_Training)
rmse_Training=rmse(Actual_Training,Predictawal_Training)
mape_Training=mape(Actual_Training,Predictawal_Training)
smape_Training=smape(Actual_Training,Predictawal_Training)
Evaluasi_Training_Awal=cbind(mse_Training,rmse_Training,mape_Training,smape_
Training); Evaluasi_Training_Awal

#Transformasi hasil Prediksi Awal menjadi Prediksi Final
Predictfinal_Training=ifelse(Predictawal_Training<0, 0, Predictawal_Training)
#Evaluasi kebaikan Prediksi Final
mse_Trainingfin=mse(Actual_Training,Predictfinal_Training)
rmse_Trainingfin=rmse(Actual_Training,Predictfinal_Training)
mape_Trainingfin=mape(Actual_Training,Predictfinal_Training)
smape_Trainingfin=smape(Actual_Training,Predictfinal_Training)
Evaluasi_Training_Final=cbind(mse_Trainingfin,rmse_Trainingfin,mape_Trainingfin,
smape_Trainingfin); Evaluasi_Training_Final

#Plot hasil prediksi
summary(Predictawal_Training)
summary(Predictfinal_Training)
win.graph()
par(mfrow=c(1,2))
plot(Actual_Training, type="l")
points(Predictawal_Training, col="blue")
text(110,160,font=1,adj=0,cex=1.2,paste0("RMSE=",round(rmse_Training,digits=4)))
plot(Actual_Training, type="l")
points(Predictfinal_Training, col="red")
text(110,160,font=1,adj=0,cex=1.2,paste0("RMSE*=",round(rmse_Trainingfin,digits=
4)))

```


Lampiran 20. *Syntax* NN

```

#Membuka package yang dibutuhkan
library(neuralnet)
library(DataCombine)
library(Metrics)
library(zoo)
library(dplyr)
library(lubridate)
library(openxlsx)
library(readxl)

#Meng-import data
DATA = read.csv ("E:/DATA.csv",sep="," ,header=T)

#Penambahan variabel dummy
Data_d = data.frame(DATA)
Data_d = mutate(Data_d, dummy = ifelse(month(Tanggal) %in% 05:10, 0, 1))

#Penambahan Lag
Data=slide(Data_d, Var = "CH_Keputih", slideBy = -
c(1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,365))
colnames(Data)=c("Tanggal",
                "Juanda_Tavg","Juanda_RHavg","Juanda_RR","Juanda_ffavg",
                "Perak2_Tavg","Perak2_RHavg","Perak2_RR","Perak2_ffavg",
                "Perak1_Tavg","Perak1_RHavg","Perak1_RR","Perak1_ffavg",
                "CH_Keputih", "CH_KdCowek", "CH_Simo", "CH_Gubeng",
                "CH_Wonorejo", "CH_Wonokromo", "CH_GnSari",

"dummy","lag1","lag2","lag3","lag4","lag5","lag6","lag7","lag8","lag9","lag10","lag1
1","lag12",
                "lag13","lag14","lag15","lag365")
write.csv(Data, file="E:/Model_CH_Keputih.csv")

#Mengambil data yang dibutuhkan
data = Data[366:3652,-1]
Actual_Training=data[1:3256,13]
Actual_Testing=data[3257:3287,13]

#Scale data untuk Neural Network
max=apply(data,2,max)
min=apply(data,2,min)
scaled=as.data.frame(scale(data,center=min,scale=max-min))
head(scaled)

#Membagi data menjadi training dan testing
nrow(data)
training=scaled[1:3256,]
testing=scaled[3257:3287,]

#Membangun model Neural Network

```

```

set.seed(1)
Model2=neuralnet(CH_Keputih~Perak1_Tavg+Perak1_RHavg+Perak1_ffavg+dummy
+
      lag1+lag2+lag3+lag4+lag5+lag6+lag7+lag8+lag9+lag10+lag11+
      lag12+lag13+lag14+lag15+lag365, training, stepmax=1e+06, rep=5,
hidden=c(2,2,2),
      act.fct="tanh", linear.output=T)
summary(Model2)

#Plot model Neural Network
win.graph()
plot(Model2,rep="best")
estimasi=Model2$result.matrix; estimasi
bobot=Model2$weights; bobot

#Hasil prediksi model Neural Network terbaik (fitting model)
Predictscale_Training=as.data.frame(Model2$net.result[1])
Predictawal_Training=(Predictscale_Training*(max(data$CH_Keputih)-
min(data$CH_Keputih)))+min(data$CH_Keputih)
Predictawal_Training=Predictawal_Training[,]
#Evaluasi kebaikan prediksi awal
mse_Training=mse(Actual_Training,Predictawal_Training)
rmse_Training=rmse(Actual_Training,Predictawal_Training)
mape_Training=mape(Actual_Training,Predictawal_Training)
smape_Training=smape(Actual_Training,Predictawal_Training)
Evaluasi_Training_Awal=cbind(mse_Training,rmse_Training,mape_Training,smape_
Training); Evaluasi_Training_Awal

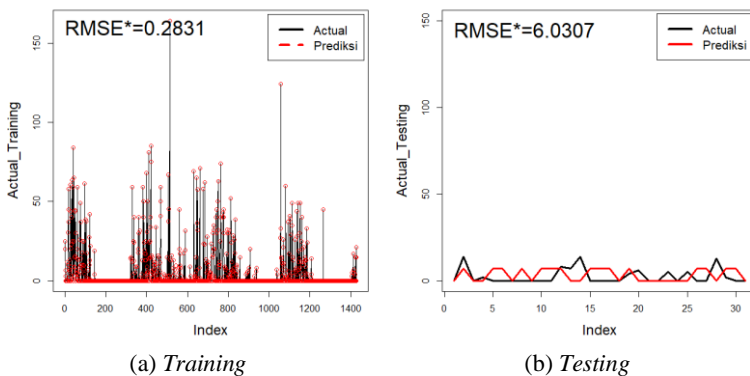
#Transformasi hasil prediksi awal menjadi prediksi final
Predictfinal_Training=ifelse(Predictawal_Training<0, 0, Predictawal_Training)
#Evaluasi kebaikan prediksi final
mse_Trainingfin=mse(Actual_Training,Predictfinal_Training)
rmse_Trainingfin=rmse(Actual_Training,Predictfinal_Training)
mape_Trainingfin=mape(Actual_Training,Predictfinal_Training)
smape_Trainingfin=smape(Actual_Training,Predictfinal_Training)
Evaluasi_Training_Final=cbind(mse_Trainingfin,rmse_Trainingfin,mape_Trainingfin,
smape_Trainingfin); Evaluasi_Training_Final

#Plot hasil prediksi
summary(Predictawal_Training)
summary(Predictfinal_Training)
win.graph()
par(mfrow=c(1,2))
plot(Actual_Training, type="l")
points(Predictawal_Training, col="blue")
text(110,160,font=1,adj=0,cex=1.2,paste0("RMSE=",round(rmse_Training,digits=4)))
plot(Actual_Training, type="l")
points(Predictfinal_Training, col="red")
text(110,160,font=1,adj=0,cex=1.2,paste0("RMSE*=",round(rmse_Trainingfin,digits=
4)))

```

Lampiran 21. Alternatif Solusi untuk Mengatasi Prediksi Konstan

Analisis tambahan mengenai cara mengatasi prediksi yang konstan salah satunya dapat dilakukan dengan menghilangkan semua observasi tanggal 29 Februari. Penghilangan observasi di tanggal tersebut bertujuan untuk membuat jumlah observasi untuk setiap tahunnya sama persis berjumlah 365. Berikut hasil analisis ulang prediksi curah hujan di daerah Keputih dengan menghilangkan observasi tanggal 29 Februari.



Dari hasil analisis diatas dapat diketahui bahwa solusi lain untuk mengatasi prediksi yang konstan dengan cara menghilangkan observasi tanggal 29 Februari tidak berpengaruh signifikan terhadap nilai RMSE dan hasil prediksi yang dihasilkan. Hasil yang diperoleh menunjukkan prediksi pada data *testing* tetap konstan meskipun terjadi sedikit penurunan pada nilai RMSE.

Lampiran 22. Surat Pernyataan Data

SURAT KETERANGAN

Saya yang bertanda tangan di bawah ini menerangkan bahwa :

1. Mahasiswa Statistika FMKSD-ITS dengan identitas berikut :

Nama : ICHA TIRHISS FEBRIANA

NRP : 062115 4000 0049

Telah mengambil data di instansi/perusahaan kami :

Nama Instansi : Dinas Pekerjaan Umum, Bina Marga, dan
Pematusan Kota Surabaya

Divisi/ bagian : *Sub Bagian Penyusun Program dan Laporan*

sejak tanggal 1 Maret 2019 sampai dengan 30 April 2019 untuk
keperluan Tugas Akhir/ ~~Thesis Semester Gasal/Genap~~* 2018/ 2019.

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.

Surabaya, 2 Juli 2019

Pimpinan Perusahaan
*Kepala Sub Bagian Penyusun
Program dan Laporan*



*(coret yang tidak perlu)

Lampiran 23. Surat Pernyataan Data (Lanjutan)**SURAT PERNYATAAN**

Saya yang bertanda tangan di bawah ini, mahasiswa Departemen Statistika FMKSD ITS :

Nama : ICHA TIRHISS FEBRIANA

NRP : 062115 4000 0049


menyatakan bahwa data prediktor 3 Stasiun BMKG (Juanda, Perak 2 & Perak 1) yang digunakan dalam Tugas Akhir ini merupakan data sekunder yang diambil dari publikasi lainnya yaitu:

Sumber : data online BMKG

Keterangan : <http://dataonline.bmkg.go.id/home>

Surat Pernyataan ini dibuat dengan sebenarnya. Apabila terdapat pemalsuan data maka saya siap menerima sanksi sesuai aturan yang berlaku.

Mengetahui
Pembimbing Tugas Akhir



Dr. rer. pol. Dedy Dwi P., S.Si., M.Si.
NIP. 19831204 200812 1 002

Surabaya, 1 Juli 2019



Icha Tirhiss Febriana
NRP. 062115 4000 0049

(Halaman ini sengaja dikosongkan)

BIODATA PENULIS



Penulis dengan nama lengkap Icha Tirhiss Febriana merupakan anak pertama dari dua bersaudara yang dilahirkan di Kabupaten Jember pada tanggal 19 Februari 1997. Pendidikan yang telah diselesaikan adalah SDN Kebonagung I (2003-2009), SMPN 1 Sukorambi (2009-2012), dan SMAN 1 JEMBER (2012-2015). Kemudian melanjutkan pendidikan di Institut Teknologi Sepuluh Nopember Departemen Statistika melalui jalur SNMPTN. Selama masa perkuliahan penulis aktif di beberapa organisasi seperti Himpunan Mahasiswa Statistika ITS (HIMASTA-ITS) sebagai Sekretaris II pada periode 2016/2017, kemudian dilanjutkan menjadi Sekretaris I di HIMASTA-ITS periode kepengurusan 2017/2018. Selain itu, penulis juga aktif dalam mengikuti kepanitiaan yang diadakan oleh tingkat jurusan, ITS, maupun nasional. Diantaranya menjadi mentor dalam kegiatan GERIGI ITS 2017, Tim Soal pada *big event* Statistika ITS yang biasa dikenal dengan Pekan Raya Statistika (PRS), serta sering menjadi sekretaris dalam kepanitiaan-kepanitiaan di tingkat jurusan. Di bidang akademik, penulis aktif dalam mengikuti perlombaan tingkat Nasional dan berhasil memperoleh berbagai prestasi diantaranya menjadi semifinalis lomba *National Statistics Challenge 2017* yang diadakan oleh Universitas Brawijaya, semifinalis *Statistics Team Competition* LOGIKA UI, dan semifinalis Pekan Analisis Statistika yang diadakan oleh Universitas Mulawarman Samarinda. Bukan hanya itu, selama menjalani perkuliahan penulis berkesempatan dalam menjalani program *internship* di Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) Pusat Jakarta, menjadi asisten dosen di beberapa mata kuliah seperti Pengantar Metode Statistika, Pengendalian Kualitas Statistika, Pemrograman Komputer dan Data Mining. Penulis juga pernah mengikuti kegiatan survei sebagai pengaplikasian ilmu statistika. Penulis terbuka dalam menerima kritik dan saran yang membangun demi kebaikan kedepan. Apabila pembaca ingin berdiskusi lebih lanjut, penulis dapat dihubungi melalui email ichatirhissfebriana@gmail.com.